

The dynamic role of mRNA m^{6}A methylation in *Arabidopsis thaliana* and *Solanum lycopersicum*

By

Samuel Erwin Chak Seng Chung

May 2023

Supervisors:
Professor Rupert Fray
Professor Graham Seymour

Thesis submitted to the University of Nottingham for Degree of Doctor of Philosophy

Acknowledgements

First and foremost, I would like to thank my supervisor, Rupert Fray, for his guidance, encouragement, and unwavering support throughout my research journey. His invaluable feedback and passion for science is reflected on his supervision and has been instrumental in shaping my PhD journey.

I would also like to thank the faculty and staff of Plant Science and SB Glasshouse team, especially Zoe Phillips, Nicky Leftley, Cade, Hailey, Mark, Catherine for their assistance and support throughout my time in the lab and also sometimes snacks too. Their knowledge, expertise, and willingness to help have been instrumental in my research. I would like to express my gratitude to Professor Nigel Mongan and Dr David Gopaulchan for their assistance with Bioinformatics analysis.

I am also thankful to the members of A49, especially Dr Alexander Bridgen for his assistance with lab work, tennis, and raves. Dr Mi Zhang and Dr Eleanor Bellows for their help with radiation and RNA work. Dr Cameron Grundy for his support at the start of my PhD and with my tomato work. I would also like to thank Zuzanna Lorenz, Rachel Thompson, and Anna Harris under my supervision for their help throughout my project in the lab and taking photos in the glasshouse.

I am also grateful to HollyAnna Ellis for her constant support throughout my PhD journey, be it going to the Peaks or taking care of mousey.

Above all, I would like to thank my parents, Wing On Chak and Min Song Kyoung, for their sacrifices, encouragement, love, and unwavering belief in me.

Thank you all for your support, guidance, and encouragement throughout my academic journey at the University of Nottingham.

Declaration

I declare that all work presented in this thesis has been produced by myself and has not been submitted for any other degree.

Table of Contents

Acknowledgements ..... **I**
Declaration ..... **II**
List of Figures ..... **VIII**
List of Tables ..... **XII**
List of Abbreviations ..... **XIII**
Abstract **1**
Chapter 1 ..... **2**
1 **Introduction** **3**
1.1 Central dogma of molecular biology 3
1.2 Nucleic Acid Structure 4
1.3 Gene expression and modifications 5
1.4 Splicing 9
1.5 N^6-methyladenosine RNA modification 11
1.6 Writer proteins of m^6A 14
1.6.1 MTA/METTL3 15
1.6.2 MTB/METTL14 15
1.6.3 FIP37/WTAP 16
1.6.4 VIRLIZER/KIAA1429 16
1.6.5 HAKAI 17
1.6.6 HIZ1 and HIZ2 17
1.7 Other identified methyltransferases 18
1.8 Eraser proteins of m^6A 18
1.9 Reader proteins of m^6A 19
1.10 Molecular functions of m^6A in plants 21
1.11 Biological roles of m^6A in plants 22
1.11.1 Root development 23
1.11.2 Trichome development 24
1.11.3 Stress response 25
1.12 Other functions of m^6A 27
1.12.1 m^6a plays a role in cancer growth and progression 27
1.12.2 Neuronal development 27
1.13 Conclusion 28
Chapter 2 ..... **30**
2 **Materials and Methods** **31**

2.1 Materials 31
2.1.1 Arabidopsis thaliana seeds 31
2.1.2 *A. thaliana* growth conditions 31
2.1.3 Tomato seeds 31
2.1.4 Reagents 32
2.1.5 Enzymes 35
2.1.6 Kits 35
2.1.7 Primers 36
2.1.8 Buffers 38
2.1.9 Plant media 41
2.1.10 Growth Conditions 42
2.1.11 Bacteria work 42
2.1.12 Radioactive nucleotides 43
2.2 Methods 43
2.2.1 Germination of seeds 43
2.2.2 Harvesting of seeds 44
2.2.3 DNA extraction 45
2.2.4 Agarose gel electrophoresis 47
2.2.5 Purification of PCR and DNA products 47
2.2.6 Purification of PCR products from gel extraction 47
2.2.7 DNA and RNA quantification 48
2.2.8 A-tailing PCR products 48
2.2.9 T4 DNA ligation 48
2.2.10 Plasmid Construction 49
2.2.11 Plant transformation via tissue culture 51
2.2.12 GUS staining 52
2.2.13 DEX induction 53
2.2.14 Phenotyping of root architecture 53
2.2.15 Pericycle activation for lateral root initiation 53
2.2.16 Total RNA extraction 54
2.2.17 Differential Expression Analysis 54
2.2.18 Code for DESeq2 RNA-seq quality assessment and differentially expressed gene files 55
2.2.19 Code for DESeq2 Heatmap 57
2.2.20 Integrative Genome Viewer 60
2.2.21 Statistical analysis 60
2.2.22 Bulked segregant analysis sequencing 60
2.2.23 cDNA synthesis 61

2.2.24 RT-qPCR 62
2.2.25 Northern Blotting 62
2.2.26 Thin Layer Chromatography 65
2.2.27 DNA restriction digestion 68
Chapter 3....70
3 Identification and characterisation of *fip37-4* suppressor mutants 71
3.1 Chapter Aims....71
3.2 Introduction 71
3.3 Results....72
3.3.1 Characterisation of *fip37-4* suppressor mutants 73
3.3.2 Screening of P1-14, P10-23, and P3-07....73
3.3.3 Height variability and segregation ratio of suppressor line P3-07 76
3.3.4 Phenotypic observations of suppressor line P10-23 and P3-07 76
3.3.5 Root development of suppressor lines in normal and abiotic stress growth medium ...82
3.3.6 Suppressor lines revert trichome branching and lateral root initiation back to WT-like85
3.3.7 Measurement of m^6A methylation 87
3.3.8 Northern blotting shows increase in transcript levels in suppressor mutants 89
3.3.9 Bulked segregant analysis and Next Generation Sequencing produces candidate genes 92
3.3.10 P3-07 filtering step 1 of candidate genes....92
3.3.11 P10-23 filtering of genes to produce candidate list 93
3.3.12 P3-07 filtering step 2 of candidate genes....102
3.3.13 P10-23 filtering step 2 of candidate genes....102
3.4 Discussion....107
3.4.1 Suppressor mutant identification reveals two lines that successfully rescues *fip37-4*.107
3.4.2 Suppressor mutants restores root architecture in normal and abiotic stress conditions 108
3.4.3 Trichome branching and lateral root initiation 110
3.4.4 An increase in m^6A levels in suppressor lines P3-07 and P10-23 111
3.4.5 Northern blotting and suppressor mutant efficiency of splicing 112
3.4.6 BSA and possible candidate genes....114
3.5 Summary and future work....116
Chapter 4....117
4 Transcriptomic analysis of *fip37-4* and suppressor mutants P3-07 and P10-23 .118
4.1 Chapter Aims....118
4.2 Introduction 118
4.3 Results....119
4.3.1 Total RNA preparation for RNA-sequencing....119

4.3.2 RNA-sequencing pipeline for processing of raw data....119
4.3.3 Coverage analysis shows that transcript expression has increased after T-DNA insertion in *fip37-4* and suppressor lines....124
4.3.4 DESeq2 RNA-sequencing sample analysis....126
4.3.5 Heatmap analysis shows clustering of genes for P3-07 and P10-23....130
4.3.6 Visualisation of spliced junction reads....133
4.3.7 DEXseq expression analysis....133
4.3.8 Ratio of counts of correctly spliced junctions....134
4.3.9 Visualisation of overlapping genes in P3-07 and P10-23 and GO analysis....138
4.4 Discussion....152
4.4.1 Transcription factor activity....153
4.4.2 Response to stress....155
4.4.3 Summary and future work....156
Chapter 5....158
5 Characterisation of MTA writer protein in *Solanum lycopersicum*....159
5.1 Chapter Aims....159
5.2 Introduction....159
5.3 Results....161
5.3.1 CRISPR/Cas9 sgRNA and vector design for *MTA, MTB*....161
5.3.2 Design of an hpRNAi vector for generating knock-down lines of MTA....162
5.3.3 Generation of *MTA* knock-down lines....164
5.3.4 GUS staining for confirmation of vector expression....166
5.3.5 Physiological characterisation of RNAi-*MTA* lines in growth media....168
5.3.6 Physiological characterisation of *RNAi-MTA* lines in Levington compost....173
5.3.7 RT-qPCR measurements on *RNAi-MTA* lines show increase in *MTA* transcripts....177
5.3.8 Two-dimensional Thin Layer Chromatography showed minimal changes in m^6A level after DEX induction....182
5.4 Discussion....184
5.4.1 Development of CRISPR/Cas9 and hpRNAi pOpOff2(kan) system....184
5.4.2 Analysis of *RNAi-MTA1* and *RNAi-MTA2* DEX induced mutant characterisation....186
5.4.3 RT-qPCR analysis on DEX induced *RNAi-MTA* lines....187
5.4.4 Summary and future work....189
Chapter 6....191
6 Characterisation of YT521-B double knock-out protein mutants in *Solanum lycopersicum*....192
6.1 Chapter Aims....192
6.2 Introduction....192
6.3 Results....199

6.3.1 Identification of double mutants of YTHDF gene knock-out 199
6.3.2 Gene expression patterns for *YTHDF1-YTHDF4* across different tissues and developmental stages 207
6.3.3 Physiological characterisation of YTHDF single knock-out mutant 209
6.4 Discussion 210
6.4.1 Identification of double homozygous YTHDF1-YTHDF4 knock-outs 210
6.4.2 Summary and future work 213
Chapter 7 215
7 General Discussion 216
7.1 Suppression of *fip37-4* 216
7.2 RNA-seq analysis on suppressor mutants 219
7.3 Knock-down of key writer complex protein MTA in tomato 221
7.4 Single and double knock-outs of YTHDF proteins in tomato 222
7.5 Conclusion 223
8 References 225
9 Appendix 236

List of Figures

Figure 1.1 Process of transcription from DNA to RNA and translation RNA to protein.....3

Figure 1.2 Nucleotide structure of DNA and RNA.....4

Figure 1.3 Structure of nitrogenous bases that make up nucleotides present in DNA and RNA.....5

Figure 1.4 Representation of Adenosine modifications. 9

Figure 1.5 Schmatic showing the process of splicing from DNA to mature mRNA.....10

Figure 1.6 Diagram showing the three key components associated with m^6A methylation in plants on mRNA. 14

Figure 2.1 Schematic showing the assembly for Northern blot. 64

Figure 3.1 Schematic representation of EMS mutagenesis of *fip37-4* 75

Figure 3.2 P3-07 representation of plant height variability across two backcrosses 78

Figure 3.3 Phenotype of P3-07 F_2 BC1 segregating line showing rescuing of *fip37-4* to WT like phenotype. 79

Figure 3.4 Phenotype of P3-07 F_2 BC2 showing suppressor line rescuing *fip37-4* to WT like phenotype. 80

Figure 3.5 Suppressor line P10-23 shows clear reversion back to WT-like phenotype.....81

Figure 3.6 Root architecture of WT, *fip37-4* and suppressor mutants in normal and abiotic stress conditions. 84

Figure 3.7 Trichome branching and lateral root initiation.....86

Figure 3.8 Thin layer chromatography indicating m^6A level in *fip37-4* and suppressor mutants.....88

Figure 3.9 Northern blotting to show transcript abundance of mutant *FIP37* and WT *FIP37*. 91

Figure 3.10 Schematic representation of BSA genes in suppressor line P3-07 and P10-23. 94

Figure 3.11 Schematic representation of suppressor mutation in BSA of Dominant, Semi-dominant, and Recessive ratios. 103
Figure 4.1 Flow chart showing the pipeline used for processing RNA-seq data to differentially expressed genes. 122
Figure 4.2 Quality check statistics on processing of raw data showing trimmed and aligned data. 123
Figure 4.3 Coverage analysis of WT, *fip37-4*, P307 and P1023 showing increased expression after T-DNA insertion for suppressor lines. 125
Figure 4.4 RNA-Seq data quality assessment.....127
Figure 4.5 Diagnostic plots on normalised counts produced by DESeq2. 128
Figure 4.6 Volcano plot of all differentially expressed genes in all samples. 129
Figure 4.7 Heatmap analyses illustrating differential expression of significant genes in different comparison. 132
Figure 4.8 Sashimi plot displaying splice junction reads and alternative spliced events in WT and *fip37-4*.....135
Figure 4.9 Sashimi plot displaying splice junction reads and alternative spliced events in P3-07 and P10-23.....136
Figure 4.10 Splicing efficiency in suppressor line shows increases in exon and intron splicing. 137
Figure 4.11 DEG across different comparison groups specific to P3-07 showing significant genes overlapping.....139
Figure 4.12 DEG across different comparison groups specific to P10-23 showing significant genes overlapping.....140
Figure 4.13 GO analysis of significant DEG in P3-07 compared to *fip37-4*. 149
Figure 4.14 GO analysis of significant DEG in P10-23 compared to *fip37-4*. 150
Figure 4.15 GO context on GO:MF, GO:BP, and GO CC. 151
Figure 5.1 Schematic of *RNAi-MTA* vector design for *MTA* gene silencing in *S. lycopersicum*. 163

Figure 5.2 Tomato transformation showing from germination of seed to producing positive transgenic plants. 165
Figure 5.3 Histochemical GUS staining of *RNAi-MTA1* and *RNAi-MTA2* DEX inducible lines. 167
Figure 5.4 Phenotype observation of DEX induced *RNAi-MTA1* on growth media. 170
Figure 5.5 Phenotype observation of DEX induced *RNAi-MTA2* on growth media. 171
Figure 5.6 Characterisation of DEX induced *RNAi-MTA1* and *RNAi-MTA2* mutant lines grown on media. Histogram representation of key plant growth measurements. 172
Figure 5.7 Phenotype observation of *RNAi-MTA1* on Levington compost. 174
Figure 5.8 Phenotype observation of *RNAi-MTA2* on compost. 175
Figure 5.9 Characterisation of DEX induced *RNAi-MTA1* and *RNAi-MTA2* mutant lines grown on compost. Histogram representation of key plant growth measurements. 176
Figure 5.10 Primer check efficiency for all *MTA* specific primers and two housekeeping genes.....180
Figure 5.11 *MTA* transcript repression by DEX inducible *RNAi-MTA* show an increase in transcript abundance on DEX induced lines. 181
Figure 5.12 Two-dimensional Thin Layer Chromatography on *RNAi-MTA2* shows minimal changes in m^6A levels on DEX induction. 183
Figure 6.1 Phylogeny comparison of YTH domain-containing proteins and the different naming conventions for tomato across different research groups. 197
Figure 6.2 Phylogeny tree of similarity showing YTHDC and YTHDF proteins. 198
Figure 6.3 Single CRISPR/Cas9 knock-out mutation in *YTHDF1 – YTHDF4* showing mutation detection.....201
Figure 6.4 Double knock-out of *YTHDF1* to *YTHDF4* combinations of CRISPR/Cas9 mediated mutations showing mutation detection. 202
Figure 6.5 Phenotype displaying single homozygous knock-out of *ythdf* mutant plant.....204

Figure 6.6 Phenotype displaying double homozygous knock-out of *ythdf* mutant plants. ...205

Figure 6.7 Characterisation of single and double mutant knock-out lines of *ythdf1* - *ythdf4.* Histogram representation of key plant growth measurements. 206

Figure 6.8 Normalised expression of *YTHDF1*-*YTHDF4* on different tomato physiological parts. 208

Figure A9.1 Double KO plants showing increased branching. 255

List of Tables

Table 2.1 Tomato plant seed used in this study. 32

Table 2.2 List of reagents of analytical grade or higher that were used in this study. 32

Table 2.3 List of enzymes of analytical grade or higher that were used in this study. 35

Table 2.4 List of kits of analytical grade or higher that were used in this study. 35

Table 2.5 List of oligonucleotides primers used in this study.....36

Table 3.1 List of candidate genes for P3-07 suppressor mutant showing high impact variants.....95

Table 3.2 List of candidate genes for P10-23 suppressor mutant showing high impact variants.....99

Table 3.3 Mutation display term and the consequences associated with the impact factor for BSA analysis. 101

Table 3.4 List of candidate genes for P3-07 suppressor mutant showing filtered ratios of VAR:WT. 104

Table 4.1 Up-regulated genes in P3-07 compared to *fip37-4*.....141

Table 4.2 Down-regulated genes in P3-07 compared to *fip37-4*. 142

Table 4.3 List of Up-regulated DEG in P10-23 compared to *fip37-4*. 145

Table 6.1 Nomenclature comparison between this study and other research groups for tomato and *A. thaliana* YTH proteins. 196

Table 6.2 Overview of mutation present in CRISPR/Cas9 mutants 203

Table A9.1 BSA analysis identified genes from P10-23 using filtering step 2.....236

Table A9.2 P10-23 x *fip37-4* down-regulated genes.....249

List of Abbreviations

ATP	Adenosine triphosphate
ALKBH5	Alpha-Ketoglutarate-Dependent Dioxygenase AlkB Homolog 5
AMV	Alfalfa Mosaic Virus
AS	Alternative splicing
bp	Base pairs
BSA	Bulked segregant analysis
CBLL1	Casitas B-lineage lymphoma-transforming sequence-like protein 1
cDNA	Complementary DNA
CDS	Coding sequence
CPSF	Cleavage and polyadenylation specificity factor
CTD	C-terminal YSPTSPS heptad repeat domain
DE	Differentially expressed
DEG	Differentially expressed gene
DEX	Dexamethasone
dsRNA	Double stranded RNA
DNA	Deoxyribonucleic acid
DRG	Dorsal root ganglion
EMS	Ethyl methanesulfonate
EMT	Epithelial-mesenchymal transition
FIP37	FKBP12-interacting protein of 37 kDa
FP10	*fip37-4* x P10-23
FP3	*fip37-4* x P3-07
FTO	Fat Mass and Obesity-Associated protein
gDNA	Genomic DNA
GO	Gene ontology
HET	Heterozygous
HOM	Homozygous
hpRNAi	Hair-pin RNAi
IGV	Integrative genomics viewer
IGV	Integrative genome viewer
kDa	Kilo-dalton
KD	Knock-down
KO	Knock-out
LRD	Lateral root density

MA	Manhattan plot
miRNA	Micro RNA
mRNA	Messenger RNA
MTC	Methyltransferase complex
MTC	Methyltransferase complex
NASC	Nottingham Arabidopsis stock centre
ncRNA	Non-coding RNA
NGS	Next generation sequencing
NLR	Nucleotide-binding, leucine-rich repeat
NMD	Nonsense mediated decay
PIC	Pre-initiation complex
Pol II	RNA polymerase II
Pri- miRNA	Primary miRNA
PRL	Primary lateral root
PTC	Premature termination codons
RBPs	RNA-binding proteins
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RNAi	RNA interference
RNAP	RNA polymerases
rRNA	Ribosomal RNA
RT-qPCR	Real-time quantitative polymerase chain reaction
RT-qPCR	Reverse transcription quantitative Polymerase Chain Reaction
SAM	S-Adenosyl Methionine
SEM	Standard error of the mean
sn	Small nuclear
snRNA	Short interfering RNA
snRNPs	Small nuclear ribonucleoproteins
SRSF	Serine- and arginine-rich splicing factors
STM	Shoot Meristemless
TEC	Transcription elongation complex
TLC	Thin Layer Chromatography
tRNA	Transfer RNA
UPF1	Up-frameshift 1
UTR	Untranslated region
VEP	Variant Effect Predictor

WT	Wild-type
WTAP	Wilms' tumor 1-associating protein
YTH	YT521-B homology
YTHDC	YTH Domain Containing
YTHDF	YTH Domain Family
ZC3HI3	Zinc finger CCCH domain-containing protein 13

Abstract

The N6-methyladenosine (m^6A) modification in mRNA is the most abundant internal chemical modification found in eukaryotic mRNA, and it is also present in a variety of bacteria and RNA viruses. This modification is highly conserved across most eukaryotic organisms and their components and interactions is becoming a popular interest in the field of RNA research. The addition of m^6A in mRNA is catalysed by the methyltransferase complex termed “Writer,” which is a multi-protein subunit complex consisting of various proteins. The modification of m^6A is recognised by the “Reader” proteins, while the removal of this modification is facilitated by proteins termed the “Erasers.” We have identified two *Arabidopsis thaliana* suppressor lines from EMS mutagenesis which revert *fip37-4 (SALK insertion line)*, a knockdown of FIP37 (Writer protein) low methylation phenotype to Wild type (WT) like phenotype. Bulk segregant analysis and RNA-Sequencing have identified candidate genes for both lines which could explain the reversion back to WT-like phenotype. This study also explores the reader proteins in *Solanum lycopersicum*, as we have identified single and double knockout of each of the four YTHDFs in tomato. Single knockout of each of the four YTHDFs reader proteins in *S. lycopersicum* did not result in embryo lethality. Furthermore, we have generated an inducible RNAi knockdown line targeting the *MTA* gene in *S. lycopersicum*, which resulted in an increase in the number of lateral branches along with an increase in MTA transcript abundance upon induction and activation of the hpRNAi construct. This work provides an avenue to produce knockdown or knockout lines for the writer protein MTA and reader proteins consisting of YTHDFs, enabling a better understanding of how m^6A methylation is deposited and recognised.

Chapter 1

Introduction

1 Introduction

1.1 Central dogma of molecular biology

The characteristics of living organisms are derived from their genetic information stored within the nucleus in the form of polymers. Nucleic acids, including deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), serve as the building blocks that store and transmit the genetic information, referred to as the organism's genome. The 'central dogma' of molecular biology, as described by Francis Crick, depicts the one-way flow of information from DNA to RNA to protein (Crick, 1970). DNA is transcribed into messenger RNA (mRNA) template, which is then translated into the sequence of amino acids specific to a protein (Figure 1.1). Subsequent discoveries have revealed that gene expression is regulated by several complex mechanisms. The ability to adjust RNA and protein content in response to environmental cues is achieved through a highly regulated process. DNA and RNA modifications play crucial roles in regulating key internal cell processes, for example the methylation of cytosine residues in DNA to form 5-methylcytosine, which can have diverse effects on gene expression (Lewis et al., 1992).

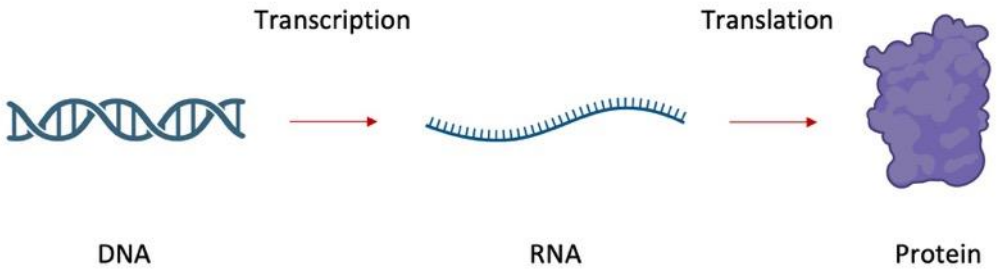


Figure 1.1 Process of transcription from DNA to RNA and translation RNA to protein.

1.2 Nucleic Acid Structure

DNA and RNA are both linear chain molecules composed of building blocks called nucleotides. The main difference between DNA and RNA is the presence of a hydroxyl group on C-2 of ribose, a sugar with the chemical formula $(CH_2O)_5$, whereas DNA is composed of deoxyribonucleotides lacking the hydroxyl group (Figure 1.2). The reactive 2'-hydroxyl groups in RNA make it less stable compared to DNA.

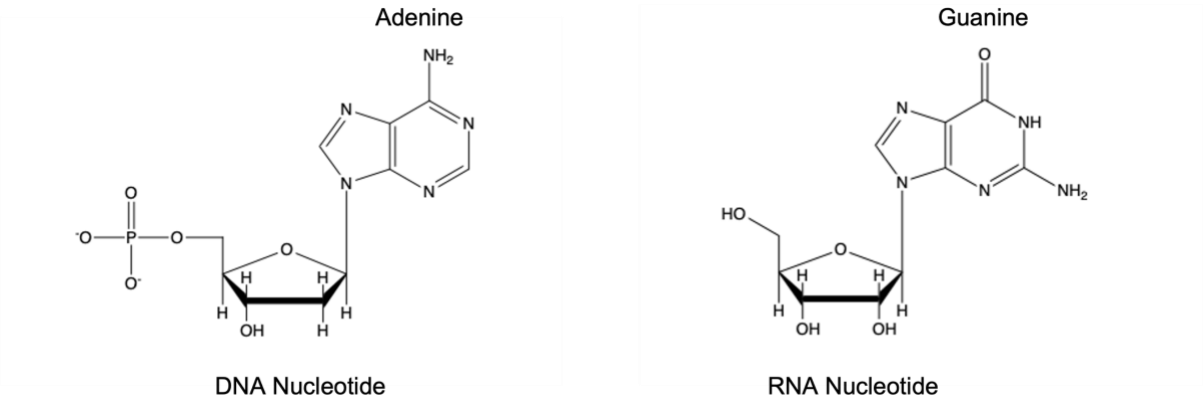
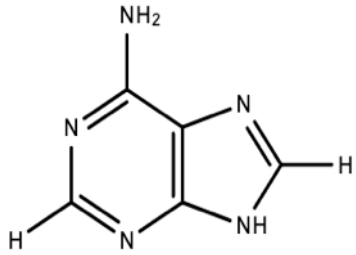


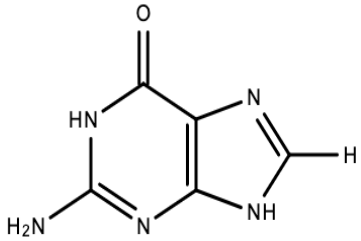
Figure 1.2 Nucleotide structure of DNA and RNA.

The purine bases adenine and guanine are present in both DNA and RNA, along with the pyrimidine base cytosine (Figure 1.3). However, the pyrimidine base thymine is only present in DNA, while uracil replaces thymine in RNA (Figure 1.2). In a nucleic acid, nucleotides are linked together in a polynucleotide chain through covalent phosphodiester bonds, which join the 5' carbon of one nucleotide to the 3' carbon of the next, forming a sugar-phosphate backbone. The structure of DNA is a double-stranded, double helical structure formed by two polynucleotide chains in an antiparallel direction through specific base pairings. Adenine pairs with thymine through two hydrogen bonds, while guanine pairs with cytosine through three hydrogen bonds.

Purine

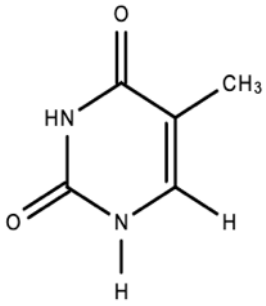


Adenine

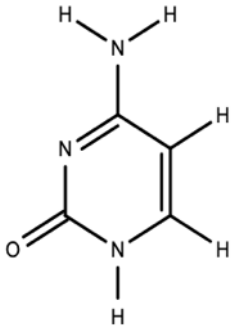


Guanine

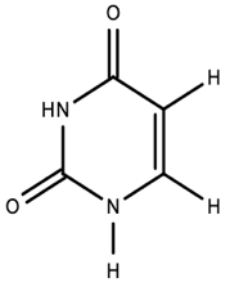
Pyrimidines



Thymine



Cytosine



Uracil

Figure 1.3 Structure of nitrogenous bases that make up nucleotides present in DNA and RNA.

1.3 Gene expression and modifications

The primary product of gene expression is RNA, which plays an essential role in protein synthesis. The formation of mRNA is a critical mechanism for transferring information for protein synthesis. For example, in addition, there are other RNAs called non-coding RNAs (ncRNAs) that do not code for proteins but serve various functions within an organism. The most abundant ncRNAs include (i) ribosomal RNA (rRNA), which forms complex three-dimensional structures that combine with polypeptides to create ribosomes responsible for protein synthesis; (ii) transfer RNA(tRNA), which acts as mediators to translate the codons.

(three-nucleotide sequence) of mRNA to produce amino acids; and (iii) small nuclear RNA (snRNAs), which can play roles in mRNA and rRNA processing.

RNA synthesis is carried out by RNA polymerases (RNAP), and in eukaryotes, a large protein complex known as RNA polymerase II (Pol II) transcribes all protein-coding genes in the nucleus. During this process, the 5' guanine cap, 3' polyadenylated tail, and removal of introns produce a mature mRNA, which can then translocate from the nucleus to the cytoplasm for translation to occur. mRNA transcription is initiated by recruiting RNA polymerase to bind to the promoter region of the target gene and assembling a large complex of proteins known as the pre-initiation complex (PIC). The PIC recognises the transcription start site, which leads to RNA polymerase synthesising RNA (Roeder, 1996). The efficiency and specificity of mRNA transcription initiation are regulated by various factors, including DNA sequence, chromatin structure, and regulatory proteins. Further to this, transcript elongation occurs for the generation of mRNA transcript, which involves the sequential addition of NTPs via phosphodiester bonding. During elongation, Pol II functions as part of an elongation complex (known as transcription elongation complex - TEC) together with the target DNA and nascent transcript. In order to coordinate transcription elongation and co-transcriptional processes, Pol II activity is influenced by a diverse range of factors. These include transcription factors, chromatin modifiers, chromatin assembly factors and chaperones, RNA processing factors, and histone variants, all of which act both independently and in combination across gene bodies.

In the process of transcription, there are often interruptions during periods of productive elongation. These pauses, sometimes termed as interruptions, were initially noticed *in vitro* with *Escherichia coli* polymerase transcribing the lac operon and lambda DNA (Dahlberg & Blattner, 1973; Kassavetis & Chamberlin, 1981; Kingston & Chamberlin, 1981). When the TEC's structural orientations alter or exhibit interactions with external protein components, the normal catalytic activity of RNAP is disrupted. (Landick, 2006; Adelman & Lis, 2012). Pausing has been shown to allow for maximum binding of termination and 3' end processing factors to

Pol II (Glover-Cutter *et al.*, 2008). Termination of transcription is a complex process that involves Pol II recognising a termination signal and subsequently releasing the nascent RNA transcript. In mammalian cells, the central hexamer AAUAAA motif is a key cis-acting element within the polyadenylation signals that is required for 3' cleavage and polyadenylation of pre-mRNA transcripts. This polyadenylation signal motif is recognised by CPSF (cleavage and polyadenylation specificity factor), which is a multi-subunit complex responsible for polyadenylation-site definition, pre-mRNA cleavage, and poly-(A)-polymerase recruitment (Proudfoot & Brownlee, 1976; Tian & Graber, 2012; Shi & Manley, 2015). CPSF is comprised of six subunits: CPSF160, CPSF100, CPSF73, CPSF30, Fip1, and Wdr33 (Mandel *et al.*, 2008; Shi *et al.*, 2009; Hunt *et al.*, 2012).

CPSF30 is a crucial polyadenylation factor in plants, and it is the ortholog of the 30 kDa subunit of mammalian CPSF. Recent studies on CPSF30 in the model plant *Arabidopsis thaliana* have revealed its unique biochemical properties. In contrast to mammals, yeast, and *Drosophila melanogaster*, the CPSF30 gene in plants, specifically in *A. thaliana* (AT1G30460), generates two transcripts using alternative poly(A) sites, leading to the production of two different polypeptides (Delaney *et al.*, 2006; Hunt *et al.*, 2012).

In *A. thaliana*, the 28 kDa protein encoded by the smaller transcript is likely analogous to other eukaryotic CPSF30 proteins. On the other hand, the larger transcript encodes a 65 kDa polypeptide in which the majority of the 28 kDa polypeptide is fused with a YTH domain-containing polypeptide (Delaney *et al.*, 2006). YTH domain family proteins have recently been shown to bind to N^6-methyladenosine (m^6A)-modified mRNAs (Wang *et al.*, 2014; Arribas-Hernández *et al.*, 2018). This provides an interesting avenue for research, as understanding of this complex of proteins specific to m^6A methylation is still in the early stages.

The first co-transcriptional modification during Pol II transcription is capping of the pre-mRNA transcript. A C-terminal YSPTSPS heptad repeat domain (CTD) on Pol II's largest subunit (RPB1) differentiates it from other Pol complexes. CTD phosphorylation at the serine 5 residue

facilitates to capping enzyme recruitment, which adds the 5' m^7G cap to nascent mRNAs, also known as cap 0 (Rodriguez *et al.*, 2000). The m^7G cap includes functions such as cap-dependent translation initiation, which protects the mRNA from 5'-3' cleavage by exonucleases, and recruits protein factors involved in pre-mRNA processing, polyadenylation, and nuclear export (Ramanathan *et al.*, 2016).

In animals, but not plants or fungi, the first two nucleotides of the mRNA transcript can be methylated on the ribose sugar, forming 2'-O-ribose methyladenosine, 2'-O-ribose methylcytosine, 2'-O-ribose methylguanine, and 2'-O-ribose methyluridine (Am, Cm, Gm, Um), which are known as cap1 on the first nucleotide and cap2 on the second. It has been revealed that cap1 is pivotal in the non-self-discrimination of the innate immune response against foreign RNA (Daffis *et al.*, 2010). If the cap1 nucleotide is adenosine, the nitrogenous base can be further modified to N^6, 2'-O-dimethyladenosine (m^6A_m) (Figure 1.4) (Wei *et al.*, 1975). Quantitative studies have indicated that the proportion of m^6A_m present in mRNA varies across different organisms and cell types, ranging from 10% to nearly 50% (Kruse *et al.*, 2011; Wang *et al.*, 2019). In addition, the role of m^6A_m has been proposed to function in regulating decapping, stability, and translation of mRNA due to its proximity to the mRNA cap structure. However, the exact role of m^6A_m in gene expression remains a topic of controversy and further investigation is needed. Furthermore, Mauer *et al.*, (2017) has shown that Fat Mass and Obesity-Associated protein (FTO) could demethylate m^6A_m.

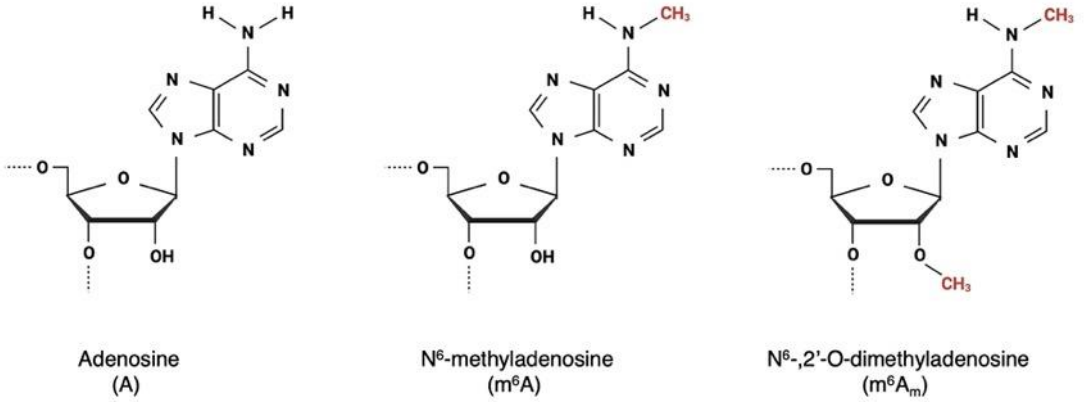


Figure 1.4 Representation of Adenosine modifications.

1.4 Splicing

Gene regulation is further extended into another layer of regulation, where splicing reactions process pre-mRNA to mature mRNA (Figure 1.5). Protein-coding sequences in eukaryotic genes often contain regions of noncoding introns (Berget *et al.*, 1977; Chow *et al.*, 1977). The molecular process by which splicing takes place has been investigated in various systems and is now considerably well-defined. Cleavage of introns from primary transcripts occurs through recognition and cleavage at conserved splice sites located at the 5′ and 3′ ends of introns. A complex known as the spliceosome, composed of small nuclear ribonucleoproteins (snRNPs) U1, U2, U4/U6, and U5, along with approximately 150 associated proteins, catalyses splicing reactions (Wahl *et al.*, 2009). The post-transcriptional regulation of gene expression also plays a pivotal role in plant responses to environmental changes. Alternative splicing (AS) is a fundamental process involved in post-transcriptional gene regulation and provides a means for plants to regulate their environmental adaptation and fitness.

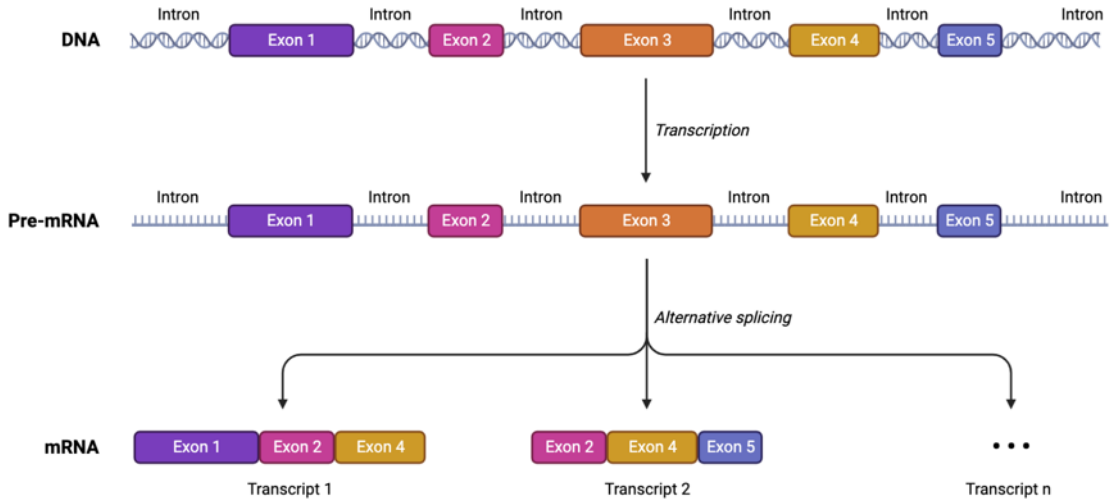


Figure 1.5 Schmatic showing the process of splicing from DNA to mature mRNA.

Although the 5′ and 3′ splice sites in plant and animal introns are similar (GU-AG), plant introns possess unique features such as distinct size, nucleotide composition, branch point sequence, and polypyrimidine tract, which distinguish them from animal introns. These differences suggest that the initial events involved in splice site recognition may be distinct for plant introns (Reddy, 2007; Wang *et al.*, 2012). AS not only contributes to proteome diversity but also regulates gene expression levels by producing mRNA isoforms with premature termination codons (PTCs) (Reddy, 2007; Wang *et al.*, 2012).

Nonsense-mediated mRNA decay (NMD) is an evolutionarily conserved mRNA surveillance pathway in eukaryotes that recognises and selectively degrades aberrant transcripts containing PTCs. Up-frameshift 1 (UPF1) is an Adenosine triphosphate (ATP)-dependent RNA helicase that is critical for NMD (Kim & Maquat, 2019; Kurosaki *et al.*, 2019). Transcriptome and phenotypic studies conducted across various organisms have revealed that NMD has roles that extend beyond RNA quality control, suggesting its involvement in regulating gene expression in diverse physiological processes (Watabe *et al.*, 2021). Links

between m^6A methylation and NMD have expanded recently, with a study demonstrating that depletion of METTL3, an m^6A writer protein, leading to reduced m^6A modification levels on serine- and arginine-rich splicing factors (SRSF) transcripts, leading to YTHDC1-dependent NMD of SRSF transcripts (Li *et al.*, 2019). In plants, examples of NMD gene regulation by regulation of AS in favour of transcripts with NMD target features have been reported, and recently, NMD has also been proposed to work as a viral restriction mechanism (Palusa & Reddy, 2010; Garcia *et al.*, 2014; Martínez-Pérez *et al.*, 2017).

Post-transcriptional modifications add another layer of gene regulation, termed the "epitranscriptome," which is involved in many processes and modifications. This layer of gene regulation plays several critical roles, such as controlling mRNA stability, localisation, splicing, and translation efficiency. These modifications are involved in fine-tuning gene expression in response to developmental cues, environmental changes, and cellular signalling, making it a highly dynamic and complex mechanism involving various enzymes, regulatory proteins, and RNA-binding molecules. Further research leading to the understanding of post-transcriptional modifications and their functions is vital for developing new technologies that could advance vaccine designs and improve global food security through advancements in crop production.

1.5 N^6-methyladenosine RNA modification

The presence of RNA modifications was first observed by Perry *et al.*, (1974) and Desrosiers *et al.*, (1974). However, with the advancement of technologies such as high-throughput sequencing and new detection methods, many more modifications have been discovered. There are over 180 different types of post-transcriptional modifications that have been identified across RNA species in all kingdoms of life and viruses, and this list is growing due to rapid advancements in the field (Boccaletto *et al.*, 2022). These modifications are predominantly found in tRNA and rRNA, where they play crucial roles in regulating RNA function. However, mRNA, which serves as the intermediary between DNA and proteins, also

undergoes chemical modifications, including N^7-methylguanosine (m^7G), N6-methyladenosine (m^6A), 5-methylcytosine (m^5C), N1-methyladenosine (m^1A), pseudouridine (Ψ), inosine (I), and m^6A_m (Li *et al.*, 2017).

Among the 180 different types of post-transcriptional modifications, m^6A is the most abundant internal chemical modification found in eukaryotic mRNA, although it is also present in a variety of bacteria and RNA viruses. The m^6A modification is formed by the addition of a methyl group (-CH3) from S-Adenosyl Methionine (SAM) to the nitrogen- 6 position of an adenosine base at any internal position along an RNA transcript (Zhong *et al.*, 2008). m^6A was first identified in the 1970s through enzyme treatments and chromatography (Perry & Kelley, 1974; Desrosiers *et al.*, 1974). Since then, m^6A has been identified more thoroughly across viruses (Beemon & Keith, 1977; Aloni *et al.*, 1979), mammals (Perry *et al.*, 1975; Adams & Cory, 1975), *Drosophila* (Levis & Penman, 1978), monocots including maize *Zea mays* (Nichols, 1979), wheat *Triticum aestivum* (Cuming *et al.*, 1979), oat *Avena sativa* (Haugland & Cline, 1980), and the dicot *A. thaliana* (Zhong *et al.*, 2008). Research on m^6A was not popular after its initial discovery due to limitations in detecting and quantifying m^6A sites in mRNA, until its important biological function was reported to cause embryo lethality in *A. thaliana* (Zhong *et al.*, 2008). Since the discovery of this biological function, advancements in technology have resulted in a resurgence of interest in this modification. The development of a new antibody based method for identifying m^6A containing messages in a small number of m^6A transcripts by the Fray group paved the way for the quantification of m^6A modification (Bodi *et al.*, 2010). Furthermore, two independent groups in 2012 reported the first m^6A mapping methods involving pairing antibody pull downs with high-throughput sequencing (Dominissini *et al.*, 2012; Meyer *et al.*, 2012). Further advancements in m^6A mapping have led to the development of new strategies for recognising and detecting m^6A, either by antibody-dependent or independent methods, such as methylated RNA immunoprecipitation (MeRIP-Seq) (Dominissini *et al.*, 2012), UV cross-linking and immunoprecipitation (miCLIP-Seq) (Linder *et al.*, 2015), super-low-input m^6A (SLIM-Seq) (Yin *et al.*, 2022a), deamination adjacent to RNA

modification targets (DART-Seq) (Meyer, 2019), m^{6}A-selective chemical labelling (m^{6}A-SEAL-Seq) (Wang *et al.*, 2020), m^{6}A-selective allyl chemical labelling (m^{6}A-SAC-Seq) (Hu *et al.*, 2022b), and more recently, nanopore-based direct RNA sequencing (Liu *et al.*, 2019).

The main focus of m^{6}A studies has been on humans and other mammalian systems, while less attention has been directed towards researching m^{6}A modification in plants at the molecular level. However, with the increasing attention surrounding plant science and m^{6}A methylation, it has led to many advancements in m^{6}A knowledge and potentially cross-transferable findings to mammalian systems. The frequency of m^{6}A distribution is primarily clustered around the stop codons and 3' untranslated regions (UTRs), 3'-end of coding sequence (CDS), and the first quarter of 3' UTR (Dominissini *et al.*, 2012, 2013; Meyer *et al.*, 2012; Ke *et al.*, 2015). In plants, m^{6}A distribution is enriched in the 3' UTRs but is also present in the start codon (Slobodin *et al.*, 2017). According to studies, start codons of genes linked to photosynthesis and proteins associated with chloroplasts are where more than 60% of m^{6}A modifications are found (Li *et al.*, 2014c; Luo *et al.*, 2014).

Another layer of gene regulation such as DNA and histone modification can have effects on gene expression (Egger *et al.*, 2004). A study published recently has demonstrated that histone H3 trimethylation at Lys36 (H3K36me3), which serves as a marker for transcription elongation, is involved in global m6A deposition (Huang *et al.*, 2019). The authors have observed that m^{6}A modifications are concentrated in the regions surrounding H3K36me3 peaks, and their levels decrease globally when the cellular H3K36me3 is depleted. The findings of this histone mark H3K36me3 as an element associated with m^{6}A through METTL14 adds another layer of complexity in gene regulation by m^{6}A modification.

1.6 Writer proteins of m^{6}A

The m^{6}A modification involves multiple components, including methylases known as "Writers", demethylases known as "Erasers", and m^{6}A-recognising proteins known as "Readers". In plants and other eukaryotes, the binding of m^{6}A writers occurs through a highly conserved consensus sequence, RRACH (R = G/A; H: U/A/C) (Dominissini *et al.*, 2012; Linder *et al.*, 2015; Shen *et al.*, 2016). The result of m^{6}A methylation of mRNA is due to the commonly called m^{6}A writer complex which compose of a multi protein complex (Figure 1.6) (Bokar *et al.*, 1997).

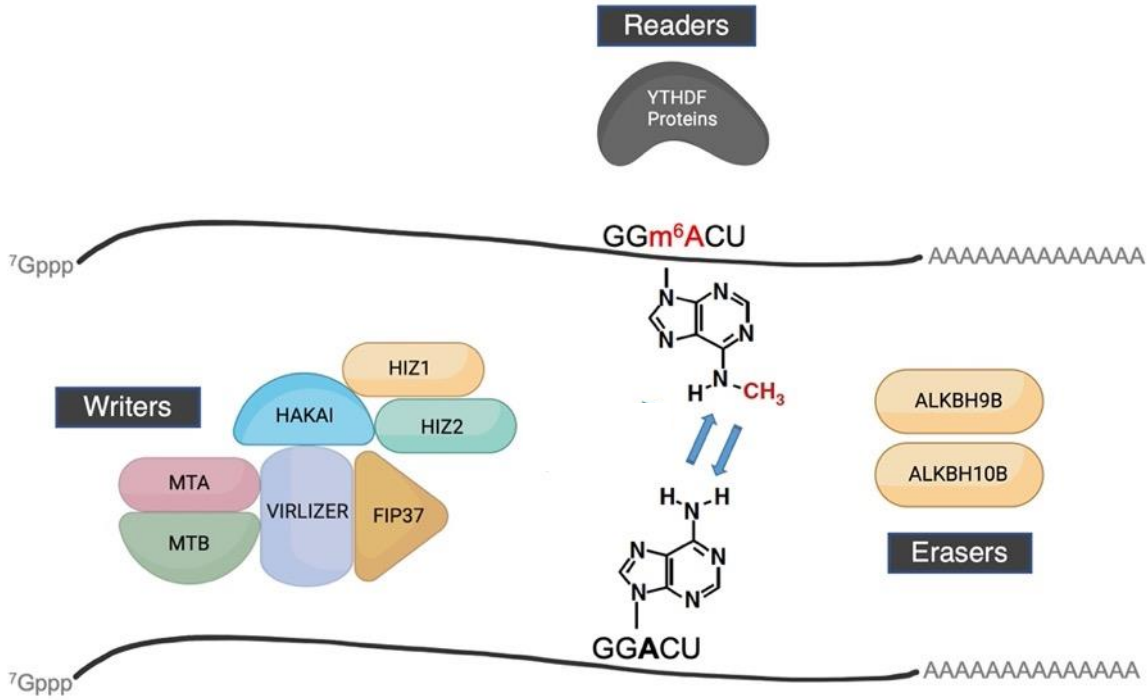


Figure 1.6 Diagram showing the three key components associated with m^{6}A methylation in plants on mRNA.

1.6.1 MTA/METTL3

The first mammalian m^{6}A methyltransferase to be identified was named METTL3 and was cloned from the 200 kDa methylase complex (Bokar *et al.*, 1997). METTL3 is a member of the putative SAM-dependent methyltransferase family, which is highly conserved in both plants and mammals (Bujnicki *et al.*, 2002). In *A. thaliana*, these components include the mRNA adenosine methyltransferase MTA protein (a homologue of the mammalian METTL3) that is involved in the main catalytic component of the complex. The functional importance of MTA was demonstrated, whereby the disruption of the *MTA* gene led to an arrest at the globular stage of embryo development (Zhong *et al.*, 2008). Further studies in plants showed that the direct reduction of m^{6}A levels to 5-15% exhibited reduced apical dominance, increased trichome branching, defects in protoxylem and gravitropism, and reduced root growth (Bodi *et al.*, 2012; Růžička *et al.*, 2017)

1.6.2 MTB/METTL14

The protein MTB (METTL14 human homolog protein) has also been identified as part of the m^{6}A writer complex in *A. thaliana*. METTL14, which is phylogenetically related to METTL3, is highly conserved in mammals and closely related at the amino acid level (Bujnicki *et al.*, 2002; Liu *et al.*, 2014). It was also identified that the recognition and binding of H3K36me3 by METTL14, which is a crucial component of the m^{6}A MTC to result in an enrichment of m^{6}A in the vicinity of H3K36me3 (Huang *et al.*, 2019). This binding subsequently enables the m^{6}A MTC to associate with the adjacent Pol II, leading to co-transcriptional deposition of m^{6}A on actively transcribed nascent RNAs. In plants, MTB (homolog of METTL14) associates with MTA to form heterodimers and interacts with other MTC proteins (Růžička *et al.*, 2017). Both METTL3 and METTL14 are necessary for specific and efficient methyltransferase activity (Wang *et al.*, 2016). Complete knockout of MTB is also embryo lethal (Růžička *et al.*, 2017).

1.6.3 FIP37/WTAP

The function of other proteins interacting with METTL3 was unknown until it was shown in *A. thaliana* by the Fray lab that FKBP12 Interacting Protein 37 (FIP37) was a partner protein of MTA and co-localises in the nucleus with other components of the MTC (Zhong *et al.*, 2008). FIP37 is a homolog of human Wilms' Tumour1-Associating Protein (WTAP). Depletion of the pre-mRNA splicing regulator WTAP has been shown to result in a significant decrease in m^6A levels, suggesting that WTAP is the third major binding partner of the methylation complex. WTAP plays a crucial role in initiating and controlling the localisation of nuclear foci that are enriched with pre-mRNA splicing factors necessary for the activity of the METTL3-METTL14 complex (Liu *et al.*, 2014; Ping *et al.*, 2014; Schwartz *et al.*, 2014).

In *A. thaliana*, overexpression of FIP37 lines exhibit SAM over proliferation, delayed growth display, and serrated leaves (Shen *et al.*, 2016). The mutant plants with reduced FIP37 showed phenotypes with increased over-proliferation of shoot apical meristem along with an 85% reduction in m^6A levels (Shen *et al.*, 2016). Knockout (KO) mutants of FIP37 are embryo lethal due to a strong delay in endosperm development and embryo arrest (Vespa *et al.*, 2004; Zhong *et al.*, 2008; Růžička *et al.*, 2017). These three proteins; MTA, MTB, and FIP37 are considered to form the core complex of the m^6A MTC; however, they are not the only proteins in the complex.

1.6.4 VIRLIZER/KIAA1429

The identification of KIAA1429 as the fourth key component of the m^6A MTC was shown to be a strong interactor with WTAP, and was found to cause a significant loss of m^6A when KO mutations were introduced in mammals (Schwartz *et al.*, 2014). KIAA1429 mediates preferential mRNA methylation in 3′UTR and near the stop codon (Yue *et al.*, 2018). Studies performed in *Drosophila* have expanded the knowledge of this specific protein in the MTC complex. In *Drosophila*, fl(2)d (homolog of WTAP) and vir (homolog of KIAA1429) are required

for sex-dependent regulation of alternative splicing of the sex determination factor Sex lethal (Sxl). It was also shown that null mutants of Ime4 (homolog of METTL3) and absence of m^6A resulted in the inability to fly and generating male sex bias. The female-specific splicing of *Sxl* requires m^6A found in 5'UTR (Haussmann *et al.*, 2016; Lence *et al.*, 2016).

In *A. thaliana*, the known interaction with MTA and FIP37 was further expanded to include VIRLIZER (homolog of KIAA1429) and HAKAI, which were found to be included in the *A. thaliana* MTC complex (Zhong *et al.*, 2008; Růžička *et al.*, 2017). The VIRLIZER mutant displayed pleiotropic phenotypes, including aberrant root cap formation, altered gravity response, lateral root development, and defective cotyledonal development (Růžička *et al.*, 2017). Full KOs of VIRLIZER do not progress past the embryonic globular stage.

1.6.5 HAKAI

HAKAI, an E3 ubiquitin ligase, was initially identified for its role in mediating the post-translational down-regulation of the E-cadherin complex (Fujita *et al.*, 2002). HAKAI is present and named both in plants and mammals as part of the m^6A writer complex. In animals, it is also known as Casitas B-lineage lymphoma-transforming sequence-like protein 1 (CBLL1) (Růžička *et al.*, 2017; Yue *et al.*, 2018). In animals, HAKAI interacts with and requires RNA binding motif protein 15 (RBM15) and the zinc finger CCCH domain-containing protein 13 (ZC3H13). HAKAI in *A. thaliana* was identified simultaneously with VIRLIZER. HAKAI is required for full m6A methylation, and its knockout leads to a 35% reduction in m^6A abundance (Růžička *et al.*, 2017). Notably, disruption of HAKAI in *A. thaliana* does not result in lethality or cause any aberrant phenotypes, in contrast to MTA, MTB, FIP37, and VIR, which fail to progress beyond the globular stage.

1.6.6 HIZ1 and HIZ2

Recently, two new members of the writer complex named HIZ1 and HIZ2 were identified in the Fray lab, and they were found to bind to HAKAI (Zhang *et al*., 2022). HAKAI was found to interact with HIZ1 and MTA. Mutants of *hiz1* showed no observable phenotypes, although its constitutive over expression led to a reduction in m^6A levels. Knockout of HIZ2 resulted in decreased global m^6A levels by approximately 85% and produced severe developmental defects (Zhang et al. 2022). HIZ2 is likely to be the functional homolog of *Drosophila* Flacc and structurally more similar to Flacc than the mammalian ZC3H13 protein. It has been shown that disruption of Flacc resulted in reduced m^6A levels and defects in sex determination due to aberrant splicing of *Sxl* (Knuckles *et al.*, 2018).

1.7 Other identified methyltransferases

Apart from the main identified methyltransferase proteins, several others have been described, including METTL16 (Pendleton *et al.*, 2017), METTL5 (van Tran *et al.*, 2019), and the zinc finger CCHC type-containing 4 (ZCCHC4) (Ma *et al.*, 2019). Currently, there is evidence regarding the orthologues of METTL5 and ZCCHC4 in plants. However, in *A. thaliana*, FIONA1 (FIO1) an orthologue of METTL16 was shown to be involved in the regulation of period length in the plant circadian clock (Kim *et al.*, 2008; Pendleton *et al.*, 2017). The mechanism of FIO1 was shown to deposit m^6A into U6 small nuclear (sn)RNA and a small subset of poly(A) RNA (Wang *et al.*, 2022). Another independent methyltransferase named MTC (orthologue of METTL4) was found in *A. thaliana*, which revealed similar features with the METTL3-METTL14 heterodimer, suggesting it functions as an m^6A writer (Luo *et al.*, 2022b). Overall, there is still a lack of studies that confirm each of these proteins identified as methyltransferases are part of the MTC or whether they are acting independently, and therefore further investigation on this topic is required.

1.8 Eraser proteins of m^6A

A key discovery that led to m^6A being known as a dynamic modification was the identification of proteins that catalyse the removal of m^6A. Initially, the m^6A modification was presumed to be a static modification, until the findings of the first identified m^6A demethylase, FTO, in mammals (Jia *et al.*, 2011). It was reported that knockdown of FTO led to increased amounts of m^6A in mRNA, and overexpression of FTO resulted in decreased m^6A in human cells (Jia *et al.*, 2011). The connection of the FTO protein with obesity was identified through GWAS, and has been shown to also affect energy homeostasis, thereby creating increased interest in the research field (Frayling *et al.*, 2007; Scott *et al.*, 2007; Church *et al.*, 2010). The study by Mauer *et al.*, (2017) showed that m^6A_m at the 5′ end of certain mRNAs, next to the 7-methylguanosine cap structure, can positively influence mRNA stability by preventing DCP2-mediated decapping. This modification is itself, regulated by FTO, which exhibits a preference for m^6A_m over m^6A.

Another demethylase protein, Alpha-Ketoglutarate-Dependent Dioxygenase alkB homolog 5 (ALKBH5), was identified and, along with FTO, belongs to the ALKBH family (Zheng et al., 2013). There are currently just two known m^6A erasers in mammals, and they are FTO and ALKBH5. Although FTO has no identified homolog in the plant system, in *A. thaliana*, there are 13 ALKBH family members with subcellular localisations (Mielecki *et al.*, 2012). Among the ALKBH family, ALKBH9B was suggested to have demethylase activity; however, *alkbh9b* and *alkbh9c* mutants did not display any changes in m^6A levels (Duan *et al.*, 2017). The other homologous protein, ALKBH10B, was identified to demethylate m^6A *in vivo* and *in vitro*, and the *alkbh10b* mutant showed increased m^6A levels (Duan *et al.*, 2017). Although there are studies indicating the functional roles of eraser proteins in relation to m^6A methylation, they are limited to specific tissues and conditions, and further investigation is needed.

1.9 Reader proteins of m^6A

Eukaryotes have the ability to recognise m^6A marks and recruit m^6A binding proteins known as 'Readers' proteins. One of the most studied RNA-binding proteins is the YT521-B Homology (YTH) protein domain of approximately 140 amino acids (Imai *et al.*, 1998; Hartmann *et al.*, 1999; Stoilov *et al.*, 2002). These proteins recognise the m^6A mark through a highly conserved hydrophobic binding pocket which contains three aromatic side chains surrounding the methyl group in m^6A, resulting in 10-20-fold higher affinity for m^6A-containing RNA compared to unmethylated RNA (Li *et al.*, 2014b; Luo & Tong, 2014; Xu *et al.*, 2014; Theler *et al.*, 2014; Zhu *et al.*, 2014; Patil *et al.*, 2018).

The YTH domains can be defined into two phylogenetic classes: the nuclear YTH proteins, termed YTH domain containing (YTHDC), and the cytoplasmic YTH domain family (YTHDF) (Scutenaire *et al.*, 2018). The YTHDF-containing proteins found in mammals are capable of binding to m6A sites present in mRNA. YTHDF1-YTHDF3, the three paralogs, exhibit a high degree of amino acid similarity throughout their entire length, including the N-terminal intrinsically disordered regions. In mammals, a further two different proteins are termed YTHDC1 and YTHDC2. The YTHDC1 is nuclear and binds to certain sites on mRNA and nuclear ncRNAs, in comparison to YTHDC2 which is enriched in the perinuclear regions of the cytoplasm. The YTHDC1 has long N- and C-terminal intrinsically disordered regions (Patil *et al.*, 2016). In eukaryotes, YTHDC1 and YTHDF type proteins are found in eukaryotes, including plants, which contain larger families of both types of YTH domain proteins compared to other organisms (Li *et al.*, 2014a; Scutenaire *et al.*, 2018).

In *A. thaliana*, there are two YTHDC reader proteins: CPSF30, a 30-kD subunit of the CPSF involved in pre-mRNA cleavage and 3'-end formation, and AT4G11970 (*ECT12*). Among the YTHDF proteins, there are 11 proteins called Evolutionarily Conserved C-Terminal Region (ECT) ranging from ECT1 to ECT11 (Thomas *et al.*, 2012; Bruggeman *et al.*, 2014; Pontier *et al.*, 2019). It is currently unclear if all the ECT proteins bind to m^6A or act independently of other proteins, and if different environmental cues trigger different functions of each of the reader proteins. It has been thought up until now that the YTH domain proteins are the only

class of m^{6}A readers in plants. However, recently, a novel mRNA m^{6}A reader protein named Flowering Locus K (FLK) was identified to directly bind to the m^{6}A site in the 3'UTR of Flowering Locus C (*FLC*) transcripts to repress FLC levels by reducing its stability and splicing (Amara *et al.*, 2023). In mammals, there are other RNA-binding proteins that have been highlighted to function as m^{6}A readers, such as translation initiation factor eIF3 (Meyer *et al.*, 2015). Findings in Meyer *et al.*, (2015) reveal that m^{6}A modifications in the 5' UTR can interact with eIF3. Ribosome profiling analysis of the transcriptome shows that the translation of mRNAs containing m^{6}A in the 5' UTR is decreased when METTL3 is depleted, while mRNAs with m^{6}A modifications elsewhere within the transcript do not exhibit this effect.

1.10 Molecular functions of m^{6}A in plants

The reported molecular roles of mRNA m^{6}A methylation are numerous and varied. In addition to the previously mentioned effects of m^{6}A in splicing, this modification has also been linked to a wide variety of molecular functions in cells. Considering the significant role of gene duplication and translocation mechanisms in facilitating gene expansion in plants, it is probable that plants have evolved additional mechanisms to counteract any potential negative impacts arising from such gene rearrangements. In one study, Pontier *et al.*, (2019) elucidate how m^{6}A RNA modification and a complex network of conserved and plant-specific proteins together form a post-transcriptional regulatory pathway that plays a crucial role in enforcing polyadenylation and 3'-end formation at loci characterised by transcriptional read-through, cryptic intergenic splicing, and chimeric mRNA formation in *A. thaliana*. It is possible that this alternative polyadenylation pathway, termed the m^{6}A-assisted polyadenylation, requires the writer protein FIP37 and reader protein CPSF30-L for proper functioning (Pontier *et al.*, 2019). Further studies have also shown the involvement of m^{6}A in the alternative polyadenylation pathway as m^{6}A modification can affect poly(A) site selection in maize (Luo *et al.*, 2020).

1.11 Biological roles of m^{6}A in plants

The emergent use of plants as a model system has proved to be a valuable model for studying mRNA modifications due to several unique features and advantages. Firstly, plants possess a wide array of mRNA modification enzymes that are conserved across different kingdoms and species, making them an ideal model for investigating the diverse roles of mRNA modifications in gene expression regulation. Secondly, plants offer experimental versatility and genetic manipulability, for example, using CRISPR/Cas9-mediated gene editing and sexually crossing plants for the heritability of genetic mutations. Thirdly, plants exhibit complex developmental processes and responses to environmental stimuli, providing an opportunity to study the dynamic changes of mRNA modifications in different physiological and environmental contexts. Furthermore, the importance of global food security through crop improvement is one of the key topics discussed worldwide due to increasing global demands for crop yields.

One of the key discoveries in the field of mRNA modification was the use of *A. thaliana* to show that m^{6}A is needed for ontogenesis in eukaryotes. Mutants of MTA, the key component of the writer complex proteins for m^{6}A methylation, were embryo defective and arrested at the globular stage (Zhong *et al.*, 2008). Further studies have confirmed that postembryonic hypomorphic knockdowns of MTA, MTB, FIP37, and VIR are embryo lethal, and that low m^{6}A plants displayed abnormal organ formation, slower growth, loss of apical dominance, an increased number of trichome branches, and aberrant development of lateral roots and vasculature (Vespa *et al.*, 2004; Zhong *et al.*, 2008; Bodi *et al.*, 2012; Růžička *et al.*, 2017). Recently, the identification of HIZ1 and HIZ2 has added to the current knowledge of writer complex and m^{6}A methylation. It was found that *hiz2* mutants significantly reduce root length and lateral root number, and overexpression of *hiz1* was found to influence root hair morphology (Zhang *et al.*, 2022).

Further studies have revealed that stem cell differentiation in the shoot apical meristem is dependent on m^{6}A methylation (Shen *et al.*, 2016). *fip37* mutants exhibited an increase in shoot apical meristem size, resulting in the formation of aberrant leaf primordia accompanied with increased expression of the transcription factors Wuschel (WUS) and Shoot Meristemless (STM) (Shen *et al.*, 2016). m^{6}A methylation has also been linked to the modulation of circadian period, as mutants exhibit lengthening of the circadian period (Parker *et al.*, 2019). In addition, m^{6}A plays a role in regulating flowering, as ALKBH10B and U6 snRNA methyltransferase were found to regulate flowering time by regulating the stability of key flowering genes (Duan *et al.*, 2017; Xu *et al.*, 2022). Intriguingly, the study revealed that *ect2/ect3* and *ect2/ect3/ect4* mutants exhibit slow stem growth, delayed flower formation, and distorted flower and fruit morphology (Arribas-Hernández *et al.*, 2020). By comparing *ect* mutants with low m^{6}A plants, their phenotypic similarities become evident, suggesting that m^{6}A readers may have a varied and significant role in gene regulation.

1.11.1 Root development

m^{6}A methylation has been linked to root development, as knockdown of MTA, MTB, FIP37, and VIRLIZER, leading to reduced m^{6}A levels, result in partial gravitropic root growth, a reduced number of lateral roots, and aberrant root cap formation (Zhong *et al.*, 2008; Bodi *et al.*, 2012; Růžička *et al.*, 2017). Moreover, the overexpression of *HIZ1* leads to similar root phenotypes as low m^{6}A writer mutant plants, which includes shorter primary roots and a decreased number of lateral roots (Zhang *et al.*, 2022). HIZ2 mutants also show significantly reduced root length and lateral root number (Zhang *et al.*, 2022). The key writer protein MTA has recently been linked to miRNA precursors and auxin response. A recent study has shown that miR393b accumulates through the requirement of m^{6}A methylation (Bhat *et al.*, 2020). They demonstrated that the reduced auxin response in *mta* mutant plants is partially caused by the regulatory defects of miR393b resulting from the lack of MTA activity (Bhat *et al.*, 2020). In addition, reader proteins have also been linked to roots as *ect2/ect3* double and

ect2/ect3/ect4 triple mutants, exhibit slower root growth and defective directionality of root growth (Arribas-Hernández *et al.*, 2020). Overall, m^6A writer and reader protein mutants have been shown to consistently affect auxin response and root growth across different lines of mutants. Further studies exploring the association between m^6A methylation and auxin response related to root formation could potentially provide insights into advantageous root development under stress environments.

1.11.2 Trichome development

m^6A methylation has been found to play a crucial role in trichome branching development. Trichomes are specialised, large, single cells that protrude from the epidermis of leaves, petioles, sepals, and stems, exhibiting a unique branched cellular architecture. Typically, *A. thaliana* trichomes have three or four branches on leaves (Schellmann & Hülskamp, 2005). *A. thaliana* trichomes, in particular, have been extensively studied as a model system for exploring cell fate determination, cell morphogenesis, and endoreplication in plant biology (Pesch & Hülskamp, 2009). Previous research reported that overexpression of *FIP37* increased trichome branching in *A. thaliana* (Růžička *et al.*, 2017). Intriguingly, mutants of MTA with reduced m^6A levels exhibited similar phenotypes, showing increased trichome branching of four or more branches (Bodi *et al.*, 2012). Additionally, *hiz2* mutants exhibited an increased in trichome branching phenotype, with more than three branches found in 65% of mutants compared to 15% in WT plants. In addition to the effects of writer proteins on trichome branching, reader proteins have also been shown to play a role. Studies on *ect2/ect3/ect4* single, double, and triple mutants have revealed that these mutants exhibit hyperbranched trichomes, similar to the *mta* and *hiz2* mutants. However, *ect4* mutants had little effect on trichome branching (Scutenaire *et al.*, 2018; Wei *et al.*, 2018). Investigating the molecular mechanisms and genetic factors that regulate trichome branching can provide valuable insights into plant morphology, physiology, and adaptation to various environmental

conditions. Notably, both writer and reader proteins of m^{6}A methylation have been shown to impact trichome branching (Růžička *et al.*, 2017; Arribas-Hernández *et al.*, 2018)

1.11.3 Stress response

Studies on m^{6}A methylation have emphasised the crucial role of this modification in the plant stress response, wherein plants undergo intricate molecular and physiological alterations to adapt to unfavourable environmental conditions. Deciphering the function of m^{6}A methylation in stress response pathways in plants can offer insights into the regulatory mechanisms that enable plants to cope with challenging conditions, and may pave the way for the development of stress-tolerant crops with enhanced agricultural performance. The involvement of m^{6}A in plant responses to environmental stimuli has been widely studied, particularly its role in mRNA m^{6}A methylation and salt stress tolerance in *A. thaliana*. Mutants of m^{6}A writer components MTA, MTB, VIRLIZER, and HAKAI displayed salt-sensitive phenotypes in an m^{6}A-dependent manner (Hu *et al.*, 2021). The authors observed that, except for HAKAI, the levels of m^{6}A writer components were increased in response to salt stress, and m^{6}A levels in mRNA were significantly upregulated in salt-treated seedlings. Intriguingly, VIRLIZER-mediated m^{6}A methylation under salt-stressed conditions was found to negatively regulate three negative regulators of stress response by altering 3'UTR lengths via alterations of poly(A) site usage (Hu *et al.*, 2021).

Recent research in the field of m^{6}A methylation has demonstrated that ECT2, a reader protein, undergoes relocation from the cytoplasm to stress granules in response to heat stress, which suggests a possible involvement in stress-induced inhibition of translation initiation (Scutenaire *et al.*, 2018). Additionally, ECT2, ECT4, and to a lesser extent ECT3, have been shown to relocate to distinct cytoplasmic foci distinct from P-bodies under osmotic stress conditions (Arribas-Hernández *et al.*, 2018). Furthermore, it has been proposed that ALKBH10B, an eraser protein, plays a crucial role in m^{6}A demethylation and thereby

influences stress-responsive genes transcript levels that are vital for the growth, germination, and survival of *A. thaliana* under salt stress and abscisic acid treatment (Shoaib *et al.*, 2021).

Research has also highlighted the importance of m^6A modification in biotic stress responses in plants. Specifically, in the case of the multifunctional Alfalfa Mosaic Virus (AMV), it was shown that the virus coat protein recruits the m^6A eraser ALKBH9B, leading to decreased m^6A levels in viral RNA (Martínez-Pérez *et al.*, 2017). Mutants of *alkbh9b* were found to negatively affect virus accumulation and systemic infection, signifying that m^6A has strong antiviral effects in the case of AMV (Martínez-Pérez *et al.*, 2017). These findings provide insights into the regulatory mechanisms that enable plants to defend against viral infections and suggest that manipulating m^6A levels could potentially lead to the adoption of antiviral strategies in crops.

The present research on m^6A and stress response provides initial evidence of relationships between critical writer and reader protein associations with stress response. However, further investigation is required to discern the exact role of how methylation can adapt, and subsequently cause changes, in the plant to respond to environmental stress. Additionally, different durations of stress induction may be crucial in identifying alterations in m^6A methylation. Combining loss-of-function mutants of both m^6A writer and readers will be necessary to fully comprehend the mechanisms of the m^6A regulatory pathway in stress adaptation.

1.12 Other functions of m^{6}A

1.12.1 m^{6}a plays a role in cancer growth and progression

Apart from m^{6}A methylation and the effects it has on plants this modification has been intensely studied in other model systems. One focus of m^{6}A methylation is the relationship with cancer. Recently, it has been shown that maintaining an optimal level of RNA methylation is critical for regulating the expression of key transcripts, and any stress or oncogenic trigger that alters the expression of m^{6}A writers-readers-erasers (Panneerdoss *et al.*, 2023). The alteration of methylation levels in genes associated with cell cycle, epithelial-mesenchymal transition (EMT), and angiogenesis uncontrolled expression or activity could lead to tumour growth and progression (Panneerdoss *et al.*, 2023).

Recent developments have also shown that m^{6}A methylation is an important target for immunotherapy technologies. These technologies include, small extracellular vesicles (sEVs) with high CD47 expression and cyclic arginine–glycine–aspartic (c(RGDyC)) modification, which have been developed for the effective delivery of siRNA against m^{6}A reader YTH m^{6}A RNA binding protein 1 (YTHDF1) to treat gastric cancer through epigenetic and immune regulation (You *et al.*, 2023). The engineered extracellular-vesicle-based platform possesses the potential to be a low-toxic, accurate tumour-targeted, and versatile nanoparticle for the delivery of RNA drugs. It could potentiate immunotherapy by restoring sensitivity to antitumor immunity (You *et al.*, 2023).

1.12.2 Neuronal development

The depth of m^{6}A modification has further expanded to the field of neural pathology. A recent study investigated the abundance of m^{6}A-modified RNAs and YTHDF1/YTHDF3 m^{6}A reader expression in four regions of the human brain, including individuals with Parkinson's disease and non-affected individuals, dementia with Lewy bodies, or mild cognitive impairment (MCI)

(Martinez De La Cruz *et al.*, 2023). The author's results indicated that m^{6}A-modified RNAs were consistently more abundant in dementia with Lewy bodies tissue and significantly reduced or misplaced in brain tissue from individuals with Parkinson's disease. These findings propose that targeting epitranscriptomic processes that influence translational control may lead to the development of new therapies that have the possibility to prevent synaptic failure and neuronal pathology in the field of neurodegenerative diseases (Martinez De La Cruz *et al.*, 2023).

In addition, the m^{6}A modification is associated with the regulation of axon regeneration in the adult mammalian nervous system. According to a recent study, sciatic nerve injury increases the levels of m^{6}A-modified transcripts that encode many regeneration-associated genes and protein translation machinery components in the adult mouse dorsal root ganglion (DRG) (Weng *et al.*, 2018). The KO of the m^{6}A MTC component METTL14 or reader protein YTHDF1 resulted in a decrease in functional axon regeneration and a general attenuation of injury-induced protein translation in the peripheral nervous system *in vivo* (Weng *et al.*, 2018).

1.13 Conclusion

In conclusion, the epigenetic modification of m^{6}A is proving to be a major modification that affects a vast variety of cellular function across different model systems that are researched across many scientific fields. The continuous development of different detection along with the quantification of RNA modification, will vastly accelerate the discovery of new proteins and their functions in relation to the m^{6}A modification. The dynamic complexity of m^{6}A modification having writers, readers, and eraser proteins makes this a more complicated system encompassing various mechanisms with a variety of functions for one modification. In plants, m^{6}A is proving to be a key focus of research as increasing evidence suggests the importance of it in plant development and stress response. Further investigation into writer and reader proteins in plants could provide insights that may translate to other systems such as mammals

and accelerate related discovery associated with m^6A modification. Furthermore, m^6A has also shown to be an important modification in cancer and neuronal development and pathology, further highlighting its significance in the field of epitranscriptomics. The following thesis explores the m^6A modification in the plant system of *A. thaliana* and *S. lycopersicum* using a variety of molecular, bioinformatics and physiological approach to research the writer and reader proteins.

Chapter 2

Materials and Methods

2 Materials and Methods

2.1 Materials

2.1.1 Arabidopsis thaliana seeds

WT *Arabidopsis thaliana* (Columbia) and all mutant plant seeds used in this study were obtained from Professor Rupert Fray (University of Nottingham). The Bulked Segregant Analysis (BSA) was conducted using *A. thaliana* seeds obtained from three independent seed pools labelled P3-07 and P10-23, and P1-14 at the segregating F_2 generation derived from an EMS mutagenesis screen on a *fip37-4* Salk line (SALK_018636). The *fip37-4* seeds used in this study were obtained from the Nottingham Arabidopsis Stock Centre (NASC).

2.1.2 *A. thaliana* growth conditions

The plants were cultivated in 9cm plastic pots in growth chambers with long-day conditions consisting of 16-hour day and 8-hour night photoperiod, 23°C temperature, 60% relative humidity, and light intensity at 180 ± 20 µmol/s/m2. Short-day length (12-hour day and 12-hour night) conditions were also maintained at 23°C and a light intensity of 180±20 µmol/s/m2. The plants were watered every 4 days or earlier if drying occurs on topsoil to maintain regular watering of plants. Seeds were collected once the siliques had fully matured and turned brown, and were stored in a dry environment at room temperature.

2.1.3 Tomato seeds

Tomato seeds (*Solanum lycopersicum* variety Ailsa Craig) WT were obtained from fruit of a homozygous line from Professor Rupert Fray at Sutton Bonington. Plant seed collection is shown in the Table 2.1.

Table 2.1 Tomato plant seed used in this study.

Plant seed ID	Description
pAGM-MTA	CRISPR knock-out of *MTA*
pAGM-MTB	CRISPR knock-out of *MTB*
pAGM-VIRLIZER	CRISPR knock-out of *VIR*
YTHDF1-CRISPR	CRISPR knock-out of *YTHDF1*
YTHDF2-CRISPR	CRISPR knock-out of *YTHDF2*
YTHDF3-CRISPR	CRISPR knock-out of *YTHDF3*
YTHDF4-CRISPR	CRISPR knock-out of *YTHDF4*
YTHDF1-CRISPR	Double CRISPR knock-out of *YTHDF1* + *YTHDF3*
YTHDF1-CRISPR	Double CRISPR knock-out of *YTHDF1* + *YTHDF4*
YTHDF1-CRISPR	Double CRISPR knock-out of *YTHDF2* + *YTHDF3*
YTHDF1-CRISPR	Double CRISPR knock-out of *YTHDF2* + *YTHDF4*
YTHDF1-CRISPR	Double CRISPR knock-out of *YTHDF3* + *YTHDF4*
RNAi-MTA	RNAi knock-down of *MTA*

2.1.4 Reagents

Table 2.2 List of reagents of analytical grade or higher that were used in this study.

Reagent	Manufacturer
10mM ATP	New England Biolabs
10x TBE	Roche
6-Benzylaminopurine (BAP)	Sigma Aldrich (Merck)
AE phenol	Sigma Aldrich (Merck)
Agar	Sigma Aldrich (Merck)
Agarose	Sigma Aldrich (Merck)
Ampicillin	Sigma Aldrich (Merck)
$CaCl_2$	Fisher Scientific (Merck)
Cefotaxime	Sigma Aldrich (Merck)

Cellulose TLC plates	Merck
Chloramphenicol	Merck
Chloroform	Fisher Scientific (Merck)
Chloroform isoamyl alcohol	Sigma Aldrich (Merck)
Cutsmart Buffer	New England Biolabs
Dexamethasone	Test
DMSO	Sigma Aldrich (Merck)
dNTPs	New England Biolabs
EDTA	Sigma Aldrich (Merck)
Ethanol	Fisher Scientific (Merck)
Ethidium bromide	Sigma Aldrich (Merck)
Gel Loading Dye, Purple (6x)	New England Biolabs
Glucose	Fisher Scientific (Merck)
Glycerol	Fisher Scientific (Merck)
Indole-3-acetic acid (IAA)	Sigma Aldrich (Merck)
Isopropanol	Fisher Scientific (Merck)
Kanamycin	Sigma Aldrich (Merck)
KCl	Fisher Scientific
HyperLadder™ 1kb	Bioline
LiCl	Sigma Aldrich (Merck)
$MgCl_2$	New England Biolabs
Murashige and Skoog Salts	Sigma Aldrich (Merck)
N-N Dimethylformamide (DMF)	Sigma Aldrich (Merck)
NaCl	Sigma Aldrich (Merck)
NaOAc	Fisher Scientific (Merck)
NP-40	Sigma Aldrich (Merck)
Oligo d(T)25 Magnetic Beads	New England Biolabs
Phenol Chloroform Isoamyl-alcohol	Sigma Aldrich (Merck)
Phusion Buffer	New England Biolabs
Silvet L-77	Thermo Fisher
SOC media	New England Biolabs
Sodium Dodecylsulfate (SDS)	Applichem GmbH
Spectinomycin	Sigma Aldrich (Merck)
Sucrose	Sigma Aldrich (Merck)
Taq Polymerase Buffer	New England Biolabs

Thermopol Buffer	New England Biolabs
Trizma base	Sigma Aldrich (Merck)
Tween-20	Sigma Aldrich (Merck)
5-Bromo-4-Chloro-3Indolyl Beta-D-Galactopyranoside (X-Gal)	Fisher Scientific (Merck)
6x Purple Loading Buffer	New England Biolabs
Amylose Magnetic Beads	New England Biolabs
Bromophenol Blue	BDH Laboratory Supplies
dATP	New England Biolabs
Dextran	Sigma Aldrich
Dextran Sulphate	New England Biolabs
Formaldehyde	Fisher Scientific (Merck)
HCl	Sigma Aldrich
HyperLadder 1kb	Bioline
LB	University of Notitngham
Na2CO3	Sigma Aldrich
Na2HPO4	Sigma Aldrich
NaoH	Sigma Aldrich
Na2HP04	Sigma Aldrich
Na3PO4	Sigma Aldrich
NaH2PO4	Sigma Aldrich
NaHCO3	Sigma Aldrich
NaOAc	Fisher Scientific (Merck)
NH4OH	Sigma Aldrich
Salmon Sperm	Thermo Scientific
SOC media	Thermo Scientific
Sodium Hypochlorite	Sigma Aldrich
Xylenecynol FF	Sigma Aldrich
Beta - Mercaptoethanol	Sigma Aldrich
Cellulose TLC Plates	Merck
ssRNA ladder	New England Biolabs

2.1.5 Enzymes

Table 2.3 List of enzymes of analytical grade or higher that were used in this study.

Enzymes	Manufacturer
BpiI	New England Biolabs
BSaI	New England Biolabs
P1 nuclease	Sigma Aldrich (Merck)
Phusion High-Fidelity DNA Polymerase	New England Biolabs
Poly Nucleotide Kinase (PNK)	Fisher Scientific (Merck)
Reverse Transcriptases	New England Biolabs
RNAase Inhibitor Murine	New England Biolabs
T1 RNase	New England Biolabs
T4 DNA Ligase	New England Biolabs
T7 DNA Ligase	New England Biolabs
Taq DNA Polymerase	New England Biolabs
Q5 DNA Polymerase	New England Biolabs
TURBO Dnase	Invitrogen (Thermo Fisher Scientific)

2.1.6 Kits

Table 2.4 List of kits of analytical grade or higher that were used in this study.

Method	Kit Name	Manufacturer
DNA Extraction	MagMAX Plant DNA Isolation Kit	Thermo Fisher Scientific
DNA Purification	Monarch Genomic DNA Purification Kit	New England Biolabs
RNA Purification	GeneJET RNA Cleanup + Concentration Micro Kit	Thermo Fisher Scientific
Plasmid Extraction	GeneJET™ Plasmid Miniprep Kit	Thermo Fisher Scientific
Cloning gene of interest into entry vector	pCR8™8/GW/TOPO™ TA Cloning Kit	Thermo Fisher Scientific
Removal of genomic DNA	TURBO DNase	Thermo Fisher Scientific
RNA extraction	RNAqueous Total RNA Isolation Kit	Thermo Fisher Scientific

2.1.7 Primers

Table 2.5 List of oligonucleotides primers used in this study.

Primer ID	Sequence (5' to 3')	Description
RNAi Primer		
RNAi MTA - Fwd	CCCTAATGCCCCCAACCATA	Amplification of genomic DNA of MTA
RNAi MTA - Rev	ACACTTCCCTCAGTCTGCTT	Amplification of genomic DNA of MTA
PCR8 attL2 - Fwd	GTTCGTGCCTTCATCCGTTT	Amplification of MTA gene in PCR8
PCR8 attL2 - Rev	ACAACGTTCAAATCCGCTCC	Amplification of MTA gene in PCR8
pJET attL2 - Fwd	GAGATTTTGAGACACGGGCC	Amplification of MTA gene in pJET
pJET attL2 - Rev	GTTGTAAAACGACGGCCAGT	Amplification of MTA gene in pJET
popOff2 Int-35S - Fwd	TCTTCTTCGTCTTACACATCACT	MTA amplification to confirm insert into popOff2
popOff2 Int-35S - Rev	GGGAAGCTGGGTTCGAAATC	MTA amplification to confirm insert into popOff2
NPTII FOR (Pri 17)	GATGGATTGCACGCAGGTTC	Sequence check for NPTII
NPTII REV (Pri 18)	AAAAGCGGCCATTTTCCACC	Sequence check for NPTII
PDK Intron F	TTCCCAACTGTAATCAATCC	Sequence check direction of insert
PDK Intron R	TGACAAGTGATGTGTAAGAC	Sequence check direction of insert
CRISPR Primers		
Tom MTA - sgRNA	TGTGGTCTCAATTGCAGAAATCGAAAAC CGGAGGTTTTAGAGCTAGAAATAGCAAG	Amplification of MTB sgRNA guide from pICH86966
Tom MTB - sgRNA	TGTGGTCTCAATTGCATGAGCAGGAGCT CTCCCGTTTTAGAGCTAGAAATAGCAAG	Amplification of MTB sgRNA guide from pICH86966
Cas9 - Fwd	AGGAAATAGGCAAGGCCACC	Cas9 screening for presence of t-DNA insertion
Cas9 - Rev	TTATCGAGGTTAGCGTCGGC	Cas9 screening for presence of t-DNA insertion
NPTII - Fwd	GATGGATTGCACGCAGGTTC	NPTII screening for presence of t-DNA insertion
NPTII - Rev	AAAAGCGGCCATTTTCCACC	NPTII screening for presence of t-DNA insertion
pICH47751 - Fwd	AGGATATATTGGCGGGTAAAC	Level 1 CRISPR construct confirmation of insertion
pICH47751 - Rev	CTGCATCCACCCCAGTACAT	Level 1 CRISPR construct confirmation of insertion
pAGM - Fwr	ATAAGCCCATCAGGGAGCAG	Level 2 CRISPR construct confirmation of insertion
pAGM - Rev	CGGATAAACCTTTTCACGCC	Level 2 CRISPR construct confirmation of insertion
NeoKan 47732 - Fwd	CGCGTTCAAAAGTCGCCTAA	Kanamycin check for piCH47732
NeoKan 47732 REV - Rev	CGATGTTTCGCTTGGTGGTC	Kanamycin check for piCH47732
Vir A F	CACTATATCGATGTTCGGCGT	Virlizer mutation check
Vir A R	AAGATCTGATCGATAATGAGC	Virlizer mutation check
Vir D2 F	ATCATTCGCATTGTGCCAGG	Virlizer mutation check
Vir D2 R	GCGGGGAAGCTTACAATAAT	Virlizer mutation check
Vir D4 F	AGCATCGTATGTTCGCTGGCA	Virlizer mutation check
CRISPR Screen Check		
MTA - screen-Fwd	GGCGATAGTGGCAGTGGAAA	MTA mutation genome screen and sequencing
MTA - screen-Rev	CTACCCTCCCCAAACCCCTA	MTA mutation genome screen and sequencing
MTB (10) - screen-For	TGGGCCTAAAGATGGTG	MTB mutation genome screen and sequencing
MTB (10) - screen-Rev	TGTCAGTGTCTGCAACCCTA	MTB mutation genome screen and sequencing
MTB (20) - screen-For	ACACATTACCCTCCCAGACC	MTB mutation genome screen and sequencing
MTB (20) - screen-Rev	ATCGACCCCAACACCAACTT	MTB mutation genome screen and sequencing
VIR - tom-1-F	GCTCTCAAATTGCTGCACCT	VIRILIZER mutation genome screen and sequencing

VIR - tom-1-R	AATTGCCCACTGTCACCAAG	VIRILIZER mutation genome screen and sequencing
VIR 2 screen fwd	TTCCCTTCGACATGCATCTGT	VIRILIZER mutation genome screen and sequencing
VIR 2 screen rev	GGGAACAAGGGAAACTCTGC	VIRILIZER mutation genome screen and sequencing
Vir -tom-3-F	ACACGGAGAGAGTTACTGCA	VIRILIZER mutation genome screen and sequencing
Vir -tom-3-R	GGTAGGTACAGATCCCCGTC	VIRILIZER mutation genome screen and sequencing
YTH Reader Primers		
ECT1 - screen-Fwd	CTTCGCCTGTCCCTACTGTG	ECT1 mutation genotyping and sequencing
ECT1 - screen-Rev	GACCTGTAGCGCGAGTATCA	ECT1 mutation genotyping and sequencing
ECT2 - screen-Fwd	CTTGCCCCATGTTACATCAG	ECT2 mutation genotyping and sequencing
ECT2 - screen-Rev	GGATGAGTCCAGGCTACATTGC	ECT2 mutation genotyping and sequencing
ECT3 - screen-Fwd	AGCCCCTGAGTTTGTTGTTGA	ECT3 mutation genotyping and sequencing
ECT3 - screen-Rev	ATCGGATCTGCAGCATTCTC	ECT3 mutation genotyping and sequencing
ECT4 - screen-Fwd	CTGTGGACTTGTATGTGTTTGCC	ECT4 mutation genotyping and sequencing
ECT4 - screen-Rev	AAGGCTGAGCTTGATGGAAAG	ECT4 mutation genotyping and sequencing
ECT1 XbaI - Fwd	CTAGTCTAGATCCGAGCTAGATCTTCAACAG	ECT1 promoter amplification to check insert into pGWB3
ECT1 SacI - Rev	CTAGGAGCTCGAAACAGAGAAGTGAGTTTCTC	ECT1 promoter amplification to check insert into pGWB3
ECT 2 HindIII - Fwd	CTAGAAGCTTCAATTGAACTGAATTGAAAATG	ECT2 promoter amplification to check insert into pGWB3
ECT 2 HindIII - Rev	CTAGTCTAGACTAAATAGCAAACCCACAA	ECT2 promoter amplification to check insert into pGWB3
ECT3 HindIII - Fwd	CTAGAAGCTTAGCTTGAGAGTTCTGAAGCAGGG	ECT3 promoter amplification to check insert into pGWB3
ECT 3 XbaI - Rev	CTAGTCTAGAGGGTTGAGGGGTGTGGTAGTTG	ECT3 promoter amplification to check insert into pGWB3
ECT 4 SpeI - Fwd	CTAGACTAGTTATCAAGCTTCAAAAGCCAAT	ECT4 promoter amplification to check insert into pGWB3
ECT4 XbaI - Rev	CTAGTCTAGAGAACATTTTTCTGTTATTG	ECT4 promoter amplification to check insert into pGWB3
ECT1 pGWB3 GUS - Fwd	CCGCAGTGGTTTGATTCGTC	ECT1 gene insertion check for plant DNA extraction
ECT1 pGWB3 GUS - Rev	TCCACAGTTTTCGCGATCCA	ECT1 gene insertion check for plant DNA extraction
ECT2 pGWB3 GUS - For	TGGCTCTAGTTTACCACGGTC	ECT2 gene insertion check for plant DNA extraction
ECT2 pGWB3 GUS - Rev	ATTGACCCACACTTTGCCGT	ECT2 gene insertion check for plant DNA extraction
ECT3 pGWB3 GUS - Fwd	ACATGGAGTCTTGGGAGGGT	ECT3 gene insertion check for plant DNA extraction
ECT3 pGWB3 GUS - Rev	TGACCCACACTTTGCCGTAA	ECT3 gene insertion check for plant DNA extraction
ECT4 pGWB3 GUS - Fwd	TAGCATTCGCCATTCAGGCT	ECT4 gene insertion check for plant DNA extraction
ECT4 pGWB3 GUS - Rev	TCTGTGTGGTTGTTCGGGTT	ECT4 gene insertion check for plant DNA extraction

2.1.8 Buffers

2.1.8.1 Edwards buffer

Tris hydrochloride 1M $NH_2C(CH_2OH)_3$. HCl – Dissolve 12.1g of Tris base in 80ml of nuclease free H_20. The pH is adjusted by adding concentrated HCl until pH of 8 is achieved.

Sodium Chloride 5M NaCl – Dissolve 292g of NaCl in 800ml of nuclease frere H_20. Adjust the volume to 1 litre and autoclave.

Ethylenediaminetetraacetic acid EDTA $[CH_2N(CH_2CO_2H_2]_2$ – Dissolve 186.1g of EDTA to 800ml of nuclease free H_20 on a stirrer. Adjust pH to 8.0 with NaOH pellets and autoclave.

Sodium Dodecyl Sulfate $C_{12}H_{25}NaO_4S$ – Dissolve SDS power in 100ml of nuclease free H_20 and add 900ml of nuclease free H_20 under a stirrer.

2.1.8.2 GUS staining buffers

Sodium Phosphate Monobasic 0.2M NaH_2PO_4 – Dissolve 12.4g of NaH_2PO_4 in sterile distilled water and make up to volume of 500ml.

Sodium Phosphate Dibasic 0.2M Na_2HPO_4 – Dissolve 14.195g of Na_2HPO_4 in 300ml of sterile distilled water and make up to volume of 500ml.

Phosphate Buffer 0.2M – Put 300ml of 0.2M Na_2HPO_4 in a 500ml beaker, then gradually pour 0.2M Sodium Phosphate Monobasic monohydrate while stirring under pH meter until it is pH 7.

2.1.8.3 Northern Blotting

10% SDS – Dissolve 80g SDS in 800ml of sterile water

Sodium Phosphate Buffer 1M pH 6.5 – Mix 137ml of 1M NaH_2PO_4 and 63ml of 1M Na_2HPO_4 stock solutions and make up volume to 200ml.

1M of NaH2P04 – Dissolve 24g of NaH_2PO_4 (MW = 120 $gmol^{-1}$) in 200ml of deionised sterile water.

1M of Na_2HPO_4 – Dissolve 14.2g of Na_2HPO_4 (MW = 142 $gmol^{-1}$) in 100 ml of deionised sterile water.

0.5M EDTA pH8 – Dissolve 18.61g of $Na_2EDTA{\cdot}2H_2O$ (MW = 372.24 $g{\cdot}mol^{-1}$) in 100 ml of deionised sterile water and add 1.8-2 g of NaOH pellets to dissolve $Na_2EDTA{\cdot}2H_2O$ to adjust pH to pH 8.

Ethidium bromide buffer – 1ml of deionised formamide, 550 µl of sterile water, 330 µl of formaldehyde (37%), 40 µl of EDTA (0.5 M, pH 8.0), 40 µl of 10 $mg{\cdot}ml^{-1}$ ethidium bromide and 40 µl of sodium phosphate buffer (1 M, pH 6.5).

5× Bromophenol blue loading dye - 0.25% (w/v) bromophenol blue, 0.25% (w/v) Xylenecynol FF and 40% (w/v) sucrose in H_2O.

5M NaCl – Dissolve 22.9g NaCl in 100ml sterile deionised water.

20×SSC - Dissolve 140.26 g NaCl (MW = 58.44 $g \cdot mol^{-1}$) and 70.584 g sodium citrate dehydrate (MW = 294.1 $g \cdot mol^{-1}$) in 800 ml of deionised water. Autoclave and store at room temperature.

2.1.8.4 TLC Running Buffers

First dimension buffer - 5:3, Isobutyric acid:0.5M NH_4OH

Second dimension buffer - 70:15:15, Isopropanol:HCl:Water

2.1.8.5 Stock Solutions

Stock solution used to make up buffers are detailed below:

50mM Ammonium Hydrogen carbonate (ABC) buffer - Dissolve 1.58g in 400ml sterile distilled water (SDW).

Ampicillin (100 mg/mL)- Dissolve 1g Ampicillin in 10ml water, filter sterilise and store at -20°c.

Cefotaxime (500mg/mL) Dissolve 5g Cefotaxime in 10mL water, filter sterilse and and store at -20°c.

Chloramphenicol (25 mg/mL)- Dissolve 0.26g in 10ml Ethanol, filter sterilise and store at -20°c.

100mM CaCl2- Dissolve 1.11g of CaCl2 into 100ml of water. Autoclave to sterilise.
1M DTT- Dissolve 1.542g DTT in 10ml of 50mM ABC Buffer, filter sterilise and store at -20°c.

100mM EDTA pH8- Dissolve 29.224g EDTA in 400ml, adjust to pH8 with NaOH pellets. Make up to 500ml, autoclave to sterilise.

500mM EDTA pH8- Dissolve 93.05g EDTA in 400ml, adjust to pH8with NaOH pellets. Make up to 500ml, autoclave to sterilise.

50% Glycerol- Mix 20ml of Glycerol and 20ml of water, place on rollers till mixed. Filter sterilise.

0.5M Hepes-HCL pH 7.5- Dissolve 23.83g of Hepes in 60ml of distilled water, adjust pH to 7.5 with NaOH. Make up to 100ml with distilled water. Filter sterilise and store at -20°c

0.5M IAA- Dissolve 0.925g IAA in 10 ml of 50mM ABC buffer, filter sterilise and store at -20°c.

1M KCl- Dissolve 7.45g of KCl in 100ml of distilled water. Autoclave to sterilise
49

3M KOAc pH4.8- Dissolve 9.81g of KOAc in 60ml of water, adject pH using Glacial Acetic Acid to pH 4.8. Autoclave to sterilise.

2M LiCl- Dissolve 8.46g of LiCl in 100ml of distilled water. Autoclave to sterilise.

200mM MgCl2- Dissolve 1.9g of MgCl2 in 100ml of distilled water. Autoclave to sterilise.

1M MgS04- Dissolve 12.03g of MgSO4 in 100ml of distilled water. Autoclave to sterilise.

20mM MnCl2- Dissolve 2.51g of MnCl2 in 100ml of distilled water. Autoclave to sterilise.

2M NaCl- Dissolve 11.68 of NaCl in 100ml of distilled water. Autoclave to sterilise.

5M NaCl- Dissolve 29.22g of NaCl in 100ml of distilled water. Autoclave to sterilise.

3M NaOAc pH5.3- Dissolve 24.61g of NaOAc in 60ml of water, adjust pH using
Acetic Acid to pH 4.8. Autoclave to sterilise.

100mM NaOAc pH5.3- Dissolve 0.82g of NaOAc in 60ml of water, adjust pH using glacial Acetic Acid to pH5.3. Autoclave to sterilise.

0.1N NaOH- Dissolve 1g NaOH in 250ml.Filter sterilise.

10% NP40- Mix 1ml of NP40 with 9ml of water. Mix until dissolved, filter sterilise.
1x PBS- Put 1 tablet in 500ml water. Autoclave to sterilise.

10% Sarkosyl solution- Dissolve 1g in 10ml of distilled water. Filter sterilise.

10% SDS- Mix 10ml of SDS into 90ml distilled water. Filter sterilise.

Spectinomycin (100mg/mL) Dissolve 1g Spectinomycin in 10ml water, filter sterilise and store at -20°c.

1x TBE- Mix 100ml of 10x TBE into 900ml distilled water.
Tetracycline (12.5mg/ml)- Dissolve 0.125g tetracycline in 10ml distilled water. Filter sterilise.

1M Tris-HCL- Dissolve 60.5g of Tris base in 400ml of distilled water. Adjust the PH accordingly (ph7.5, 8.5, 9.5) with HCL. Autoclave to sterilise.

2.1.9 Plant media

2.1.9.1 Sterilisation of seeds

Tomato seeds were obtained from fruit of a homozygous line from Rupert Fray at Sutton Bonington. Seeds were surfaced sterilised by being soaked in 20ml of 5% (w/v) sodium hypochlorite in 50ml falcon tube for 20 minutes. Seeds were then washed with 20ml of dH_20 three times before placing on media.

2.1.9.2 Germination medium

Germination of tomato seeds were done on solid media under sterile conditions. Germination medium consisted of 2.2g/L Murashige and Skoog salts, adjusted to Ph 5.8 and autoclaved to 121°C. After autoclaving media was poured into sterile plastic pots.

2.1.9.3 Pre callusing medium

Pre callusing medium consisted of 2.2g/L Murashige and Skoog salts, 20g/L Sucrose, adjusted to Ph 5.8 with 1M NaOH. Following the addition of NaOH, 5g/L Agar was added and sterilised by autoclaving at 121°C and spread equally into petri dishes.

2.1.9.4 Callusing medium

Callusing medium is made up with the same components of the pre-callusing medium but with the addition of 0.1mg/L IAA, 1.5mg/L BAP and 1.5mg/L Zeatin. Addition of the hormones were done after the liquid media was removed from the autoclave and reduced to 60°C.

2.1.9.5 Antibiotic Selection medium

The antibiotic selection medium is made with the same components of the callusing medium with the addition of 1ml 50mg/ml Kanamycin to 1 litre of media.

2.1.9.6 Rooting medium

Rooting medium consisted of 2.2g/L Murashige and Skoog salts, 20g/L Sucrose, adjusted to Ph 5.8 with 1M NaOH and 2.5g/L Agar. The media was autoclaved to 120°C and left to cooldown to 60°C before the addition of 1mg/L IAA.

2.1.9.7 Liquid Resuspension Media

Liquid resuspension media is used to resuspend pelleted *A. tumefaciens* and during transformation of cut tomato leaves. Resuspension media is made by sterile H_20 with the addition of 40g/L Sucrose and is autoclaved to 121°C.

2.1.10 Growth Conditions

2.1.10.1 Media

Seeds and tomato plants sowed on the required media were placed in growth rooms at 16-hour light and 80-hour dark interval at 23°C and 18°C respectively.

2.1.10.2 Compost

Seeds that were sowed on compost were on Levington M3 compost. Plants that were grown on compost were placed in glasshouse compartment rooms at 16-hour light and 8-hour dark interval at 23°C and 18°C respectively. To ensure that day lengths were consistent with 16-hour mercury vapour lamps were used to provide additional lighting in the glasshouse.

2.1.11 Bacteria work

2.1.11.1 Bacteria liquid and solid culture growth conditions

All *E. coli* and *A. tumefaciens* strains are grown in a 37°C and 28°C respectively in shaking incubator at 200 rpm. Bacteria that were grown on petri dishes were placed in a 37°C and 28°C standing incubator. Every culture growth was monitored and measured to an optical density at 600nm (OD 600) by a BioMate Spectrophotometer (Thermo Fisher Scientific).

2.1.11.2 Storage of bacteria

All *E. coli* and *A. tumefaciens* bacteria culture was stored long term in 50% glycerol / 50% bacteria culture in 2ml safe lock Eppendorf tubes in -80°C.

2.1.11.3 Growth media

Several different types of growth media were used in this study. Standard Luria-Bertani media (LB) was used in all *E.coli* and *A.tumefaciens.* Liquid LB consisted of 5g/L yeast extract, 10g/L tryptone, 10g/L NaCl and 10g/L NaOH adjusted to pH 7.0. Solid LB agar was prepared using the same components as liquid LB but with the addition of 15g/L Bactoagar No. 3 to the liquid LB and agar. The solution is then autoclaved at 121°C.

2.1.12 Radioactive nucleotides

ATP, [γ-32P] – 3000Ci/mmol 10mCi/ml EASYTIDE, 250 μCi, Perkin Elmer (Waltham, Massachusetts, United States) was used to label exposed nucleotides for TLC analysis. CTP, [γ-32P] – 3000Ci/mmol 10mCi/ml EASYTIDE, 250 μCi, Perkin Elmer (Waltham, Massachusetts, United States) was used to label the RNA-probe for Northern blot.

2.2 Methods

2.2.1 Germination of seeds

For *A. thaliana* and tomato seeds all plants were sowed on M3 Levington compost in 9cm plastic pots. After watering the soil, seeds were scattered over the soil. To ensure that seeds germinated uniformly throughout all trays, the tray was then covered with a transparent plastic tray lid and left in the cold room at 5^0C for two days. Plants were grown in growth chambers with long day conditions at 23°C with 16-hour day and 8-hour night photoperiod, 60% relative humidity and light intensity at $180\pm20\mu mol/s/m^2$. The plants were not watered once they reached 8-weeks old. Sowing of seeds for tomato was done on the same M3 Levington compost but placed in a 3x6 cell tray. Seeds were sown on wet soil and placed in the cold room for 2 days for stratification. After two days trays were transferred to their respective growth rooms.

To germinate seeds on media for *A. thaliana,* there are three steps involved. Firstly, surface sterilisation of seeds involved placing it in 1.5ml Eppendorf tubes. Secondly, 700ul of 5% sodium hypochlorite was added to the tube and vortexed. Three sterile water washes was done after and seeds were transferred and spread evenly across plates. The plates were then taped up with Micropore tape and placed in a cold room for 2 days for stratification. Tomatoes seeds were done using the same technique but in larger 50ml falcon tubes. Germination on antibiotic medium were the same as above mentioned.

2.2.2 Harvesting of seeds

The *A. thaliana* seeds were harvested after 10 weeks of growth and plants were placed in a drying glasshouse for 1 week to ensure plants are sufficiently dried. Seeds were collected using a sieve and placed in paper bags and stored in a dry atmosphere at room temperature. Tomato fruits were harvested at red-ripening stage of appropriate age and was rinsed with H20 under a sieve to clean the seeds. Seeds were then soaked in 0.1M HCl in 500ml beakers for 16 hours to remove excess tomato flesh. After 16 hours, seeds were then rinsed with H_20 and left to dry on Whatman paper.

2.2.3 DNA extraction

General genomic DNA extraction for *A. thaliana* and *S. lycopersicum* were extracted from tissues using Edwards Buffer. The plant material was homogenised into a fine powder, and 700µl of Edwards buffer was added to each sample. After centrifugation, an equal volume of isopropanol was added to the supernatant, and the DNA was precipitated at -20°C overnight. The DNA was then pelleted and washed using 70% ethanol. The pellet was dried and re-suspended in water. DNA quality and quantity were determined using a NanoDrop Spectrophotometer. Samples were used immediately or stored at -20°C.

DNA was extracted from 5-week-old plant tissue samples of WT, *fip37-4*, P3-07, and P10-23 for whole genome resequencing for BSA analysis. The tissue was collected by cutting young healthy leaves and storing them in liquid nitrogen at -80°C. The samples were then grounded in a mortar and pestle, and DNA was extracted using the MagMAX plant DNA isolation kit according to the manufacturer's protocol. To ensure optimum conditions the mortar and pestle was dipped in liquid nitrogen and tissue samples were then grinded and transferred to a fresh Eppendorf tube.

2.2.3.1 Tomato DNA extraction

Tissue samples from tomato from two different types of knockout and knockdown line by CRISPR/Cas9 and RNAi lines were collected at the required age and frozen in liquid nitrogen until DNA extraction. For collection of tissue in compost in the glasshouse young healthy leaves were cut at the base and placed in aluminium foil and into liquid nitrogen. Samples were stored in the -80^0C freezer until DNA extraction was performed. At the time of extraction samples were taken out of the freezer and placed in a liquid nitrogen dewer. Mortar and pestle were then used for DNA extraction. To ensure optimum conditions the mortar and pestle was dipped in liquid nitrogen and tissue samples were then grinded and transferred to a fresh

Eppendorf tube. DNA extraction protocol was modified from Shure *et al.*, (1983). Pipette 700µl of Edwards buffer in the Eppendorf tube per 1g of tissue. Powder and buffer were vortexed and centrifuged at max speed for 10 minutes. Supernatant was collected and transferred to a fresh Eppendorf tube. Equal volume of isopropanol was added and precipitated overnight at 20°C. After precipitation samples were centrifuged for 10 minutes at max speed and supernatant was removed. To ensure further purification 70% EtOH was added vortexed and centrifuged. The EtOH was then removed, and tubes were placed in a vacuum compartment for 3 minutes and resuspended in 40µl of nuclease free H_20. If high-purity DNAs were needed, the MagMAX plant DNA isolation kit was used according to the manufacturer's instructions.

2.2.3.2 Polymerase chain reaction

Standard Polymerase Chain Reaction (PCR) with genomic DNA or cDNA template was used to prepare DNA fragments for cloning or sequencing. Oligonucleotides used for polymerase chain reaction (PCR) was ordered from Sigma-Aldrich without any modifications and suspended in nuclease free H_20 to a final concentration of 100 pmol/µl. PCR was used to amplify the tomato CRISPR/Cas9 constructs and also to generate RNAi-*MTA* constructs plants. PCR was also used to amplify fragments for cloning and to check correct insertion if present in plasmids and bacteria.

PCRs were performed using Standard Taq Polymerase (NEB), Q5® DNA Polymerase (NEB), Phusion® DNA Polymerase (NEB). Standard Taq Polymerase was used for screening of plant genomic DNA for presence of mutation and screening of clones in transformation. Phusion and Q5 DNA Polymerases was used for amplification of sequences for bacteria cloning due to the high fidelity and proofreading capabilities of these polymerases. All reactions for all polymerases were done with the standard reaction volume and following manufacturer's protocols from NEB. Annealing temperatures for primer pairs were calculated using NEB Tm calculator (https://tmcalculator.neb.com). PCR amplifications were performed using

Eppendorf 5332 Mastercycler Gradient Thermal Cyclers. To detect positive clones of reconstructed plasmids in bacterial colonies, colony PCR was performed by adding a portion of the colony to the PCR reaction tubes using a 2ul pipette tip without touching the media surface. The standard PCR protocol was used except for the initial denaturation step, which was increased to 10 minutes to disrupt bacterial cells and allow the template plasmid DNA to be released into the PCR reaction.

2.2.4 Agarose gel electrophoresis

DNA electrophoresis was performed to confirm the presence of correct size amplified fragments using 1% (w/v) agarose gels Tris/Borate/EDTA (TBE) in gel trays with specified combs to produce gel wells. 5μl of Hyperladder 1kb (0.2 – 10kb, Bioline) was used as a molecular weight marker. Volumes of up to 20μl was pipetted in each well and gels were run on at 100 volts for 45 minutes and imaged with a Biorad imager.

2.2.5 Purification of PCR and DNA products

Purification of PCR products was performed to achieve accurate sequencing results and also efficient downstream cloning steps. This method was also used prior to golden gate assembly and cloning into desired vectors for efficient transformation. Purification was done using the Monarch® Genomic DNA Purification Kit (NEB), as per the manufacturer protocols.

2.2.6 Purification of PCR products from gel extraction

Amplified PCR products or restriction digestion of plasmids containing multiple bands was processed by using gel extraction method to obtain the desired band. Bands were excised from 1% (w/v) agarose gels when needed by scalpel blades and placed into an Eppendorf tube. 400-500μl of gel dissolving solution (NEB) was placed in tubes and incubated at 50°C for 5 minutes while vortexing every minute. The dissolve gel in solution was then passed

through the Monarch® Genomic DNA Purification Kit (NEB) following the manufacturer's protocol after step 2.

2.2.7 DNA and RNA quantification

DNA and RNA concentration and purity was performed using the Thermo Fisher Scientific Nanodrop 2000 Spectrophotometer by first blanking the reading using nuclease free H_20 and subsequently pipetting 2μl of required DNA or RNA.

2.2.8 A-tailing PCR products

To produce amplified fragments with a single overhanging A at both the 5'and 3' ends of each PCR product , A tailing was done using amplification with standard Taq Polymerase (NEB). A tailing allows cloning into gateway vectors. Resulting products was then cloned into entry vector pCR™8/GW/TOPO™ for subsequent gateway reaction into popOff2(Kan) destination vector.

2.2.9 T4 DNA ligation

DNA ligation were carried out following the manufacturer's protocol using T4 DNA Ligase (NEB). The molar ratio of 1:3 vector to insert was used in the ligation reactions, and the NEBioCalculatorTMv1.10.1 was used to calculate the molar ratios. To prepare the reaction mix, a PCR tube was used, and it contained 2μl T4 DNA Ligase Buffer (10X), 50ng (0.020 pmol) vector DNA (4kb), 37.5ng (0.060 pmol) insert DNA (1kb), 1μl T4 DNA Ligase, and sterile deionised water up to a total volume of 20μl. The reaction mix was gently mixed, incubated at 16°C overnight, and heat-inactivated at 65°C for 10 minutes. Then, 10μl of the reaction was used for transformation into competent cells.

2.2.10 Plasmid Construction

2.2.10.1 CRISPR/Cas9 selection of target sites

Genes of interest were screened for 20 bp sequences ending in a PAM (NGG) site, following the pattern GNNNNNNNNNNNNNNNNNNNGG, using the CRISPR-P design website (http://crispr.hzau.edu.cn/cgi-bin/CRISPR2/SCORE). The target sequence was designed from the gene of interest. The gene of interest were, MTA, MTB, VIRLIZER genomic sequences that were taken from the tomato SOL genomics database. Potential target sequences were assessed for scores and efficiency of the target site. Subsequently, specific primers spanning the cut site for target amplification were designed for confirmation of mutation.

2.2.10.2 Golden Gate level 1 for CRISPR/Cas9

The designed gene specific sgRNA guide fragments were used as template DNA in a standard PCR reaction with plasmid pICH86966. The reaction was run on a gel and was then put through the gel extraction protocol and the DNA purified using NEB Purification kit. DNA concentration was quantified by Nanodrop and then diluted to 20 ng/µL. Each PCR product (20 ng/µL) was used in a BSaI golden gate reaction with plasmid pICH47751 and pICSL01009. Ligated level 1 plasmids were transformed into DH5α, and plated using agar plates containing 100 mg/L Ampicillin, 200 mg/L X-gal, and 1 mM IPTG. White and blue colonies were present but white colonies were selected and screened for correct sequence.

2.2.10.3 Golden Gate level 2 for CRISPR/Cas9

After performing positive selection and screening on the level 1 golden gate reaction, the positively identified pICH47751 was used in the level 2 golden gate reaction. The reaction consisted of using pICH47751-MTA-*sgRNA*, pICH47732, pICH47742, pICH41766, and ligated into pAGM4723. The ligated products were then transformed into DH5α, as described

previously, and plated on solid plates containing 50 mg/L Kanamycin. White and red colonies were observed, but white colonies were selected and screened for correct assembly. Colonies were subsequently confirmed by Eurofins sequencing to ensure proper assembly.

2.2.10.4 Construction of hpRNAi pOpOff2(kan) vector for RNAi silencing

The present study utilised a dexamethasone-inducible RNAi system, employing the vector pOpOff2(kan) as previously described by Wielopolska *et al.*, (2005). Genomic DNA from WT tomato plants at 2 weeks of age was extracted using the MagMAX Plant DNA Isolation Kit (Thermo Fisher Scientific) to ensure high-quality DNA. A 456 bp fragment was then amplified using RNAi MTA-Fwd and RNAi MTA-Rev primers, following the methods outlined in section 2.2.8 for A-tailing. The resulting PCR product were visualised on a gel, and gel excision was performed as described in section 2.2.6. The purified PCR product was sent for Eurofins sequencing to confirm the absence of base-pair errors and ensure the correct sequence was amplified.

The PCR product was then ligated into the pCR8/GW/TOPO® TA Cloning Kit, following the manufacturer's instructions. The ligation product was transformed into One Shot *E. coli* cells. Positive clones of the resulting PCR8-MTA construct were screened using PCR8 attL2-Fwd and PCR8 attL2-Rev primers, and subsequently sent for Eurofins sequencing to confirm correct insertion of the PCR product. Since both PCR8 and pOpOff2(kan) are spectinomycin resistant, the PCR8-MTA construct was amplified with both attL2 sites using the same PCR8 attL2-Fwd and PCR8 attL2-Rev primers, and the purified product was further cloned into pJET 1.2 vector (Thermo Fisher Scientific). The resulting construct was screened for positive colonies and sent for Eurofins sequencing.

The LR Clonase reaction was then performed using Invitrogen Gateway LR Clonase II enzyme mix, following the manufacturer's instructions, with pJET1.2-MTA and pOpOff2(kan) as substrates, according to the LR Clonase kit manufacturer's instructions. The resulting product

was transformed into DH5α cells and colonies were screened using primers popOff2 Int-35S-Fwd and popOff2 Int-35S-Rev to confirm positive reactions. The confirmed vector was then purified and stored at -80°C for subsequent use in *Agrobacterium*-mediated transformation of *S. lycopersicum* tomato cotyledons.

2.2.11 Plant transformation via tissue culture

The following protocol was used for transformation of *Agrobacterium* into cotyledons and stems of *S. lycopersicum* from tissue culture.

2.2.11.1 Excision and pre-callusing

Seeds were surfaced sterilised (2.1.8.1) and sown on sterile germination media (2.1.8.2). After seedlings reached 7-10 days (prior to first true leaves), cotyledons and hypocotyls were excised using a sterile scalpel blade adaxially (Bird *et al.*, 1988). Excisions were performed with the seedlings in SDW to minimise damage and osmotic water loss. Both the cotyledons and hypocotyls were placed carefully on callusing medium on sterile plates. Plates were sealed with micropore tape (3M) and placed in growth rooms under standard growing conditions with no light for a minimum of 48 hours, to pre-callus.

2.2.11.2 *Agrobacterium* growth for inoculation

After 48 hours of pre-callusing, colonies from *Agrobacterium* strains, containing transformation constructs of interest were inoculated into 5mL of liquid LB media and grown under standard conditions. On the day of co-cultivation, 200mL of LB with appropriate antibiotic were inoculated with 5mL of *Agrobacterium* into a baffled flask. Cultures were grown under standard conditions at 28°C to an OD600 of 0.2-0.3. Cultures were then transferred equally into 50mL falcon tubes and centrifuged at 3000rpm for 7 minutes. The supernatant was decanted and gently resuspended in liquid resuspension media.

2.2.11.3 Tissue incubation on *Agrobacterium*

Using the liquid resuspension containing the *Agrobacterium* it was then placed into a sterile magenta box and excised plant tissues, cotyledons and hypocotyls were placed in the magenta box containing the cultures and incubated at room temperature for 20 minutes stirring occasionally. Tissues were dry blotted onto sterile filter paper to remove excess cell culture before replacing the tissues on pre-callusing media plates. The plates were sealed with micropore 3M tape and returned to the dark growth room for a further 24-48 hours.

2.2.11.4 Selection media for tissue callus

After incubation in growth room in the dark for 48 hours the tissues were gently transferred taking care not to damage the tissue to callusing medium on plates. The plates were then sealed with micropore tape and placed under normal growth conditions (2.1.9.1). Callusing plates were refreshed every 10-14 days, until callus were developed, and meristematic shoots were grown from callusing tissues.

2.2.11.5 Excision of shoots and rooting

Once shoots were grown to a suitable height for excision, it was then excised with a scalpel blade and placed on rooting medium in sterile plastic media pots. Once root and branches developed the resulting shoot was then transferred to compost and placed in plastic bags to maintain humidity. After several days the plants were then transferred to a glasshouse and grown normally.

2.2.12 GUS staining

GUS staining was performed on whole tomato seedlings that were grown on their respective media. Seedlings were immersed in GUS staining reageant mix (0.2M phosphate buffer, potassium ferricyanide, potassium ferrocyanide, 0.5M EDTA, Triton X-100, and X-Gluc). The stained roots were then taken out and soaked in 100% EtOH until the seedlings were clear.

2.2.13 DEX induction

Tomato seeds were sterilised and sown on ½ MS pots and grown in a Conviron growth room. After 2 weeks seedlings were then carefully transferred to a pot with ½ MS + Dexamethasone 30μM and seedlings were imaged and phenotyped after 4 weeks. For dexamethasone induction on compost, seeds were sown on Levington compost and germinated in a glasshouse. After 2 weeks, seedlings were subjected to induction by dexamethasone solution (30μM) every 5 days to ensure constant induction applied and plants were not subjected to low watering conditions.

2.2.14 Phenotyping of root architecture

A. thaliana seeds were sterilised and placed on normal media for phenotyping on normal conditions. For phenotyping on high salinity conditions, plates were made using standard ½ MS media with the addition of NaCl to the media before it was poured onto plates. *A. thaliana* seeds were spread evenly across the top row of the vertical growth plate to ensure sufficient room was available for roots to grow. In drought conditions sorbitol was added to the ½ MS media before it was poured on the plate. The plates were imaged from 14-day-old seedlings and primary root lengths and lateral root number were measured using ImageJ software (v1.8.0_172).

2.2.15 Pericycle activation for lateral root initiation

The pericycle activation for lateral root initiation protocol were performed according to the method described in Himanen *et al.*, (2002). *A .thaliana* seedlings were grown vertically on ½ MS plates and after 4 days were transferred to vertical plates containing ½ MS and 10 µM NPA. On the 3rd day of culturing on NPA, they were transferred onto ½ MS plus 10 µM NAA. The roots were photographed under a Stereo Dissecting Microscope (Zeiss Stemi SV6). The frequency and lengths of root hairs were measured using ImageJ software (v1.8.0_172).

2.2.16 Total RNA extraction

Total RNA was extracted from *A. thaliana* plant tissue using an adapted acidified phenol yeast RNA extraction protocol (Schmitt *et al.*, 1990). The plant tissue was homogenised into a fine powder with liquid nitrogen to prevent RNA degradation. Ground plant material was then transferred to a microcentrifuge tube and mixed with AE buffer (50mM NaOAc [pH 5.3] and 10mM EDTA) and 10% SDS. After vortexing, AE saturated phenol was added, heated, and centrifuged. The upper phase was transferred to a new tube and mixed with phenol:chloroform:isoamyl alcohol (25:24:1), centrifuged, and the upper aqueous layer was collected. RNA was precipitated overnight at -20°C with 1/10 volume 3M NaOAc and 2.5 volume of absolute ethanol, pelleted, washed (80% EtOH), and dried. The RNA pellet was dissolved in sterile deionized water and its quality and quantity were measured using a NanoDrop Spectrophotometer. Total RNA samples were either used immediately or stored at -80°C. Total RNA extraction for RNAi lines in tomato was done using the RNAqueous™ Total RNA Isolation Kit as per the manufacturer's instructions.

2.2.17 Differential Expression Analysis

To generate differentially expressed genes in the RNA-seq experiment the total RNA was extracted from two-week old seedlings of WT, *fip37-4*, and suppressor mutants P3-07 and P10-23. The RNA samples were sent to Novogene Co. Ltd in Cambridge for Plant and Animal

Eukaryotic mRNA Sequencing service. The RNA samples went through QC checks and mRNA library preparation via Poly-A enrichment. Illumina NovaSeq PE150 sequencing was conducted on all RNA samples at a depth of 20M reads. In order to process raw RNA-seq data the datasets were uploaded and pre-processed using Galaxy (Jalili *et al.*, 2021). To evaluate if raw RNA-seq data is sufficient in quality, the FastQC program was used to assess the quality scores of each base sequence. A score of 30 or higher indicates a 1 in 1000 chance of an incorrect base call. MultiQC was then used to collate all FastQC results in one graph. Adapters and low-quality reads were filtered using Trim Galore, a wrapper script was used to automate quality and adapter trimming (Krueger, 2021). After quality checking and trimming, the trimmed data was aligned to the *A. thaliana* TAIR10 reference genome using the Hierarchical Indexing for Spliced Alignment of Transcripts (HISAT2) program, which is a fast and sensitive alignment tool (Kim *et al.*, 2015). The BAM files generated after alignment to the *A. thaliana* reference genome (TAIR10) were processed using featureCounts, a program for counting RNA-seq reads (Liao *et al.*, 2014). To identify differentially expressed genes (DEG), the count tables produced by featureCounts were inputted into DESeq2. DESeq2 was then subsequently used to produce all RNA-seq data quality assessment graphs. DESeq2 was also used to produce heatmaps and differentially expressed gene files.

2.2.18 Code for DESeq2 RNA-seq quality assessment and differentially expressed gene files

The code below was run on R studio environment on a Macintosh computer. Libraries were pre-installed following installations instructions from the creator of each program.

//

```
library("DESeq2")
library("BiocParallel")
library("apeglm")
library("ashr")
library("ggplot2")
library("vsn")
library("pheatmap")
library("RColorBrewer")
library(dplyr)
library(GenomicFeatures)
```

```
library(readr)
library(gplots)
library(AnnotationDbi)
library(org.At.tair.db)
library(pathview)
library(gage)
library("ggrepel")

setwd("D:/UoN/SamuelChak")
getwd()
register(SnowParam(6))
cts = read.csv("counts_table.csv")
coldata = read.csv("samples.csv")

#Define the row names from the Gene_id column
row.names(cts) <- cts$Gene_id

# remove the column Gene_id since it is now used as a row name
cts <- cts %>% dplyr::select (-Gene_id)
names(cts) <- gsub(x = names(cts), pattern = "X", replacement = "")#Remove the X from column names

#Define the row names from the Gene_id column
row.names(coldata) <- coldata$sample

# remove the column Gene_id since it is now used as a row name
coldata <- coldata %>% dplyr::select (-sample)

#View count matrix
head(cts,2)

#View sample data
Coldata

#Check if row names in sample data are in the same order as coloum names in count matrix
all(rownames(coldata) == colnames(cts))

#Construct the DESeqDataSet
dds <- DESeqDataSetFromMatrix(countData = cts, colData = coldata, design = ~ condition)
dds
colData(dds)

#Pre-filtering: remove rows(genes) that have counts < 10
keep <- rowSums(counts(dds)) >= 10
dds <- dds[keep,]

#Pre-filter: remove rows(genes) that have no counts
#dds <- dds[ rowSums(counts(dds)) > 0, ]

#removing rows (genes) that have <0.5 TMP counts in at least 3 (biological rep) samples
#dds <- dds[ rowSums(counts(dds)  > 0.5 ) > 3, ]

#Differential expression analysis
dds <- DESeq(dds)

# plot Cooks distances to spot individual sample outliers
par(mfrow=c(1,1), mar=c(8,4,2,1))
tiff("cook.tiff", width = 25, height = 8, units = 'in',res = 300)
boxplot(log10(assays(dds)[["cooks"]]), range=0, las=2)
dev.off()

# plot of p-values
hist(res$pvalue, breaks=50, col="grey")

# MA plot
DESeq2::plotMA(dds, ylim=c(-2,2))

## MA plot version2
```

```
# Get differential expression results
res <- results(dds)
table(res$padj<0.05)
## Order by adjusted p-value
res <- res[order(res$padj), ]
## Merge with normalized count data
resdata <- merge(as.data.frame(res), as.data.frame(counts(dds, normalized=TRUE)), by="row.names",
sort=FALSE)
names(resdata)[1] <- "Gene"
head(resdata)
## change labelsig=FALSE to labelsig=TRUE to label genes
maplot <- function (res, thresh=0.05, labelsig=FALSE, textcx=1, ...) {
  with(res, plot(baseMean, log2FoldChange, pch=20, cex=.5, log="x", ...))
  with(subset(res, padj<thresh), points(baseMean, log2FoldChange, col="red", pch=20, cex=1.5))
  if (labelsig) {
    require(calibrate)
    with(subset(res, padj<thresh), textxy(baseMean, log2FoldChange, labs=Gene, cex=textcx, col=2))
  }
}
png("diffexpr-maplot.png", 1500, 1000, pointsize=20)
maplot(resdata, main="MA plot")
dev.off()


# Dispersion plot
png("Dispersion plot.png", 1000, 1000, pointsize=20)
plotDispEsts(dds, main="Dispersion plot")
dev.off()

# Volcano plot
# order results table by the smallest adjusted p value:
res <- res[order(res$padj),]
results = as.data.frame(mutate(as.data.frame(res), sig=ifelse(res$padj<0.05, "FDR<0.05", "Not Sig")),
row.names=rownames(res))
#plot
p = ggplot(results, aes(log2FoldChange, -log10(pvalue))) + geom_point(aes(col = sig)) +
  scale_color_manual(values = c("red", "black")) + ggtitle("Volcano plot")
p
## plot with top "significant" genes labeled
p + geom_text_repel(data=results[1:10, ], aes(label=rownames(results[1:10, ])))

comparison_1=results(dds, contrast=c("WT","FIP37"), parallel=TRUE) %>% as.data.frame
comparison_2=results(dds, contrast=c("P3-07","WT",), parallel=TRUE) %>% as.data.frame
comparison_3=results(dds, contrast=c("P3-07","FIP37"), parallel=TRUE) %>% as.data.frame
```

2.2.19 Code for DESeq2 Heatmap

```
comparison_1$Id <- rownames(comparison_1)
comparison_2$Id <- rownames(comparison_2)
comparison_3$Id <- rownames(comparison_3)

comparison_1$Contrast <- rep("comparison_1",nrow(comparison_1))
comparison_2$Contrast <- rep("comparison_2",nrow(comparison_2))
comparison_3$Contrast <- rep("comparison_3",nrow(comparison_3))

#merge contrast results
Res_fer <- rbind(comparison_1,
                 comparison_2,
                 comparison_3,)

#Filter the results table by significant results
#make a list of significant results
Res_fer$na <- ifelse(Res_fer$padj<0.05,TRUE, FALSE)
a <- filter(Res_fer, na == "TRUE")
```

```
b <- data.frame()
B=c(7)
b <- data.frame(a[,B])
colnames(b) <- c("Id")
b=b %>% distinct(Id, .keep_all = TRUE)
#keep only the significant results in the contrast data, by keeping only those on the sig. list
Res_fer=Res_fer[Res_fer$Id %in% b$Id,]

#Filter the results table by log2FoldChange
#make a list of log2FoldChange >1.5
Res_fer$log2FoldChange.abs <- abs(Res_fer$log2FoldChange)
Res_fer$na.FC <- ifelse(Res_fer$log2FoldChange.abs>2,TRUE, FALSE)
a <- filter(Res_fer, na.FC == "TRUE")
b <- data.frame()
B=c(7)
b <- data.frame(a[,B])
colnames(b) <- c("Id")
b=b %>% distinct(Id, .keep_all = TRUE)
#keep only log2FoldChange >2 in the contrast data, by keeping only those on the sig. list
Res_fer=Res_fer[Res_fer$Id %in% b$Id,]

# Run clustering analysis for genes(y axis) and samples (x axis)
#Select date for the clustering dataframe
B=c(7,2)

comparison_1.1 <- comparison_1[,B] %>% as.data.frame
colnames(comparison_1.1) <- c('Id','comparison_1')
comparison_2.1 <- comparison_2[,B] %>% as.data.frame
colnames(comparison_2.1) <- c('Id','comparison_2')
comparison_3.1 <- comparison_3[,B] %>% as.data.frame
colnames(comparison_3.1) <- c('Id','comparison_3')
#merge contrast results
tree.matrix=cbind(comparison_1.1,
                  comparison_2.1,
                  comparison_3.1,)

#keep only the significant results in the contrast data, by keeping only those on the sig. list
tree.matrix=tree.matrix[tree.matrix$Id %in% b$Id,]
#Format the dataframe
tree.matrix2 <- tree.matrix %>% dplyr::select (-Id)

#Run clustering for genes
dd <- dist(scale(tree.matrix2), method = "euclidean")# create distance matrix
tree <- hclust(dd, method = "ward.D2")# Hierarchical Clustering

#plot clstering with ggtree
p=ggtree(tree, color="black", size=0.2, branch.length='none')
mypal =
c("#781156","#A51876","#D21E96","#E43FAD","#EA6CC0","#F098D3","#114578","#185EA5","#1E78D2","#3F91
E4","#6CABEA","#98C4F0","#117878","#18A5A5","#3FE4E4","#6CEAEA","#98F0F0",
"#117845","#18A55E","#1ED278","#3FE491","#6CEAAB","#98F0C4","#787811","#A5A518","#D2D21E","#E4E43
F","#EAEA6C","#F0F098","#F7F7C5","#784511","#A55E18","#D2781E","#E4913F","#EAAB6C","#F0C498","#78
1122","#A5182F","#D21E2C","#E43F5B","#EA6C81","#F098A7")
p
p +geom_tiplab(aes(label=label), size=0.5, hjust=-0.1) +
  coord_cartesian(clip = 'off') +
  theme(panel.background = element_blank(),
        legend.title = element_text(size = 2),
        legend.key.size = unit(0.2, "cm"),
        legend.text=element_text(size=2),
        plot.margin=margin(0, 120, 0, 6)) +
  scale_color_manual(values = mypal) +
  guides(color=FALSE, shape=FALSE, fill=FALSE)
ggsave("clusering_gene.tiff", height=30, width=3, units='in', dpi=600)

#get the order of the tip labels from the tree
d = fortify(tree)
```

```
d = subset(d, isTip)
a=with(d, label[order(y, decreasing=F)])
#get the order of the tip labels from the tree (by creating the gene_order vector which can be inputted directly)
gene_order=unique(a)
#Reorder y axis by phylogenetic tree (enter the list using the phylogenetic_tree_order vector)
Res_fer$Id <- factor(Res_fer$Id, levels = gene_order)

#Run clustering for samples
tree.matrix3=t(tree.matrix2)#transpose the dataframe

dd <- dist(scale(tree.matrix3), method = "euclidean")# create distance matrix
tree <- hclust(dd, method = "ward.D2")# Hierarchical Clustering

#plot clstering with ggtree
p=ggtree(tree, color="black", size=0.4, branch.length='none')
mypal =
c("#781156","#A51876","#D21E96","#E43FAD","#EA6CC0","#F098D3","#114578","#185EA5","#1E78D2","#3F91
E4","#6CABEA","#98C4F0","#117878","#18A5A5","#3FE4E4","#6CEAEA","#98F0F0",
"#117845","#18A55E","#1ED278","#3FE491","#6CEAAB","#98F0C4","#787811","#A5A518","#D2D21E","#E4E43
F","#EAEA6C","#F0F098","#F7F7C5","#784511","#A55E18","#D2781E","#E4913F","#EAAB6C","#F0C498","#78
1122","#A5182F","#D21E2C","#E43F5B","#EA6C81","#F098A7")
p
p +geom_tiplab(aes(label=label), size=0.75, hjust=0.5,vjust=1) +
  coord_cartesian(clip = 'off') + coord_flip() +  scale_x_reverse() +
  theme(panel.background = element_blank(),
        legend.title = element_text(size = 2),
        legend.key.size = unit(0.2, "cm"),
        legend.text=element_text(size=2),
        plot.margin=margin(0, 0, 0, 0)) +
  scale_color_manual(values = mypal) +
  guides(color=FALSE, shape=FALSE, fill=FALSE)
ggsave("clusering_samples.tiff", height=2, width=6, units='in', dpi=600)

#get the order of the tip labels from the tree
d = fortify(tree)
d = subset(d, isTip)
a=with(d, label[order(y, decreasing=F)])
#get the order of the tip labels from the tree (by creating the sample_order vector which can be inputted directly)
sample_order=unique(a)
#Reorder y axis by phylogenetic tree (enter the list using the phylogenetic_tree_order vector)
Res_fer$Contrast <- factor(Res_fer$Contrast, levels = sample_order)

#Rescale fold change value
Res_fer$value.rescaled <- rescale(Res_fer$log2FoldChange, to = c(-2.1, 2.1))

#Plot heatmap:red-blue
p = ggplot(data=Res_fer) +
  geom_raster(aes(x=Contrast, y=Id, fill=value.rescaled)) +
  scale_fill_gradient2(low="blue", mid = "white", high="red", na.value="grey", name="log2FoldChange") +
  theme(panel.border = element_rect(colour = "black", fill=NA, size=0.1)) +
  scale_y_discrete(expand = c(0,0)) + scale_x_discrete(expand = c(0,0))
p
p + guides(color=FALSE, shape=FALSE, fill=FALSE) +
  theme(axis.text.x=element_text(angle=90, hjust = 1, vjust = 0.5, size=4, color="black"),
        axis.text.y=element_text(angle=0, hjust = 1, vjust = 0.5, size=1, color="black"),
        panel.background = element_blank(),
        axis.ticks = element_blank(),
        plot.title = element_blank(),
        axis.title.x  = element_blank(),
        axis.title.y  = element_blank(),
        legend.title = element_blank(),
        legend.text=element_blank()) +
  labs(x = "Fermentation time (hr)")
ggsave("DESeq_heatmap.rb.tiff", height=7, width=6, units='in', dpi=600)


//
```

2.2.20 Integrative Genome Viewer

The RNA-seq paired-end reads were aligned to the Arabidopsis thaliana TAIR10 reference genome using HISAT2 on Galaxy. The *A. thaliana* TAIR10 Fasta file was used in history to create an index file for alignment, with default parameters settings. The resulting BAM files and their respective BAM index were downloaded and imported into Integrative Genomics Viewer 2.11.1, with the reference genome set as *A. thaliana* TAIR10, to visualise the coverage of reads across genes.

2.2.21 Statistical analysis

All statistical data recording and subsequent analysis in this thesis were conducted using GraphPad Prism 9 software. The corresponding statistical tests employed are indicated in the figure legend.

2.2.22 Bulked segregant analysis sequencing

Bulked segregant analysis (BSA) was performed on EMS mutagenised positively screened lines for P3-07 and P10-23. Segregant lines were divided into two pools of short and tall plants and DNA extraction was performed as previously described in Methods section 2.2.3. Around 200ng total genomic DNA from tall and short suppressor plants were sequenced by Novogen Co Ltd. using the Animal and Plant Whole Genome Sequencing service. All sequencing for BSA was performed at a depth of 25 million pair-end reads.

After sequencing completion from Novogene the raw paired-end data was downloaded from their database and was subjected to processing through sequencing pipeline of Professor Nigel Mongan at the University of Nottingham. The paired-end FastQ files obtained from Novogene were subjected to quality filtering and adapter sequence trimming using the $-TRIM GALORE! wrapper with default parameters. The resulting trimmed FastQ files were then

aligned to the TAIR10 using the Burrows-Wheeler Alignment - MEM tool. The aligned sequencing files were pre-processed using the PICARD tools, including samtools View, sortSam, rmdup, and MarkDuplicates. SAM files were converted to BAM files using $ samtools view and sorted using $ samtools sort. Duplicate alignments were removed using $ samtools rmdup, followed by $ samtools MarkDuplicates. All samtools operations were carried out using default parameters.

To identify variant alleles within the aligned FastQ files, a Sequence dictionary for the *A. thaliana* genome was generated using $ samtools CreateSequenceDictionary on TAIR10 genome file, resulting in the output file TAIR10genome.dict. Read groups were then added to the aligned FastQ files using $ addorreplacegroups. The identified variants that differed from the *A. thaliana* genome were called using $ GATK Haplotype caller, which simultaneously calls SNPs and indels through local de novo assembly of haplotypes in active regions. The VCF output file was uploaded to EnsemblPlants using their online tool Variant Effect Predictor to annotate the SNPs and indels in a txt file. The Variant Effect Predictor txt files for the mutant (tall) and non-mutant (short) populations were manually filtered in Microsoft Excel to create a list of unique mutation sites specifically in the tall suppressor plant population, excluding the short plant population.

2.2.23 cDNA synthesis

cDNA synthesis was performed on RNAi tomato plant lines after total RNA extraction was done using the RNA extraction kit. To synthesise first strand cDNA, the SuperScript® IV Reverse Transcriptase kit (Invitrogen) was used according to the manufacturer's instructions. Briefly, 1 µg of total RNA was mixed with 50 µM oligo d(T)20 primer and 1 µl of 10 mM dNTP in a PCR tube, and incubated at 65°C for 5 minutes. After cooling the tube on ice for 1 minute, a reverse transcription reaction mix containing 4 µl of 5x SSIV buffer, 1 µl of 100 mM DTT, 1 µl of RNaseOUT™ Recombinant RNase Inhibitor, and 1 µl of SuperScript® IV Reverse

Transcriptase (200 U/μL) was added. The tube was incubated at 55°C for 10 minutes, followed by heat inactivation at 80°C for 10 minutes. The resulting cDNA was diluted 1:4 for use in standard PCR and RT-qPCR reactions, and stored at -20°C.

2.2.24 RT-qPCR

Real-time quantitative polymerase chain reaction (RT-qPCR) was conducted to analyse gene expression. SensiMix™ SYBR® Low-ROX Kit was used following the manufacturer's instructions. The RT-qPCR reaction mix composed of 1 μL 25 μM forward primer, 1 μL 25 μM reverse primer, 2 μL of dH2O, 1 μL template cDNA, and 5 μL 2x SensiMix SYBR Low-ROX was loaded into Corning Axygen PCR Microplate size 384 wells plates. No-RT cDNA was employed as a negative control. Transcript levels were quantified using a Light Cycler 480 Real-Time PCR system (Roche Applied Science). The amplification programme consisted of 95°C for 10 minutes (pre-incubation) and 40 cycles of 95°C for 15 seconds, 60°C for 60 seconds, 72°C for 20 seconds (amplification). At the end of the amplification step, a melting curve was generated by holding samples at 95°C for 5 seconds, followed by 60°C for 60 seconds, and then increasing the temperature by 1°C per cycle from 60-97°C. All samples were run in triplicate and housekeeping gene *SAND* (Solyc03g115810) and *SITIP41* (Solyc10g049850) was used as the reference gene for normalising the amplification in gene expression analysis. *MTA* (Solyc08g066730) gene specific primers was also used to quantify gene expression.

2.2.25 Northern Blotting

2.2.25.1 Electrophoresis

In order to prevent any contamination that could affect RNA quality preparation of the denaturing gel, the gel tray, comb and electrophoresis tank were soaked in a 1% (w/v) solution of SDS for 30 minutes and rinsed with sterile, deionised water. The gel was prepared by dissolving 1.2g of agarose in 2ml of 1M sodium phosphate buffer (pH 6.5) and 90ml of sterile

deionised water by microwaving. Once fully dissolved, the gel was allowed to cool to 50°C and 8ml of formaldehyde (37%, pH 7.0) was added and mixed thoroughly before being poured into the gel tray in a fume hood. The gel was left to set for at least one hour. Meanwhile, total RNA samples and running buffer were prepared. The running buffer was prepared by mixing 13ml of 1M sodium phosphate buffer (pH 6.5), 50ml of formaldehyde (37%), and 587ml of sterile, deionised water. A total of 8ug of total RNA was mixed with an equal volume of ethidium bromide buffer, denatured at 65°C for 5 minutes, and immediately placed on ice. The ssRNA ladder (New England Biolabs) was prepared in the same way, and 1/5 volume 5x Bromophenol Blue was added to total RNA samples and the ssRNA ladder. The gel was run at 90V for 123 minutes with a circulator used to keep the buffer even.

2.2.25.2 Blotting

Following electrophoresis, the gel was imaged using the Gel Doc XR+ Gel Documentation System (BIO-RAD) in accordance with the manufacturer's instructions and Image LabTM Software (BIO-RAD). The gel was then cut to the desired size for blotting. One sheet of nylon membrane (PerkinElmer) and two sheets of 3M Whatman paper were cut to the same size as the gel. A plastic tray was used to hold 500ml of 10× SSC, which served as the blotting buffer. The tray was cleaned with 1% (w/v) SDS and rinsed with sterile deionised water. A stack of paper towels was soaked with 200 ml of 2× SSC, and the nylon membrane and two sheets of 3M Whatman paper were soaked in this solution. The blotting was assembled as shown in Figure 2.1. The gel was placed on the 3M Whatman paper with the bottom side up, and the blotting process was left for 24 hours. After blotting, the total RNA-containing nylon membrane was rinsed with 2xSSC and UV crosslinked for 1200ms.

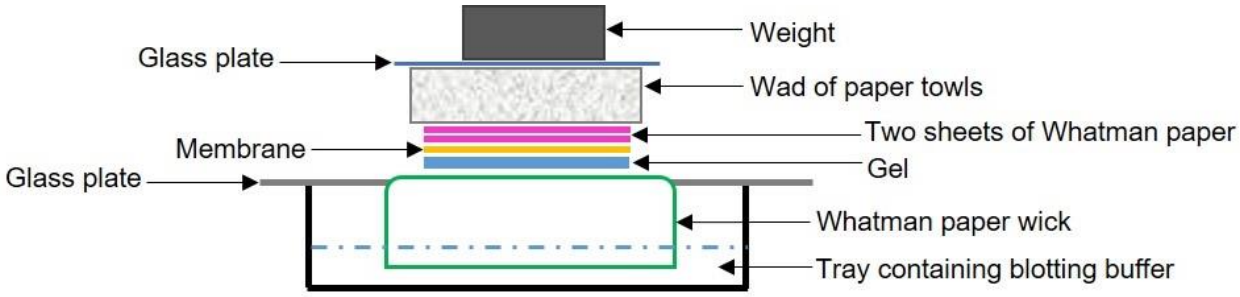


Figure 2.1 Schematic showing the assembly for Northern blot.

2.2.25.3 Prehybridisation

The UV-crosslinked RNA-containing nylon membrane was placed with the RNA side facing inward in a glass cylinder containing 20 ml of prehybridization buffer. The prehybridization buffer was composed of 4 ml of 5 M NaCl, 3 ml of sterile water, 10 ml of formamide, 2 g of Dextran sulfate, 1 ml of heat-denatured salmon sperm DNA (10 mg/ml), and 2 ml of 10% (w/v) SDS. The cylinder was incubated with rotation at 55°C for 2 hours.

2.2.25.4 RNA probe preparation

To generate the RNA probe, *A. thaliana* cDNA was used as a template for PCR under standard conditions with primers containing the T7 promoter sequence. The Riboprobe® system-T7 kit (Promega) transcription reaction was set up in accordance with the manufacturer's instructions. A reaction mixture consisting of 1× Transcription optimised buffer, 10 mM dithiothreitol (DTT), 40U Recombinant RNasin® Ribonuclease Inhibitor, 12 µM rCTP, 0.5 mM rGTP, 0.5 mM rATP, 0.5 mM rUTP, 50 µCi [α-32P]rCTP, 20 U T7 RNA Polymerase, and ~ 500ng DNA template was incubated at 37°C for 1 hour. Following this, 2µl of RQ1 RNase-Free DNase was added to the reaction mixture, followed by incubation at 37°C for 15 minutes. The RNA probe was purified using Bio-Spin® 30 columns as recommended by the

manufacturer, resulting in purified RNA probe ready for fragmentation. To fragment the RNA probe, a carbonate/bicarbonate buffer system was used as described by Bodi *et al.*, (2012). The RNA probe volume was made up to 80µl using sterile deionised water (SDW) then 10µl of 400mM $NaHCO_3$ and 10µl of 600mM Na_2CO_3 and the RNA probe was fragmented at 60°C, and the length of time for fragmentation was calculated using the formula: (L0-Lt)/(kL0Lt), L0 = initial length of transcript (kb), Lt = desired RNA fragment length (kb, normally 0.05-0.06 kb), k = constant = 0.11 kb min^{-1} , t = time (minute).

2.2.25.5 Hybridisation

The resulting fragmented RNA probe was added to the pre-hybridisation buffer directly after the membrane had been pre-hybridised for 2-hours. The hybridisation of the RNA probe was performed at 65°C with rotation for 20-hours.

2.2.25.6 Washing Membrane

The RNA probe hybridized membrane underwent a series of washes with SSC buffers containing 0.1% (w/v) SDS. The membrane was first washed with 2xSSC and 0.1% SDS at room temperature with rotation for 5 minutes. This was followed by subsequent washes with 2xSSC, 1xSSC, 0.2xSSC, and 0.1xSSC buffers all containing 0.1% SDS at 65°C with rotation for 15-minutes each. Following washing, the membrane was heat-sealed in a plastic sleeve and placed under a phosphor screen (Fuji-Screen) for 3 days before being scanned using the Bio-Rad Molecular Imager FX system in combination with Quantity One software.

2.2.26 Thin Layer Chromatography

2.2.26.1 Poly-A purification

The total RNA used for Poly-A purification was extracted using the method described in this chapter. The method utilised to isolate Poly-A RNA from Total RNA was a modified version of the NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB). The 2X NEB Next RNA Binding Buffer was prepared by adding 1M LiCl, 40mM Tris HCl (pH 7.5), 2mM EDTA, and autoclaving before adding 0.1% (v/v) NP-40. Similarly, 1X NEB Next Wash Buffer was prepared by adding 150mM LiCl, 20mM Tris HCl (pH 7.5), 1.0mM EDTA, and autoclaving before adding 0.01% (v/v) NP-40. Prior to use, Oligo d(T)25 beads were resuspended by rolling at room temperature for 30 minutes. An aliquot of 20µl of beads was transferred to a 1.5ml Eppendorf Microcentrifuge tube and placed on a magnetic stand for two minutes until clear and packing buffer was discarded. After washing with 100µl of 2X NEB Next RNA Binding Buffer. The tube was placed on a magnetic stand for 2-minutes until the solution was clear and the supernatant was discarded. This wash step was repeated twice. The beads were re-suspended in 100µl of 2X NEB Next RNA Binding Buffer and 20µg of total RNA in a total volume of 100µl was added to the tube and mixed thoroughly by pipetting.

Upon incubation at 65°C for 5 minutes, the tube was placed on ice until cooled and placed inside a 50ml falcon tube and rolled for 20 minutes at room temperature. After rolling, the tube was put on the magnetic stand for 2 minutes. The supernatant containing Total RNA -Poly(A) was transferred to a new 1.5ml Eppendorf microcentrifuge tube and stored at -80°C. The tube containing beads bound with Poly-A RNA was removed from the magnetic stand, 200µl of 1x NEB Next Wash Buffer was added and mixed by pipetting. Following this, the tube was placed on the magnetic stand for 2-minutes and supernatant was removed. This was step was repeated and 50µl of 55°C SDW was added and mixed by pipetting which was then incubated at 55°C for 5-minutes and placed on the magnetic stand. The supernatant was transferred to a new tube. The tube containing the magnetic beads were resuspended by pipetting with 50µl of 2X NEB Next RNA Binding Buffer and transferred to a new tube. The tube was placed in a 50ml falcon tube and rolled for 15-minutes and subsequently was placed on the magnetic stand and washed with 200µl of 1X NEB Next Wash Buffer and washed with 200µl of ice-cold

water. Poly(A) RNA was then eluted with 10µl of 55°C SDW and mixed thoroughly by pipetting. It was then incubated at 55°C for 5-minutes and placed on the magnetic stand for 2-minutes. The supernatant containing poly(A) RNA was transferred to a new tube. The quality and quantity of Poly-A RNA samples were measured using a NanoDrop Spectrophotometer (NanoDrop Technologies, USA). Poly(A) RNA samples were used immediately or stored at -80°C.

2.2.26.2 Labelling Reaction

To measure internal mRNA m^6A levels TLC was performed. The detection of m^6A is in the consensus sequence GAC. The T1 enzyme used in TLC experiments cuts RNA after G exposing the next nucleotide including the m^6A modified adenosine. A 1.5ml Eppendorf microcentrifuge tube was used to mix 25ng poly(A) RNA with 1µl T1 Ribonuclease (1000U/µl), 1µl 10x PNK buffer, and sterile deionized water to a total volume of 10µl. The mixture was incubated at 37°C for 15 minutes. To label the exposed nucleotides, 1µl PNK and 0.5µl γ32P-ATP were added, and the reaction was incubated for 1 hour at 37°C. The 5'-labeled fragments were precipitated overnight at -20°C after adding 1µl 3M NaOAc and 40µl 100% ethanol. The labelled nucleotide fragments were pelleted by centrifuging at 13,000rpm for 20 minutes. The pellet was then dried for 5 minutes on a 37°C heating block. To the RNA pellet, 1µl 10x P1 buffer, 0.5µl P1 nuclease, and 8.5µl sterile deionised water were added, and the RNA was digested by incubating at 37°C for 2 hours.

2.2.26.3 Running Thin Layer Chromatography

A cellulose TLC plate (20cm x 20cm) was marked with a pencil 3cm from the bottom left corner. The TLC plate was then placed flat on a 37°C heating block, and 1µl of labelled nucleotides was pipetted onto the pencil mark, creating a spot no larger than 3mm in diameter, which was then allowed to dry. The plate was subsequently placed into a chromatography

tank containing 200-300ml of first dimension buffer (isobutyric acid, 500mM NH_4OH (5:3, v/v)) and left overnight until the solvent was 2cm away from the top of the TLC plate. The plate was then removed, allowed to air dry, and subsequently placed in a second chromatography tank containing 200-300ml of the second-dimension buffer (isopropanol, concentrated HCl (specific gravity 1.18), H_2O (70:15:15, v/v/v)). The dried TLC plate was rotated 90° to the direction of the first dimension and placed in the second-dimension chromatography tank, left overnight until the solvent front was 2cm away from the top of the TLC plate. The TLC plate was removed, allowed to air dry, and heat-sealed in plastic film. The plates were then placed under a phosphor screen (Fuji-Screen) in an X-ray cassette, and the phosphor screens were imaged using a Molecular Imager FX Phospho scanner (BIO-RAD) in combination with QuantityOne software using the manufacturer's default settings for ^{32}P.

2.2.26.4 Analysis of TLC screen

The imaged phosphor screens were analysed with QuantityOne (BIO-RAD) software. The software detects the saturation of the pixels in each area. A ratio was created between the saturation of the A spot and the m^6A spot. To normalise the ratio a background spot was created on the corner of the TLC plate.

2.2.27 DNA restriction digestion

Plasmid DNA was digested using restriction endonucleases, following the manufacturer's instructions. In a reaction, 50 µl total volume containing 1 µg of plasmid DNA was incubated at 37°C for 1 hour in the presence of the cutting buffer. The reaction was heat-inactivated at 85°C for 20 minutes. To prevent re-ligation of cut plasmid DNA, 5' ends were sometimes dephosphorylated using Quick CIP (New England Biolabs). Quick CIP (2.5 µl) was added to the heat-inactivated digestion reaction and incubated at 37°C for 20 minutes, then heat-inactivated at 80°C for 20 minutes. Digested plasmid DNA products were analysed by

electrophoresis on a 1% agarose gel, and if required for cloning, bands were excised from the gel and purified.

Chapter 3

Identification and Characterisation of *fip37-4* Suppressor Mutants

3 Identification and characterisation of *fip37-4* suppressor mutants

3.1 Chapter Aims

The aim of this chapter is to identify *fip37-4* suppressor mutants and possible new molecular players involved in m^{6}A methylation to explain the suppression of *fip37-4* phenotype. The rationale for performing the screen and subsequent Next-Generation Sequencing (NGS) is to generate a list of candidate genes that may facilitate the identification of new proteins and potentially uncover novel mechanisms associated with m^{6}A methylation. The screen will also provide an opportunity to obtain multiple independent suppressor lines, which could uncover multiple novel mechanisms related to m^{6}A modification. This chapter will explore low methylation line *fip37-4* and various suppressor mutants with the aim to:

- Identify *fip37-4* suppressor mutants that are segregating in the F_2 generation.
- Characterise the suppressor mutants by phenotyping and employing various other methods such as m^{6}A measurements, northern blotting, root architecture and trichome branching measurements.
- Performed bulk segregant analysis (BSA) coupled with next generation sequencing (NGS) to produce candidate list of genes.

3.2 Introduction

Previous research has demonstrated the crucial role of *A. thaliana* m^{6}A writer complex in the methylation of transcripts during plant development. The requirement of adenosine methylase A (MTA) for m^{6}A modification, and its interaction with FIP37 was first shown by Zhong *et al.*, (2008). KO mutants of core writer complex proteins, including MTA, MTB, FIP37, and VIRLIZER, result in embryonic lethality, presumably due to their absolute requirement for m^{6}A formation and to the essential nature of at least a basal level of m^{6}A. Another core protein

associated with the writer complex was identified by Růžička *et al.*, (2017) termed HAKAI however, knockout of this protein is not embryo lethal, gave a ~35% reduction in m^6A levels, and displayed similar phenotype to WT. Recent findings have revealed that two additional proteins are part of the methyltransferase complex (HIZ1 and HIZ2), and one of these (HIZ2) is required for full levels of m^6A methylation in *A. thaliana* (Zhang *et al.*, 2022). This discovery prompts further investigation into potential additional molecular components involved in the methyltransferase complex or in m^6A methylation. The dynamic nature of m^6A modification, including the processes of reading, writing, and erasing m^6A methylation, highlights the complexity of this gene regulation mechanism. To identify novel molecular components, a forward genetics screen was employed. This approach aimed to isolate mutants that suppress the *fip37-4* phenotype, effectively reverting it back to the WT phenotype. The suppression of *fip37-4* may uncover novel regulators of FIP37, or other factors implicated in m^6A methylation.

Upon starting this work, Professor Rupert Fray, Dr Helin Tan and colleagues performed a suppressor mutant screen on a hypomorphic FIP37 mutant *fip37-4* where they identified mutants which revert phenotypes of *fip37-4* back to WT. EMS mutagenesis was performed on approximately 5000 *fip37-4* seeds using an adapted version of Rédei & Li, (1969). These seeds were germinated (M1 generation), seeds collected from pools of ~5 plants, and those seeds germinated (M2 generation) and plants were screened for suppressor characteristics associated with *fip37-4* phenotype, including serrated leaves, short plant height compared to WT and increased trichome branching. Any plants showing suppression of the *fip37-4* phenotype in the M1 generation (dominant suppressors) were also identified and seed collected. Three EMS mutants from three independent seed pools were obtained for confirmation and identification of suppressor mutants.

3.3 Results

3.3.1 Characterisation of *fip37-4* suppressor mutants

The identification of unknown molecular players involved in m^6A methylation and in direct association to FIP37 or other writer complex proteins was conducted using EMS mutagenesis on *fip37-4* (SALK_018636) seeds. The *fip37-4* phenotype is characterised by a loss of apical dominance, stunted growth such as reduced plant height, rosette size, and increased in shoot density as shown in Figure 3.3B.

Due to embryonic lethality in complete FIP37 KO mutants, the knock-down version of FIP37 was required. The T-DNA insertion line, *fip37-4* was chosen for this study as the T-DNA insertion is located in the intron between exons 7 and 8 of *FIP37*. *T*-DNA insertion into an intron can still result in the production of WT transcripts although at reduced efficiency and lower transcript levels being produced therefore resulting in a hypomorphic mutant rather than a complete knockout. Intron containing T-DNA sequence can be spliced out, producing functional transcripts as demonstrated in insertion mutants *AGL104, ASP2* (Miesak & Coruzzi, 2002; Verelst *et al.*, 2007). Previous studies have shown that T-DNA in introns can be viable in genes where homozygous insertions in coding sequence gives rise to lethal mutations (Sato *et al.*, 2005). The three EMS suppressor lines were designated P1-14, P10-23, and P3-07, coming from M1 seed pools 1, 10 and 3 respectively.

3.3.2 Screening of P1-14, P10-23, and P3-07

The term suppressor mutant in this chapter is referred to as a reversion of *fip37-4* phenotype back to WT. The F_2 generation of each of the respective pool of seed were sowed on compost and grown in a controlled Conviron growth chamber to ensure uniform growth conditions for accurate phenotypic evaluation. Unfortunately, P1-14 line showed poor germination rates and was thus discarded from further analysis. BSA previously identified P10-23 F_2 segregating line as a dominant mutation line. P10-23 was subsequently backcrossed to *fip37-4* and the F_1 plant was allowed to self-pollinate, generating a second segregating F_2 population (Figure

3.1). This backcrossing to *fip37-4* was repeated and produced P10-23 BC3 F_2 which was used in this chapter. Repeated backcrossing sequentially enabled background mutations to be filtered out allowing for less candidate genes to analyse. P3-07 was received as the original F_2 segregating population. This line was backcrossed to *fip37-4* twice to produce P3-07 BC2 F_2 (Figure 3.1).

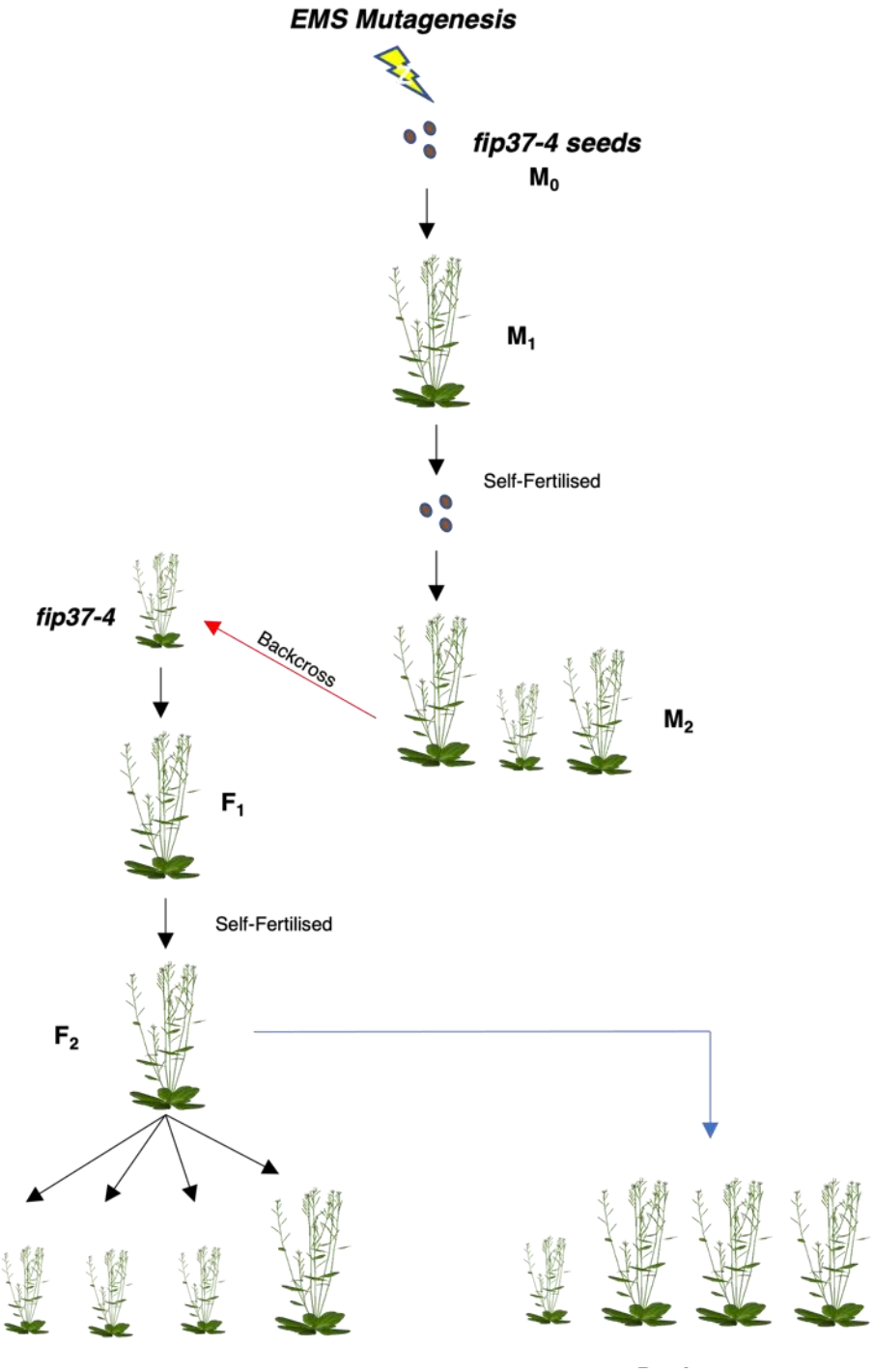


Figure 3.1 Schematic representation of EMS mutagenesis of *fip37-4*

Schematic diagram showing the generation of suppressor mutants using EMS mutagenesis on *fip37-4*. The resulting M_2 generation was backcrossed to *fip37-4* and the F_2 segregating generation was produced and used in this chapter.

3.3.3 Height variability and segregation ratio of suppressor line P3-07

To understand the variation in plant height within the segregating population, measurements were taken on all plants from P3-07 F_2 BC1 and P3-07 F_2 BC2. However, no measurements were conducted on the P10-23 line due to its dominant nature of mutation displaying a segregation ratio of 1:3 for short to tall plants. Analysis of plant height variability in P3-07 F_2 BC1 revealed clear segregation of short, medium, and tall plants across all individuals, as shown in Figure 3.2A. This segregation pattern, characterised by a range of height segregation, suggests that P3-07 may exhibit a semi-dominance mutation, with the heterozygous plant displaying some suppression of *fip37-4*.

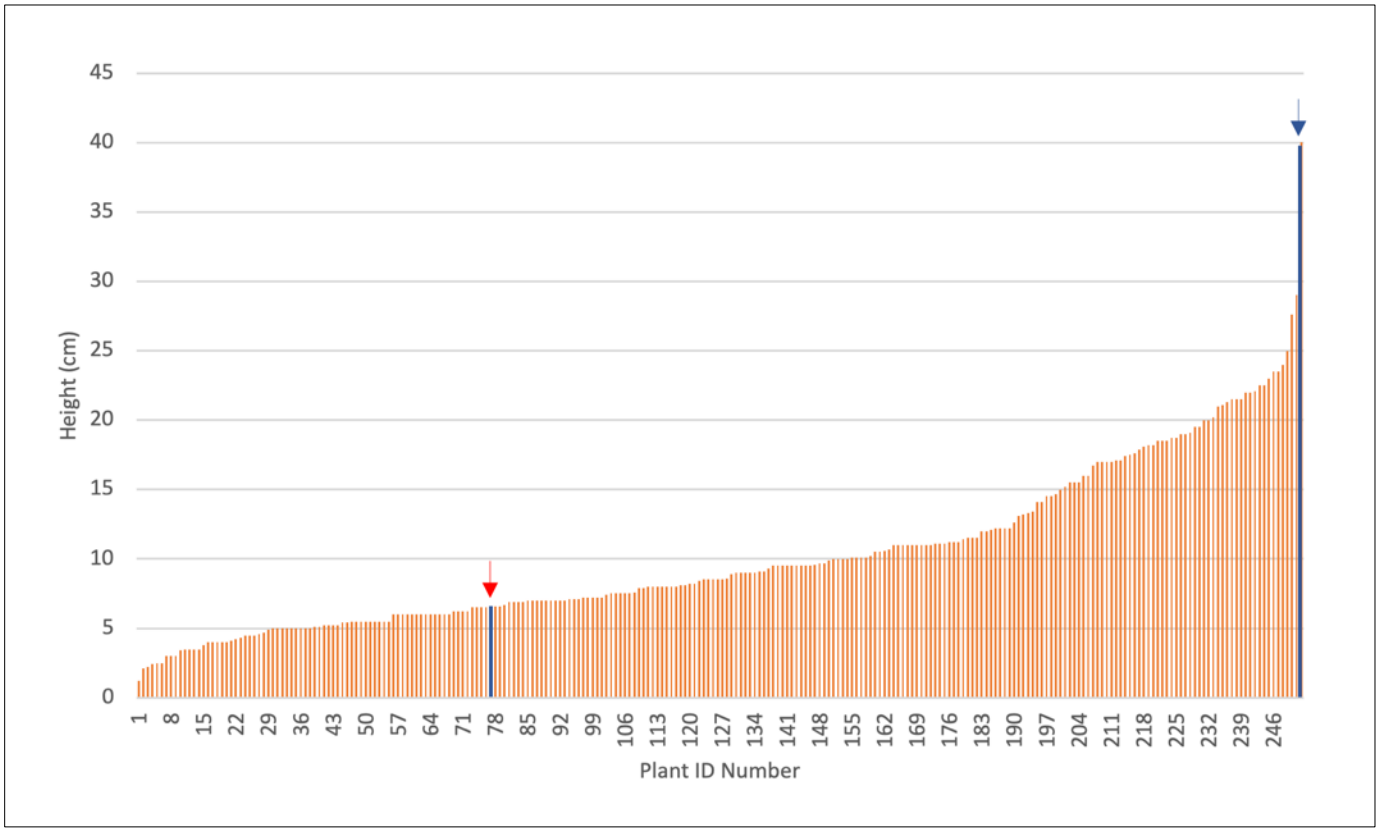
The measurement of plant height was also conducted for P3-07 F_2 BC2 on all plants. The results indicated that shorter plants in the population have a minor increase in plant height when compared to P3-07 F_2 BC1, (Figure 3.2B). Despite this increase in height, the tall plant population remained relatively unchanged, suggesting that this segregation of height variability could be indicative of a semi-dominant mutation.

3.3.4 Phenotypic observations of suppressor line P10-23 and P3-07

The *fip37-4* is a hypomorphic mutant, exhibiting reduced transcript abundance of *FIP37.* This results in developmental defects such as loss of apical dominance, serrated leaves, and increased trichome branching, as shown in Figure 3.3B. To characterise the suppression of *fip37-4* phenotype, observations involving plant height, leaf shape, and overall plant structure were conducted on P10-23 and P3-07. The phenotype shown in Figure 3.3C, shows P3-07 line and the plants referred to as “Short 1.0,1.1,” displayed no suppression of *fip37-4* phenotype with severely reduced plant height, short stems, and loss of apical dominance. In contrast the phenotype shown in Figure 3.3A, referred to as “Tall 2.0,” exhibited clear suppression of *fip37-4* phenotype, with plant height and rosette leaf shape resembling a WT

phenotype. The P10-23 line also exhibited a suppression of fip37-4 as the plant, “Tall 2.1,” showed clear suppression displaying a height similar to WT (Figure 3.5).

(A)



(B)

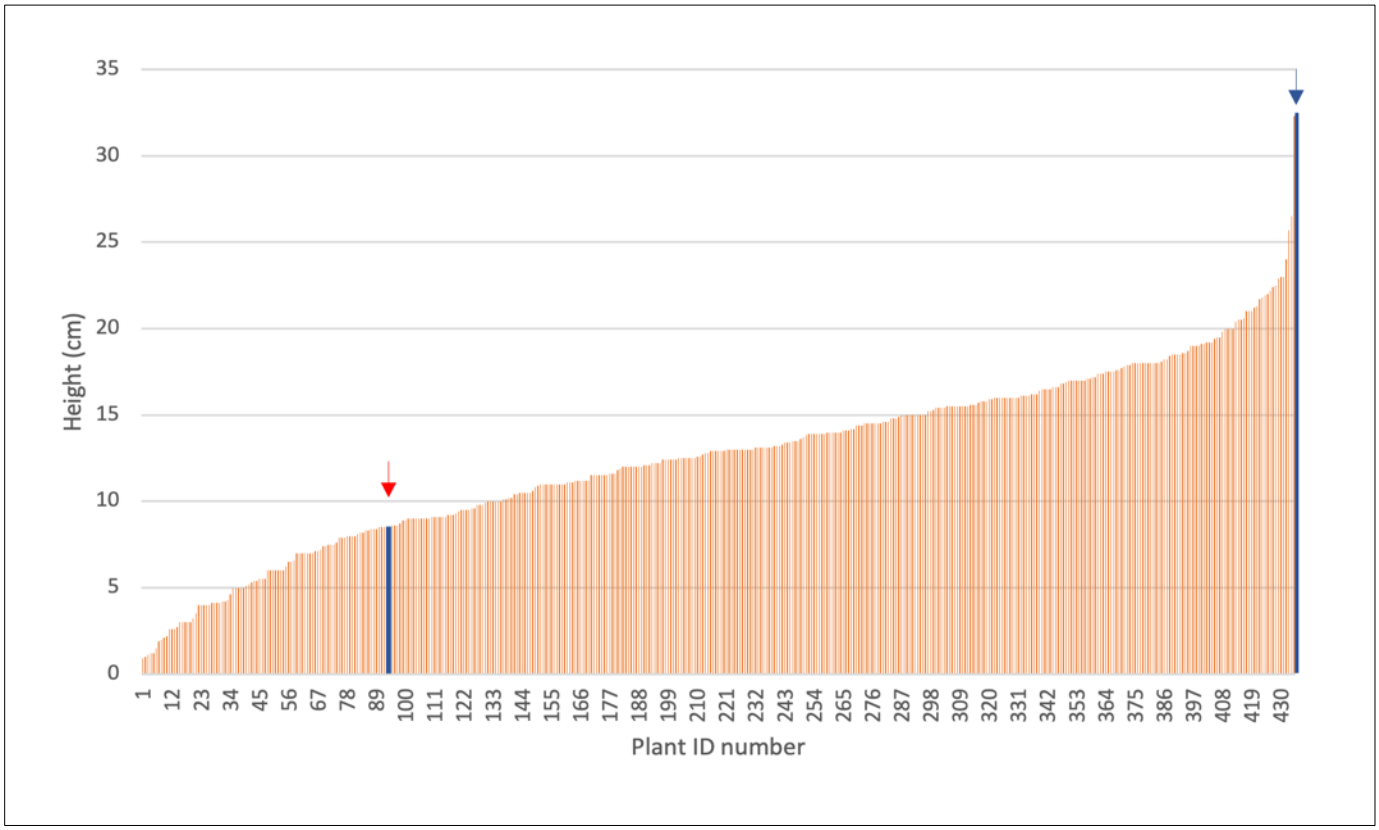


Figure 3.2 P3-07 representation of plant height variability across two backcrosses

(A) Height variability of P3-07 BC1 showing differences in height phenotype in backcrossed number 1. (B) Plants are P3-07 BC2 which was backcrossed to *fip37-4* and self-fertilised to produce segregating F2 line. Red arrow shows *fip37-4* average height while blue arrow shows WT average height. In (A), for the WT group, the mean and SD are 39.85 cm and 1.78 cm, respectively, while for the *fip37-4* group, the mean and SD are 5.98 cm and 3.01 cm, respectively. In (B), for the WT group, the mean and SD are 33.30 cm and 2.96 cm, respectively, while for the fip37-4 group, the mean and SD are 9.04 cm and 3.84 cm, respectively. Plants are showing differences in height variability. Plants were sampled at 5-week-old grown on compost in Conviron growth cabinet with 16-hour day length n = 260 – 445.

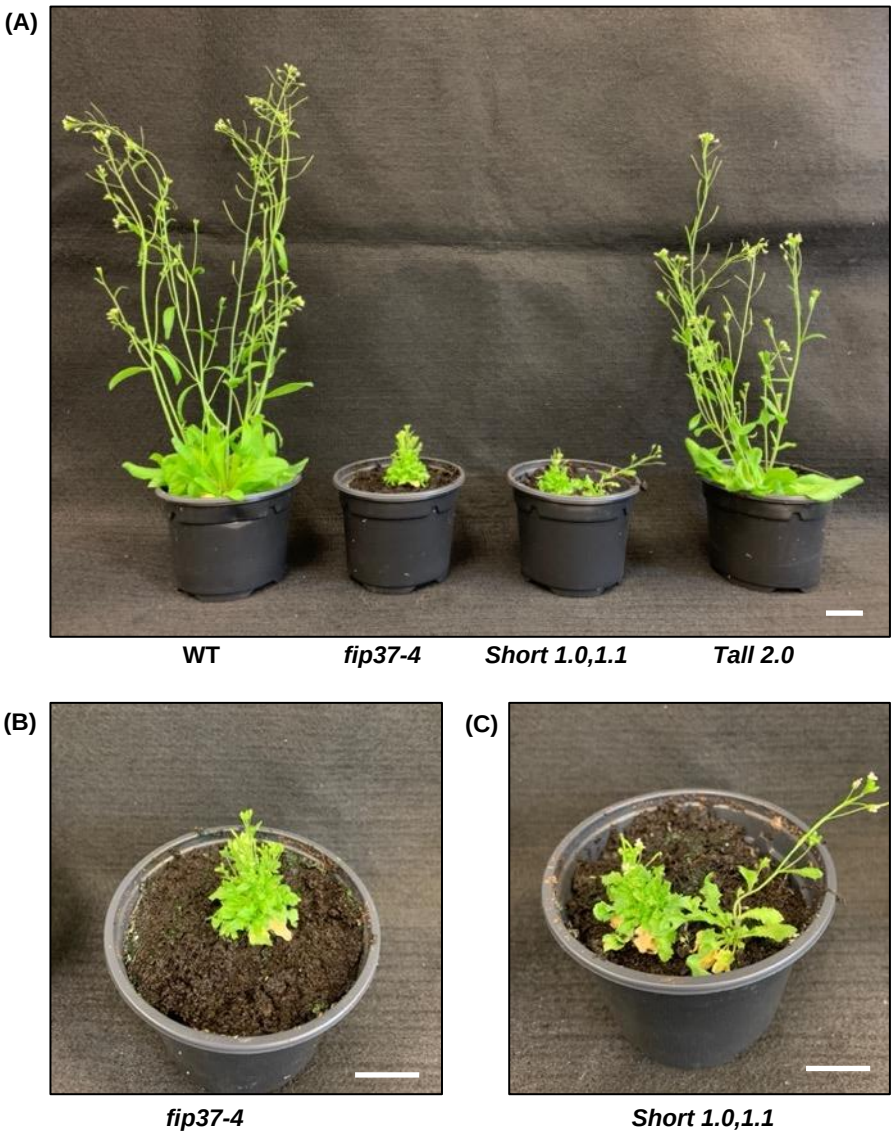


Figure 3.3 Phenotype of P3-07 F_2 BC1 segregating line showing rescuing of *fip37-4* to WT like phenotype.

(A) Plant line going from left to right WT, *fip37-4*, P3-07 F_2 BC1 short 1.0,1.1 and tall 2.0. Suppression of phenotype of *fip37-4* for Tall 2.0. (B) Close up of *fip37-4* phenotype showing developmental defects. (C) Close up phenotype of two short plants 1.0, 1.1 showing similar phenotype to *fip37-4*. Plants are 5-week-old grown on compost in Conviron growth room with 16-hour day length. Scale bar for (A) = 3cm, (B) / (C) = 1.7 cm.

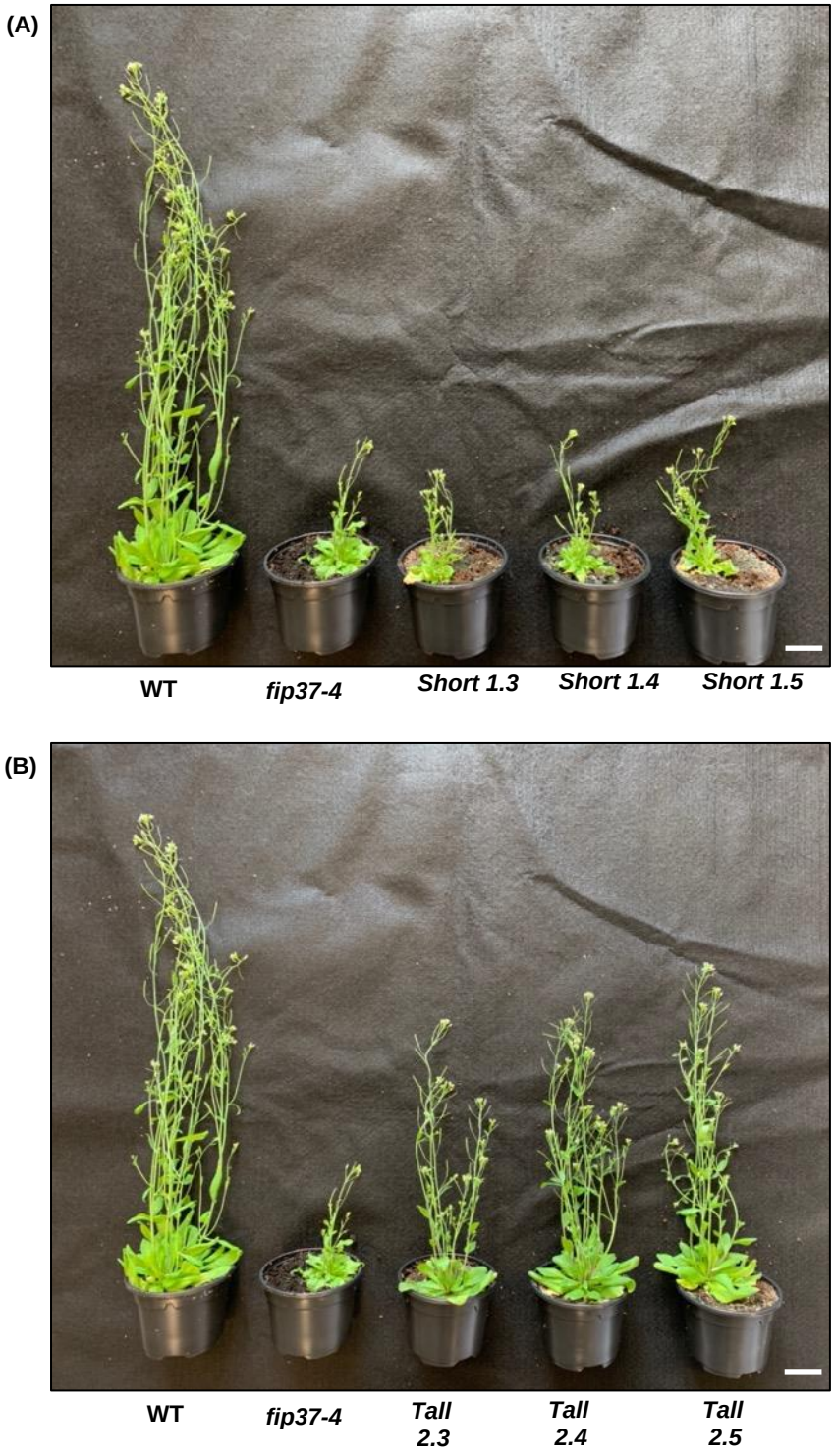


Figure 3.4 Phenotype of P3-07 F_2 BC2 showing suppressor line rescuing *fip37-4* to WT like phenotype.

(A) Phenotype of P3-07 F_2 BC2 showing three short plants displaying severe developmental defects. (B) Phenotype of P3-07 F_2 BC2 showing clear suppression of *fip37-4* phenotype in 3 tall plants. Plants are 5-week-old grown on compost in Conviron growth room with 16-hour day length. Scale bar represents 3.5cm.



Figure 3.5 Suppressor line P10-23 shows clear reversion back to WT-like phenotype.

(A) Phenotype of P10-23 F_2 BC2 showing WT, *fip37-4*, Short 1.1, and Tall 2.1 short plant displaying severe developmental defects and suppression of *fip37-4* phenotype. Plants are 5-week-old grown on compost in Conviron growth room with 16-hour day length. Scale bar represents 7cm.

3.3.5 Root development of suppressor lines in normal and abiotic stress growth medium

In order to assess the impact of suppressor mutants on root architecture, root phenotype observations were conducted on vertical growth media plates. Previous research has shown that hypomorphic mutants of m^6A writer display aberrant root phenotypes associated with root cap formation, gravity response, main and lateral root development, and auxin response defects (Růžička *et al.*, 2017; Bhat *et al.*, 2020; Zhang *et al.*, 2022). The suppression of *fip37-4* phenotype in P10-23 and P3-07 could add further evidence to the possibility that candidate genes called from BSA are influencing root developmental response.

The phenotypic differences across plant height and leaf shape are evident, however, the impact on root formation in the suppressor mutants is also notable. Measurements were taken on the main and lateral roots under three different growth conditions. Normal growth conditions were with 1/2 MS media, while abiotic stress conditions, such as high salinity and drought, were simulated by adding NaCl (75mM) and Sorbitol (150mM), respectively, to the ½ MS media as described in the methods. Investigating the role of m^6A methylation in environmental stress tolerance is fundamental, as previous studies have demonstrated its significance in salt and drought stress response (Hu *et al.*, 2021; Hou *et al.*, 2022). The observation of root development in suppressor lines P10-23 and P3-07 under conditions of high salinity and drought will provide insights into whether suppression of *fip37-4* is maintained under abiotic stress conditions.

3.3.5.1 ½ MS growth conditions

The results indicate a significant reduction in the primary root length (PRL) of *fip37-4* as compared to the WT control, with a 40% decrease observed. However, in the case of P3-07 line, a statistically significant increase of 24% in PRL compared to *fip37-4* was observed, resulting in some suppression of phenotype, but still significantly less than that of the WT. The

lateral root density (LRD) for this growth medium showed that *fip37-4* was significantly lower compared to the WT control, with a 69% decrease observed. In contrast, P3-07 compared to *fip37-4* showed an increase of 56%, suggesting a slight reversion to the WT LRD. For line P10-23, the PRL and LRD analyses demonstrated a complete reversion back to the WT phenotype, as seen by the statistically significant differences compared to *fip37-4* as well. This strongly suggest a suppression of the phenotype for P10-23.

3.3.5.2 ½ MS + NaCl growth conditions

The results demonstrate a significant decrease in PRL in *fip37-4* compared to the WT, with a 40% decrease observed. However, compared to *fip37-4*, P3-07 showed a significant increase in PRL of 37%, indicating a slight reversion back towards the WT phenotype. The LRD did not display a significant difference across the WT, *fip37-4*, and P3-07. For line P10-23, the PRL did not show full reversion back to WT but similar length to *fip37-4* when compared to *fip37-4*. However, the LRD did show a full reversion back to WT, indicating strong suppression of phenotype for this trait in P10-23.

3.3.5.3 ½ MS + Sorbitol growth conditions

Under sorbitol stress conditions, a significant reduction in PRL of 51% was observed in *fip37-4* compared to WT phenotype. However, the suppressor line P3-07 showed a significant increase in PRL compared to *fip37-4*, resulting in a full reversion to a root phenotype similar to WT. P10-23 also showed a full reversion back to WT in PRL but not for LRD.

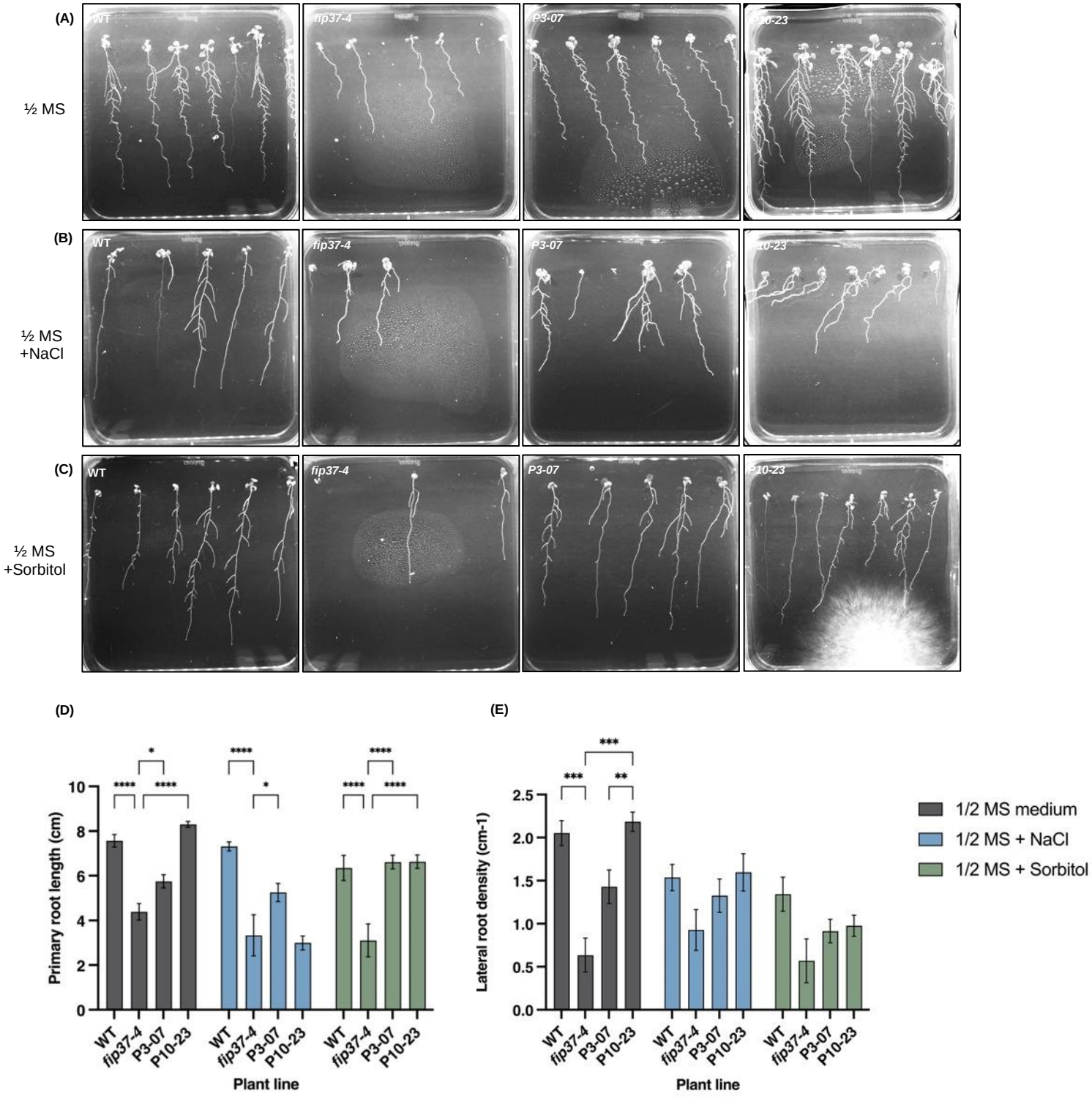


Figure 3.6 Root architecture of WT, *fip37-4* and suppressor mutants in normal and abiotic stress conditions.

(A) to (C) showing WT, *fip37-4*, and suppressor mutant P3-07 and P10-23 in ½ MS media, ½ MS + NaCl (75mM), and ½ MS + Sorbitol (150mM) respectively from top to bottom. (D) Bar charts showing statistical difference in primary root length and lateral root density for three growth conditions. Plants were sample at 14 day-old grown on vertical plates in a 16-hour day length. Error bars represent SEM; n = 12-20. * Indicates significant difference, * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$, **** $P < 0.0001$ using two-way ANOVA with Tukey's multiple comparisons test. Scale bar = 1cm

3.3.6 Suppressor lines revert trichome branching and lateral root initiation back to WT-like

In order to test if the *fip37-4* suppressor mutants P3-07 and P10-23 revert trichome branching number back to the WT phenotype, trichome branching counts were measured on the first rosette leaf pair. In WT, trichome branching predominantly consists of three and four branches, as shown in Figure 3.7A+C. Compared to WT, *fip37-4* exhibits a change in trichome branching, with a majority of four (60%) and five (25%) branches and occasionally six (4%). P3-07 shows a potential for reversion back to WT, with 40-50% of trichome branches ranging from three to four, which is a significant decrease in trichome branch number compared to *fip37-4*, where only 11% of trichome branches have three branches. The P10-23 line exhibited closer reversion back to WT-like trichome branching, with 58% of trichome branches having 3 branches, approximately 37% difference compared to P3-07. The P10-23 line also showed a close reversion of 4 trichome branch percentage, having the same as WT but less than *fip37-4*.

To investigate if *fip37-4*, P3-07, and P10-23 are responsive to auxin, lateral root initiation was conducted by transferring four-day-old seedlings grown on ½ MS to a medium containing N-1-naphthylphthalamic acid (NPA, 10μM) for one day, followed by transfer to a medium containing 1-naphthaleneacetic acid (NAA, 10μM) for five days. To minimise the impact of root hairs on lateral root imaging, roots were briefly soaked in 100% ethanol for five seconds. The results indicate that *fip37-4* exhibits reduced lateral root hair and density compared to WT. However, both P3-07 and P10-23 exhibit clear suppression as indicated by increased lateral root length and density relative to *fip37-4*.

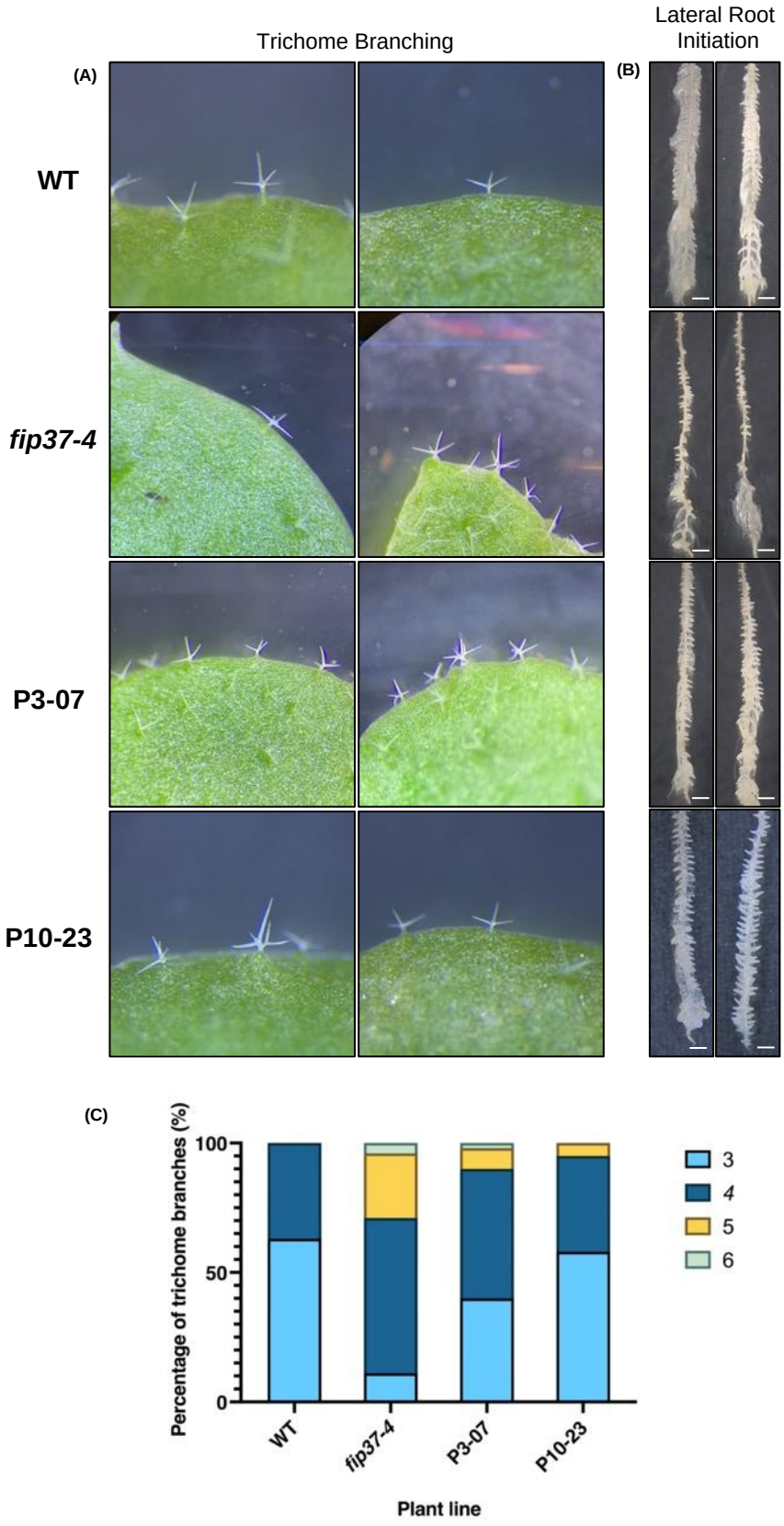


Figure 3.7 Trichome branching and lateral root initiation.

(A) Trichome branching of WT, *fip37-4*, P3-07, and P10-23 from top to bottom showing increased trichome branching in *fip37-4* and suppressor lines. (B) Roots under Stereo Dissecting Microscope after treatment with NPA, (10 μM) and then transferred onto 1/2MS with NAA (10 μM) for 5 days. Scale bar = 500μm. (C) Percentage of trichome branches on the first pair of rosette leaves from 2-week-old plants on compost. Twenty leaves were used for counting of trichome branching in each line.

3.3.7 Measurement of m^{6}A methylation

To determine if restoration of m^{6}A levels has occurred in the suppressor mutant line, the two-dimensional thin layer chromatography (TLC) method, adapted from Keith, (1995), was utilised to detect, and quantify m^{6}A levels. Thin layer chromatography is a method that is capable of identifying modified nucleotides present in RNA (Keith, 1995). RNA with methylated and non-methylated nucleotides are able to separate in two dimensions whereby the nucleotides disperse across the cellulose substrate according to their charge and hydrophobicity. Advances in isolation of RNA have enabled this method to detect m^{6}A modification in a sensitive, quantitively, and reliably reproducible fashion. The results from the TLC analysis indicated a significant reduction of 78% in m^{6}A levels in the *fip37-4* line compared to the WT, reflecting the successful knockdown of m^{6}A methylation in the hypomorph (Figure 3.8F). However, no statistically significant increase in m^{6}A levels was observed in the P3-07 line compared to *fip37-4* with only a slight increase in methylation levels. In contrast, a significant increase in m^{6}A levels was found in P10-23 compared to *fip37-4*, and demonstrating only a 28% difference in m^{6}A levels compared to WT, indicating a strong reversion to WT-like m^{6}A levels (Figure 3.8F).

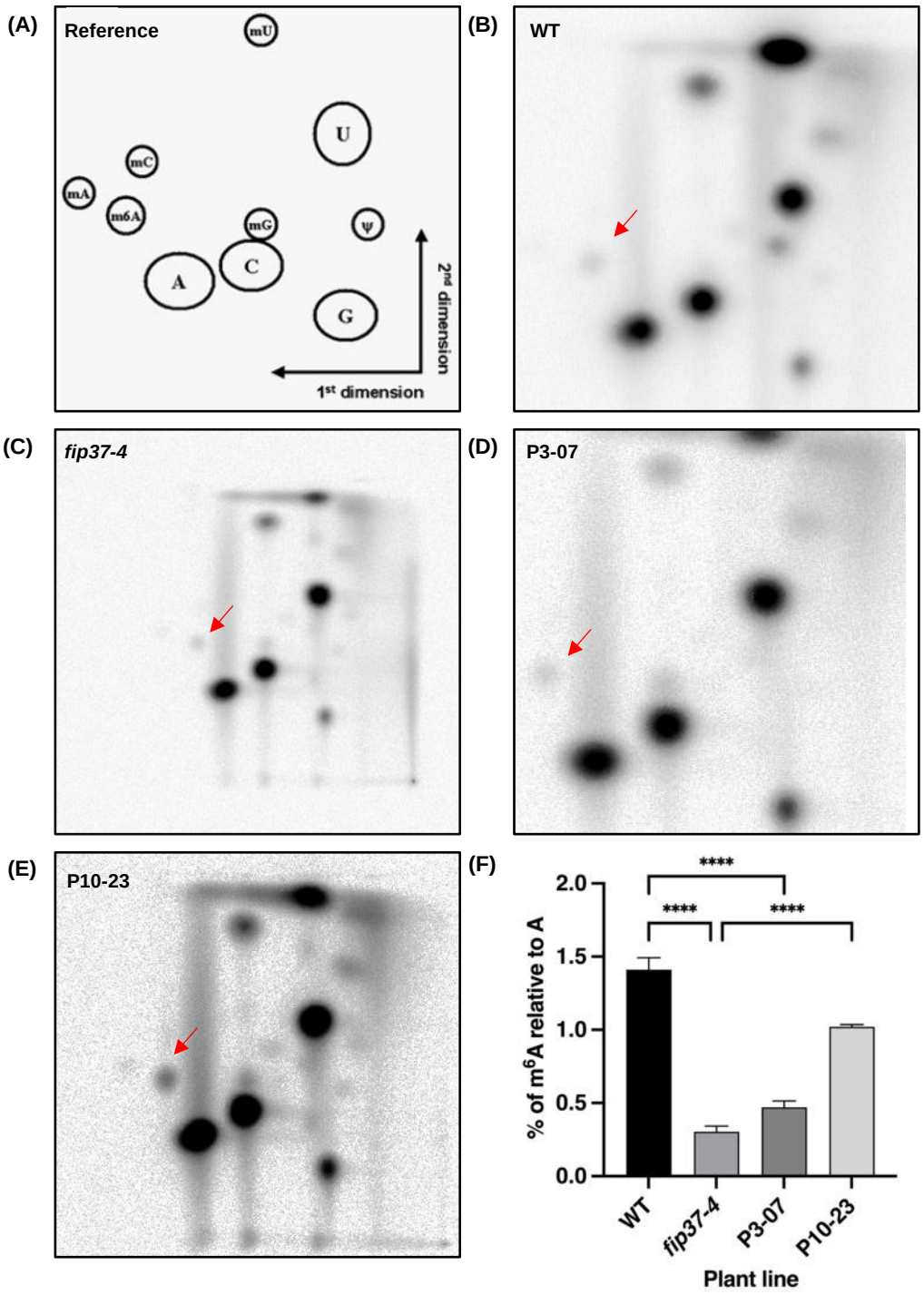


Figure 3.8 Thin layer chromatography indicating m^{6}A level in *fip37-4* and suppressor mutants.

(A) Schematic diagram of 1st and 2nd dimension TLC indicating the relative position of m^{6}A and non-modified nucleotides on a TLC plate. (B) to (E) shows images of TLC with labelled m^{6}A in their respective line labelled at top left corner. (F) Column graph displaying quantification of m^{6}A levels by mean percentage of m^{6}A relative to A nucleotide ratio. Three independent biological repeats measuring TLC ratios were conducted from 2-week-old seedlings grown on MS plates. Error bars represent SEM and **** shows significant difference $P < 0.0001$ by ANOVA. Red arrow on TLC plates indicates location of m^{6}A nucleotide.

3.3.8 Northern blotting shows increase in transcript levels in suppressor mutants

In order to analyse transcript abundance and possible splicing effects of *fip37-4* and the suppressor mutants Northern blotting was carried out to determine expression level of *FIP37*. Northern blot provides a powerful tool to visualise transcript abundance performed using a denaturing RNA gel which separates RNA by size prior to transferring and fixing the RNA to a nylon membrane. Several washes of the membrane with a radiolabelled probe complementary to the gene of interest (FIP37) hybridises to the fixed RNA and the non-specific bound probes are washed away. The *fip37-4* Salk line is a T-DNA insertion mutant in which the T-DNA is located between the seventh and eighth intron (Figure 3.9B). The T-DNA insertion results in the production of two types of transcripts: a WT version of *FIP37* and mutated transcripts containing the T-DNA insertion. The WT version is generated through the splicing out of the intron along with the inserted T-DNA sequence (Miesak & Coruzzi, 2002; Verelst *et al.*, 2007). However, not all intron-containing T-DNA may be spliced out, resulting in the production of a truncated, non-functional protein. As the T-DNA is inserted within the intron, northern blot analysis can determine if the suppressor mutant is linked to changes in splicing and expression.

Northern blot analysis was conducted to assess the expression levels of both correctly spliced T-DNA containing *FIP37* transcripts and transcripts containing the T-DNA in the mutant lines *fip37-4*, P3-07, and P10-23 in comparison to WT. The results showed a reduction in correctly spliced *FIP37* transcript levels in the *fip37-4* line, accompanied by elevated levels of *FIP37* transcripts containing T-DNA as expected (Figure 3.9A). The P10-23 line did display a significant increase in the expression of the correctly spliced *FIP37* transcript compared to *fip37-4*, while still showing increased levels of mutant *FIP37* transcripts (Figure 3.9A). The P3-07 line showed a slight reduction in mutant *FIP37* transcript levels compared to *fip37-4*, but a slight increase in the correctly spliced *FIP37* transcript compared to *fip37-4* (Figure 3.9A). The results of P10-23 and P3-07 differ significantly from each other, with P10-23 exhibiting higher

levels of correctly spliced *FIP37* transcript compared to P3-07 (Figure 3.9A). These results suggest that there may be increased splicing efficiency in the production of correctly spliced *FIP37* transcript in both P10-23 and P3-07 lines.

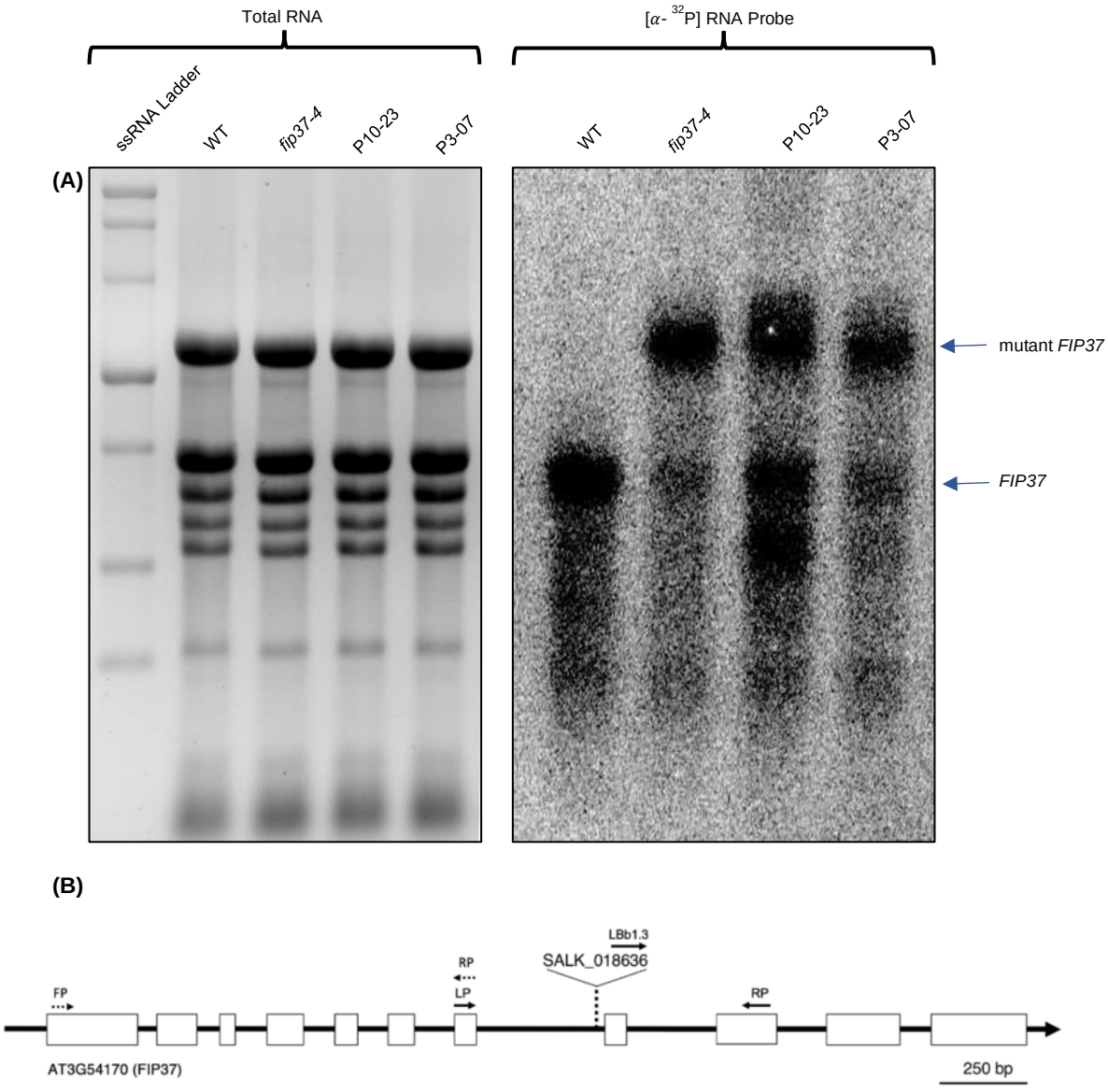


Figure 3.9 Northern blotting to show transcript abundance of mutant *FIP37* and WT *FIP37*.

(A) Northern blot on Total RNA from WT, *fip37-4*, P10-23 and P3-07 with ssRNA ladder displaying the bands corresponding to 3000, 2000, 1000 and 500bp. Transcripts were visualised using [α32P]-dCTP-labelled RNA probe. (B) Schematic diagram of *FIP37* showing T-DNA insertion labelled SALK_018636. Primers used for amplifying RNA probe from cDNA is shown by dotted black line on exon one and seven. Total RNA extracted from 14-day old seedlings from rosette leaves grown on MS plates in Conviron growth room.

3.3.9 Bulked segregant analysis and Next Generation Sequencing produces candidate genes

To further understand the cause of the suppressor mutant reversion of phenotype back to WT, BSA coupled with NGS was done on both suppressor mutant lines. BSA provides an approach for rapidly identifying a causal gene linked to a phenotype (Giovannoni *et al.*, 1991; Michelmore *et al.*, 1991). Identifying a suppressor phenotype starts with the construction of a segregating population with two bulked DNA samples that are made with progenies with contrasting phenotype.

Two bulked populations of contrasting phenotype, with one exhibiting short stature similar to the *fip37-4* and the other exhibiting tall stature similar to the WT, were subjected to DNA extraction. NGS was utilised to map the causal DNA changes, where two pools of DNA were sequenced from each line. The initial BSA was performed on the plant line P3-07 BC2 F_2 grown in a controlled Conviron growth room using compost as a growth medium. The segregating line was screened and tall plants with a WT-like phenotype displaying clear signs of suppression were selected, as shown in Figure 3.4. A total of 25 tall WT-like plants were chosen. Similarly, 25 short (*fip37-4*-like) plants exhibiting similar phenotypic characteristics as depicted in Figure 3.3C were selected. For line P10-23 BC3 F_2, contrasting short (n=28) and tall (n=25) plants were chosen. DNA extraction was performed on all four pools of plants using the Qiagen Plant DNA pro kit. Equal amounts of DNA were pooled and send to Novogene for standard Illumina barcoded DNA library preparation. The preparation was done on both DNA samples with a sequence depth of 25 million reads with paired end read length of 150bp on Illumina NovaSeq 6000 sequencer. The pipeline for processing of raw data from NGS was described in Chapter 2.

3.3.10 P3-07 filtering step 1 of candidate genes

Analysis of BSA in the P3-07 line revealed the presence of 7110 heterozygous (HET) and homozygous (HOM) genes exclusively in short DNA pool (Figure 3.10A). Further to this, 15727 HET and HOM genes were found in both short and tall DNA pool, while 7953 HET and HOM genes were uniquely detected in tall DNA pool (Figure 3.10A). Out of the 7953 genes found in tall DNA pool, 91 were HOM only in tall DNA pool (Table 3.1). Based on the P3-07 analysis of results, height variations and phenotypic observations indicate the presence of a semi-dominant mutation. To further investigate this, BSA analysis was conducted where short and tall plants were separated, and their DNA was extracted and sequenced. Filtering step 1 involved selecting mutations that were uniquely found in the tall DNA pool and assuming that the tall plants chosen for DNA extraction were HOM (Tallest plant) for the mutation and the shortest plant chosen did not have the mutation. Further filtering identified mutations that were unique and HOM in the tall DNA pool, resulting in a list of 91 candidate genes (Table 3.1).

3.3.11 P10-23 filtering of genes to produce candidate list

Upon analysis, P10-23 displayed a dominant mutation as evidenced by the 1:3 short to tall segregation ratio. The first filtering step identified 3864 genes that were unique to the tall DNA pool, consisting of HOM and HET mutations (Figure 3.10C). Subsequently, the list was filtered further by assessing their variance consequence using the Ensembl Variant Effect Predictor (VEP) on NGS sequencing data processing. The high impact category was selected resulting in a list of 55 unique genes in the tall plant DNA pool (Table 3.2).

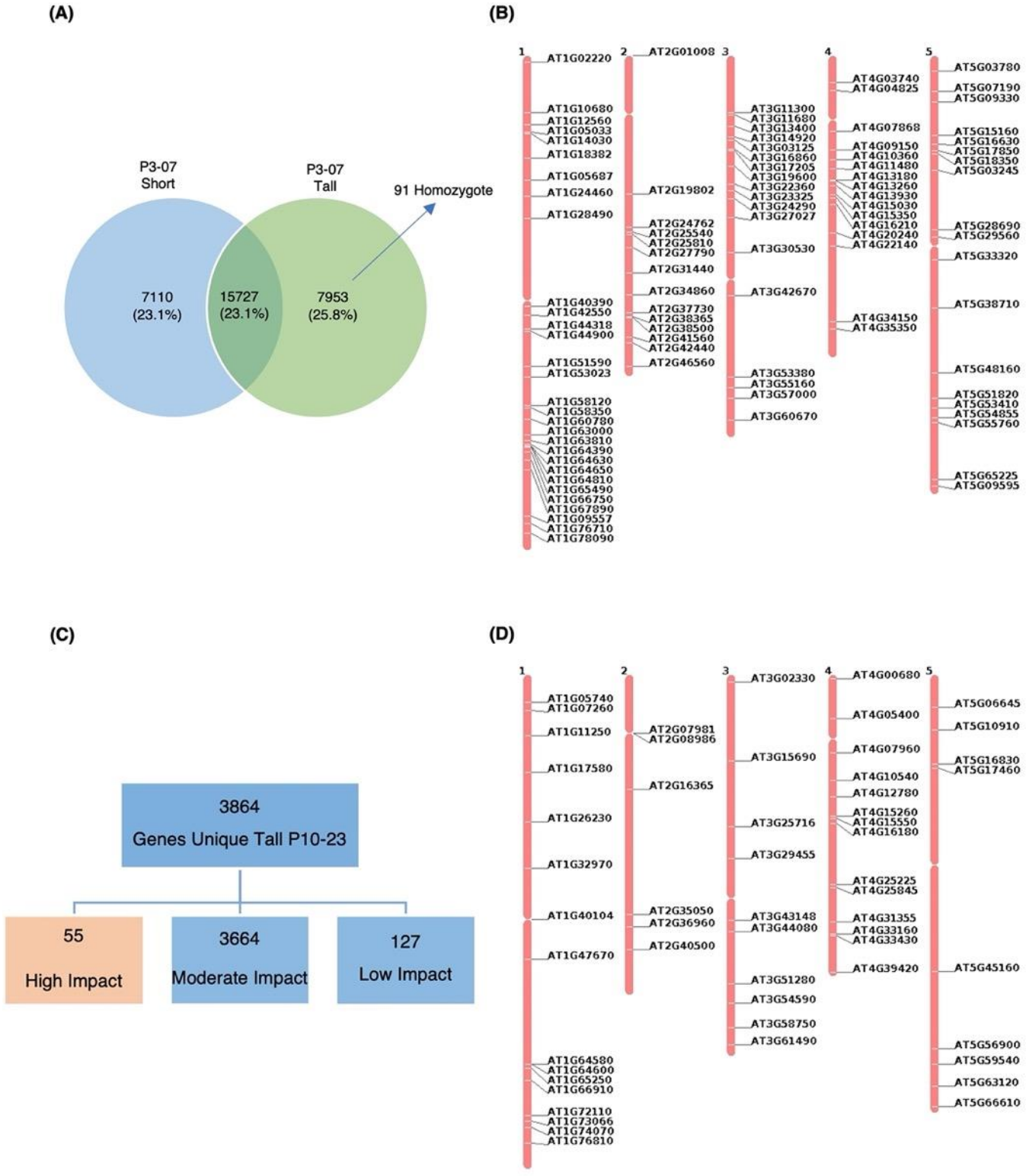


Figure 3.10 Schematic representation of BSA genes in suppressor line P3-07 and P10-23.

(A) Venn diagram of shared and unique genes in BSA and NGS for P3-07. (B) Chromosome mapping of 91 homozygote genes in P3-07. (C) Flow chart showing three impact list of candidate gene in suppressor line P10-23. (D) Chromosome mapping of 55 high impact gene from P10-23.

Table 3.1 List of candidate genes for P3-07 suppressor mutant showing high impact variants.

Gene	Gene description	Consequence	Function
AT1G02220	NAC domain transcription factor which functions as a negative regulator of the TDIF-PXY module and fine-tunes TDIF signaling in vascular development.	Upstream gene variant	Protein coding
AT1G05033	Natural antisense transcript overlaps with AT1G13740	Downstream gene variant	Antisense long noncoding RNA
AT1G05687	Natural antisense transcript overlaps with AT1G21880	Downstream gene variant	Antisense long noncoding RNA
AT1G09557	Natural antisense transcript overlaps with AT1G75450	Downstream gene variant	Antisense long noncoding RNA
AT1G10680	P-glycoprotein 10	Upstream gene variant	Protein coding
AT1G12560	Member of Alpha-Expansin Gene Family. Containing a conserved root hair-specific cis-element RHE. Expressed specifically in root hair cell and involved in root hair elongation.	Downstream gene variant	Protein coding
AT1G14030	Encodes a lysine methyltransferase whose main soluble physiological substrates are chloroplastic fructose 1,6-bisphosphate aldolases, FBA1, FBA2, and FBA3. Lysines near the C-terminal end of the target proteins are trimethylated.	Upstream gene variant	Protein coding
AT1G18382	Natural antisense transcript overlaps with AT1G18380	Upstream gene variant	Antisense long noncoding RNA
AT1G24460	Encodes a novel protein that co-immunoprecipitates with SYP41. It is involved in vacuolar trafficking and salt tolerance, potentially via a role in vesicle fusion and in maintaining TGN structure or identity	Downstream gene variant	Protein coding
AT1G28490	Encodes SYP61, one of 24 Arabidopsis syntaxins. Its mRNA has been shown to be expressed. SYP61 and SYP121 coordinate the trafficking of plasma membrane aquaporin PIP2;7 to modulate the cell membrane water permeability.	Downstream gene variant	Protein coding
AT1G40390	DNAse I-like superfamily protein	Upstream gene variant	Protein coding
AT1G42550	Encodes a plant-specific protein of unknown function that appears to be conserved among angiosperms	Downstream gene variant	Protein coding
AT1G44318	Aldolase superfamily protein	Upstream gene variant	Protein coding
AT1G44900	Encodes MCM2 (MINICHROMOSOME MAINTENANCE 2), a protein essential to embryo development. Overexpression results in altered root meristem function.	Downstream gene variant	Protein coding
AT1G51590	Encodes an alpha-mannosidase I enzyme responsible for N-glycan maturation.	Upstream gene variant	Protein coding
AT1G53023	Ubiquitin-conjugating enzyme family protein	Upstream gene variant	Protein coding
AT1G58120	Hypothetical protein	Downstream gene variant	Protein coding
AT1G58350	Putative serine esterase family protein	3_prime_UTR_variant	Protein coding
AT1G60780	adaptin component of heterotetrameric protein complex that regulates protein sorting at the trans-Golgi network/early endosome.	Upstream gene variant	Protein coding
AT1G63000	Nucleotide-rhamnose synthase/epimerase-reductase	Downstream gene variant	Protein coding
AT1G63810	Nucleolar protein	Upstream gene variant	Protein coding
AT1G64390	Glycosyl hydrolase 9C2	Downstream gene variant	Protein coding
AT1G64630	with no lysine (K) kinase 10	Downstream gene variant	Protein coding
AT1G64650	Major facilitator superfamily protein. GOLGI S-ADENOSYL METHIONE TRANSPORTER 1, GOSAMT1	Upstream gene variant	Protein coding
AT1G64810	Encodes a chloroplast localized RNA binding protein that is involved in group II intron splicing. Splicing defects can account for the loss of photosynthetic complexes in apo1 mutants.	Downstream gene variant	Protein coding

AT1G65490	Secreted peptide which functions in plant growth and pathogen defense.	Downstream gene variant	Protein coding
AT1G66750	Encodes a CDK-activating kinase that interacts with SPT5, a regulator of transcription and histone methylation.	Downstream gene variant	Protein coding
AT1G67890	Encodes a protein with similarity to RAF MAP kinases. Based on loss of function phenotype, RAF11 appears to be involved in mediating ABA responses such as dormancy and environmental stress	Downstream gene variant	Protein coding
AT1G76710	SET domain group 26	Upstream gene variant	Protein coding
AT1G78090	homologous to the C-terminal part of microbial trehalose-6-phosphate phosphatases	Upstream gene variant	Protein coding
AT2G01008	Maternal effect embryo arrest protein	Downstream gene variant	Protein coding
AT2G19802	Transmembrane protein	Upstream gene variant	Protein coding
AT2G24762	Encodes a member of the GDU (glutamine dumper) family proteins involved in amino acid export	Upstream gene variant	Protein coding
AT2G25540	Cellulose synthase	Upstream gene variant	Protein coding
AT2G25810	Tonoplast intrinsic protein 4	Downstream gene variant	Protein coding
AT2G27790	RNA-binding (RRM/RBD/RNP motifs) family protein	Upstream gene variant	Protein coding
AT2G31440	Encodes a gamma-secretase subunit. Associates with other subunits in intracellular membrane compartments.	Upstream gene variant	Protein coding
AT2G34860	DnaJ-like zinc finger domain-containing protein which regulates the assembly of photosystem I (PSI) and seed development.	Downstream gene variant	Protein coding
AT2G37730	Glycosyltransferase (DUF604)	Upstream gene variant	Protein coding
AT2G38365	Endonuclease/glycosyl hydrolase	Downstream gene variant	Protein coding
AT2G38500	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	Upstream gene variant	Protein coding
AT2G41560	Encodes a calmodulin-regulated Ca(2+)-ATPase that improves salt tolerance in yeast	5_prime_UTR_variant	Protein coding
AT2G42440	Lateral organ boundaries (LOB) domain family protein	Upstream gene variant	Protein coding
AT2G46560	transducin family protein / WD-40 repeat family protein	Intron Variant	Protein coding
AT3G03125	Natural antisense transcript overlaps with AT3G15518	Downstream gene variant	Antisense long noncoding RNA
AT3G11300	Hypothetical protein	Upstream gene variant	Protein coding
AT3G11680	aluminum activated malate transporter family protein	Upstream gene variant	Protein coding
AT3G13400	Encodes a pollen/pollen tube-specific gene that is essential for pollen tube growth in the transmitting tract by mediating the biosynthesis of JA that modifies the components of pollen tube cell walls	Downstream gene variant	Protein coding
AT3G14920	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A protein	Downstream gene variant	Protein coding
AT3G16860	COBRA-like protein 8 precursor	Upstream gene variant	Protein coding
AT3G17205	Ubiquitin protein ligase 6	Intron Variant	Protein coding
AT3G19600	Encodes a Ser-2-specific RNAPII CTD phosphatase with two tandem-repeated CTD phosphatase domains that belongs to the group III CTD phosphatase-like (CPL) family.	Upstream gene variant	Protein coding
AT3G22360	encodes an alternative oxidase whose expression is limited to flowers and floral buds.	Downstream gene variant	Protein coding
AT3G23325	Splicing factor 3B subunit 5/RDS3 complex subunit 10	Downstream gene variant	Protein coding
AT3G24290	ammonium transporter 1	Upstream gene variant	Protein coding

AT3G27027	GPI-anchored-like protein (DUF 3339)	Downstream gene variant	Protein coding
AT3G30530	basic leucine-zipper 42	Downstream gene variant	Protein coding
AT3G42670	Encodes a nuclear localized SNF domain containing protein involved in RNA silencing.	Upstream gene variant	Protein coding
AT3G53380	Concanavalin A-like lectin protein kinase family protein	Upstream gene variant	Protein coding
AT3G55160	THADA is an orphan gene in Arabidopsis thaliana. It is the only DUF2428 domain containing protein in the genome.	Upstream gene variant	Protein coding
AT3G57000	nucleolar essential protein-like protein	Downstream gene variant	Protein coding
AT3G60670	PLATZ transcription factor family protein	Upstream gene variant	Protein coding
AT4G03740	Hypothetical protein	Downstream gene variant	Protein coding
AT4G04825	microRNA ath-MIR3932a precursor	Downstream gene variant	miRNA primary transcript
AT4G07868	Transmembrane protein	Downstream gene variant	Protein coding
AT4G09150	T-complex protein 11	Downstream gene variant	Protein coding
AT4G10360	TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein	Downstream gene variant	Protein coding
AT4G11480	Encodes a cysteine-rich receptor-like protein kinase.	Downstream gene variant	Protein coding
AT4G13180	NAD(P)-binding Rossmann-fold superfamily protein	Upstream gene variant	Protein coding
AT4G13260	Encodes YUC2. Catalyzes conversion of IPA (indole-3-pyruvic acid) to IAA (indole-3-acetic acid) in auxin biosynthesis pathway.	Downstream gene variant	Protein coding
AT4G13930	Encodes a serine hydroxymethyltransferase maximally expressed in root; involved in S-adenosyl-L-methionine biosynthesis, recycling of S-ademosylmethionine-dependent methylation, methionine formation.	Upstream gene variant	Protein coding
AT4G15030	folate-sensitive fragile site protein	Downstream gene variant	Protein coding
AT4G15350	Member of CYP705A	Downstream gene variant	Protein coding
AT4G16210	Enoyl-CoA hydratase/isomerase A	Upstream gene variant	Protein coding
AT4G20240	Putative cytochrome P450	Upstream gene variant	Protein coding
AT4G22140	Encoding a chromatin remodeling factor that regulates flowering time.	Upstream gene variant	Protein coding
AT4G34150	Calcium-dependent lipid-binding (CaLB domain) family protein	Downstream gene variant	Protein coding
AT4G35350	tracheary element vacuolar protein. A clubroot pathogen effector targets cruciferous cysteine proteases to suppress plant immunity.	Downstream gene variant	Protein coding
AT5G03245	Natural antisense transcript overlaps with AT5G20910	Downstream gene variant	Antisense long noncoding RNA
AT5G03780	Encodes a protein whose sequence is similar to human telomere proteins	Upstream gene variant	Protein coding
AT5G07190	Gene is expressed preferentially in the embryo and encodes a unique protein of unknown function.	Downstream gene variant	Protein coding
AT5G09330	NAC domain containing protein 82	Upstream gene variant	Protein coding
AT5G09595	Natural antisense transcript overlaps with AT5G66280	Downstream gene variant	Antisense long noncoding RNA
AT5G15160	BNQ2 belongs to a family of atypical non-DNA binding basic helix-loop-helix (bHLH) proteins that heterodimerize with and negatively regulate bHLH transcription factors.	Upstream gene variant	Protein coding
AT5G16630	DNA repair protein Rad4 family	Downstream gene variant	Protein coding

AT5G17850	CCX2 is a putative cation/Ca2+ exchange protein. It is located in the endoplasmic reticulum. It plays a role in salt induced calcium signaling. Loss of function results in decreased cytosolic and increased ER Ca2+ concentrations.	Downstream gene variant	Protein coding
AT5G18350	Disease resistance protein (TIR-NBS-LRR class) family	Upstream gene variant	Protein coding
AT5G28690	carboxylate clamp-TPR protein (DUF1685)	Downstream gene variant	Protein coding
AT5G29560	lipid droplet and endoplasmic reticulum-associated protein.	Downstream gene variant	Protein coding
AT5G33320	Encodes a plastid inner envelope protein PPT (phosphoenolpyruvate/phosphate translocator) that catalyzes the transport of phosphoenolpyruvate and phosphate across the inner envelope membrane of plastids.	Downstream gene variant	Protein coding
AT5G38710	Methylenetetrahydrofolate reductase family protein	Intron Variant	Protein coding
AT5G48160	Encodes a nuclear PHD finger protein that is functionally redundant with OBE1 and plays an important role in the maintenance and/or establishment of the root and shoot apical meristems.	Upstream gene variant	Protein coding
AT5G51820	Encodes a plastid isoform of the enzyme phosphoglucomutase involved in controlling photosynthetic carbon flow.	Upstream gene variant	Protein coding
AT5G53410	F-box associated ubiquitination effector family protein	Downstream gene variant	Protein coding
AT5G54855	Pollen Ole e 1 allergen and extensin family protein	Downstream gene variant	Protein coding
AT5G55760	Encodes SRT1, a member of the SIR2 (sirtuin) family HDAC (histone deacetylase) (SRT1/AT5g55760, SRT2/AT5G09230).	Downstream gene variant	Protein coding
AT5G65225	Hypothetical protein	Downstream gene variant	Protein coding

Table showing list of the 91 candidate genes identified from the P3-07 F_2 generation. The table include the gene ID, gene description, consequence, and biological function for each candidate gene.

Table 3.2 List of candidate genes for P10-23 suppressor mutant showing high impact variants.

Gene ID	Gene Description	Consequence	Mutation	Amino Acids	Codons	Biological Function
AT1G05740	Mitochondrial	Splice Donor Variant	-	-	-	Protein Coding
AT1G07260	UDP-GLUCOSYL TRANSFERASE 71C3	Frameshift Variant	act/acAt	T/TX	320-321	Protein Coding
AT1G11250	SYNTAXIN OF PLANTS 125	Frameshift Variant	att/aTtt	I/IX	859-860	Protein Coding
AT1G17580	Arabidopsis Myosin	Frameshift Variant	gtG/gt	V/X	24	Protein Coding
AT1G26230	Chloroplasts chaperonins	Frameshift Variant	tcC/tc	S/X	384	Protein Coding
AT1G32970	Subtilisin like serine endopeptidase	Stop Gained	gca/gGca	A/GI	1633-1634	Protein Coding
AT1G40104	Mitochondrial	Stop Gained	Aag/Tag	K/*	817	Protein Coding
AT1G47670	Transmembrane amino acid transporter family	Frameshift Variant	gGc/gc	G/X	374	Protein Coding
AT1G64580	Pentatricopeptide repeat (PPR) superfamily protein	Frameshift Variant	Tgc/gc	C/X	898	Protein Coding
AT1G64600	Copper ion binding, methyltransferase	Stop Gained	Cag/Tag	Q/*	154	Protein Coding
AT1G65250	ZRK14	Frameshift Variant	agt/agAt	S/RX	803-804	Protein Coding
AT1G66910	Protein kinase superfamil	Frameshift Variant	tATAGGCAAGAG /t	YRQE/X	776-786	Protein Coding
AT1G72110	O-acyltransferase (WSD1-like) family protein	Frameshift Variant	tTa/ta	L/X	572	Protein Coding
AT1G73066	Leucine-rich repeat family protein	Stop Gained	tCa/tGa	S/*	1481	Protein Coding
AT1G74070	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	Frameshift Variant	aaA/aa	K/X	150	Protein Coding
AT1G76810	Eukaryotic translation initiation factor 2 (eIF-2) family protein	Frameshift Variant	-/T	-/X	1041-1042	Protein Coding
AT2G07981	Hypothetical protein	Stop Gained	Aaa/Taa	K/*	1123	Protein Coding
AT2G08986	Hypothetical protein	Stop Gained	Aaa/Taa	K/*	181	Protein Coding
AT2G16365	ATFDA10, F-BOX/DUF295 ANCESTRAL 10	Stop Gained	tat /taAAAGTAGTGt	Y/*K*C	1610-1611	Protein Coding
AT2G35050	Kinase superfamily with octicosapeptide/Phox/Bem1p domain-containing	Frameshift Variant	gaa/gAaa	E/EX	2290-2291	Protein Coding
AT2G36960	TKI1, TSL-KINASE INTERACTING PROTEIN 1	Frameshift Variant	Ggg/gg	G/X	640	Protein Coding
AT2G40500	Protein kinase superfamily	Frameshift Variant	aaa/aaGAAGa	K/KKX	326-327	Protein Coding
AT3G02330	MEF13, MITOCHONDRIAL EDITING FACTOR 13	Frameshift Variant	gcTTGTGCTCACATGGGTC TGATCGAC /gc	ACAHMGLID/ X	2139-2163	Protein Coding
AT3G15690	Single hybrid motif superfamily protein	Frameshift Variant	Aaa/aa	K/X	565	Protein Coding
AT3G25716	Transmembrane protein	Frameshift Variant	agG/ag	R/X	45	Protein Coding
AT3G29455	Hypothetical protein	Frameshift Variant	-/CAGACACATGTTT	-/QTHVX	804-805	Protein Coding
AT3G43148	Myosin heavy chain-like	Frameshift Variant	tcc /tTTTTTGTTATCACATTTTT ACCATTTTTTTAcc	S/FFVITFLPF FYX	694-695	Protein Coding
AT3G44080	F-box family protein	Frameshift Variant	aat/aGat	N/RX	382-383	Protein Coding
AT3G51280	Tetratricopeptide repeat (TPR)-like superfamily p	Stop Gained	Gag/Tag	E/*	670	Protein Coding

AT3G58750	CITRATE SYNTHASE 2, CSY2	Stop Gained	tgG/tgA	W/*	1356	Protein Coding
AT3G61490	Pectin lyase-like superfamily protei	Frameshift Variant	aaa/aaACa	K/KX	284-285	Protein Coding
AT4G00680	Actin depolymerizing factor 8	Frameshift Variant	GTt/t	V/X	25-26	Protein Coding
AT4G05400	Copper ion binding protein	Frameshift Variant	ccA/cc	P/X	21	Protein Coding
AT4G07960	XyG glucan synthase	Frameshift Variant	tTg/tg	L/X	527	Protein Coding
AT4G10540	SBT3.8, SUBTILASE 3.8	Frameshift Variant	-/TGTAT	-/CX	903-904	Protein Coding
AT4G12780	AUXILIN-LIKE1 Negative regulation of growth and endocytosis	Frameshift Variant	-/TC	-/X	756-757	Protein Coding
AT4G15260	UDP-Glycosyltransferase superfamily protein	Frameshift Variant	gct/gTct	A/VX	1075-1076	Protein Coding
AT4G15550	IAGLU, INDOLE-3-ACETATE BETA-D-GLUCOSYLTRANSFERASE	Frameshift Variant	aaA/aa	K/X	318	Protein Coding
AT4G16180	Transmembrane protein	Frameshift Variant	aag/aaTCg	K/NX	1541-1542	Protein Coding
AT4G25225	Transmembrane protein	Frameshift Variant	Atc/tc	I/X	82	Protein Coding
AT4G25845	Oxysterol-binding 4B-like protei	Frameshift Variant	Aaa/aa	K/X	355	Protein Coding
AT4G31355	Hypothetical protein	Frameshift Variant	aag/aaCg	K/NX	26-27	Protein Coding
AT4G33160	F-box family protein	Frameshift Variant	ctTCtc/cttc	LL/LX	162-163	Protein Coding
AT4G33430	AK1, BRI1-ASSOCIATED RECEPTOR KINASE	Frameshift Variant	ggG/gg	G/X	657	Protein Coding
AT4G39420	Spatacsin carboxy-terminus protein	Stop Gained	tGg/tAg	W/*	6338	Protein Coding
AT5G06645	Gibberellin regulated like protei	Splice Donor Variant	-	-	-	-
AT5G10910	CHLOROPLAST MRAW-LIKE, CMAL	Frameshift Variant	gtA/gt	V/X	1191	Protein Coding
AT5G16830	Member of SYP2 gene family involved in intracellular protein transport	Frameshift Variant	aaA/aa	K/X	426	Protein Coding
AT2G36960	TKI1, TSL-KINASE INTERACTING PROTEIN 1	Frameshift Variant	Ggg/gg	G/X	640	Protein Coding
AT5G17460	Glutamyl-tRNA (Gln) aminotransferase subunit C	Stop Gained	Gga/Tga	G/*	403	Protein Coding
AT5G45160	RHD-LIKE2, RL2 Root hair defective 3 GTP-binding protein	Splice Donor Variant	-	-	-	-
AT5G56900	CwfJ-like family protein / zinc finger (CCCH-type) family protein	Splice Donor Variant	-	-	-	-
AT5G59540	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein;	Stop Gained	tCg/tAg	S/*	383	Protein Coding
AT5G63120	RH30, RNA HELICASE 30	Splice Donor Variant	-	-	-	-
AT5G66610	DA1-related protein 7	Stop Gained	Gga/Tga	G/*	1285	Protein Coding

Table showing list of the 55 High Impact candidate genes identified from the P10-23 F_2 generation. The table include the gene ID, gene description, consequence, mutation with exact base pair change, amino acids change, codons position in Tair, and biological function for each candidate gene.

Table 3.3 Mutation display term and the consequences associated with the impact factor for BSA analysis.

Display Term	SO Description	SO Accession	Impact
Upstream gene variant	A sequence variant located 5' of a gene	SO:0001631	MODIFIER
Downstream gene variant	A sequence variant located 3' of a gene	SO:0001632	MODIFIER
Frameshift variant	A sequence variant which causes a disruption of the translational reading frame	SO:0001589	HIGH
Intron variant	A transcript variant occurring within an intron	SO:0001627	MODIFIER
Splice Donor variant	A splice variant that changes the 2-base region of the 5' end of an intron	SO:0001575	HIGH
Stop gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	HIGH

3.3.12 P3-07 filtering step 2 of candidate genes

Filtering step 1 was utilised to filter genes based on the confident assumption that the short plants chosen for DNA extraction were absent of the mutation and the tall plants had HOM variants present. As P3-07 is a semi-dominant line, the segregation ratio of short, medium, and tall plants could have resulted in medium height plants being classified as tall and included in the tall DNA pool. To avoid any confusion, backcrossing to *fip37-4* was performed and grown out for segregation analysis. The percentage of the mutation that would be present in each pool for all three dominant, semi-dominant, and recessive mutations was explained in Figure 3.11. Filtering step 2 involved the analysis of allele depth counts for the WT and mutant allele for the mutation. HOM and HET variants were filtered in the Tall DNA pool based on read counts of the WT allele and Variant Allele, further filtered through a ratio of Variant:WT, and only genes that had at least 66% or more of the variant allele were selected. This filtering step produced a candidate list of 82 possible genes (Table 3.4).

3.3.13 P10-23 filtering step 2 of candidate genes

In filtering step 1 of the P10-23, variants from the tall DNA pool were filtered based on their high impact consequence. Filtering step 2 was performed using the allele depth read for the variant present in the tall DNA pool. As shown in Figure 3.11, the F_2 generation produced a segregation ratio of 1:3 for short to tall plants after backcrossing to *fip37-4* and self-fertilisation. Therefore, a dominant mutation must be present in the tall DNA pool and absent in the short DNA pool. Only HET and HOM variants with a Variant:WT ratio of at least 66% were selected, resulting in a list of 382 possible candidate genes (Table A9.1)

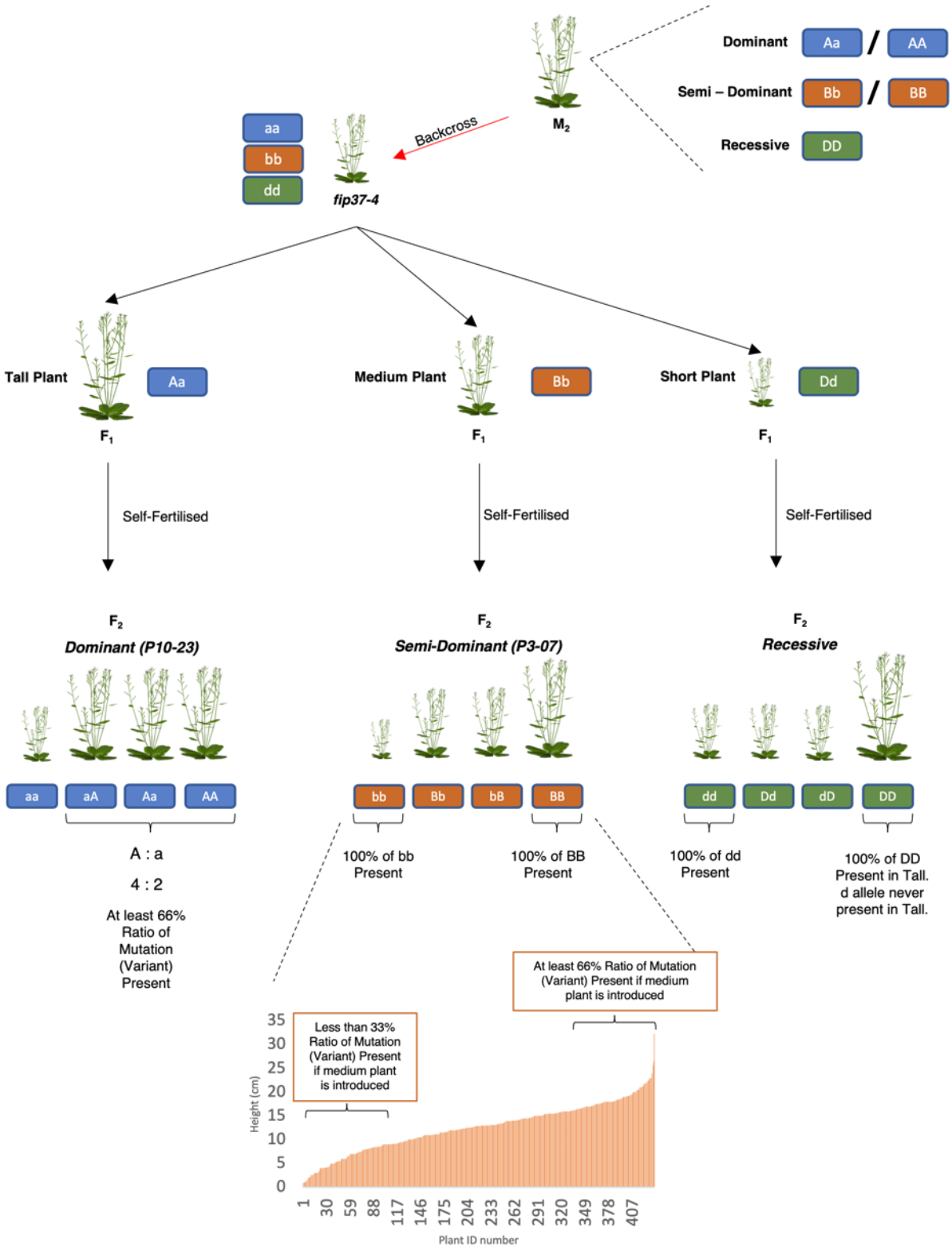


Figure 3.11 Schematic representation of suppressor mutation in BSA of Dominant, Semi-dominant, and Recessive ratios.

This schematic representation depicts the mutation of alleles in the backcrossing of suppressor lines P3-07 and P10-23, which display Dominant, Semi-dominant, and Recessive heritability. The M_2 generation of the plant was utilised for backcrossing with *fip37-4*. The blue box indicates the Dominant line, the orange box indicates the Semi-dominant line, and the green box indicates the Recessive line. The mutation present on the allele ratio is also displayed and further explained in relation to the variability of plant height for the semi-dominant line.

Table 3.4 List of candidate genes for P3-07 suppressor mutant showing filtered ratios of VAR:WT.

Gene ID	Gene Description	Consequence	AD WT	AD VAR	VAR:WT
AT1G01010	Member of the NAC domain containing family of plant specific transcriptional regulators.	Upstream Gene Variant	8	7	0.88
AT1G05190	Encodes the plastid 50S ribosomal protein L6.	Downstream Gene Variant	35	31	0.89
AT1G05270	TraB family protein that localizes to the outer mitochondrial membrane and tether to ER.	Downstream Gene Variant	9	7	0.78
AT1G05380	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein	Upstream Gene Variant	12	10	0.83
AT1G08163	microRNA ath-MIR5649b precursor	Upstream Gene Variant	12	8	0.67
AT1G47510	Encodes a phosphatidylinositol polyphosphate 5-phosphatase.	Upstream Gene Variant	14	12	0.86
AT1G53010	RING/U-box superfamily protein	Upstream Gene Variant	6	4	0.67
AT1G56260	Required for the maintenance of stem cells through a reduction in DNA damage.	Downstream Gene Variant	7	5	0.71
AT1G56590	Involved in vesicle trafficking between the trans - Golgi network and vacuoles.	Downstream Gene Variant	12	9	0.75
AT1G56660	MAEBL domain protein	Upstream Gene Variant	13	11	0.85
AT1G59500	Encodes an IAA-amido synthase that conjugates Asp and other amino acids to auxin in vitro.	Downstream Gene Variant	12	8	0.67
AT1G60900	Putative U2A65 splicing factor which functions in abscisic acid mediated flowering via regulating the precursor messenger RNA splicing of ABI5 and FLC in shoot apex.	Upstream Gene Variant	7	5	0.71
AT1G61160	Retrotransposon gag	Downstream Gene Variant	6	5	0.83
AT1G61420	S-locus lectin protein kinase family protein	Downstream Gene Variant	6	5	0.83
AT1G61950	Member of Calcium Dependent Protein Kinase, regulated by Bacillus amyloliquefaciens under high calcium stress.	Upstream Gene Variant	9	6	0.67
AT1G66045	Reverse transcriptase-like protein	Upstream Gene Variant	3	2	0.67
AT1G67330	Arabinogalactan methylesterase,involved in arabinogalactan glucuronic acid methylation.	Upstream Gene Variant	12	8	0.67
AT1G68440	Transmembrane protein. Expression induced by abiotic stressors such as ABA, drought, heat, light, NaCl, osmotic stress and wounding.	Upstream Gene Variant	13	9	0.69
AT1G68450	VQ motif-containing protein	Downstream Gene Variant	16	11	0.69
AT1G70430	Protein kinase superfamily protein	Upstream Gene Variant	4	3	0.75
AT1G71400	Encodes a CLAVATA2 (CLV2)-related gene. Complements the clv2 mutant when expressed under the control of the CLV2 promoter. The mRNA is cell-to-cell mobile.	Upstream Gene Variant	4	3	0.75
AT1G72900	Toll-Interleukin-Resistance (TIR) domain-containing protein	Upstream Gene Variant	9	7	0.78
AT1G74090	Encodes a desulfoglucosinolate sulfotransferase, involved in the final step of glucosinolate core structure biosynthesis.	Upstream Gene Variant	4	3	0.75
AT1G74456	snoRNA	Downstream Gene Variant	13	11	0.85
AT1G75560	Zinc knuckle (CCHC-type) family protein	Downstream Gene Variant	12	8	0.67
AT1G77100	Peroxidase superfamily protein	Downstream Gene Variant	4	3	0.75
AT1G77770	Forkhead box protein, putative (DUF1644)	Downstream Gene Variant	3	2	0.67
AT1G78090	Homologous to the C-terminal part of microbial trehalose-6-phosphate phosphatases	Upstream Gene Variant	11	10	0.91
AT1G80390	Member of a multigene family of Auxin responsive genes.	Downstream Gene Variant	13	10	0.77
AT1G80990	XH domain-containing protein	Downstream Gene Variant	10	8	0.80
AT2G01045	Hypothetical protein	Upstream Gene Variant	13	9	0.69
AT2G01050	Zinc ion binding / nucleic acid binding protein	Downstream Gene Variant	5	4	0.80

AT2G03855	microRNA ath-MIR5642b precursor	Upstream Gene Variant	15	14	0.93
AT2G48160	HUA and HUA-LIKE (HULK) genes act redundantly to regulate a subset of essential genes	Missense Variant	7	6	0.86
AT3G01060	DUF domain plastic localized protein of unknown function.Mutants germinate faster than wt. One of the 'green cut proteins' found only in photosynthetic organisms.	Downstream Gene Variant	5	4	0.80
AT3G42050	Member of V-ATPase family. Vacuolar-type H + -ATPase (V-ATPase) is a multisubunit proton pump located on the endomembranes.	Downstream Gene Variant	60	45	0.75
AT3G43153	cAMP-dependent protein kinase inhibitor-like protein	Upstream Gene Variant	8	7	0.88
AT3G43300	AtMIN7 is an immunity associated Arabidopsis protein targeted by HopM1, a conserved Pseudomonas syringae virulence protein.	3 Prime UTR Variant	11	9	0.82
AT3G43940	Hypothetical protein	Downstream Gene Variant	15	11	0.73
AT3G44100	MD-2-related lipid recognition domain-containing protein	Downstream Gene Variant	9	6	0.67
AT3G44405	Hypothetical protein	Downstream Gene Variant	8	6	0.75
AT3G47430	Member of the peroxin11 (PEX11) gene family, located on the peroxisome membrane, controls peroxisome proliferation. The mRNA is cell-to-cell mobile.	Upstream Gene Variant	16	14	0.88
AT3G47700	Involved in transportation of seed storage proteins from the ER to the vacuole.	Downstream Gene Variant	4	3	0.75
AT3G48640	Transmembrane protein	Downstream Gene Variant	8	6	0.75
AT3G50110	Encodes a phosphatase with low in vitro tyrosine phosphatase activity that is NOT capable of dephosphorylating in vitro the 3'phosphate group of PI3P, PI(3,4)P2.	Downstream Gene Variant	9	6	0.67
AT3G54100	O-fucosyltransferase family protein	Downstream Gene Variant	9	6	0.67
AT3G55970	One of 4 paralogs encoding a 2-oxoglutarate/Fe(II)-dependent oxygenases that hydroxylates JA to 12-OH-JA.	Upstream Gene Variant	8	7	0.88
AT3G59700	Member of Receptor kinase-like protein family. Represses stomatal immunity induced by Pseudomonas syringae pv. tomato DC3000.	Upstream Gene Variant	5	4	0.80
AT3G61120	Encodes AGL13, a member of the AGL6 clade of the MADS-box gene family.	Downstream Gene Variant	10	7	0.70
AT3G62470	Pentatricopeptide repeat (PPR) superfamily protein	Upstream Gene Variant	8	6	0.75
AT3G63280	Encodes AtNek4, a member of the NIMA-related serine/threonine kinases (Neks) that have been linked to cell-cycle regulation in fungi and mammals. Plant Neks might be involved in plant development processes.	Upstream Gene Variant	6	5	0.83
AT4G00355	Encodes an Atg8-interacting protein that is partially associated with the ER during favorable growth conditions and becomes mainly associated with a spherical compartment that dynamically moves along the ER network.	Upstream Gene Variant	12	11	0.92
AT4G05071	This gene encodes a small protein and has either evidence of transcription or purifying selection.	Upstream Gene Variant	9	7	0.78
AT4G08560	Encodes a member of the Arabidopsis Pumilio (APUM) proteins containing PUF domain (eight repeats of approximately 36 amino acids each). PUF proteins regulate both mRNA stability and translation through sequence-specific binding to the 3' UTR of target mRNA transcripts.	Upstream Gene Variant	5	4	0.80
AT4G10820	F-box family protein	Upstream Gene Variant	7	6	0.86
AT4G12545	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	Upstream Gene Variant	9	7	0.78
AT4G13190	Protein kinase superfamily protein	Downstream Gene Variant	8	7	0.88
AT4G16765	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	Downstream Gene Variant	14	13	0.93
AT4G40100	Encodes PRSL1 (PP1 regulatory subunit2-like protein1). Functions as a regulatory subunit of PP1 (protein phosphatase 1) and regulates blue light signaling in stomata.	Downstream Gene Variant	11	10	0.91
AT5G49900	Beta-glucosidase, GBA2 type family protein	Downstream Gene Variant	5	4	0.80
AT5G50790	Encodes a member of the SWEET sucrose efflux transporter family proteins. Transcriptionally activated by long photoperiods	Downstream Gene Variant	13	9	0.69

AT5G52540	Keratin-associated protein, putative (DUF819)	Downstream Gene Variant	39	38	0.97
AT5G53290	Encodes a member of the ERF (ethylene response factor) subfamily B-5 of ERF/AP2 transcription factor family.	Upstream Gene Variant	17	15	0.88
AT5G53610	Carbohydrate-binding X8 domain superfamily protein	Upstream Gene Variant	8	7	0.88
AT5G54300	Cotton fiber-like protein (DUF761)	Downstream Gene Variant	7	6	0.86
AT5G54940	Translation initiation factor SUI1 family protein	Downstream Gene Variant	16	13	0.81
AT5G55290	Member of V-ATPase family. Vacuolar-type H + -ATPase (V-ATPase) is a multisubunit proton pump located on the endomembranes.	Downstream Gene Variant	10	9	0.90
AT5G56270	Encodes WRKY transcription factor 2, a zinc-finger protein. In wrky2 mutants, egg cells polarize normally but zygotes fail to reestablish polar organelle positioning from a transient symmetric state, resulting in equal cell division and distorted embryo development.	Upstream Gene Variant	9	8	0.89
AT5G56500	Encodes a subunit of chloroplasts chaperonins that are involved in mediating the folding of newly synthesized, translocated, or stress-denatured proteins.	Upstream Gene Variant	16	12	0.75
AT5G57520	Encodes a zinc finger protein containing only a single zinc finger.	Upstream Gene Variant	7	5	0.71
AT5G58575	Component of the deubiquitination module of the SAGA complex.	Upstream Gene Variant	6	5	0.83
AT5G60980	Nuclear transport factor 2 (NTF2) family protein with RNA binding (RRM-RBD-RNP motifs) domain-containing protein	Upstream Gene Variant	9	8	0.89
AT5G61440	Encodes a member of the thioredoxin family protein. Located in the chloroplast. The mRNA is cell-to-cell mobile.	Upstream Gene Variant	9	7	0.78
AT5G61660	Glycine-rich protein	Downstream Gene Variant	10	8	0.80
AT5G61910	DCD (Development and Cell Death) domain protein	Downstream Gene Variant	7	6	0.86
AT5G62050	essential factor for protein sorting and assembly into membranes	Downstream Gene Variant	14	10	0.71
AT5G65330	AGAMOUS-like 78	Upstream Gene Variant	16	12	0.75
AT5G65900	DEA(D/H)-box RNA helicase family protein	Upstream Gene Variant	12	9	0.75
AT5G66150	Glycosyl hydrolase family 38 protein	Downstream Gene Variant	10	8	0.80
AT5G66990	RWP-RK domain-containing protein	Downstream Gene Variant	11	9	0.82
AT5G67030	Encodes a single copy zeaxanthin epoxidase gene that functions in first step of the biosynthesis of the abiotic stress hormone abscisic acid (ABA).	Downstream Gene Variant	3	2	0.67
AT5G67640	Hypothetical protein	Downstream Gene Variant	16	15	0.94

Table showing list of the 82 candidate genes identified from the P3-07 F_2 generation. The table include the gene ID, gene description, consequence, Allele depth (AD) for WT, AD for variant, and Variant:WT ratio.

3.4 Discussion

RNA modifications can serve as a posttranscriptional regulatory process, adding another layer of gene regulation in plants. Identifying and characterising the components of the m^6A writer complex, along with potential new interactors, is fundamental in understanding the regulatory functions of this modification. FIP37 is a core component of the m^6A writer complex and has interactions with multiple key components of the methyltransferase complex. In this chapter, the identification of two lines that can suppress the phenotype of *fip37-4* suggests the presence of a new molecular player linked to FIP37 or other cellular processes involved in m^6A regulation or that can result in more efficient splicing of the T-DNA transcript from the transcribed *fip37-4* RNA.

3.4.1 Suppressor mutant identification reveals two lines that successfully rescues *fip37-4*

The successful implementation of EMS mutagenesis and analysis of the segregating F_2 population resulted in the identification of two lines, P3-07 and P10-23, that rescued the *fip37-4* phenotype, displaying a phenotype similar to WT. Observations of both suppressor mutant lines revealed that the plants exhibited an increased rosette size, WT-like leaf shape, and restored apical dominance (Figure 3.4). The P10-23 line showed a segregation ratio of 1:3 (short to tall) in the F_2 population indicating a dominant mutation, but the P3-07 line displayed a segregation ratio of short, medium, and tall plants, suggesting it to be a semi-dominant mutation. The suppression of two independent mutant line phenotype, P3-07 and P10-23, to a WT phenotype was studied using BSA and NGS to generate a candidate gene list. The F_2 BC1 segregation ratio for P3-07 indicated a semi-dominant mutation that follows Mendel's law of segregation (Figure 3.2A). Theoretically, a recessive suppressor mutation should result in 25% of the plants being tall (homozygous, 1:2:1). However, following further backcrossing to *fip37-4*, the segregation ratio could suggest a semi-dominant mutation, as the height of intermediate plants increased relative to the average height of *fip37-4*. The phenotypic

observation also showed that the tall phenotype did not revert to WT as strongly as in the first backcross. Other factors, such as differences in growth conditions in different rooms, could also contribute to the changes in phenotype. For the dominant suppressor line P10-23, the mutation should not be present in the short plants, and the ratio of tall plants in the DNA pool should be 66% while it should be 0% in the short DNA pool.

3.4.2 Suppressor mutants restores root architecture in normal and abiotic stress conditions

In order to further characterise *fip37-4* and the suppressor mutants, root architecture analysis was conducted. The results showed that *fip37-4* mutant displayed significant impairments in root architecture compared to WT. The analysis also revealed that P3-07 mutant exhibited a significant increase in PRL and LRD compared to *fip37-4*, demonstrating strong suppression and reversion to WT-like structure.

The plant growth regulator auxin is a fundamental hormone that participates in nearly all aspects of plant growth and development, in specific root formation. A recent study by Zhang *et al.*, (2022) employed methylated RNA immunoprecipitation (MeRIP-Seq) on writer complex member mutants, *vir-1* and *fip37-4*, and performed gene ontology (GO) enrichment analysis on protein-coding methylated transcripts. The results indicated that root development (GO: 0048364) was one of the top categories for the *fip37-4* mutant. This suggests a strong association between writer complex proteins, m^6A methylation, and root development. One potential explanation for the reversion of suppressor mutants back to a WT root phenotype is an increase in global m^6A levels. However, the data did not show a statistically significant increase in m^6A levels in P3-07 compared to *fip37-4*.

The lack of a statistical increase in m^6A levels in P3-07 to explain the suppression of *fip37-4* suggests that other factors may be involved. Recent research has established the role of miRNAs, in regulating auxin response pathways (Luo *et al.*, 2022a). miRNAs are small non-

coding RNA molecules, averaging 22 nucleotides in length. They are transcribed from DNA sequences into primary miRNAs (pri-miRNAs) and are further processed into mature miRNA that can interact with mRNAs, leading to suppression of expression (Ha et al. 2014). The miRNA regulation could play a significant role in the regulation and rescue of suppressor lines. Previous studies have reported that MTA-deficient mutants display decreased levels of miRNAs but increased accumulation of pri-miRNA transcripts (Bhat *et al.*, 2020). MTA interaction with RNA Polymerase II during early stages of transcription and potential co-transcriptional methylation of pri-miRNAs have been documented (Bhat *et al.*, 2020). A potential mechanism for the causal mutation in suppressor lines is the activation of MTA to re-localise and bind preferentially to RNA Polymerase II, thereby triggering key genes involved in pri-miRNA processing and activating gene expression in response to stress conditions. One such condition could be a non-functional writer complex, leading to a different role for MTA that rescues plant development through the miRNA regulation pathway.

Suppressor mutants demonstrate the possibility of reverting back to the WT root phenotype under abiotic stress conditions, as observed in NaCl and Sorbitol medium, potentially linking to stress response genes (Figure 3.6). A recent study by Luo *et al.*, (2022) shows that under salt stress, NaCl increases the expression of miR393 which inhibits F-box genes TIR1/AFB2 through the enhancement of miR393a transcription and inhibition of lateral roots which is consistent with P3-07 as lateral root density decreases in abiotic stress conditions (Figure 3.6). As shown in Table 3.1, the P3-07 the downstream gene variant AT4G04825, a microRNA precursor MIR3932a, and a F-box associated ubiquitination effector family protein, AT5G53410, may play a role in this process, but further research is necessary to design experiments specific to this gene that show rescuing of *fip37-4*.

Additionally, recent studies have shown that mutants of m^6A writer components exhibit salt-sensitive phenotypes in a m^6A-dependent manner (Hu *et al.*, 2021). The authors revealed that the *vir* mutant leading to reduction of m^6A methylation and overaccumulation of reactive oxygen species homeostasis by affecting 3’ UTR lengthening, which is linked to alternative

polyadenylation, demonstrating a strong connection between m^6A methylation, 3'-UTR length, and mRNA stability during stress adaptation. The study also discussed that among the writer components, the relative expression of *FIP37* increased the most ~ 2-fold under 24-hour salt stress in *A. thaliana* compared to *MTA*, *MTB*, *VIR*, and *HAKAI*, further highlighting the sensitivity of FIP37 to salt stress.

3.4.3 Trichome branching and lateral root initiation

Trichomes have a fundamental role in plant development and are found in a wide variety of species. Trichomes are unicellular or multicellular structures that consist of a stalk with three to four branches. The onset of trichome development involves the exit of pluripotent epidermal cells from the mitotic program and their transition into the endocycle. Trichomes serve as a protective barrier against ultraviolet irradiation, pathogen attacks, excessive transpiration, and seed protection. *A. thaliana* has a unicellular non-glandular trichome that is an excellent model for studying the molecular mechanisms of trichome development in plants (Lloyd *et al.*, 1994; Chien & Sussex, 1996). The results indicate that the *fip37-4* mutant displays a trichome-branching defect with an increase in the number of trichomes with four to six branches, consistent with the hypomethylated plants observed in Bodi *et al.*, (2012) and the FIP37 overexpressed plants described by Vespa *et al.*, (2004). FIP37 is crucial in trichome development, as shown by consistent results of aberrant trichome formation in both *fip37-4* knock-down in this chapter and overexpressed FIP37 in previous studies. One likely explanation for the increased trichome branching could be related to an excessive number of endoreduplication cycles (Scutenaire *et al.*, 2018). Examining whether the suppressor mutants can rescue trichome branching provides an additional level of suppression in the phenotype for both lines. The suppressor mutants revert to normal trichome branching levels, with P3-07 showing 40-50% trichomes with three or four branches, an increase in percentage compared to *fip37-4* Figure 3.7A.

Auxin is a key phytohormone that regulates primary and lateral root growth. To check the response of the low methylation mutant *fip37-4* and suppressor mutants to auxin, NPA and NAA media treatments were utilised. NPA functions as an inhibitor of polar auxin transport in plants, interacting with PIN proteins that transport auxin (Abas *et al.*, 2020). In *fip37-4*, lateral root initiation was significantly reduced in response to this treatment, resulting in fewer lateral roots than the WT control, as shown in Figure 3.7B. Conversely, suppressor mutants showed a tendency to revert back to WT-like lateral roots, with both P3-07 and P10-23 exhibiting an increase in lateral root length and density. The reversal of this phenotype may suggest that suppressor mutants can restore auxin response, which may be associated with the rescue of normal plant development.

3.4.4 An increase in m^{6}A levels in suppressor lines P3-07 and P10-23

Previous studies have demonstrated the correlation between decreased m^{6}A levels and developmental growth defects in plants, as observed in knockdown mutants of m^{6}A writer components (Zhong *et al.*, 2008; Bodi *et al.*, 2012; Růžička *et al.*, 2017). However, it has been observed that a reduction in m^{6}A methylation does not necessarily result in a corresponding developmental phenotype which could be due to a threshold of m^{6}A level to produce a clear phenotype change. In *A. thaliana*, the HAKAI knockout resulted in a 35% decrease in m^{6}A methylation levels without significant phenotypic changes (Růžička *et al.*, 2017). This suggests that the effect of m^{6}A methylation on developmental growth is complex and not solely dependent on the presence or absence of some of the m^{6}A writer components or other related molecular factors. This observation was further supported in line P3-07, where suppression of the *fip37-4* phenotype was observed despite a lack of a statistically significant increase in m^{6}A levels. The P10-23 line displayed a statistically significant increase in m^{6}A levels when compared to the *fip37-4* line, consistent with a close reversion to WT levels. This significant increase in m^{6}A levels could be an explanation for the suppression of *fip37-4* as an increase in methylation of transcripts could indicate a functional methyltransferase complex. Further

support for this hypothesis was obtained from the Northern blot analysis, which showed an increase in WT functional transcripts, further indicating the presence of a proper methyltransferase complex.

Previous studies have shown that m^6A modification is a dynamic and reversible process, which involves the interaction between writer, eraser, and reader proteins. The causality of a mutation in one or more genes in this BSA analysis method could impact the regulation of eraser and reader proteins activities. The addition of RNA modification adds a complex layer of gene regulation that allows plants to respond to environmental and cellular changes. The potential role of m^6A reader proteins in conjunction with other candidate genes should not be ignored, as m^6A is a complex modification with multiple methyltransferase components that could individually have their own functions. Reader proteins, such as YTH domain proteins of the ECT family, play a critical role in the regulation of RNA stability (e.g. ECT2) (Shi *et al.*, 2017; Wei *et al.*, 2018). m^6A methylation is primarily observed at the 3' end of transcripts, but it can occur throughout the entire transcript. The underlying mechanism of methylation site selection is yet to be fully understood. m^6A is a dynamic and reversible modification with a significant impact on downstream gene regulation. m^6A reader proteins play a crucial role in this process, as they act upon methylated transcripts in various ways that can influence downstream applications. The significance of these reader proteins in m^6A methylation will be thoroughly investigated in subsequent chapters.

3.4.5 Northern blotting and suppressor mutant efficiency of splicing

Previous studies have demonstrated a link between m^6A methylation and splicing (Haussmann *et al.*, 2016; Lence *et al.*, 2016). The loss of function of FIP37 is embryonic lethal. *fip37-4* is a hypomorphic mutant with a T-DNA insertion between the 7^{th} and 8^{th} exons, which produces a small amount of correctly processed transcript. To determine if efficient splicing of the T-DNA in the *fip37-4* could contribute to the suppression of the phenotype, a northern blot

was performed. An increase in splicing efficiency could lead to the reversion of functional *FIP37* transcripts and a functional methyltransferase complex, resulting in the production of transcripts with m^6A modification. The results of the northern blot analysis indicated a slight increase in the levels of *FIP37* transcripts in P3-07 compared to *fip37-4* levels. However, this increase in functional transcripts did not correspond to a statistically significant increase in m^6A levels, suggesting that an increase in functional transcripts does not necessarily mean an increase in the level of functional FIP37 and methyltransferase complex. A western blot would be required to confirm the production of functional proteins, but this was not performed due to time constraints. The discrepancy between the increase in functional transcripts and the low m^6A level could be due to mutations being present upstream or downstream of the coding sequence, as observed in the comparison with the BSA candidate gene list of P3-07. Such mutations could cause changes in transcription factor binding, leading to activate or repress certain proteins linked to the suppression of the *fip37-4* phenotype. The P10-23 line exhibits an increase in the level of WT *FIP37* transcripts compared to the *fip37-4* levels, as depicted in Figure 3.9A. This marked increase in WT *FIP37* transcript level, along with the significant increase in global m^6A levels compared to the *fip37-4* levels, shown in Figure 3.8F, suggests a strong correlation between these changes and the reversion back to the WT phenotype.

3.4.6 BSA and possible candidate genes

In this chapter, two methods of filtering for candidate genes was used, namely filtering step 1 and filtering step 2, due to the presence of different types of mutations in the P3-07 and P10-23 suppressor lines. For P3-07, filtering step 1 was based on the premise that the selected tall plants were homozygous for the mutation, while the short plants did not carry the mutation. This approach resulted in the selection of unique homozygous mutations found only in the tall DNA pool, leading to a list of 91 candidate genes.

The majority of mutations identified in P3-07 filtering step 1 as unique to the tall DNA pool were not located in the coding region, so the exact position of the mutation within the CDS sequence could not be determined. However, BSA and NGS analysis identified a golgi S-Adenosylmethionine transporter 1 (GOSAMT 1) AT1G64650 from the P3-07 candidate gene list as a possible contributor to the rescue of the *fip37-4* phenotype (Table 3.2). This transmembrane protein has been shown to transport S-adenosyl methionine (SAM) into the golgi apparatus for polysaccharide methylation, and a double *gosamt1/ gosamt2* mutant displayed a phenotype similar to *fip37-4*, suggesting a connection between m^6A methylation and Golgi SAM transporters (Temple *et al.,* 2022). *A. thaliana* has three GOSAMT genes, and a KO of all three results in severe developmental growth defects, indicating that a triple mutant may also be embryo lethal, similar to the complete knockout of FIP37 (Temple *et al.*, 2022). The mutation consequence that was identified for *GOSAMT1* displayed an upstream gene variant, rather than a coding sequence mutation, which could mean that it is potentially caused by a mutation in a transcription factor regulating the transcript. In the RNA-seq data from Chapter 4, the comparison group "*fip37-4* x P3-07" exhibited a minimal upregulation of GOSAMT1, with a Log2 fold change (Log2FC) of 0.01 for P3-07. In the "WT x *fip37-4*" comparison group, GOSAMT1 showed expression change of 0.02 Log2FC for *fip37-4*, while in the "WT x P3-07" group, it was 0.04 Log2FC for P3-07. The RNA-seq data for GOSAMT1 along with the mutation within the upstream gene variant, it appears unlikely that transcription

factor activation plays a significant role in the upregulation of this gene, given the marginal changes observed in each comparison group. However, GOSAMT1 could remain a potential candidate due to its involvement in SAM transport within the Golgi complex which could indirectly influence m^6A modification. To validate this candidacy, further experiments are needed.

For line P3-07 the use of filtering step 2 method involved accounting for the possibility that medium height plants (Heterozygous for mutation) could have been inadvertently included in the Tall selection of plants and subsequently in the DNA extraction during BSA. The allele depth count information obtained from raw sequencing data was used to perform this filtering step. As a result, a list of 82 candidate genes was generated. Notably, many of the mutations identified in this filtering step were located outside of the coding region, which is consistent with the results using the method filtering step 1 for P3-07. This observation lends support to the notion that a mutation in a transcription factor could be responsible for the phenotype observed in P3-07.

The P10-23 line was subjected to a two-step filtering process that is similar to P3-07, as this line exhibited a dominant mutation. Filtering step 1 identified 3864 unique HET and HOM genes in the tall DNA pool. This list was further filtered based on Ensembl VEP variant consequence, resulting in 55 high-impact genes. These mutations were located within the coding region, providing the precise position in the CDS sequence. High-impact variants are expected to have a disruptive effect on the protein, potentially leading to protein truncation, loss of function, or triggering of nonsense-mediated decay. The list of 55 high-impact genes included 5 splice donor variants, 38 frameshift variants, and 3 stop-gained variants. In the second filtering step for P10-23, a minimum threshold of 66% mutation presence was used from the allele depth count for Var:WT, as dictated by the segregation ratio of the F_2 population (Figure 3.11). The resulting list contained 382 possible candidate genes, which is significantly higher than the list produced in filtering step 1. The large number of candidates presented in the filtering process makes it challenging to identify the specific mutation responsible for the

suppression of *fip37-4*. Both P3-07 and P10-23 yielded extensive lists of candidate genes using the two-step filtering process. To further narrow down the list, additional backcrossing to *fip37-4* and subsequent filtering steps could be carried out. This would allow for the confident identification of a smaller set of genes responsible for the suppression of *fip37-4*.

3.5 Summary and future work

EMS mutagenesis was successful in producing viable suppressor mutants that showed suppression of *fip37-4* on various levels as shown in the different methods of measurement in this chapter. The mutant lines P3-07 and P10-23 both showed varying degree of reversion back to WT phenotype. Further analysis into choosing candidate genes along with production of a knock-out mutant of the specified candidate gene and sexually crossing it with *fip37-4* to produce a WT-like plant will confirm direct linkage that suppression is from that gene. The following chapters will incorporate RNA-seq as a tool to understand the transcriptome wide picture of how suppression of *fip37-4* is achieved.

Chapter 4

Transcriptomic Analysis of *fip37-4* and Suppressor Mutants P3-07 and P10-23

4 Transcriptomic analysis of *fip37-4* and suppressor mutants P3-07 and P10-23

4.1 Chapter Aims

The objective of this chapter is to process and analyse RNA-sequencing (RNA-seq) data performed on WT, *fip37-4*, P3-07, and P10-23. The rationale for conducting RNA-seq following BSA in Chapter 3 was to identify differentially expressed genes associated with the suppression of *fip37-4*. RNA-seq provides an additional level of quantitative analysis, complementing the northern blot performed in Chapter 3. The information obtained from RNA-seq can guide functional analysis of the identified candidate genes through gene ontology analysis (GO), aiding in the prioritisation of genes for validation related to specific functional characteristics. By integrating RNA-seq with BSA, we can gain a more comprehensive understanding of the suppressor lines' characteristics. The main aims are to:

- Process raw RNA-seq data from all samples through a pipeline to produce differentially expressed (DE) genes between each sample.
- Identify significant genes from DE files and to analyse these genes in relation to explaining the suppression of *fip37-4*.
- Perform Heatmap and Gene Ontology (GO) analysis on significant genes.

4.2 Introduction

Measuring gene expression on a genome-wide scale has become a common practice in the past two decades, providing a deeper level of analysis. This technology enables us to measure and understand changes in gene expression in complex systems, such as the mechanisms of m^6A methylation in gene regulation. Several steps are involved in processing and analysing RNA-seq data, and specific pipelines are used to minimise bias in the results obtained. In Chapter 3, we characterised dominant (P10-23) and semi-dominant (P3-07) suppressor lines,

which closely and partially restored the WT phenotype, respectively. Further experiments also demonstrated that m^6A levels, along with transcript abundance from northern blotting, partially and fully reverted back to WT levels. BSA analysis generated an extensive list for both lines. RNA-seq analysis performed on suppressor lines could narrow down potential candidate genes while providing insightful transcriptomic data to uncover how changes in these levels could explain suppression of *fip37-4*.

4.3 Results

To gain a deeper understanding of the suppression of *fip37-4* in suppressor lines P3-07 and P10-23, RNA-seq was utilised to analyse transcriptomic data. RNA-seq is a highly effective molecular tool that utilises high-throughput sequencing to quantify gene expression and identify DE genes across different comparison groups.

4.3.1 Total RNA preparation for RNA-sequencing

Differential gene expression analysis was performed on RNA extracted from WT, *fip37-4*, P3-07, and P10-23 samples to further understand the suppression of *fip37-4* in suppressor lines. Total RNA was extracted from two-week old seedlings and subjected to quality control prior to sequencing. mRNA libraries (poly-A enrichment) were prepared by Novogene. Each RNA sample was sequenced with a depth of 20 million, 150 paired-end reads.

4.3.2 RNA-sequencing pipeline for processing of raw data.

The processing of raw RNA-seq data was performed using a specific pipeline, as shown in Figure 4.1, with the use of various programs available on the Galaxy web platform. The sequencing data was uploaded to usegalaxy.org, a public server, for analysis (Jalili *et al.*,

2021). Each program in the pipeline was utilised for a specific purpose, such as quality checking and data processing, to ensure the accuracy of the transcriptomic data obtained.

4.3.2.1 Quality control of raw RNA-sequencing data

To evaluate the quality of the RNA sequencing data, the FastQC program was used to assess the quality scores of each base sequence. A score of 30 or higher indicates a 1 in 1000 chance of an incorrect base call. The FastQC results for all paired-end reads from each sample were collated using MultiQC to analyse the overall distribution of quality scores. The results indicated that the majority of RNA samples contained at least 34.2% of unique reads Figure 4.2A. Moreover, the MultiQC report revealed that the mean quality scores for all samples were above a Phred score of 35, demonstrating low chances of incorrect base calling (Figure 4.2B).

4.3.2.2 Raw data trimming of adapter sequences

After performing quality check on raw data, Trim Galore, a wrapper script was used to automate quality and adapter trimming (Krueger, 2021). This step removes adapter sequences from the raw data for downstream analysis. A further quality check was performed, and MultiQC report was generated to assess the trimmed data. Overrepresented sequences were checked using FastQC on all paired-end samples, and none were found, indicating that the library was not contaminated (Figure 4.2C). Adapter sequence percentage was also checked, and the MultiQC output indicated that all samples had adapter contamination of below 0.1% Figure 4.2C. The per sequence quality score was also tested and shown to have an average Phred score of 35 Figure 4.2D.

4.3.2.3 Alignment to reference genome

After quality checking and trimming, the trimmed data was aligned to the *A. thaliana* TAIR10 reference genome using the Hierarchical Indexing for Spliced Alignment of Transcripts

(HISAT2) program, which is a fast and sensitive alignment tool (Kim *et al.*, 2015). The HISAT2 alignment scores were obtained using FastQC and all samples showed over 90% uniquely mapped reads, indicating high quality alignment to the reference genome (Figure 4.2E).

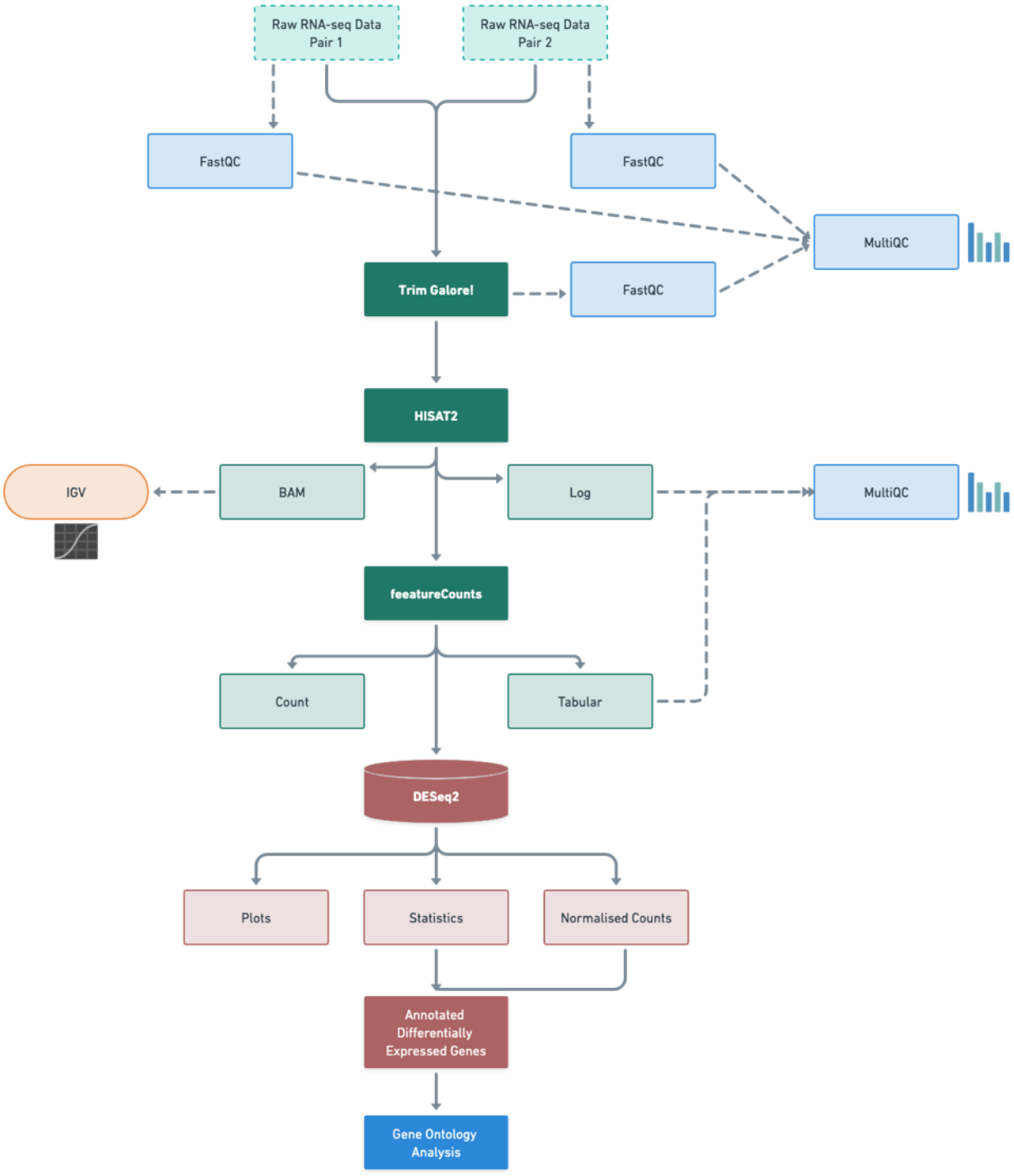


Figure 4.1 Flow chart showing the pipeline used for processing RNA-seq data to differentially expressed genes.

Raw RNA-seq data was sent to Novogene for sequencing. Processing of data was done through the use of Trim Galore, HISAT2, featureCounts, and DESeq2 while FastQC was used to quality check sequencing data at each point. MultiQC was used to collate all FastQC files into one graph. DESeq2 was performed in R studio to generate downstream files.

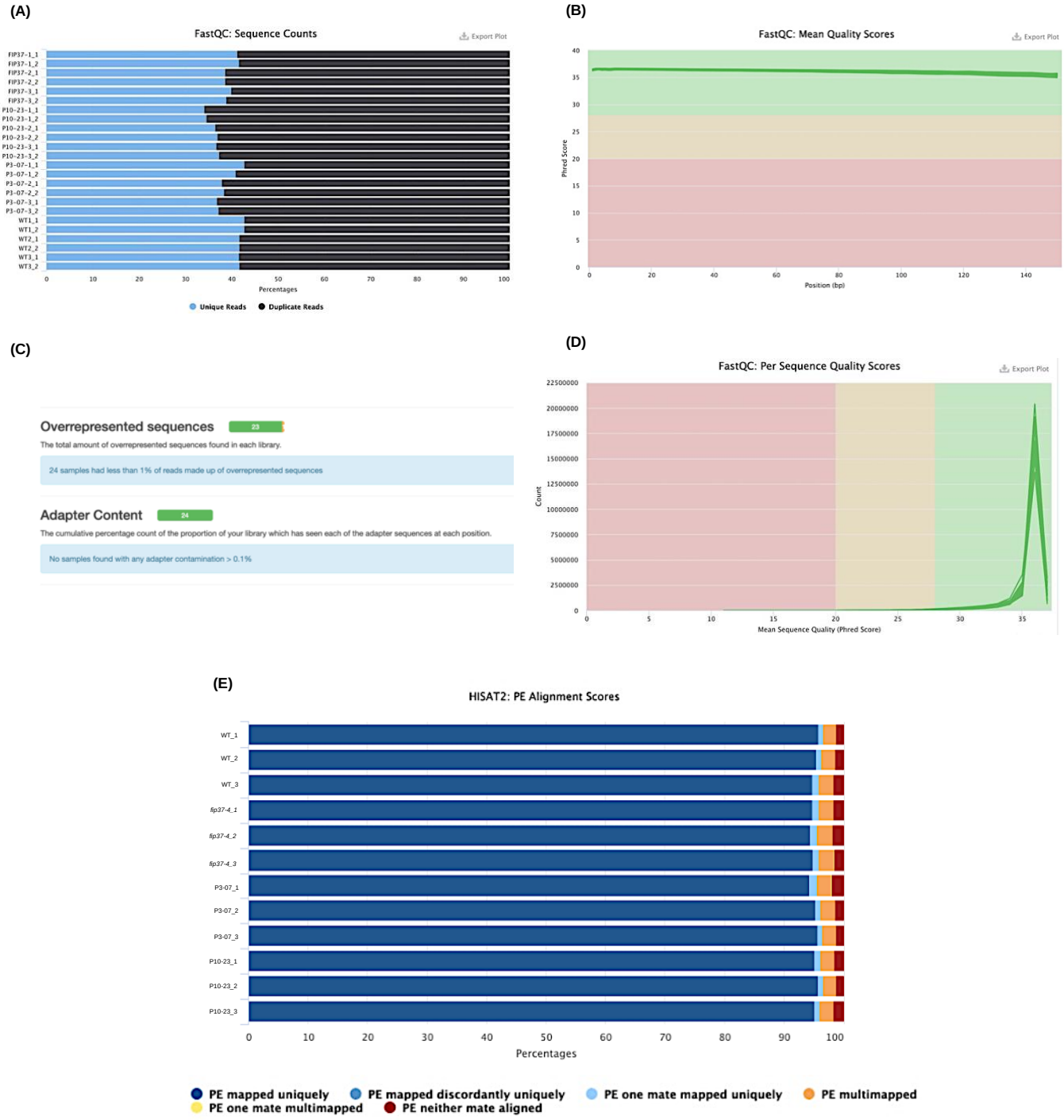


Figure 4.2 Quality check statistics on processing of raw data showing trimmed and aligned data.

(A) FastQC sequence counts statistics showing unique and duplicate reads in all paired end raw RNA-seq data. (B) FastQC mean quality scores for all samples showing Phred scores and bp positions in reads. (C) Trimmed data QC indicating all samples had less than 1% of overrepresented sequences and less than 0.1% adapter content. (D) FastQC per sequence quality scores for all paired-end trimmed samples. (E) Quality analysis on HISAT2 alignment scores across all aligned samples.

4.3.3 Coverage analysis shows that transcript expression has increased after T-DNA insertion in *fip37-4* and suppressor lines

In order to analyse the knock-down of *fip37-4*, read coverage analysis was performed to visualise *FIP37* transcript expression. After HISAT2 alignment to the *A. thaliana* reference genome (TAIR10), the BAM files generated were inputted into Integrative Genomics Viewer (IGV) to produce coverage analysis plots for WT, *fip37-4*, P3-07, and P10-23 with their respective biological replicates. The read coverage track for WT in Figure 4.3 shows strong expression of *FIP37*, indicating functional FIP37 production. Interestingly, subsequent samples for *fip37-4*, P3-07, and P10-23 showed high read coverage downstream of the T-DNA insertion located in intron between exon 7-8. For P3-07 and P10-23 suppressor lines, the results showed that *FIP37* still displayed low read coverage upstream of the T-DNA insertion, indicating that non-functional *FIP37* was produced. This further confirms that the suppression of the *fip37-4* phenotype is not due to functional *FIP37* production. In the comparison between coverage reads in Figure 4.3 and the northern blot in Chapter 3, no coverage was observed from Exon 1 to Exon 7 of FIP37 in samples *fip37-4*, P3-07, and P10-23 when in comparison to WT. However, upon closer examination of the coverage analysis, we do detect reads, although in small quantities when compared to the WT. The limited coverage observed in the samples is consistent with the northern blot results, where WT FIP37 transcripts are still produced in *fip37-4*, P3-07, and P10-23 but at reduced levels.

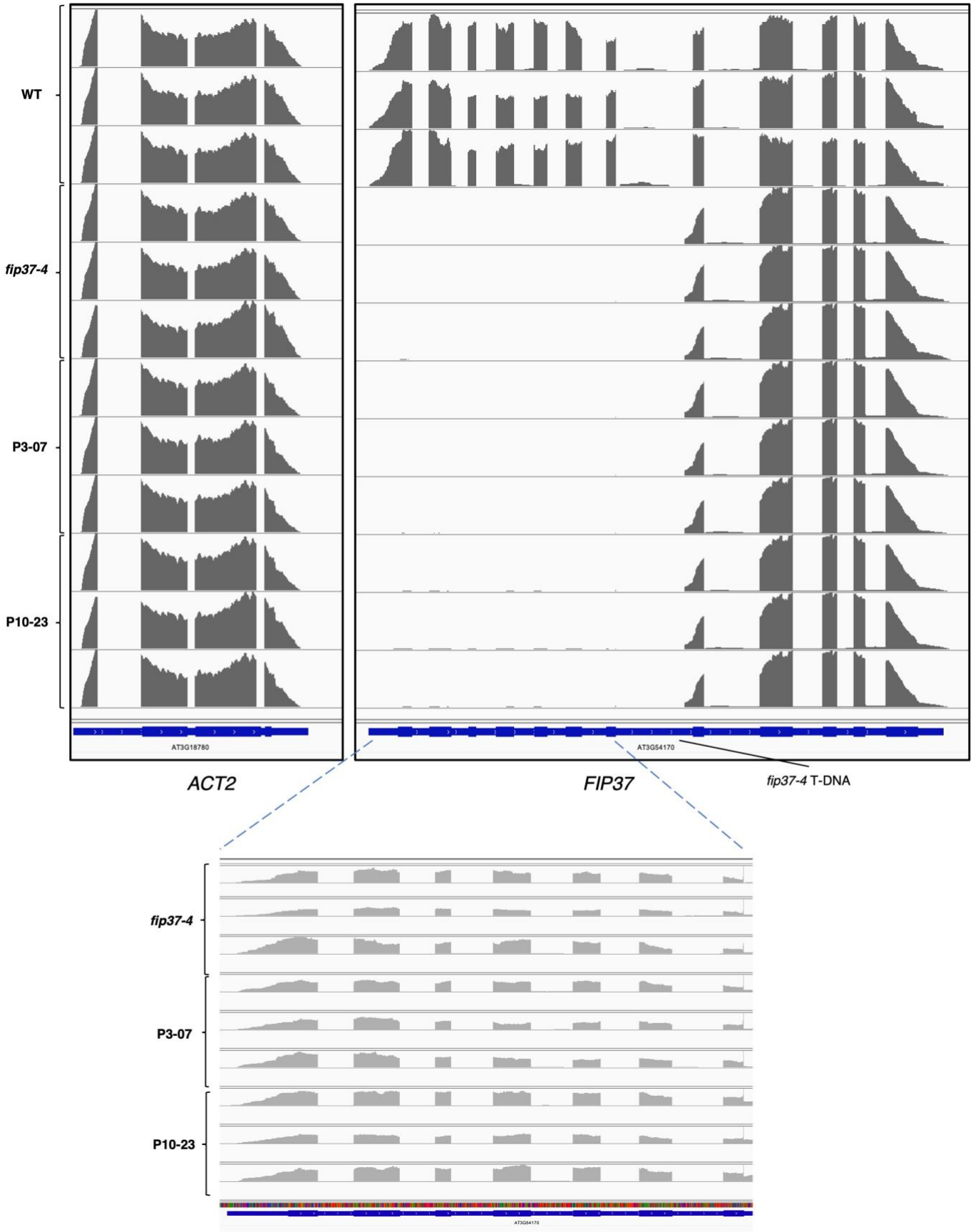


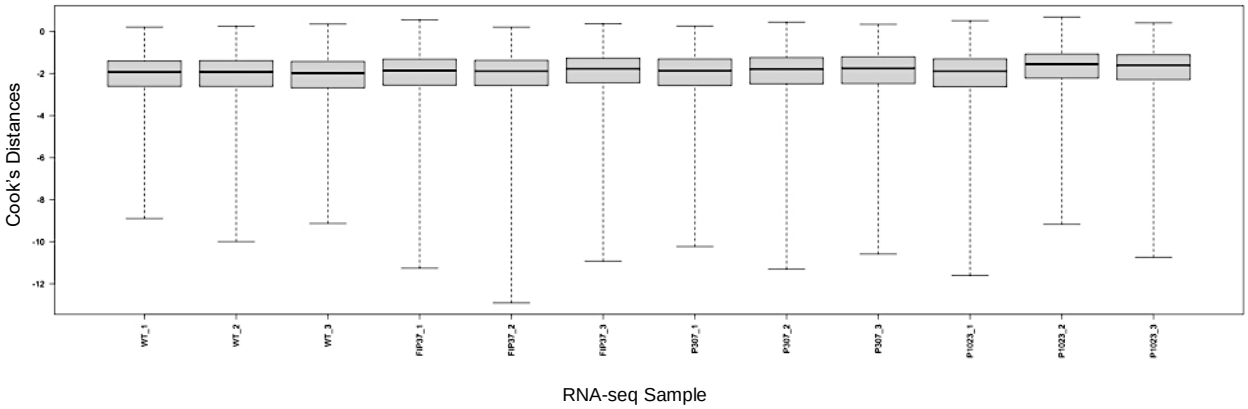
Figure 4.3 Coverage analysis of WT, *fip37-4*, P307 and P1023 showing increased expression after T-DNA insertion for suppressor lines.

The RNA-seq densities along exons are shown as histograms coloured grey. The RNA-seq densities are aligned to their respective gene *ACT2* and *FIP37* at the bottom of the plot which are coloured blue. Samples were processed using the pipeline described in Figure 4.1. The BAM files containing data and index was taken from HISAT2 alignment and coverage analysis was performed in Integrative Genomics Viewer. *FIP37* plots was normalised against a constitutively active gene *ACT2*. The below coverage shows an enlarged area from exon 1-7. Samples shown are from three biological replicate.

4.3.4 DESeq2 RNA-sequencing sample analysis

The BAM files generated after alignment to the *A. thaliana* reference genome (TAIR10) were then passed through to the next step in the pipeline which involved using featureCounts, a program for counting RNA-seq reads (Liao *et al.*, 2014). To identify differentially expressed genes (DEG), the count tables produced by featureCounts were inputted into DESeq2, a program that estimates variance-mean dependence in RNA-seq count data and produces gene-level differential expression files for sample comparisons. Cook's distance was calculated for each sample to detect outliers, but none were found (Figure 4.4A). A dispersion plot was generated to ensure a good fit for the DESeq2 model, and the results showed that the dispersion of data decreased with increasing mean expression levels, indicating a good fit for the model with no contamination or outlier samples (Figure 4.4B). The MA plot demonstrated that only genes with a large average normalised count (red dot) contain sufficient information to yield a significant call (Figure 4.5A). Additionally, a diagnostic plot showed that all samples had a high frequency of over 6,000 p values lower than 0.1.

(A)



(B)

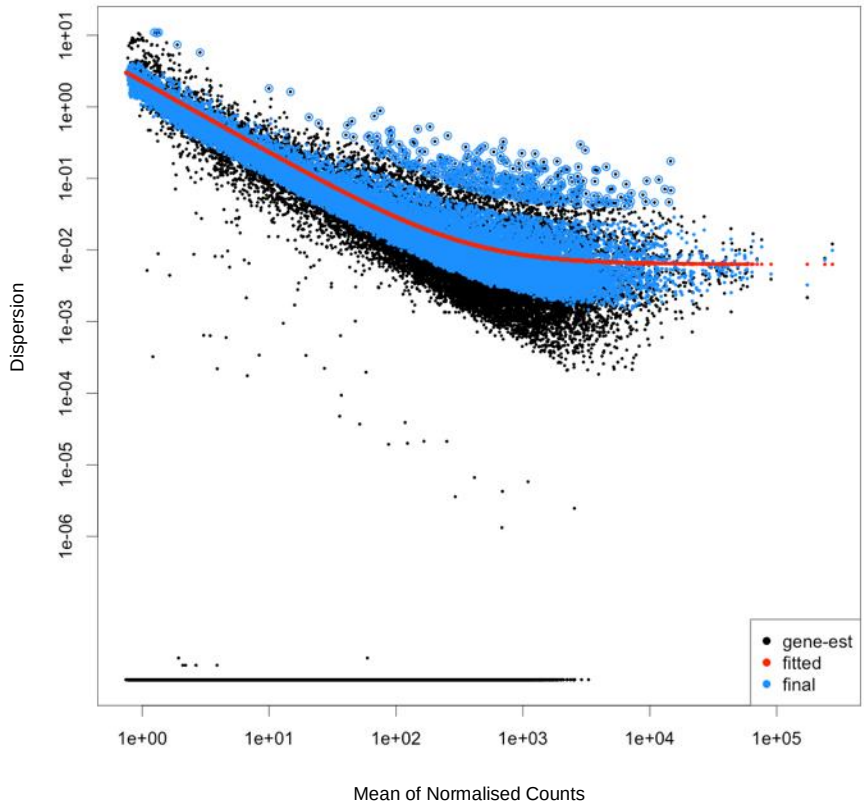
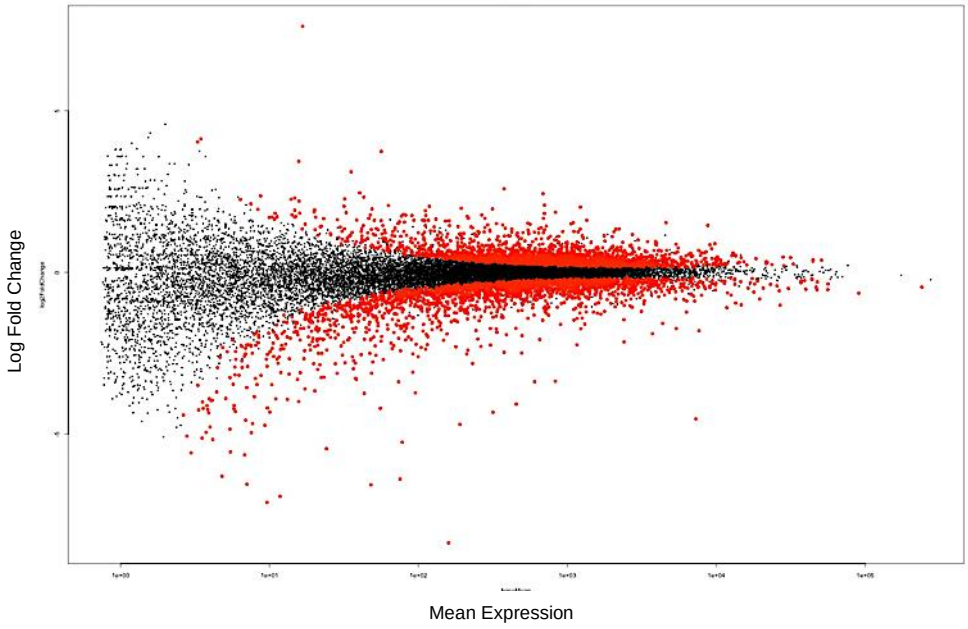


Figure 4.4 RNA-Seq data quality assessment

(A) Boxplots showing Cook's distances calculated for each sample. (B) Dispersion plot to visualise DESeq2 dispersion estimates. The black points are the dispersion estimates for each gene. Fitted red trend line shows the dispersions' dependence on the mean. The blue points show each gene's shrink towards the red line to obtain final estimates that are then used in the hypothesis test. Blue circles above the main "cloud" of points are genes which have high gene-wise dispersion estimates are labelled as dispersion outliers.

(A)



(B)

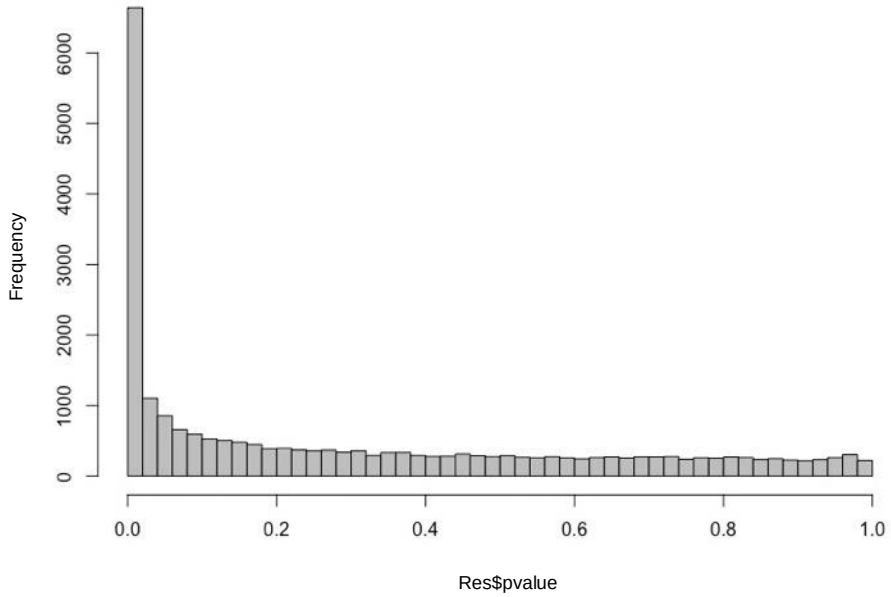


Figure 4.5 Diagnostic plots on normalised counts produced by DESeq2.

(A) MA plots of the expression level (Mean of normalised counts) and ratio (Log2FC) for each gene in all samples. Each gene is represented with a dot. Genes with an adjusted p value of < 0.05 are shown in red. (B) Histogram showing the frequency and spread of p values in all data samples.

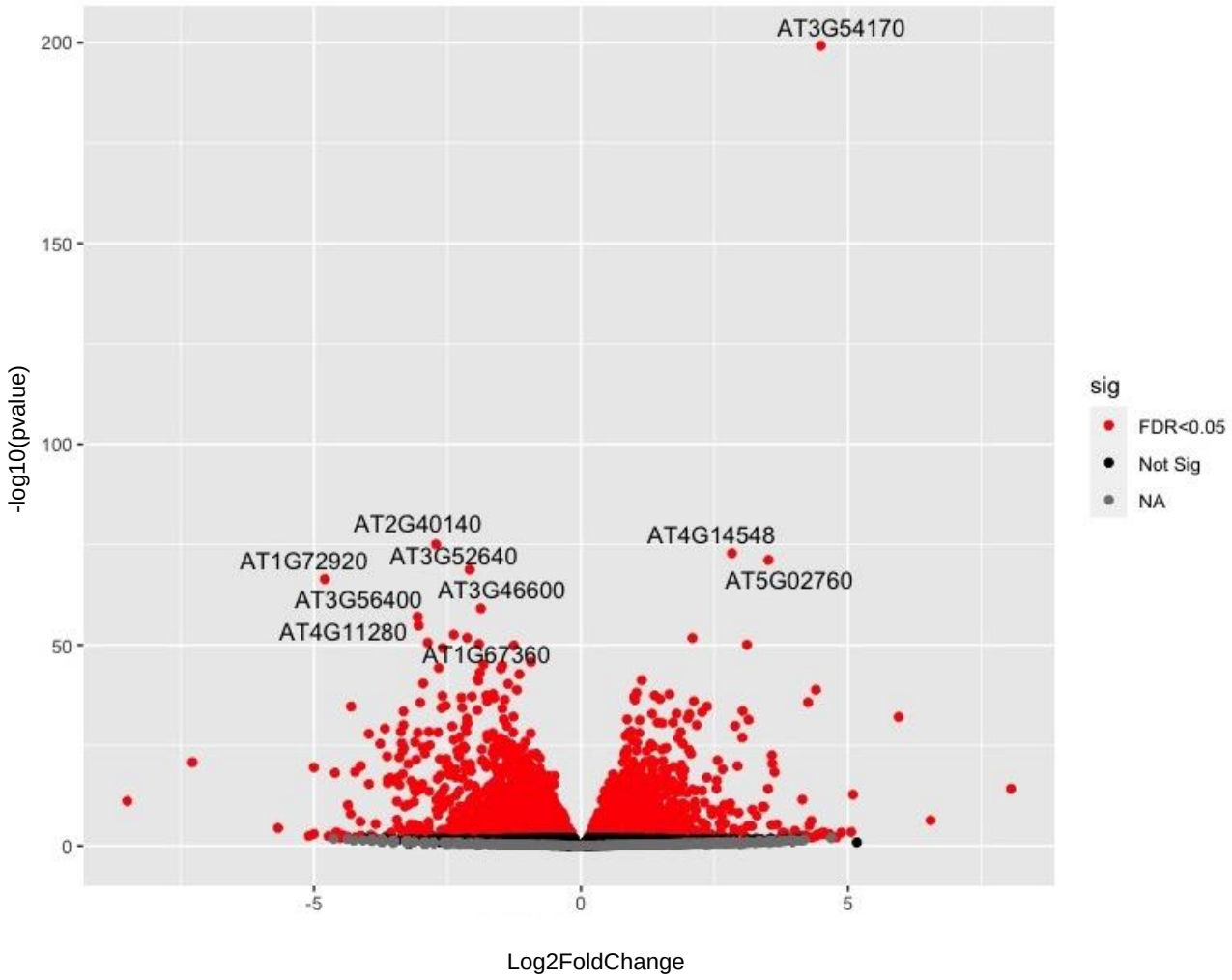


Figure 4.6 Volcano plot of all differentially expressed genes in all samples.

Volcano plot showing up-regulated and down-regulated genes by Log2FoldChange on the x-axis against –log10 (pvalue). Red dot represents significant genes (FDR < 0.05), black dot represents not significant, and grey dots represents outliers.

4.3.5 Heatmap analysis shows clustering of genes for P3-07 and P10-23

In order to visualise gene expression differences between comparison groups, heatmap analysis was performed. The heatmap in Figure 4.7A displays hierarchical clustering of significant genes extracted from the DE file of "*fip37-4* x P3-07", "WT x *fip37-4*", and "WT x P3-07" with a threshold of padj < 0.05 and log2FC > 1.5. A second heatmap, shown in Figure 4.7B, was produced for comparisons of "*fip37-4* x P10-23", "WT x *fip37-4*", and "WT x P10-23". In both heatmaps, significant genes were clustered using the Euclidean method with the red scale indicating up-regulation and blue indicating down-regulation. The P3-07 heatmap in Figure 4.7A shows that both comparison groups, "WT x *fip37-4*" and "WT x P3-07", have similar clustering of genes, indicating that P3-07 has similar expression patterns to *fip37-4* as it retains the T-DNA insertion but displays suppression of phenotype. The clustering analysis of genes located in the region labelled R1 in Figure 4.7A reveals a cluster of up-regulated genes in the suppressor line P3-07 when compared to *fip37-4*. This finding suggests that this specific region may have potential candidates involved in the suppression of *fip37-4*. Subsequently, we could extract the genes identified in this region to create a candidate list, followed by performing GO analysis to prioritise specific cellular mechanisms for further investigation. Furthermore, within the same R1 region horizontally across, both "WT x *fip37-4*" and "WT x P3-07" show down-regulation of clustered genes, in contrast to the up-regulation observed in "*fip37-4* x P3-07". This differential expression pattern suggests potential differences in gene regulation between the different comparison groups, providing valuable insights that shed light on molecular mechanisms underlying the suppression of *fip37-4*.

The heatmap for P10-23 in Figure 4.7B reveals down-regulation of a cluster of genes within the region labelled R2, which mirrors the down-regulation of genes observed in the same horizontal region in the comparison of "WT x P10-23". Similarly, the clustered region labelled R3 shows up-regulation of genes, consistent with the up-regulation observed in the same horizontal region in "WT x P10-23". Interestingly, within the region labelled R4, a set of up-regulated genes is specifically seen in "*fip37-4* x P10-23", but not evident in the comparisons

of “WT x *fip37-4*” and “WT x P10-23”. This observation presents an opportunity to explore these gene clusters further, as they may shed light on the suppression of the phenotype observed in P10-23. Overall, the two suppressor lines show different expression changes in the heatmap analysis, indicating different levels of gene expression are occurring for the suppression of *fip37-4*.

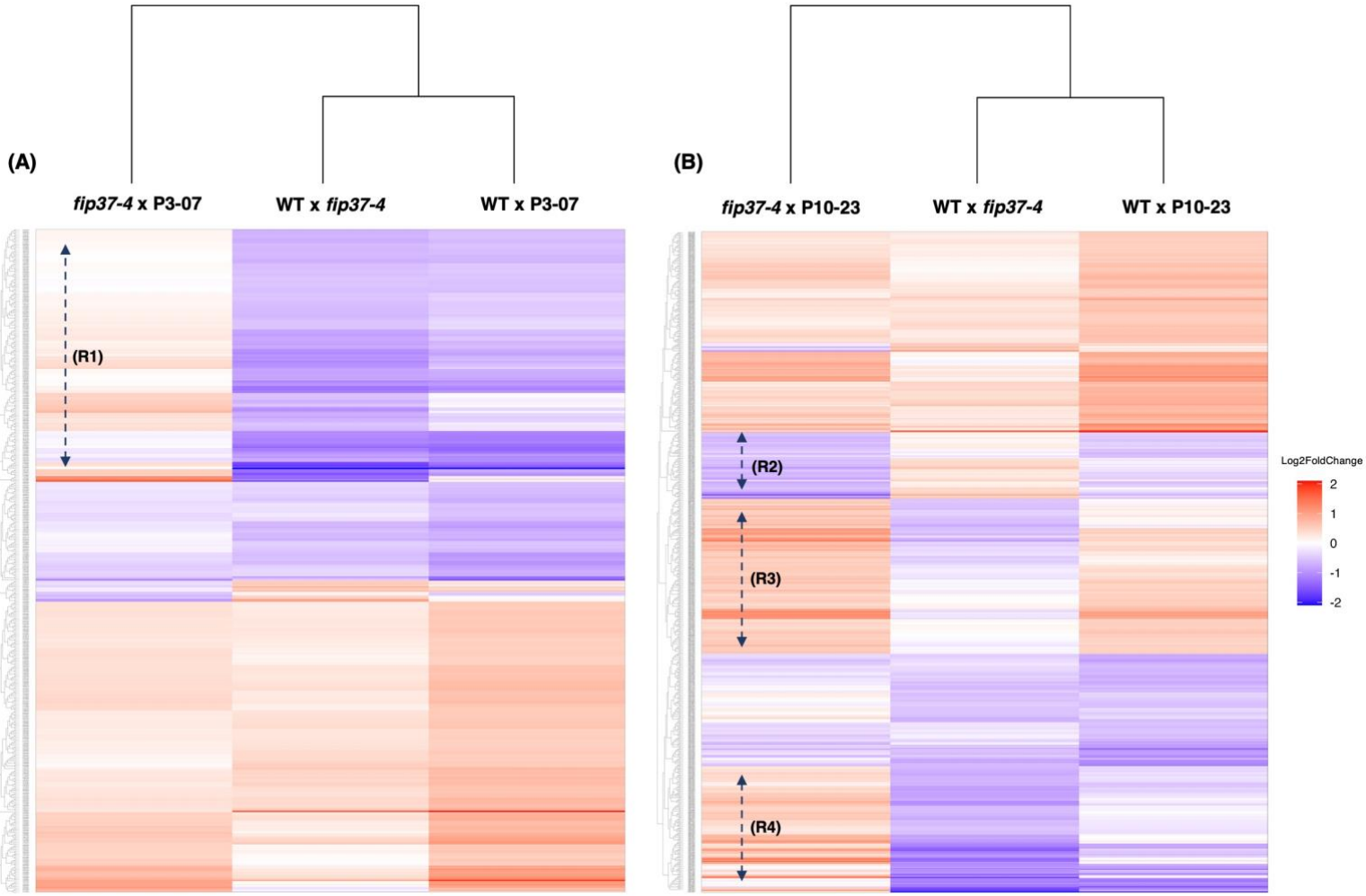


Figure 4.7 Heatmap analyses illustrating differential expression of significant genes in different comparison.

(A) Heatmap visualising DEG in three different comparison groups specific to to suppressor line P3-07. (B) Heatmap visualising DEG for three different comparison groups specific to suppressor line P10-23. Hierarchal clustering method used was Ward's minimum variance method. Distance clustering matrix measurement method used was Euclidean. Significant genes were identified with threshold of padj < 0.05, and Log2FoldChange > 1.5. Colour scale represents red being upregulated and blue being downregulated.

4.3.6 Visualisation of spliced junction reads

In order to analyse spliced junctions reads and alternative splicing events, the program IGV was utilised to generate sashimi plot to visualise individual splice junctions for each sample with three biological replicates (Figure 4.8 Figure 4.9). The T-DNA insertion is located between the 7th and 8th exons and RNA expression density is shown to increase upstream of the T-DNA. For *fip37-4*, the three biological samples show inclusion of exon 8 however, 2 samples did also show 17 junction reads to exclude exon 8. In P3-07 the same exclusion can be seen with 31 junction reads. This is consistent with P10-23 with 49 junction reads to exclude exon 8. This pattern indicates that *fip37-4* induces a different isoform with exon 8 exclusion. This exclusion of exon 8 is seen to increase in relative spliced junctions reads in P3-07 and P10-23 compared to *fip37-4* indicating increased number of exon exclusion. The splicing of the intron where the T-DNA is located could potentially influence an increase in the production of functional protein. To test if increased efficiency of splicing the T-DNA is occurring, junction reads between exon 7 and 8 from the sashimi plot were quantified. The results indicated no statistically significant change between the samples; however, junction reads did increase in both P3-07 and P10-23 by 47 and 93, respectively, compared to *fip37-4* (Figure 4.10).

4.3.7 DEXseq expression analysis

To visualise individual exon expression in *FIP37*, we utilised the program DEXseq, a method that visualises differential exon usage in RNA-seq data for three samples - P3-07, P10-23, and *fip37-4*. DEXseq implements a generalised linear model and accounts for biological variation to ensure reliable control of false discoveries (Anders *et al.*, 2012). Both P3-07 and P10-23 lines were inputted to DEXseq and found that exons 1-7 showed similar expression levels for both P3-07 and *fip37-4* Figure 4.10A. However, P3-07 exhibited increased expression compared to *fip37-4* in exons 8, 9, 10, and 11, whereas P10-23 showed an increase in expression for exons 1-7 compared to *fip37-4*, but an increase in *fip37-4* compared

to P10-23 for exons 8, 11, and 12 (Figure 4.10C). We observed a large increase in expression after the T-DNA insertion in P3-07 and P10-23, which is attributed to the presence of the 35S promoter in the intron that drives expression post-insertion (Figure 4.10A+C). The minimal increase in exon expression after exon 8 in P3-07 compared to *fip37-4* suggests an increase in splicing efficiency, although variations in promoter expression between different samples and lines should not be ruled out. Similarly, P10-23 showed an increase in exon expression between 1-7 compared to *fip37-4*, which may indicate increased splicing efficiency. In contrast, P10-23 exhibited a decrease in exon expression in exon 8,11, and 12, after the T-DNA insertion, which contradicts the hypothesis that suppression of P10-23 could be due to increase in splicing efficiency of the T-DNA.

4.3.8 Ratio of counts of correctly spliced junctions

To further analyse if splicing is causing suppression of *fip37-4*, the ratio of correctly spliced exon junctions from exon 7 to exon 8 was determined. The ratio was calculated by dividing the number of correctly spliced exon 7-8 reads by the total number of *FIP37* reads. First, all reads from chromosome three in the RNA-seq data were mapped to ENSEMBL Araport 11 and extracted into a new file. An 11bp sequence was taken from the end of exon 7 and a 12bp sequence was taken from the start of exon 8, and these were joined together to create CGTCAATGCAG-GCGAGGAGGTTG. The sequence was then transferred to the new file containing all reads on chromosome three, and the resulting count was used to calculate the ratio. If splicing efficiency is increased for the T-DNA in the intron, a higher ratio of counts would be observed in the suppressor line. However, the results show no statistically significant difference in ratio of counts for both P3-07 and P10-23 compared to *fip37-4*, suggesting that splicing efficiency is not altered (Figure 4.10D). The results did show that P10-23 has higher correctly spliced reads compared to *fip37-4*.

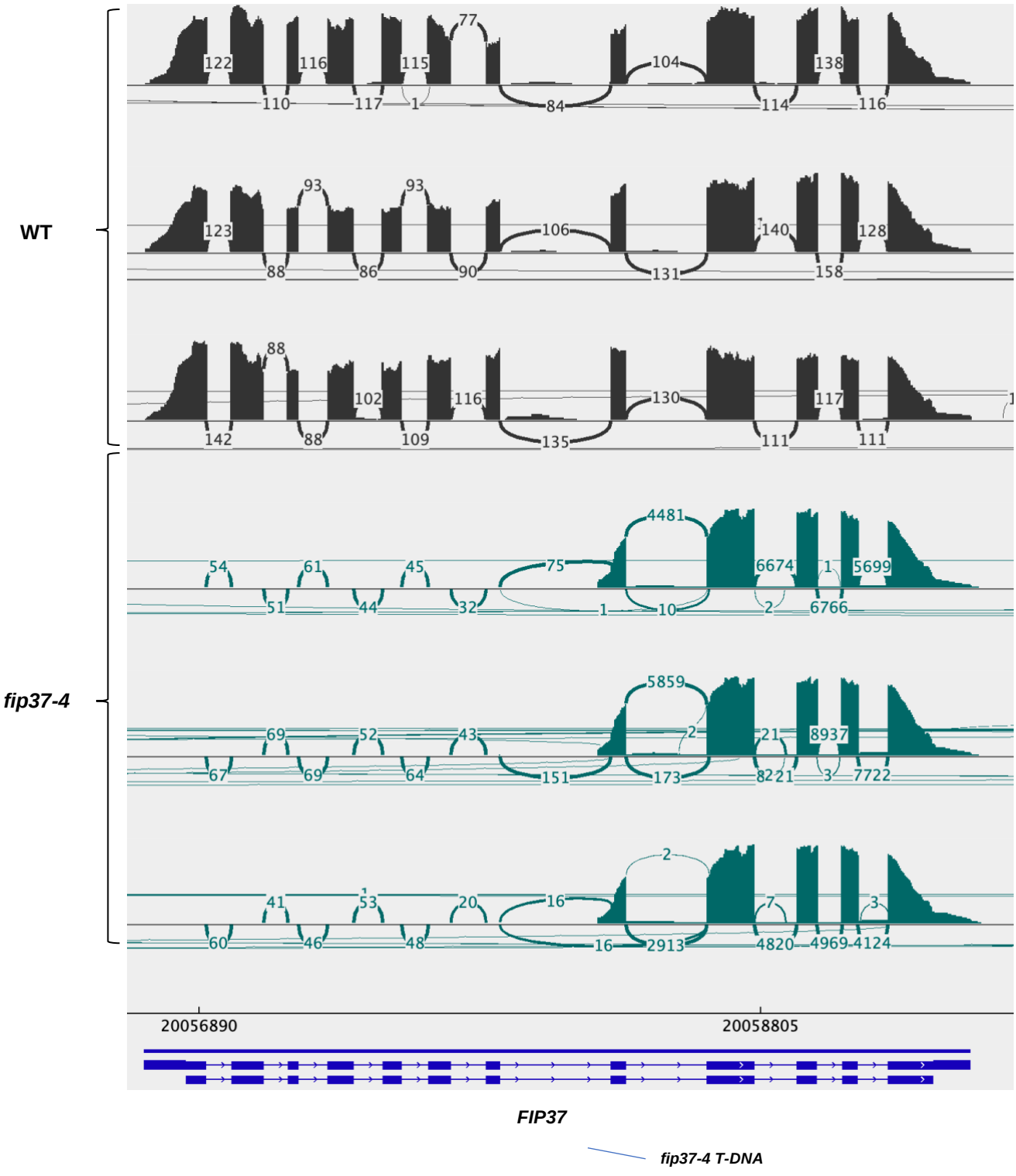


Figure 4.8 Sashimi plot displaying splice junction reads and alternative spliced events in WT and *fip37-4.*

The RNA-seq read densities along exons are shown as histograms, colour-coded by the sample for WT and *fip37-4*. The sequencing densities are aligned to the isoforms at the bottom of the plot-coloured blue. Junction reads are visualised as arcs connecting pairs of exons that the junction borders. The thickness of the arc is in proportion to the number of numbers of junction reads present in sample. Each sample is shown as three biological replicates.

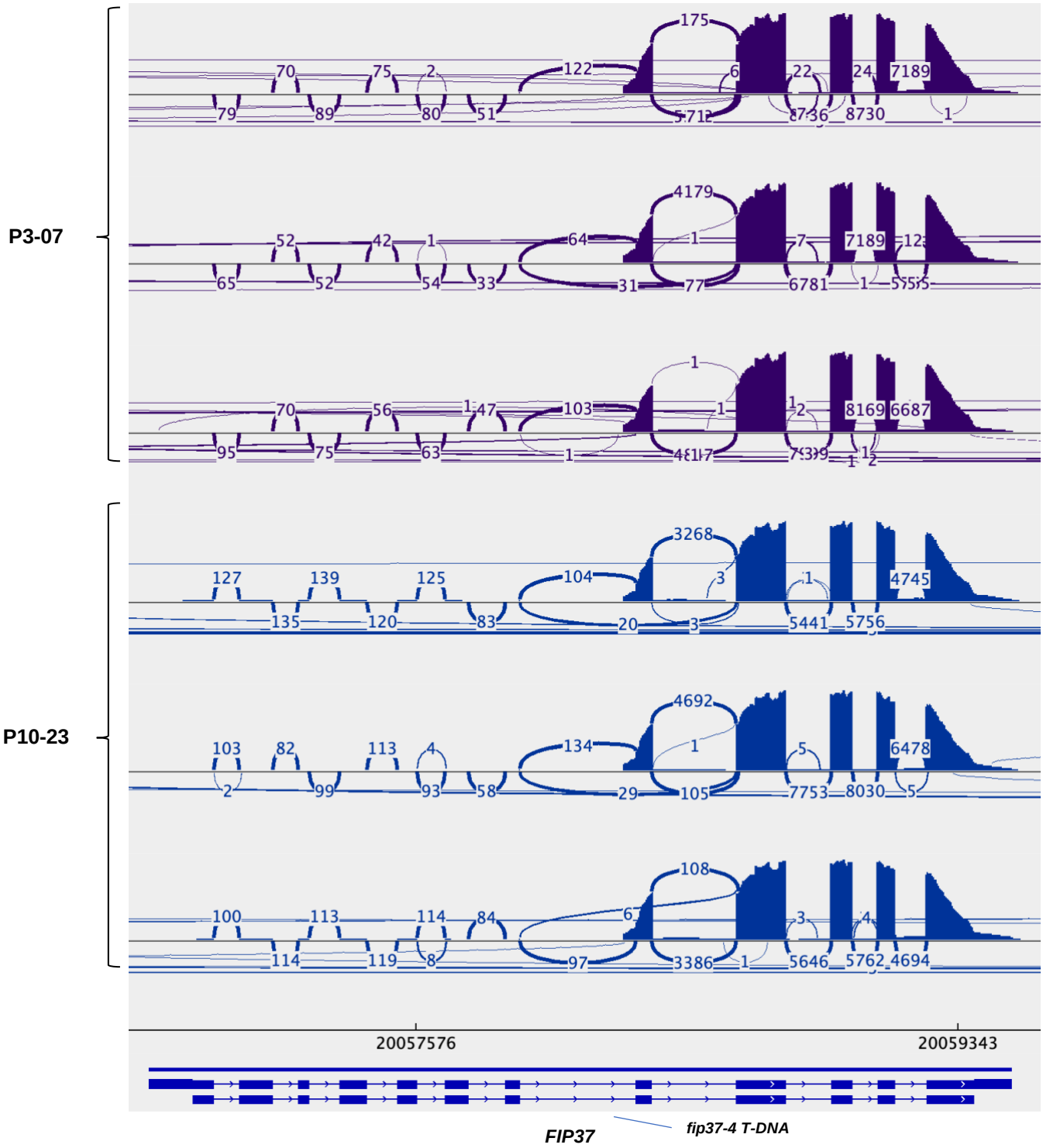


Figure 4.9 Sashimi plot displaying splice junction reads and alternative spliced events in P3-07 and P10-23.

The RNA-seq read densities along exons are shown as histograms, colour-coded by the sample for P3-07 and P10-23. The sequencing densities are aligned to the isoforms at the bottom of the plot-coloured blue. Junction reads are visualised as arcs connecting pairs of exons that the junction borders. The thickness of the arc is in proportion to the number of numbers of junction reads present in sample. Each sample is shown as three biological replicates.

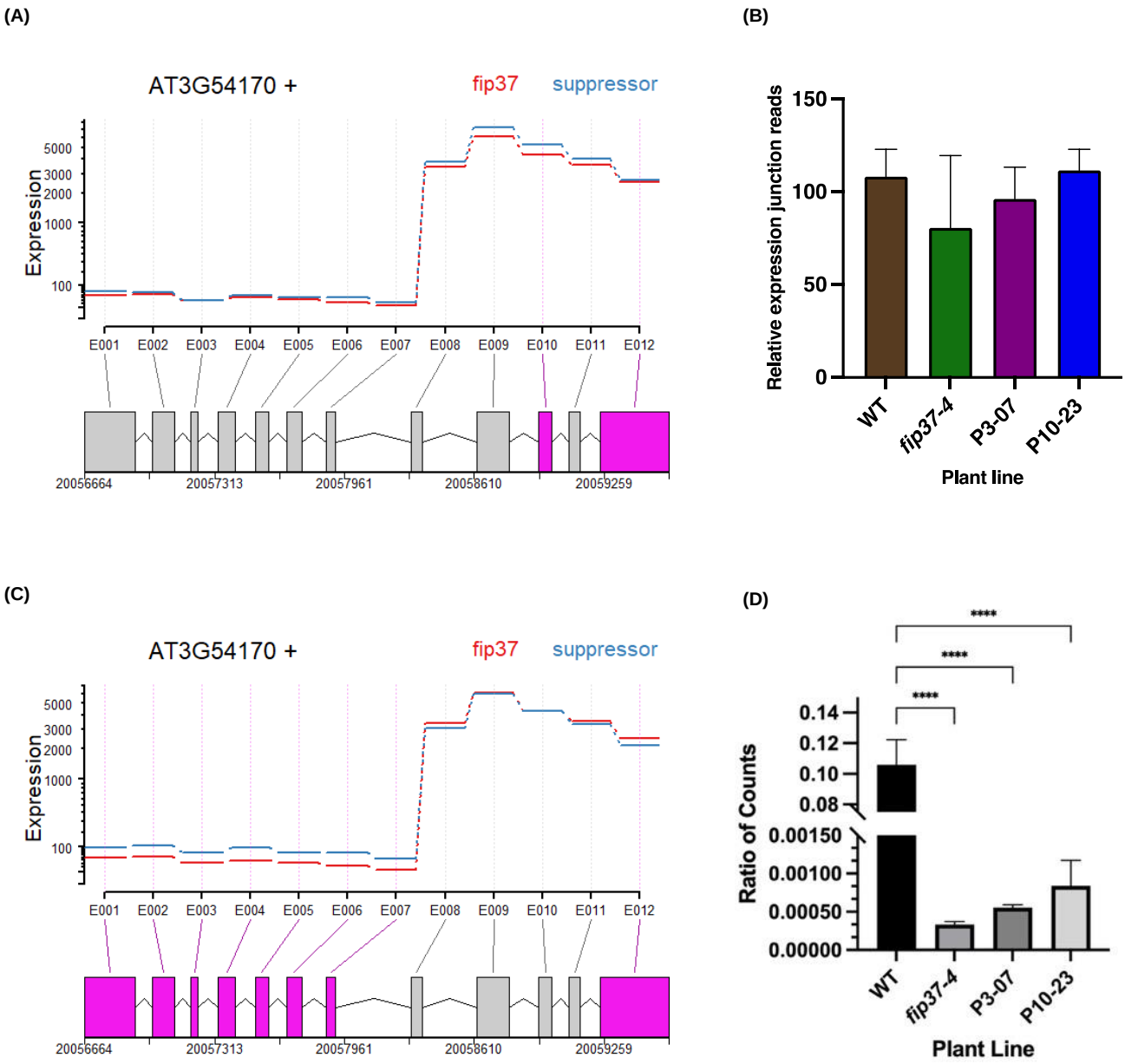


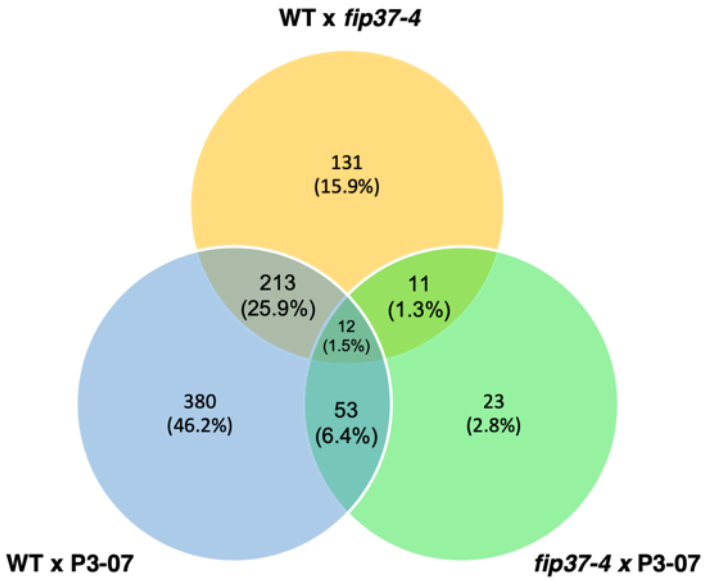
Figure 4.10 Splicing efficiency in suppressor line shows increases in exon and intron splicing.

(A) Detection of differential exon usage in AT3G54170 (*FIP37*) showing individual exon expression for P3-07 and *fip37-4*. (B) Relative expression of intron reads between exon 7-8. (C) Detection of differential exon usage in AT3G54170 (*FIP37*) in P10-23 and *fip37-4*. (D) Ratio of counts showing exon 7+8 junction connection. DEXseq was used for (A) + (B) was produced using RNA-seq data from P3-07, P10-23, and *fip37-4* using DEXseq from Ander *et al*., (2012). Error bars represent SEM; n = 3. * Indicates significant, **** $P < 0.0001$ using ANOVA.

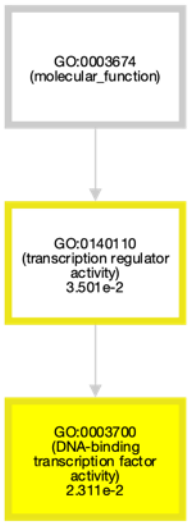
4.3.9 Visualisation of overlapping genes in P3-07 and P10-23 and GO analysis

In order to further expand the analysis from the heatmap, we analysed overlapping genes from three different groups for P3-07 and P10-23 using overlapping DEGs. Significant genes were filtered using a threshold value of padj < 0.05 and log2FC > 1.5. The analysis identified 12 overlapping genes between all three comparison groups for P3-07 which were, “WT X *fip37-4*, “WT x P3-07”, and “*fip37-4* x P3-07” shown in Figure 4.11A. Subsequent GO analysis identified DNA-binding transcription factor to be a key enrichment for the 12 genes identified (Figure 4.11C). This result was consistent with GO analysis performed on up-regulated and down-regulated significant DEGs in P3-07 compared to *fip37-4* comparison group, which also showed enrichment in DNA-binding transcription factor activity and response to stress and defence response in GO:MF and GO:BP, respectively (Figure 4.13A). For comparison groups “*fip37-4* x P10-23”, “WT x *fip37-4*”, and WT x “P10-23”, 26 overlapping genes were identified, and GO analysis performed on these 26 genes identified DNA-binding transcription factor activity as well as cellular response to hypoxia in GO:BP (Figure 4.12B). GO analysis on all significant DEGs up and down regulated in P10-23 compared to *fip37-4* showed enrichment in DNA-binding transcription factor activity and transmembrane transporter activities in GO:MF, and response to hypoxia, endogenous stimulus, and protein complex oligomerisation in GO:BP Figure 4.14. Differential expression analysis showed that in P3-07 there were 14 genes that were significantly up-regulated (log2FC > 1.5) shown in Table 4.1, whereas 84 genes were downregulated (log2FC > 1.5) shown in Table 4.2. In comparison, P10-23 had 142 genes up-regulated (log2FC > 1.5) shown in Table 4.3, and 407 genes down-regulated (log2FC > 1.5) (Table A9.2).

(A)



(B)

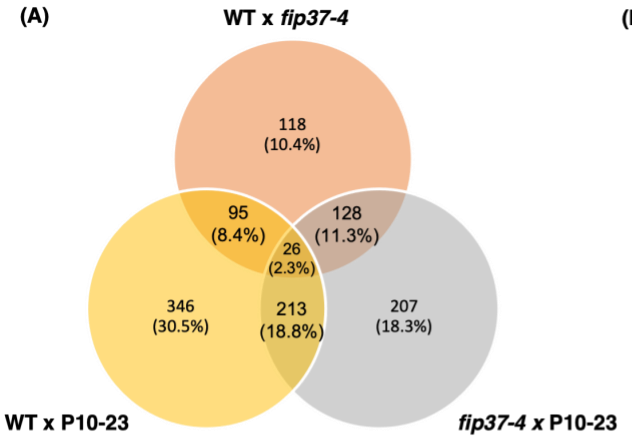


(C)

GO:MF		stats	
Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$ 0 ≤16
DNA-binding transcription factor activity	GO:0003700	2.311×10^{-2}	
transcription regulator activity	GO:0140110	3.501×10^{-2}	
transcription cis-regulatory region binding	GO:0000976	4.248×10^{-2}	
transcription regulatory region nucleic acid binding	GO:0001067	4.248×10^{-2}	
sequence-specific double-stranded DNA binding	GO:1990837	4.572×10^{-2}	

Figure 4.11 DEG across different comparison groups specific to P3-07 showing significant genes overlapping.

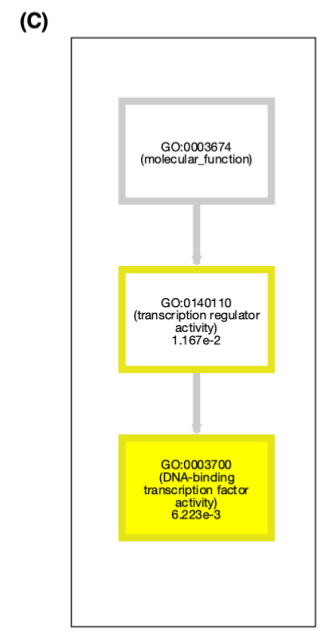
(A) Venn diagram illustrating significantly differentially expressed genes overlapping across three comparison groups. (B) GO context map of 12 DEG across all three comparisons showing groups linked and highlighted box shows lowest –log10(Padj). (C) GO analysis table showing 12 DEG across three groups. GO analysis was performed using g:SCS threshold $p < 0.05$ (Raudvere *et al*. 2019).



(B)

GO:MF		stats	
Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$
DNA-binding transcription factor activity	GO:0003700	6.223×10^{-3}	
transcription regulator activity	GO:0140110	1.167×10^{-2}	
transcription cis-regulatory region binding	GO:0000976	2.101×10^{-2}	
transcription regulatory region nucleic acid binding	GO:0001067	2.101×10^{-2}	
sequence-specific double-stranded DNA binding	GO:1990837	2.335×10^{-2}	
double-stranded DNA binding	GO:0003690	4.419×10^{-2}	
structural constituent of cell wall	GO:0005199	4.666×10^{-2}	

GO:BP		stats	
Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$
cellular response to hypoxia	GO:0071456	3.281×10^{-2}	
cellular response to decreased oxygen levels	GO:0036294	3.389×10^{-2}	
cellular response to oxygen levels	GO:0071453	3.389×10^{-2}	
defense response	GO:0006952	3.932×10^{-2}	
response to hypoxia	GO:0001666	4.959×10^{-2}	



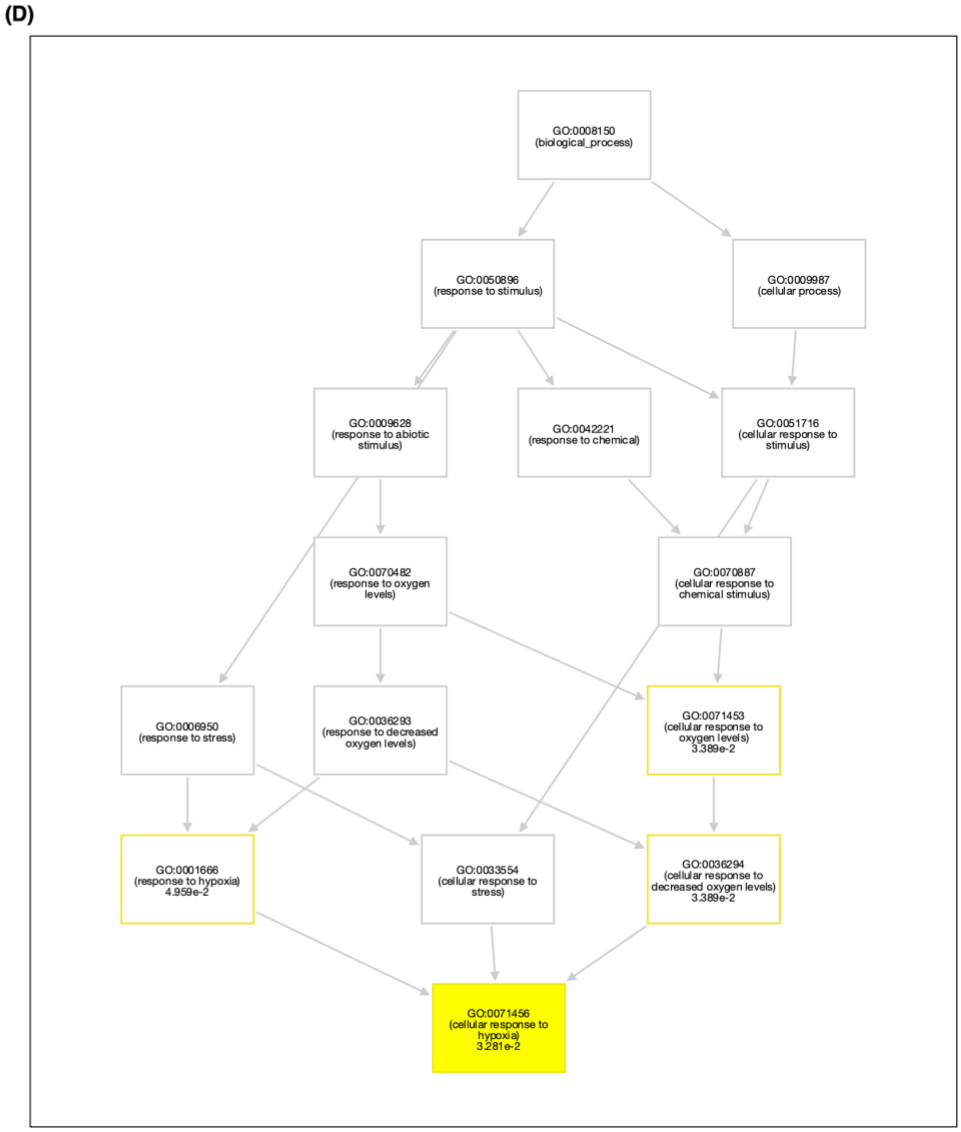


Figure 4.12 DEG across different comparison groups specific to P10-23 showing significant genes overlapping.

(A) Venn diagram illustrating significantly differentially expressed genes overlapping across three comparison groups. (B) GO analysis of 26 overlapping DEG across three comparison groups. (C) GO (MF) context showing groups linked with highlighted box showing lowest –log10(Padj). (C) GO (BF) context map showing groups linked with highlighted box showing lowest –log10(Padj). GO analysis was performed using g:SCS threshold $p < 0.05$ (Raudvere *et al*. 2019).

Table 4.1 Up-regulated genes in P3-07 compared to *fip37-4*.

Gene ID	log2FC	Adj p value	Gene Description
AT1G13608	2.61652	0.00000	Encodes a defensin-like (DEFL) family protein; iron deficiency-regulated gene expression
AT1G13609	1.54219	0.00288	Encodes a defensin-like (DEFL) family protein; iron deficiency-regulated gene expression
AT2G41240	2.10151	0.00812	Encodes a member of the basic helix-loop-helix transcription factor family protein. Functions as a key regulator of iron-deficiency responses independent of the master regulator FIT
AT3G02250	1.57560	0.00000	O-Fucosyltransferase family protein (RRT2)
AT3G21330	1.52632	0.00000	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT3G30720	1.58318	0.00000	QQS is an orphan gene that arose recently in the Arabidopsis thaliana lineage.
AT3G56970	1.98047	0.01277	Encodes a member of the basic helix-loop-helix transcription factor family protein. bHLH IVc transcription factor repressed by bHLH11. (BHLH38)
AT3G56980	1.61564	0.01252	Encodes a member of the basic helix-loop-helix transcription factor protein. bHLH IVc transcription factor repressed by bHLH11. BASIC HELIX-LOOP-HELIX 39 (BHLH39)
AT4G33970	1.51269	0.00471	Pollen expressed protein required for pollen tube growth. Along with other members of the LRX family, interacts with RALF4 to control pollen tube growth and integrity
AT4G35180	1.63632	0.04507	LYS/HIS TRANSPORTER 7 (LHT7)
AT5G02760	1.53184	0.00000	Encodes a phosphatase that functions in sustaining proper leaf longevity and preventing early senescence by suppressing or perturbing SARK-mediated senescence signal transduction
AT5G02865	1.81003	0.00273	Long non-coding RNA
AT5G48430	2.69258	0.01019	Eukaryotic aspartyl protease family protein
AT5G64750	1.81803	0.00078	Encodes a putative transcription factor containing an AP2 domain. Is a member of the ERF (ethylene response factor) subfamily B-4 of ERF/AP2 transcription factor family hypersensitive to these stresses

Table showing list of up-regulated DEG in P3-07 compared to *fip37-4* in RNA-seq analysis. The table include the gene ID, log2FC, Adj p-value, and gene description. The DEG genes was filtered with the threshold of log2FC > 1.5 and Adj p-value < 0.05.

Table 4.2 Down-regulated genes in P3-07 compared to *fip37-4*.

Gene ID	log2FC	Adj p value	Gene Description
AT1G03106	-1.68883	0.02234	Hypothetical protein
AT1G05510	-2.29405	0.00002	Protein is tyrosine-phosphorylated, and its phosphorylation state is modulated in response to ABA in *A. thaliana* seeds. OIL BODY-ASSOCIATED PROTEIN 1A
AT1G07135	-1.57880	0.01775	glycine-rich protein;
AT1G11610	-2.38829	0.05042	putative cytochrome P450 FAMILY 71, SUBFAMILY A, POLYPEPTIDE 18 (CYP71A18)
AT1G14930	-2.35645	0.02181	Polyketide cyclase/dehydrase and lipid transport superfamily protein
AT1G17810	-1.62147	0.00285	beta-tonoplast intrinsic protein (BETA-TIP)
AT1G18290	-1.63359	0.00161	PADRE protein up-regulated after infection by S. sclerotiorum
AT1G18960	-1.50656	0.00003	myb-like HTH transcriptional regulator family protein
AT1G24140	-1.62797	0.00045	Expression induced by fungal and bacterial pathogens
AT1G24145	-1.71563	0.00001	Transmembrane protein
AT1G24147	-2.05602	0.00003	Transmembrane protein
AT1G26680	-1.60896	0.01584	Transcriptional factor B3 family protein
AT1G27461	-2.77880	0.00252	Nuclear localized protein involved in osmotic stress tolerance. Transcriptional co-repressor which promotes embryonic-to-vegetative transition
AT1G27565	-3.96675	0.00033	Hypothetical protein
AT1G33760	-1.94565	0.00227	Encodes a member of the DREB subfamily A-4 of ERF/AP2 transcription factor family, ETHYLENE RESPONSE FACTOR022 (ERF022)
AT1G35210	-2.89751	0.00000	Hypothetical protein
AT1G47980	-2.31156	0.00182	Desiccation-like protein
AT1G57800	-2.49037	0.00135	predicted to encode a protein with an N-terminal PHD domain and two RING domains surrounding an SRA domain
AT1G67105	-2.84361	0.00000	Unknown
AT1G68250	-2.12439	0.00173	Hypothetical protein
AT1G70840	-2.00064	0.00306	MLP-like protein 31
AT1G72240	-1.72447	0.00033	Hypothetical protein
AT1G72910	-1.84054	0.00008	Nucleotide-binding, leucine-rich repeat (NLR) gene regulated by nonsense-mediated mRNA decay (NMD) genes UPF1 and UPF3
AT1G72920	-2.14147	0.00036	Toll-Interleukin-Resistance
AT1G75830	-2.74758	0.00044	Predicted to encode a PR pathogenesis-related protein. Belongs to the plant defensin
AT1G75945	-2.02627	0.00232	Hypothetical protein
AT2G01422	-3.26550	0.00002	Unknown
AT2G06995	-1.56765	0.00129	Unknown
AT2G15130	-1.82785	0.01392	Plant basic secretory protein

AT2G17040	-1.75571	0.00000	Member of the NAC transcription factor family and more specifically, the ONAC022 subfamily. Involved in leaf and inflorescence stem morphogenesis
AT2G21720	-1.84080	0.00580	ArgH (DUF639)
AT2G23240	-1.67790	0.00837	AtMT4b is a member of Type 4 metallothionein. It is involved in the early development of the embryo and in the accumulation of metal ions especially Zn in the seeds.
AT2G26150	-1.68418	0.00000	Member of Heat Stress Transcription Factor
AT2G27420	-1.53922	0.01530	Cysteine proteinases superfamily protein
AT2G29380	-1.77884	0.00534	Highly ABA-induced PP2C protein 3
AT2G32460	-2.16559	0.01777	Member of the R2R3 factor gene family. protein coding MYB DOMAIN PROTEIN 101
AT2G32680	-2.32874	0.00026	NLP20 LRR receptor protein involved in PAMP mediated immunity. RECEPTOR LIKE PROTEIN 23
AT2G33070	-1.75472	0.02373	Encodes a nitrile-specific protein NSP2
AT2G38465	-1.69824	0.00000	Hypothetical protein
AT2G38905	-1.68471	0.02798	Low temperature and salt responsive protein family
AT2G42000	-2.19050	0.02268	AtMT4a is a member of Type 4 metallothionein genes. It is involved in the early development of the embryo and in the accumulation of metal ions especially Zn in the seeds.
AT2G42560	-1.62012	0.01507	Late embryogenesis protein. Based on in vitro studies, likely plays a role in stabilising membranes in response to freezing stress
AT2G44840	-2.04860	0.00002	Encodes a member of the ERF ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family
AT3G01345	-2.13433	0.00000	Expressed protein;
AT3G05260	-2.00727	0.01186	NADP-binding Rossmann-fold superfamily protein
AT3G21720	-1.51601	0.00468	Encode a glyoxylate cycle enzyme isocitrate lyase
AT3G22500	-1.66992	0.04663	LATE EMBRYOGENESIS ABUNDANT PROTEIN (ATECP31)
AT3G22640	-2.09304	0.00000	Cupin family protein (PAP85)
AT3G42850	-1.83702	0.00077	Mevalonate/galactokinase family protein
AT3G46090	-1.52819	0.03446	C2H2 and C2HC zinc fingers superfamily protein (ZAT7)
AT3G56350	-1.50016	0.01082	Iron/manganese superoxide dismutase family protein
AT3G57260	-2.13426	0.00044	beta 1,3-glucanase protein coding PATHOGENESIS-RELATED PROTEIN 2
AT3G58450	-1.63679	0.00023	USP domain containing protein, member of the universal stress protein family, regulated by ABA and possibly regulated by the ABA-dependent transcription factor AREB/ABF
AT4G01140	-1.68914	0.00024	transmembrane protein, putative (DUF1191)
AT4G01970	-1.55857	0.01816	Encodes a raffinose and high affinity stachyose synthase as well as a stachyose and GoI specific galactosyl hydrolase enzyme activity
AT4G02520	-1.54008	0.00678	Encodes glutathione transferase belonging to the phi class of GSTs
AT4G04223	-1.57102	0.00000	Unknown
AT4G13395	-1.82097	0.00263	ROTUNDIFOLIA like 12
AT4G23140	-1.66095	0.01964	Arabidopsis thaliana receptor-like protein kinase

AT4G23810	-2.10637	0.00000	Member of WRKY Transcription Factor; Group III; t WRKY53 and CRK5 are antagonistic regulators of chlorophyll synthesis/degradation, senescence, and stomatal conductance
AT4G25140	-1.98916	0.00190	Encodes oleosin1, a protein found in oil bodies, involved in seed lipid accumulation. Suppression of OLEO1
AT4G25470	-1.55992	0.00218	Encodes a member of the DREB subfamily A-1 of ERF/AP2 transcription factor family
AT4G27654	-1.88835	0.00192	Transmembrane protein
AT4G28520	-2.20830	0.00306	Encodes a 12S seed storage protein that is tyrosine-phosphorylated, and its phosphorylation state is modulated in response to ABA in Arabidopsis thaliana seeds
AT4G31830	-1.50004	0.03176	Transmembrane protein
AT4G38560	-1.73023	0.00161	phospholipase-like protein
AT5G03210	-2.41260	0.02535	Encodes a small polypeptide contributing to resistance to potyvirus
AT5G03285	-1.99012	0.00000	Unknown
AT5G10760	-2.23542	0.00000	Eukaryotic aspartyl protease family protein - APOPLASTIC, EDS1-DEPENDENT 1 (AED1)
AT5G15290	-1.75183	0.00044	Uncharacterized protein family
AT5G22470	-2.13465	0.00050	PARP3 is one of three canonical PARPs in Arabidopsis
AT5G35660	-1.65243	0.01389	Glycine-rich protein family
AT5G40420	-1.58445	0.02946	Encodes oleosin2, a protein found in oil bodies, involved in seed lipid accumulation. Suppression of OLEO1
AT5G43770	-3.46116	0.00726	proline-rich family protein
AT5G45630	-2.07921	0.00025	Senescence regulator Protein of unknown function, DUF584
AT5G49525	-1.79516	0.03375	Transmembrane protein
AT5G52020	-1.58420	0.00796	Encodes a member of the DREB subfamily A-4 of ERF/AP2 transcription factor family. The protein contains one AP2 domain
AT5G52750	-1.61631	0.00000	Heavy metal transport/detoxification superfamily protein
AT5G52760	-1.54969	0.00561	Copper transport protein family
AT5G54165	-1.57368	0.00265	Avr9/Cf-9 rapidly elicited protein
AT5G54740	-3.17647	0.00131	Seed storage albumin 5
AT5G55240	-1.73680	0.02073	Catalyse hydroperoxide-dependent mono-oxygenation reactions - ARABIDOPSIS THALIANA PEROXYGENASE 2 (ATPXG2)
AT5G61600	-1.54647	0.00005	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family.
AT5G65165	-2.03045	0.00607	One of three isoforms of the iron-sulphur component of the succinate dehydrogenase complex, a component of the mitochondrial respiratory chain complex II.

Table showing list of down-regulated DEG in P3-07 compared to *fip37-4* in RNA-seq analysis. The table include the gene ID, log2FC, Adj p-value, and gene description. The DEG genes was filtered with the threshold of log2FC > 1.5 and Adj p-value < 0.05.

Table 4.3 List of Up-regulated DEG in P10-23 compared to *fip37-4*.

Gene ID	log2FC	Adj p value	Gene Description
AT1G02340	1.64160	0.00000	Encodes a light-inducible, nuclear bHLH protein involved in phytochrome signalling
AT1G04560	1.72095	0.00276	AWPM-19-like family protein
AT1G04660	1.50867	0.00000	Glycine-rich protein
AT1G05510	1.87506	0.00000	Protein is tyrosine-phosphorylated, and its phosphorylation state is modulated in response to ABA in Arabidopsis thaliana seeds
AT1G06123	3.40576	0.03955	Long non-coding RNA
AT1G06160	2.67544	0.00292	Encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family
AT1G07483	2.27933	0.05355	Long non-coding RNA
AT1G07645	1.75231	0.00034	Ortholog of HC205/ Xhdsi-1voc from Xerophyta humilis. Member of VOC metalloenzyme superfamily
AT1G14930	1.68885	0.00721	Polyketide cyclase/dehydrase and lipid transport superfamily protein
AT1G17010	2.66431	0.00004	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT1G17810	2.21582	0.00000	beta-tonoplast intrinsic protein (beta-TIP) mRNA
AT1G23200	1.56470	0.00000	ProPME pectin methyl esterase involved in embryo development, stomatal function, mucilage release
AT1G25422	1.85213	0.02190	Hypothetical protein
AT1G26233	3.55518	0.05754	Encodes a H/ACA-box snoRNA SMALL NUCLEOLAR RNA95 (SNOR95)
AT1G27990	1.65976	0.00350	Transmembrane protein
AT1G28650	3.80678	0.00002	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or hydrolase reactions with lipid and non-lipid substrates
AT1G29090	1.71831	0.01810	Cysteine proteinases superfamily protein
AT1G29450	1.61896	0.00000	SAUR-like auxin-responsive protein family (SAUR64)
AT1G29490	2.21766	0.04358	SAUR-like auxin-responsive protein family (SAUR68)
AT1G29680	2.28494	0.00043	histone acetyltransferase (DUF1264)
AT1G43780	2.25081	0.05394	SERINE CARBOXYPEPTIDASE-LIKE 44 (scpl44)
AT1G48130	1.71635	0.00025	Encodes a protein similar to the 1-cysteine (1-Cys) peroxiredoxin family of antioxidants. Expression is limited to seed (aleurone and embryo) and is not induced by ABA or drought.
AT1G48530	2.44720	0.05377	Proteasome inhibitor-like protein
AT1G54860	2.12696	0.00097	Glycoprotein membrane precursor GPI-anchored
AT1G54870	2.05088	0.00004	ChlADR is an aldehyde reductase that catalyses the reduction of the aldehyde carbonyl groups on saturated and alpha,beta-unsaturated aldehydes with more than 5 carbons in vitro
AT1G62510	1.53669	0.00008	Expressed in the root cortex
AT1G65090	1.64915	0.00106	Protein of unknown function that is found on the surfaces of lipid droplets and may function to anchor the droplets to the plasma membrane SEED LIPID DROPLET PROTEIN1 (SLDP1)
AT1G65480	2.10661	0.02287	FLOWERING LOCUS T, FT, REDUCED STEM BRANCHING 8, RSB8
AT1G69120	2.50123	0.05628	Floral homeotic gene encoding a MADS domain protein homologous to SRF transcription factors. Specifies floral meristem and sepal identity.
AT1G71250	1.85836	0.00406	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyse acyltransferase or hydrolase reactions with lipid and non-lipid substrates
AT1G72760	1.50614	0.04931	Protein kinase superfamily protein
AT1G72960	2.61456	0.01649	Root hair defective 3 GTP-binding protein (RHD3)
AT1G73190	2.49163	0.00000	ALPHA-TONOPLAST INTRINSIC PROTEIN (ALPHA-TIP) Moves to the Protein Storage Vacuole in a Golgi independent manner
AT1G73830	2.09619	0.00000	Encodes the brassinosteroid signalling component BEE3 (BR-ENHANCED EXPRESSION 3). Positively modulates the shade avoidance syndrome in Arabidopsis seedlings
AT1G74675	1.71992	0.01919	Transmembrane protein
AT1G75945	3.75720	0.00000	Hypothetical protein
AT1G77100	3.17320	0.00035	Peroxidase superfamily protein
AT1G77950	2.85424	0.00000	Cooperates with the histone mark reader EBS to modulate seed germination under high temperature AGAMOUS-LIKE 67 (AGL67)

AT2G05580	2.67931	0.00158	Glycine-rich protein family
AT2G07629	2.09923	0.00009	Hypothetical protein
AT2G07739	1.61006	0.00010	Ycf1 protein
AT2G09195	2.92478	0.00233	Long non-coding RNA
AT2G15010	1.83736	0.00006	Predicted to encode a PR (pathogenesis-related) protein. Belongs to the plant thionine (PR-13) family
AT2G19320	1.76442	0.05192	Hypothetical protein
AT2G21720	2.06192	0.00001	ArgH (DUF639)
AT2G23640	2.02075	0.00034	Encodes RTNLB13, a reticulon protein integral to the endoplasmic reticulum (ER) membrane that can shape the ER into tubules
AT2G27380	3.70789	0.00000	Encodes an extensin like gene involved in seed germination EXTENSIN PROLINE-RICH 1 (EPR1)
AT2G28490	2.02649	0.00008	RmlC-like cupins superfamily protein
AT2G30010	1.52800	0.00000	Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain.
AT2G33070	2.03069	0.00139	Encodes a nitrile-specifier protein NSP2
AT2G33520	1.95841	0.01596	cysteine-rich/transmembrane domain protein A (ATHCYSTM7)
AT2G34315	3.33330	0.01241	A virulence induced family protein
AT2G36690	1.90328	0.00189	Protein belonging to the Fe-dependent 2-oxoglutarate dioxygenase superfamily, catalyses the stereospecific hydration of GA12 to produce DHGA12
AT2G44470	2.72329	0.00953	BETA GLUCOSIDASE 29 (BGLU29)
AT2G46970	1.73213	0.00000	Encodes a novel Myc-related bHLH transcription factor, which physically associated with APRR1/TOC1 and is a member of PIF3 transcription factor family. PHYTOCHROME INTERACTING FACTOR 3-LIKE 1 (PIL1)
AT3G03620	2.03104	0.00014	MATE efflux family protein
AT3G03830	1.66618	0.00008	SAUR-like auxin-responsive protein family (SAUR28)
AT3G05260	1.54866	0.00455	NAD(P)-binding Rossmann-fold superfamily protein
AT3G06655	1.58430	0.05602	Long non-coding RNA
AT3G09590	1.64551	0.00109	CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein) superfamily protein
AT3G20210	2.92375	0.00000	Encodes a vacuolar processing enzyme with caspase-1-like activity that is specifically expressed in inner integument of developing seeds
AT3G21330	1.87001	0.00000	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT3G21370	2.64542	0.00000	BETA GLUCOSIDASE 19 (BGLU19)
AT3G21380	2.15144	0.00000	Mannose-binding lectin superfamily protein
AT3G21500	2.39904	0.01982	Encodes a protein that has very high sequence similarity to 1-deoxy-D-xylulose-5-phosphate synthase proteins but does not possess appreciable activity in vitro
AT3G22490	1.96789	0.00288	Atrab28 plays a role in the ion cell balance during late embryogenesis and germination RESPONSIVE TO ABSCISIC ACID 28 (ATRAB28)
AT3G22500	1.67829	0.00363	late embryogenesis abundant (LEA) protein. LATE EMBRYOGENESIS ABUNDANT PROTEIN ECP31 (ATECP31)
AT3G24650	1.53082	0.00237	Homologous to the maize transcription factor Viviparous-1. Full length ABI3 protein binds to the highly conserved RY motif [DNA motif CATGCA(TG)], present in many seed-specific promoters
AT3G25260	1.92930	0.00092	Major facilitator superfamily protein (ATNPF4.1); NRT1/ PTR FAMILY 4.1 (NPF4.1)
AT3G26200	1.90006	0.05738	CYTOCHROME P450, FAMILY 71, SUBFAMILY B, POLYPEPTIDE 22 (CYP71B22)
AT3G49970	1.62643	0.00280	Phototropic-responsive NPH3 family protein
AT3G52561	2.60229	0.04136	Hypothetical protein
AT3G53040	1.85674	0.00003	Late embryogenesis abundant protein, putative / LEA protein
AT3G53250	2.17547	0.00163	SAUR-like auxin-responsive protein family
AT3G55500	1.50727	0.00249	Expansin-like protein
AT3G56350	1.94946	0.00001	Iron/manganese superoxide dismutase family protein (MSD2)
AT3G58450	1.70523	0.00000	USP domain containing protein, member of the universal stress protein family, regulated by ABA and possibly regulated by the ABA-dependent transcription factor AREB/ABF
AT4G02690	2.00012	0.00063	Stress induced membrane protein. Mutants show enhanced cell death under stress LIFEGUARD 3 (LFG3)
AT4G09600	1.88776	0.00000	One of GASA gene family which is related to a GA-stimulated transcript (GAST) from tomato
AT4G10010	3.86172	0.05214	Protein kinase superfamily protein
AT4G10020	2.22541	0.00486	Encodes a putative hydroxysteroid dehydrogenase (HSD)
AT4G10160	1.68798	0.01825	RING/U-box superfamily protein

AT4G11040	2.25912	0.02004	Encodes a nuclear localized protein with sequence similarity to PP2C phosphatases that is involved in seed dormancy
AT4G11310	1.77491	0.00157	cysteine proteinase precursor-like protein CYSTEINE PROTEASE 1 (CP1)
AT4G12890	2.42163	0.00489	Gamma interferon responsive lysosomal thiol (GILT) reductase family protein
AT4G15990	2.45835	0.00000	Hypothetical protein
AT4G16160	1.80914	0.00002	Homologous to pea OEP16 and barley pPORA (OEP16), a member of Arabidopsis OEP16 family
AT4G21020	1.56345	0.00678	Late embryogenesis abundant protein (LEA) family protein
AT4G21870	1.77477	0.00006	HSP20-like chaperone
AT4G23030	1.58458	0.00493	MATE efflux family protein
AT4G25140	2.06502	0.00017	Encodes oleosin1, a protein found in oil bodies, involved in seed lipid accumulation
AT4G26740	1.62874	0.00111	Encodes caleosin, a 27-kDa protein found within seed lipid bodies
AT4G27150	2.46767	0.02878	Seed storage albumin protein SEED STORAGE ALBUMIN 2 (SESA2)
AT4G28520	2.37599	0.00033	Encodes a 12S seed storage protein that is tyrosine-phosphorylated, and its phosphorylation state is modulated in response to ABA in Arabidopsis thaliana seeds CRUCIFERIN 3 (CRU3)
AT4G31380	2.45706	0.00000	Encodes a small protein with unknown function and is similar to flower promoting factor 1 FPF1-LIKE PROTEIN 1 (FLP1)
AT4G32280	1.98831	0.00001	Indole-3-acetic acid inducible 29 protein involved in IAA signalling. Downstream target of PIF4. INDOLE-3-ACETIC ACID INDUCIBLE 29 (IAA29)
AT4G33970	3.58212	0.00000	Pollen expressed protein required for pollen tube growth. Along with other members of the LRX family, interacts with RALF4 to control pollen tube growth and integrity
AT4G34770	1.55573	0.00000	SAUR-like auxin-responsive protein family (SAUR1)
AT4G34790	1.59687	0.00905	Putative OXS2-binding DEGs were constitutively activated by OXS2. SMALL AUXIN UPREGULATED RNA 3 (SAUR3)
AT4G39130	2.43120	0.00003	Dehydrin family protein
AT5G02760	2.24439	0.00000	Encodes a phosphatase that functions in sustaining proper leaf longevity and preventing early senescence by suppressing or perturbing SARK-mediated senescence signal transduction ARABIDOPSIS PP2C CLADE D 7 (APD7)
AT5G02865	1.93301	0.00032	Long non-coding RNA
AT5G04500	2.24476	0.03153	Encodes a member of the CAZy Glycosyltransferase Family 64 that is involved in glycosylinositolphosphorylceramide and sphingolipid glycosylation
AT5G05290	3.07876	0.00000	Encodes an expansin EXPANSIN A2 (EXPA2)
AT5G05965	2.23499	0.01435	Cell wall RBR3-like protein
AT5G09780	6.07909	0.00013	Encodes a member of the REM (Reproductive Meristem) gene family, a part of the B3 DNA-binding domain superfamily REPRODUCTIVE MERISTEM 25 (REM25)
AT5G09970	1.64824	0.00000	Member of CYP78A family. Paralog of CYP78A5 and appears to function in a shoot meristem maintenance pathway with LAMP1 that parallels AMP1/CYP87A5
AT5G10120	2.78637	0.00001	Ethylene insensitive 3 family protein (EIL4)
AT5G10140	2.92883	0.00002	MADS-box protein encoded by FLOWERING LOCUS C - transcription factor that functions as a repressor of floral transition and contributes to temperature compensation of the circadian clock
AT5G14070	1.72553	0.00012	Encodes a member of the CC-type glutaredoxin (ROXY) family that has been shown to interact with the transcription factor TGA2 and suppress ORA59 promoter activity
AT5G16023	2.97830	0.00000	Encodes a plant peptide that could be involved in the coordination of socket cell development in wild-type plants ROTUNDIFOLIA LIKE 18 (RTFL18)
AT5G16460	1.70713	0.05938	Membrane protein involved in lipid droplet biogenesis primarily in embryos SEIPIN1 (SEIPIN1)
AT5G18060	1.50675	0.00000	SAUR-like auxin-responsive protein family (SAUR23)
AT5G18450	1.96345	0.00621	Encodes a member of the DREB subfamily A-2 of ERF/AP2 transcription factor family. The protein contains one AP2 domain.
AT5G23990	2.51514	0.03120	Encodes a ferric chelate reductase that is expressed at low levels in roots,shoots and flowers, but not cotyledons
AT5G27220	1.69155	0.00420	Frigida-like protein
AT5G35660	2.33084	0.00003	Glycine-rich protein family
AT5G40420	2.05694	0.00055	Encodes oleosin2, a protein found in oil bodies, involved in seed lipid accumulation
AT5G44120	2.59379	0.00000	Encodes a 12S seed storage protein. The Landsberg erecta genome contains another copy of 12S globulin gene, CRA2, which is located tandemly with CRA1. CRUCIFERINA (CRA1)
AT5G44620	2.07721	0.00168	CYTOCHROME P450, FAMILY 706, SUBFAMILY A, POLYPEPTIDE 3 (CYP706A3)
AT5G45690	2.23599	0.00000	Histone acetyltransferase (DUF1264)
AT5G45820	1.76228	0.00000	Encodes a CBL-interacting serine/threonine protein kinase comprised of an N-terminal kinase catalytic domain similar to SNF1/AMPK and a unique C-terminal regulatory domain.
AT5G46330	1.88303	0.00000	Encodes a leucine-rich repeat serine/threonine protein kinase that is expressed ubiquitously

AT5G48430	2.57481	0.00591	Eukaryotic aspartyl protease family protein
AT5G51760	1.59032	0.00217	Encodes AHG1 (ABA-hypersensitive germination 1), a putative protein phosphatase 2C (PP2C). Expressed in seeds. AHG1 functions in seed development and germination
AT5G54000	2.00228	0.00128	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT5G55135	1.96777	0.05178	Hypothetical protein
AT5G55410	2.59115	0.00000	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G55750	2.14409	0.00072	Hydroxyproline-rich glycoprotein family protein
AT5G57260	2.67193	0.00001	putative cytochrome P450
AT5G59170	2.77269	0.00000	Proline-rich extensin-like family protein
AT5G60220	1.50449	0.00009	Member of TETRASPANIN family (TET4)
AT5G62280	2.00838	0.00000	DUF1442 family protein (DUF1442)
AT5G65100	2.70970	0.00001	Ethylene insensitive 3 family protein
AT5G65800	1.86712	0.00030	1-aminocyclopropane-1-carboxylate synthase (ACS) is encoded by a multigene family consisting of at least five members whose expression is induced by hormones, developmental signals, and protein synthesis inhibition
AT5G66430	3.94722	0.05136	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
ATMG00140	2.56247	0.00330	Transmembrane protein (ORF167)
ATMG00560	1.55305	0.00428	Encodes a mitochondrial ribosomal protein L2, a constituent of the large subunit of the ribosomal complex (RPL2)
ATMG00680	4.97157	0.00641	Transmembrane protein (ORF122C)
ATMG00690	2.69241	0.00000	FO-ATPase subunit (ORF240A)
ATMG01090	1.54361	0.00003	ATP synthase 9 mitochondrial (ORF262)
ATMG01350	1.94636	0.04068	Transmembrane protein (ORF145C)

Table showing list of up-regulated DEG in P10-23 compared to *fip37-4* in RNA-seq analysis. The table include the gene ID, log2FC, Adj p-value, and gene description. The DEG genes was filtered with the threshold of log2FC > 1.5 and Adj p-value < 0.05.

(A)

GO:MF

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_{adj})$ (0 – ≤16)
DNA-binding transcription factor activity	GO:0003700	1.287×10^{-2}	
transcription regulator activity	GO:0140110	3.662×10^{-2}	

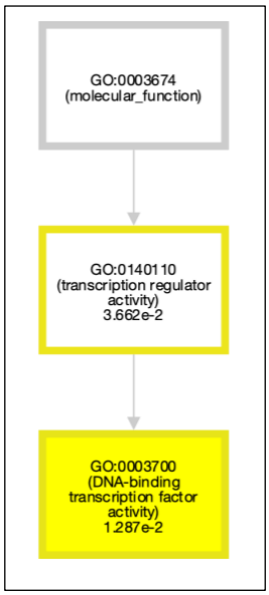
GO:BP

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_{adj})$ (0 – ≤16)
response to stress	GO:0006950	1.017×10^{-7}	
response to stimulus	GO:0050896	1.762×10^{-6}	
defense response	GO:0006952	1.792×10^{-5}	
cellular response to iron ion starvation	GO:0010106	8.238×10^{-4}	
response to external stimulus	GO:0009605	1.046×10^{-3}	
response to other organism	GO:0051707	1.676×10^{-3}	
response to external biotic stimulus	GO:0043207	1.676×10^{-3}	
seed oilbody biogenesis	GO:0010344	1.968×10^{-3}	
biological process involved in interspecies interaction b...	GO:0044419	2.028×10^{-3}	
response to biotic stimulus	GO:0009607	2.140×10^{-3}	
cellular response to hypoxia	GO:0071456	7.782×10^{-3}	
cellular response to oxygen levels	GO:0071453	8.213×10^{-3}	
cellular response to decreased oxygen levels	GO:0036294	8.213×10^{-3}	
defense response to other organism	GO:0098542	1.079×10^{-2}	
response to hypoxia	GO:0001666	1.551×10^{-2}	
response to decreased oxygen levels	GO:0036293	1.706×10^{-2}	
response to oxygen levels	GO:0070482	1.747×10^{-2}	
cellular response to chemical stimulus	GO:0070887	1.929×10^{-2}	

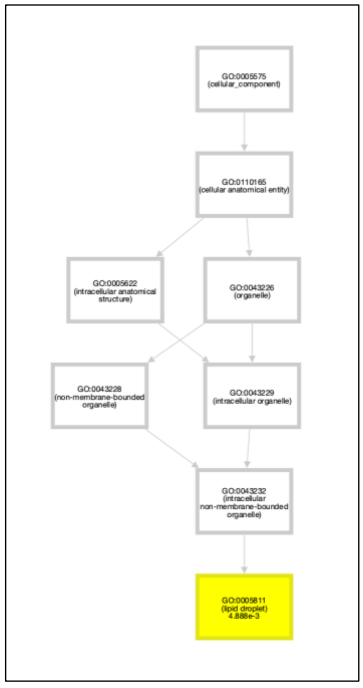
GO:CC

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_{adj})$ (0 – ≤16)
lipid droplet	GO:0005811	4.888×10^{-3}	

(B)



(C)



(D)

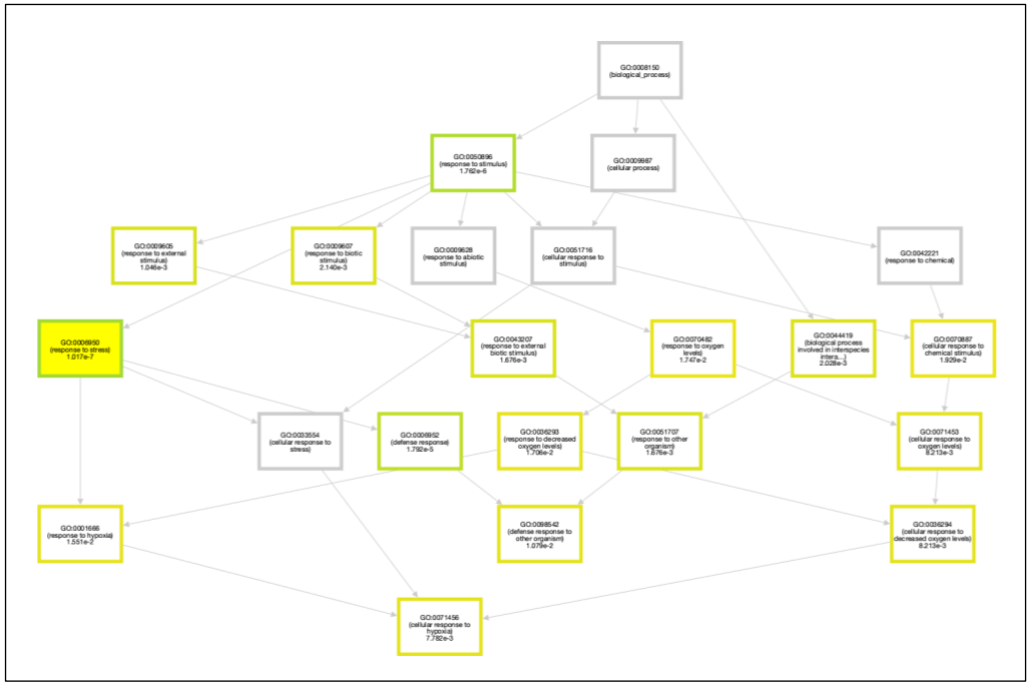
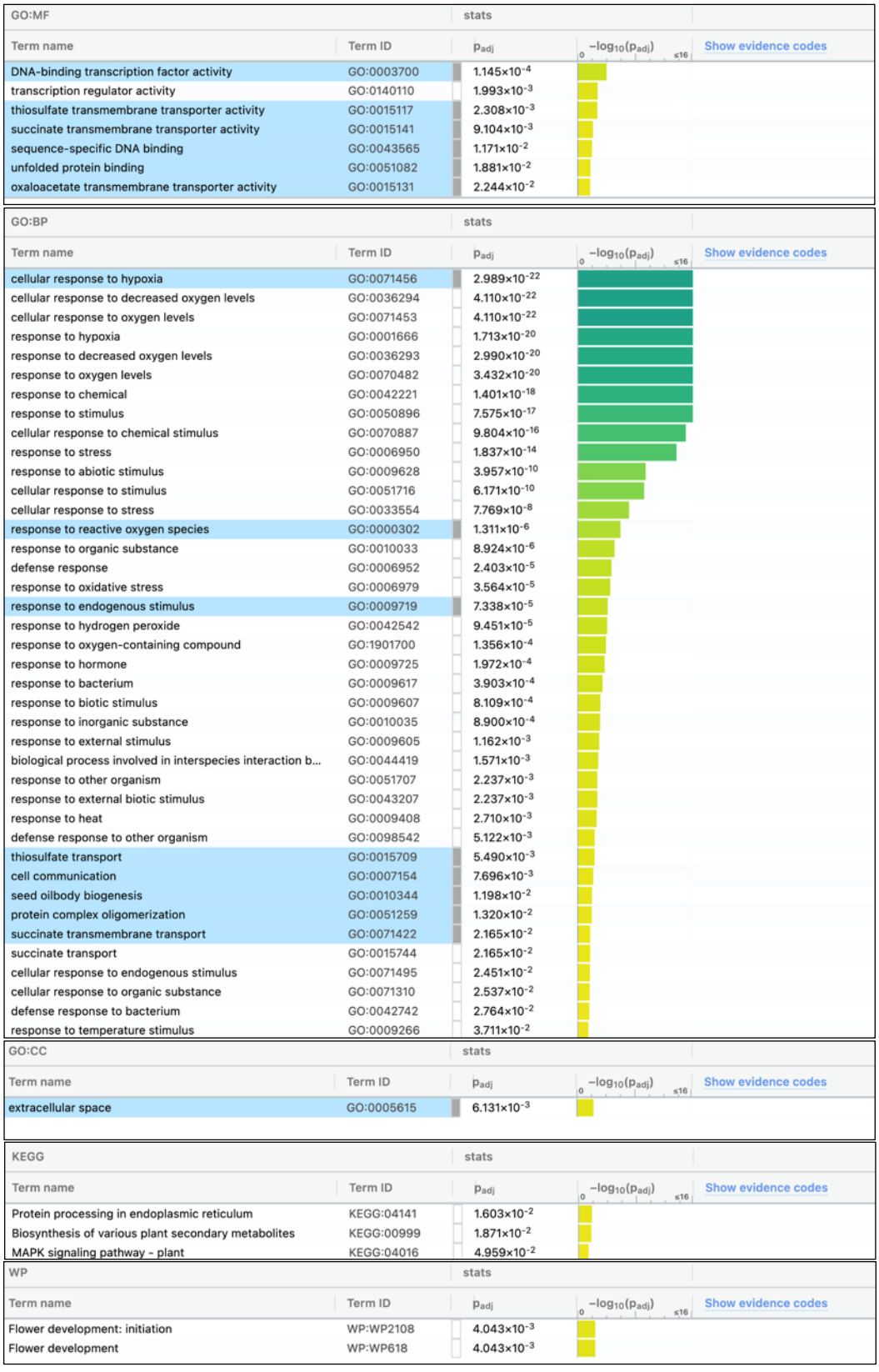


Figure 4.13 GO analysis of significant DEG in P3-07 compared to *fip37-4*.

(A) GO analysis of P3-07 compared to *fip37-4* showing GO:MF, GO:BP, GO:CC. Highlighted pathways shows driver terms in GO enrichment results in G:Profiler. This approach ensures multiple leading terms are considered rather than selecting the term with the highest significance value. (B) GO:MF context showing highlighted box as most significant. (C) GO:CC context showing highlighted box as most significant. (D) GO:BP context showing highlighted box as most significant. GO analysis was performed using g:SCS threshold $p < 0.05$ (Raudvere *et al*. 2019).



GO:MF

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	Show evidence codes
DNA-binding transcription factor activity	GO:0003700	1.145×10^{-4}		
transcription regulator activity	GO:0140110	1.993×10^{-3}		
thiosulfate transmembrane transporter activity	GO:0015117	2.308×10^{-3}		
succinate transmembrane transporter activity	GO:0015141	9.104×10^{-3}		
sequence-specific DNA binding	GO:0043565	1.171×10^{-2}		
unfolded protein binding	GO:0051082	1.881×10^{-2}		
oxaloacetate transmembrane transporter activity	GO:0015131	2.244×10^{-2}		

GO:BP

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	Show evidence codes
cellular response to hypoxia	GO:0071456	2.989×10^{-22}		
cellular response to decreased oxygen levels	GO:0036294	4.110×10^{-22}		
cellular response to oxygen levels	GO:0071453	4.110×10^{-22}		
response to hypoxia	GO:0001666	1.713×10^{-20}		
response to decreased oxygen levels	GO:0036293	2.990×10^{-20}		
response to oxygen levels	GO:0070482	3.432×10^{-20}		
response to chemical	GO:0042221	1.401×10^{-18}		
response to stimulus	GO:0050896	7.575×10^{-17}		
cellular response to chemical stimulus	GO:0070887	9.804×10^{-16}		
response to stress	GO:0006950	1.837×10^{-14}		
response to abiotic stimulus	GO:0009628	3.957×10^{-10}		
cellular response to stimulus	GO:0051716	6.171×10^{-10}		
cellular response to stress	GO:0033554	7.769×10^{-8}		
response to reactive oxygen species	GO:0000302	1.311×10^{-6}		
response to organic substance	GO:0010033	8.924×10^{-6}		
defense response	GO:0006952	2.403×10^{-5}		
response to oxidative stress	GO:0006979	3.564×10^{-5}		
response to endogenous stimulus	GO:0009719	7.338×10^{-5}		
response to hydrogen peroxide	GO:0042542	9.451×10^{-5}		
response to oxygen-containing compound	GO:1901700	1.356×10^{-4}		
response to hormone	GO:0009725	1.972×10^{-4}		
response to bacterium	GO:0009617	3.903×10^{-4}		
response to biotic stimulus	GO:0009607	8.109×10^{-4}		
response to inorganic substance	GO:0010035	8.900×10^{-4}		
response to external stimulus	GO:0009605	1.162×10^{-3}		
biological process involved in interspecies interaction b...	GO:0044419	1.571×10^{-3}		
response to other organism	GO:0051707	2.237×10^{-3}		
response to external biotic stimulus	GO:0043207	2.237×10^{-3}		
response to heat	GO:0009408	2.710×10^{-3}		
defense response to other organism	GO:0098542	5.122×10^{-3}		
thiosulfate transport	GO:0015709	5.490×10^{-3}		
cell communication	GO:0007154	7.696×10^{-3}		
seed oilbody biogenesis	GO:0010344	1.198×10^{-2}		
protein complex oligomerization	GO:0051259	1.320×10^{-2}		
succinate transmembrane transport	GO:0071422	2.165×10^{-2}		
succinate transport	GO:0015744	2.165×10^{-2}		
cellular response to endogenous stimulus	GO:0071495	2.451×10^{-2}		
cellular response to organic substance	GO:0071310	2.537×10^{-2}		
defense response to bacterium	GO:0042742	2.764×10^{-2}		
response to temperature stimulus	GO:0009266	3.711×10^{-2}		

GO:CC

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	Show evidence codes
extracellular space	GO:0005615	6.131×10^{-3}		

KEGG

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	Show evidence codes
Protein processing in endoplasmic reticulum	KEGG:04141	1.603×10^{-2}		
Biosynthesis of various plant secondary metabolites	KEGG:00999	1.871×10^{-2}		
MAPK signaling pathway - plant	KEGG:04016	4.959×10^{-2}		

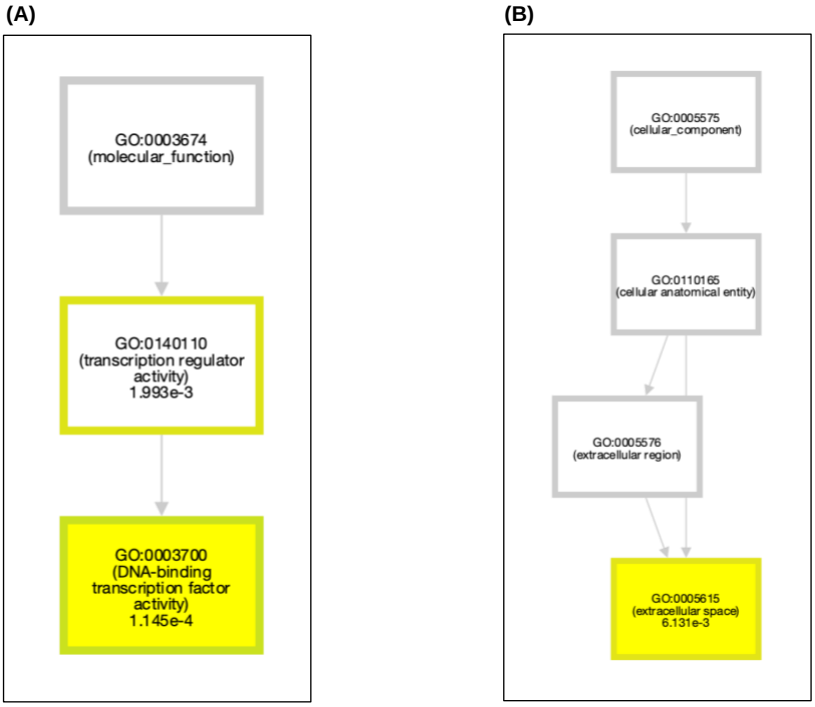
WP

stats

Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$	Show evidence codes
Flower development: initiation	WP:WP2108	4.043×10^{-3}		
Flower development	WP:WP618	4.043×10^{-3}		

Figure 4.14 GO analysis of significant DEG in P10-23 compared to *fip37-4*.

(GO analysis of P10-23 compared to *fip37-4* showing GO:MF, GO:BP, GO:CC, KEGG pathway, and WP. Highlighted pathways shows driver terms in GO enrichment results in G:Profiler. This approach ensures multiple leading terms are considered rather than selecting the term with the highest significance value. GO analysis was performed using g:SCS threshold $p < 0.05$ (Raudvere *et al*. 2019).



(C)

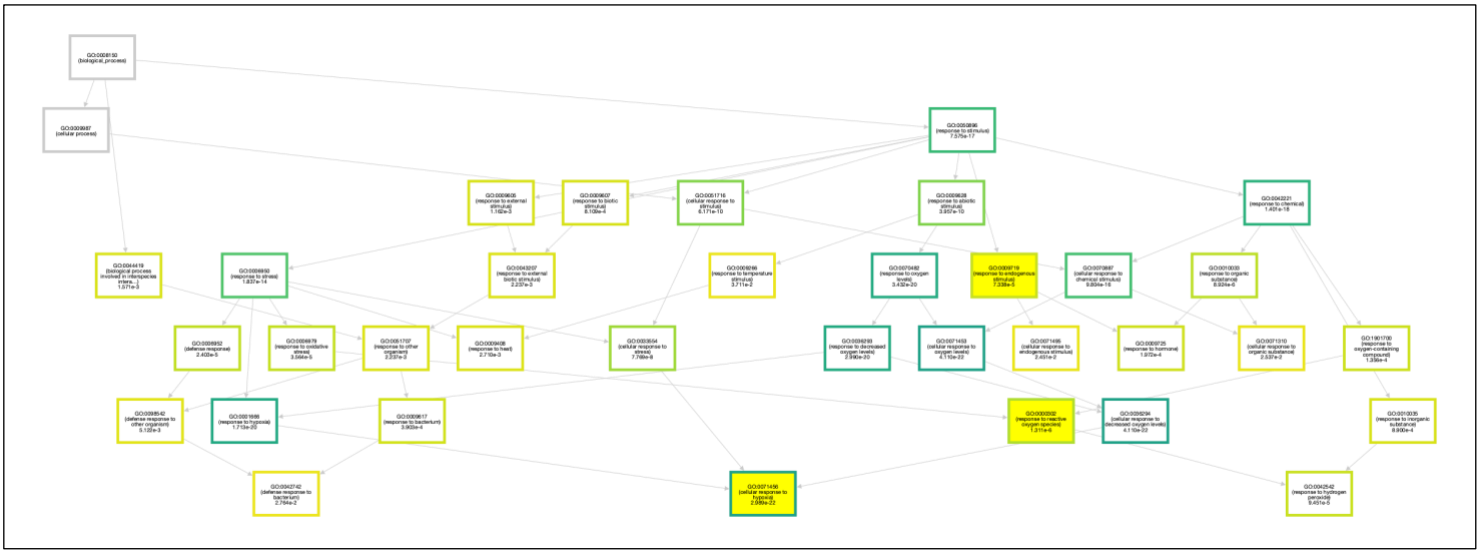


Figure 4.15 GO context on GO:MF, GO:BP, and GO CC.

(A) GO (MF) context showing groups linked with highlighted box showing lowest –log10(Padj). (B) GO (CC) context showing groups linked with highlighted box showing lowest –log10(Padj). (C) GO (BP) context showing groups linked highlighted box showing lowest -log10(Padj). GO analysis was performed using g:SCS threshold $p < 0.05$ (Raudvere *et al*. 2019).

4.4 Discussion

The process of RNA modification in specific m^6A methylation performed by the multiunit writer complex proteins adds an additional layer of complexity to the transcriptome. To obtain a deeper understanding of the results obtained in chapter three and to narrow down the candidate list generated from BSA analysis, RNA-seq was employed. The pipeline utilised for processing the raw RNA-seq data from the suppressor lines P3-07 and P10-23 was selected to ensure efficient and accurate processing of data. The sequencing quality check was performed using FastQC revealed an average Phred score of 35, indicating high quality sequencing. FastQC analysis also demonstrated that trimmed data had less than 0.1% adapter content, and that more than 92% of paired-end (PE) reads were uniquely mapped Figure 4.2.

In Figure 4.6, a volcano plot depicts all significant genes in the samples based on the RNA-seq data. Interestingly, AT3G54170 (*FIP37*) was shown to have a increase in expression of ~ 4.8 log2FC. However, this increase in expression does not reflect the T-DNA insertion line (SALK_018636) used in this study as a knock-down line. To explain this substantial increase in expression, a coverage analysis was performed to show RNA-seq densities along the FIP37 transcripts (Figure 4.3). The densities demonstrated that expression in *fip37-4* increased downstream of the T-DNA insertion between exon 7 and 8, which is explained by the presence of the 35S promoter in the T-DNA insertion that drives expression after the T-DNA (Figure 4.3). This finding is consistent with P3-07 and P10-23 as both suppressor lines have T-DNA present. For *fip37-4*, P3-07, and P10-23, expression is not present from exon 1-7 indicating efficient knock-down of FIP37 (Figure 4.3). The increase in expression density seen after exon 7 is detected by RNA-seq counts program HISAT2 as read counts are picked up by reads aligned to the genome even if only half of the transcript is expressed, resulting in *fip37-4* and suppressor lines showing up-regulation of gene. DEG analysis was performed on P3-07 samples, resulting in three comparison groups as presented in Figure 4.11A. The comparison

groups *fip37-4* x P3-07 (FP3) and *fip37-4* x P10-23 (FP10) were able to identify key genes involved in suppressing *fip37-4* expression. The Venn diagram analysis revealed that only FP3 had 23 unique genes in all three groups compared to the other two groups, while there were 12 overlapping genes among all three groups (Figure 4.11A). GO analysis was performed on these 12 genes, which identified DNA-binding transcription factors and transcription regulators as the most significant factors in GO:MF (Figure 4.11). These results are consistent with the candidate list generated in Chapter 3 BSA analysis, where most of the mutations were found to be upstream or downstream gene variants, suggesting the possible involvement of transcription regulatory pathways could be in play at suppressing *fip37-4*.

A possible candidate gene identified from the DEG analysis showed a 1.5-fold increase of AT3G02250, which is a member of the O-Fucosyltransferase family of proteins, specifically RRT2. Previous studies by Takenaka *et al*., (2018) have shown that *A. thaliana* has four RRTs, including RRT1, which is a transmembrane protein located in the Golgi apparatus. The RRTs are a novel glycosyltransferase family with plant-specific roles, such as biosynthesis of plant cell wall polysaccharides. This connection with pectin synthesis and transmembrane protein is consistent with the identification of GoSAMTs genes in the BSA analysis of P3-07 in Chapter 3, which could indicate that transmembrane proteins specific to SAM import located in the Golgi apparatus may play a role in the suppression of *fip37-4*. Although information on the transcription factors regulating the expression of RRT2 is limited, analysis on the *A. thaliana* eFP browser showed that RRT2 is predominantly expressed in the leaf and flowers (Takenaka *et al.*, 2018).

4.4.1 Transcription factor activity

Epigenetic regulation of gene expression in plants can allow adaptation to environmental changes through histone modifications (Zhao *et al.*, 2019). Although histone-modifying enzymes themselves do not have the ability to bind to DNA, they may function with various

transcription factors together with diverse histone modifiers such as histone methyltransferases (Zhou *et al.*, 2020), histone acetyltransferases (Boycheva *et al.*, 2014), and histone demethylase (Crevillén, 2020), to regulate specific target genes.

GO enrichment analysis on significant DEG in both P3-07 and P10-23 identified DNA binding transcription factor activity as the most significant GO term. There is a possibility that DNA-binding transcription factor and histone modification could be involved in the suppression of *fip37-4*. One possible idea could be that the candidate gene is involved in recruiting histone modifying to promote transcription of certain genes or increase H3K36me3 mark that suppresses *fip37-4*. Linking this idea back to current research, there are studies in human and mouse cells which suggest a possible link between histone H3 trimethylation at Lys36 (H3K36me3), m^6A deposition, and transcription elongation (Huang *et al.*, 2019). H3K36me3 is an epigenetic modification of the DNA packaging protein Histone H3, which has been found to be enriched with m^6A modification within the region of H3K36me3 peaks m^6A modifications are co-transcriptionally deposited in regions of H3K36me3 peaks, and are globally reduced when cellular H3K36me3 is depleted (Huang *et al.*, 2019). The human homologue of MTB, METTL14, is a crucial component in recognising and binding to H3K36me3, which facilitates the binding of the methyltransferase complex to adjacent RNA polymerase II and the subsequent deposition of m^6A (Huang *et al.*, 2019). This increase in m^6A methylation in both P3-07 and P10-23 shown in chapter 3 could be enough to suppress *fip37-4* phenotype. Further research is needed to understand the precise mechanism of m^6A transcript selection and regulation in plants and its relationship with histone modifications. The uncovering of H3K36me3 as an epigenetic modification in plants adds another layer of complexity to the regulation of m^6A deposition and the role of the methyltransferase complex. The finding that individual components of the methyltransferase complex have specific roles further emphasises the intricacies of this mechanism. Investigating histone modifications in the suppressor lines could potentially uncover novel pathways related to m^6A methylation and provide further insight into the regulatory mechanisms involved.

4.4.2 Response to stress

Further GO enrichment analysis revealed that both P3-07 and P10-23 significant DEGs showed terms related to defence response and response to stress. Previous research have demonstrated that m^{6}A methylation plays a crucial role in regulating mRNA stability (Wang *et al.*, 2014; Huang *et al.*, 2018), translation efficiency (Wang *et al.*, 2015), and various physiological processes in plants, such as root vascular formation (Růžička *et al.*, 2017) and apical dominance (Bodi *et al.*, 2012). Given the vast array of roles that m^{6}A methylation in plant development, the m^{6}A machinery may develop distinct mechanisms to cope with biotic and abiotic stress responses in *A. thaliana*. GO enrichment analysis of significant DEG in P3-07 compared to *fip37-4* revealed that response to stress and defence response were the top enriched terms in GO:BP (Figure 4.13A). Similarly, in P10-23 compared to *fip37-4*, the top GO enrichments were cellular response to hypoxia and response to stress (Figure 4.14). Suppressor mutants P3-07 and P10-23 could have induced stress response genes and the plants could have activated stress response pathways to suppress *fip37-4*. The stress related pathways identified were consistent with findings from Zhang *et al.*, (2022), as the author shows that GO analysis displayed enrichment in response to salt stress (GO:0009651) and regulation of defence response (GO:0031347) to be included in the dataset for *fip37-4*. Further analysis and experiments may be required to identify genes that could rescue low methylation plants back to a normal WT-like phenotype.

4.4.2.1 Alternative splicing and splicing efficiency

Emerging evidence suggests that m^{6}A modification may have a crucial role in RNA splicing (Liu *et al.*, 2015; Haussmann *et al.*, 2016; Lence *et al.*, 2016). The results from the sashimi plot analysis indicate that alternative splice events are taking place in the suppressor lines, with exon 8 exclusion observed in *fip37-4*, P3-07, and P10-23 when compared to WT (Figure

4.8 Figure 4.9). In chapter 3, the northern blot analysis revealed an increase in band intensity below the indicated WT transcripts of *FIP37* for both lines P3-07 and P10-23 compared to *fip37-4*, suggesting that different isoforms of FIP37 was picked up by the northern blot and may contribute to a functional FIP37 (Figure 3.9). Previous studies have demonstrated that m^6A, in conjunction with the nuclear binding protein YTHDC1, promotes exon inclusion in targeted mRNAs by recruiting pre-mRNA splicing factor SRSF3 and blocking SRSF10 mRNA binding (Xiao *et al.*, 2016). This suggests that reader proteins may be involved in the regulation of splicing.

4.4.3 Summary and future work

The RNA-seq analysis provided a comprehensive view of the transcriptome activity of both WT, *fip37-4* and suppressor mutants P2-04 and P10-23. The suppressor mutants exhibited significant differentially expressed genes related to DNA-binding transcription activity and response to stress. However, the specific genes or mechanisms responsible for the suppression of *fip37-4* in P3-07 and P10-23 are still unclear. Further backcrossing of suppressor lines with *fip37-4* is required to narrow down the candidate list. In Chapter 3, two lists of candidate genes were generated using different filtering methods based on the mutation type for P3-07 and P10-23. The suppressor line P3-07 produced a candidate list primarily comprising mutations located in the upstream and downstream gene variant regions. By cross-referencing the expression data from RNA-seq in Chapter 4 with the candidate list produced in Chapter 3, we can potentially eliminate genes that are involved in the suppression of fip37-4, especially if they are present in the upstream and downstream gene variant regions. This cross-referencing of both datasets can also lead to the prioritisation of certain chromosomal regions as potential regions of interest. Additionally, investigation into stress response genes and specific histone modification proteins could reveal new pathways related to individual methyltransferase proteins or m^6A methylation in general.

Chapter 5

Characterisation of *MTA* Writer Protein in *Solanum lycopersicum*

5 Characterisation of MTA writer protein in *Solanum lycopersicum*

5.1 Chapter Aims

The objective of work described in this chapter was to generate mutant writer complex lines in Tomato. The main aims were to:

- Generate knock-out / knock-down lines of *MTA* in tomato using a CRISPR/Cas9 and inducible hairpin RNA interference (hpRNAi) system.
- Characterise *mta* mutants in different growth conditions such as growth media and compost.
- Confirm transcript expression levels by performing RT-qPCR on mutant line.

5.2 Introduction

Tomato (*Solanum lycopersicum*) is a global food crop and continues to be of increasing importance for consumption as a fresh and processed crop. Tomato is a member of the Solanaceae family, which includes other key crops such as potato (Solanum tuberosum), eggplant (*Solanum melongena L*.) and pepper (*Capsicum annuum L*.). Apart from consumption, it is also a key model plant for researching the fundamental principles of growth and development due to a relatively small genome size of approximately 950Mb. The emerging economic importance, and the ability to use molecular techniques such as genetic engineering, recombinant inbred lines and along with the high-quality annotated sequence database has allowed tomato to be recognised as a key model crop for genetic studies (Tomato Genome Consortium. 2012; Fernandez-Pozo *et al.*, 2015). Moreover, tomato is also an excellent plant for studying fruit development, ripening, and quality traits, which have

important implications for food security and human health. Food security is still a challenging issue facing society as reports have stated that there are almost 770 million people undernourished in 2021 worldwide (FAO, 2022). Thus, research in tomato using molecular biology provides a platform for elucidating fundamental biological processes in plants and developing new approaches to improve crop yields and quality.

The epitranscriptomic modification on mRNA named m^6A is deposited by a multi-protein writer complex termed the methyltransferase complex (MTC), which is responsible for adding modifications to eukaryotic mRNAs. In recent years, numerous studies have investigated this modification in mammalian and plant systems, specifically in *A. thaliana* (Zhong *et al.*, 2008; Liu *et al.*, 2014; Růžička *et al.*, 2017; Zhang *et al.*, 2022). To date, there are currently no known knock-out (KO) or knock-down (KD) lines for each of the MTC proteins in tomato giving rise to the idea of producing viable KO or KD lines in tomato. The characterisation of m^6A writers is essential for improving our understanding of m^6A methylation and could have the opportunity to uncover new pathways or mechanisms related to this prevalent modification. Recently, it has been reported that global m^6A levels increase during tomato fruit expansion from immature green to mature green stage indicating that m^6A methylation could be further explored in fruit ripening (Hu *et al.*, 2022a).

RNA-binding proteins (RBPs) play a crucial role in the function of the m^6A, as they bind to m^6A modified cellular RNAs and mediate downstream biological functions of the methylation modification that the best studied m^6A readers contain a YT521-B homology (YTH) domain. Plant genomes encode both YTH Domain Containing (YTHDC) and YTH Domain Family (YTHDF) type proteins. Although some research has been undertaken regarding the 13 YTH proteins in *A. thaliana* mainly ECT1 to ECT4, little is known about the function of YTH proteins in tomato. Previous studies reported five YTHDC proteins and four YTHDF proteins in the tomato genome. Previous work by Cameron Grundy generated viable single homozygous KO and heterozygous double KO for each of the YTHDF proteins in tomato which is described in later chapters. Phylogenetic analysis performed has stated that the 11 *A. thaliana* YTHDFs

fall into 4 classes. These 4 classes are the same for other higher plants (e.g. maize and rice) and each class contains multiple members of YTHDF proteins in differing numbers which is presented in chapter 6 Figure 6.2. Tomato presents an ideal model as it has four classes but only contain one member of YTHDF in each. The 11 *A. thaliana* YTHDFs proteins could be represented by each of the tomato YTHDF1 to YTHDF4. The availability of viable tomato KO lines for each group of YTHDF1 to YTHDF4 proteins presents an opportunity to investigate a double KO/KD of both writer and reader protein functions. This mutant could further our understanding of how m^6A methylation functions. Characterisation of each writer protein could shed light on their interactions with m^6A readers and unravel the underlying molecular mechanisms for both types of proteins. The generation of both writer and reader KO lines in tomato could significantly advance our understanding of the interplay between these two types of proteins.

5.3 Results

5.3.1 CRISPR/Cas9 sgRNA and vector design for *MTA*, *MTB*

The initial approach that was taken to generate KO lines for both MTA and MTB involved the use of a CRISPR/Cas9 construct to induce frameshift mutations. The CRISPR constructs were assembled utilising a two-level golden gate cloning system (Weber *et al.*, 2011; Nekrasov *et al.*, 2013). Constructs containing *MTA* and *MTB* sgRNAs were generated and confirmed using Sanger sequencing to ensure that the correct sequences were inserted into each stage and subsequently the final vector. *Agrobacterium*-mediated transformation was performed for both *MTA* and *MTB* CRISPR/Cas9 constructs on more than 1000 cotyledons and stems of *S. lycopersicum*, resulting in the generation of 0 transgenic lines. Subsequently, alternative sgRNAs were redesigned for both *MTA* and *MTB*, and transformation was repeated for both targets of *MTA* and *MTB* with new CRISPR/Cas9 constructs. However, additional transformation of more than 1000 cotyledons and stems did not generate any transgenic lines.

5.3.2 Design of an hpRNAi vector for generating knock-down lines of MTA

A different approach was taken to generate KD of writer proteins instead of KO. An inducible RNA interference (RNAi) silencing approach was chosen, which allows for the application of dexamethasone-inducible RNAi against plant genes as reported by Wielopolska *et al.*, (2005). This vector system enables the production of RNAi-inducing hairpin RNA, which can be regulated by the application and removal of dexamethasone (DEX). A 456 bp PCR amplified fragment was generated using Taq Polymerase from the full-length genomic sequence of *MTA* (SOL Network – Solyc08g066730.3 (+ strand)) to generate an 'A' base overhang for cloning into the entry vector PCR8 (Figure 5.1). Subsequent PCR amplification was done on PCR8 with attL1 and attL2 sites intact for cloning into Pjet 2.1. This was necessary due to the spectinomycin resistance of both PCR8 and pOpOff2(kan). The *MTA* fragment was inserted into the entry vector using the Gateway™ LR Clonase reaction (Figure 5.1). Double entry of the *MTA* fragment on forward and mirrored attR sites in pOpOff2(kan) destination vector was confirmed with sequencing for correct base pair sequences and orientation. The resulting plasmid was named *RNAi-MTA*. The *RNAi-MTA* plasmid was transformed into EHA105 *Agrobacterium* and PCR confirmation along with sequencing was done to confirm correct insertion into *Agrobacterium*.

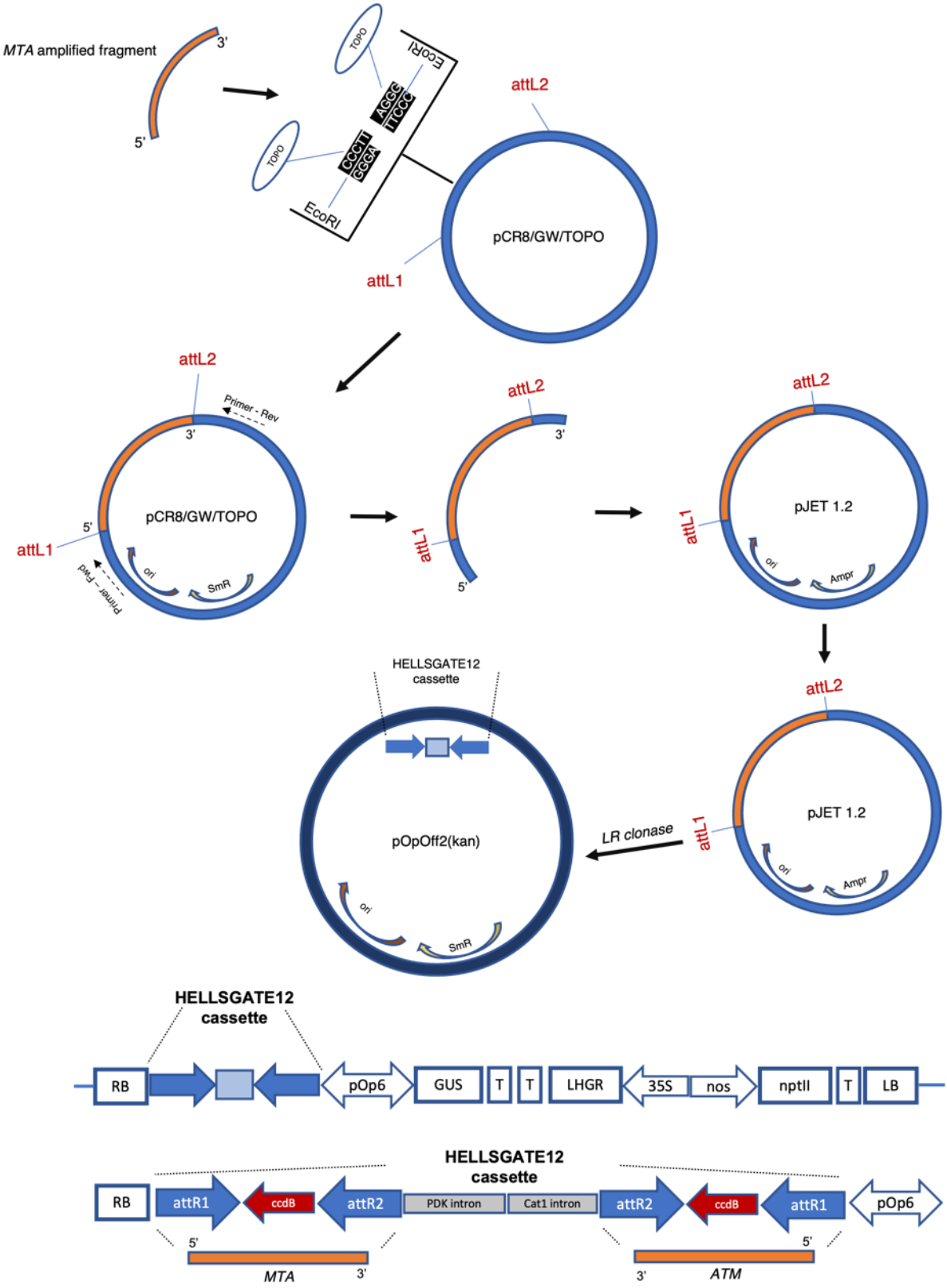


Figure 5.1 Schematic of *RNAi-MTA* vector design for *MTA* gene silencing in *S. lycopersicum*.

The *MTA* gene (orange fragment) was amplified from purified tomato genomic DNA and was cloned into PCR8/GW/TOPO. This was then PCR amplified with both attL sites and cloned into pJET1.2. From pJET1.2 LR Clonase™ was performed with pOpOff2(kan) for insertion into two sties into the HELLSGATE12 cassette. HELLSGATE12 cassette is shown within the RB and LB of pOpOff2(kan) along with the enlarged HELLSGATE12 cassette.

5.3.3 Generation of *MTA* knock-down lines

In order to generate KD lines of *MTA* in tomato *Agrobacterium* transformation of *RNAi-MTA* construct was performed. Transformed plant cells were selected on kanamycin containing media and cultured, allowing the regeneration of whole transgenic tomato plants. EHA105 was the strain of bacteria chosen for its high efficiency in tomato transformation (Chetty *et al.*, 2013). Several attempts were made to transform the tomato plants, but callus generation was not observed, indicating that the protocol needed to be optimised. Different concentrations of media and OD of the bacteria were tested to improve the transformation efficiency. The optimised protocol described in Chapter 2 yielded numerous callusing events and produced stable transgenic lines with the desired *RNAi-MTA* integration.

The transformation of *RNAi-MTA* was performed on more than 800 cotyledons of *S. lycopersicum*. The transformation resulted in the generation of 16 transgenic lines (Figure 5.2). The T_0 transgenic lines were screened for the presence of T-DNA insertion after shoots were placed to grow on compost. PCR analysis was employed to confirm the presence of the plasmid and the double *MTA* insertion in both Gateway^{TM} sites for each of the 16 transgenic plants. Out of the 16 transgenic plants, 9 plants were confirmed to have the T-DNA insertion in both Gateway sites. The resulting T_1 line of plants was used in further experiments throughout this chapter. Each plant from the T_1 line was subjected to the same PCR confirmation to confirm heritability of the produced mutations.



Figure 5.2 Tomato transformation showing from germination of seed to producing positive transgenic plants.

From left to right, the process of tomato transformation by *A. tumefaciens* is shown. Germination of WT tomato seeds was done under sterile conditions on ½MS media. The cotyledons were cut adaxially and dipped in bacteria containing *RNAi-MTA* constructs. The cotyledons were then transferred to selection medium. Once shoots were present, they were cut and placed on rooting media, and subsequently transferred to compost.

5.3.4 GUS staining for confirmation of vector expression

To confirm the expression of the pOpOff2(kan) vector in tomato plants after DEX induction, GUS staining was performed to visually analyse and screen for high GUS activity. The pOpOff2(kan) vector contains a pOp6 promoter that can drive the expression of both the gene of interest and the GUS reporter in a bidirectional manner (Figure 5.1). The results of the GUS staining displayed strong GUS activity that were present in the leaves, roots, and stem of tomato seedlings after DEX induction (Figure 5.3). However, the expression was not uniform across the different physiological parts of the plants, suggesting that some areas had higher expression than others. Transgenic plants without DEX induction were used as a control, and they showed no GUS activity, indicating that there was no leaky expression of the plasmid without the presence of DEX. GUS staining was performed on all T_1 plants prior to physiological and RT-qPCR measurements to confirm plasmid expression and ensure consistent results is obtained.

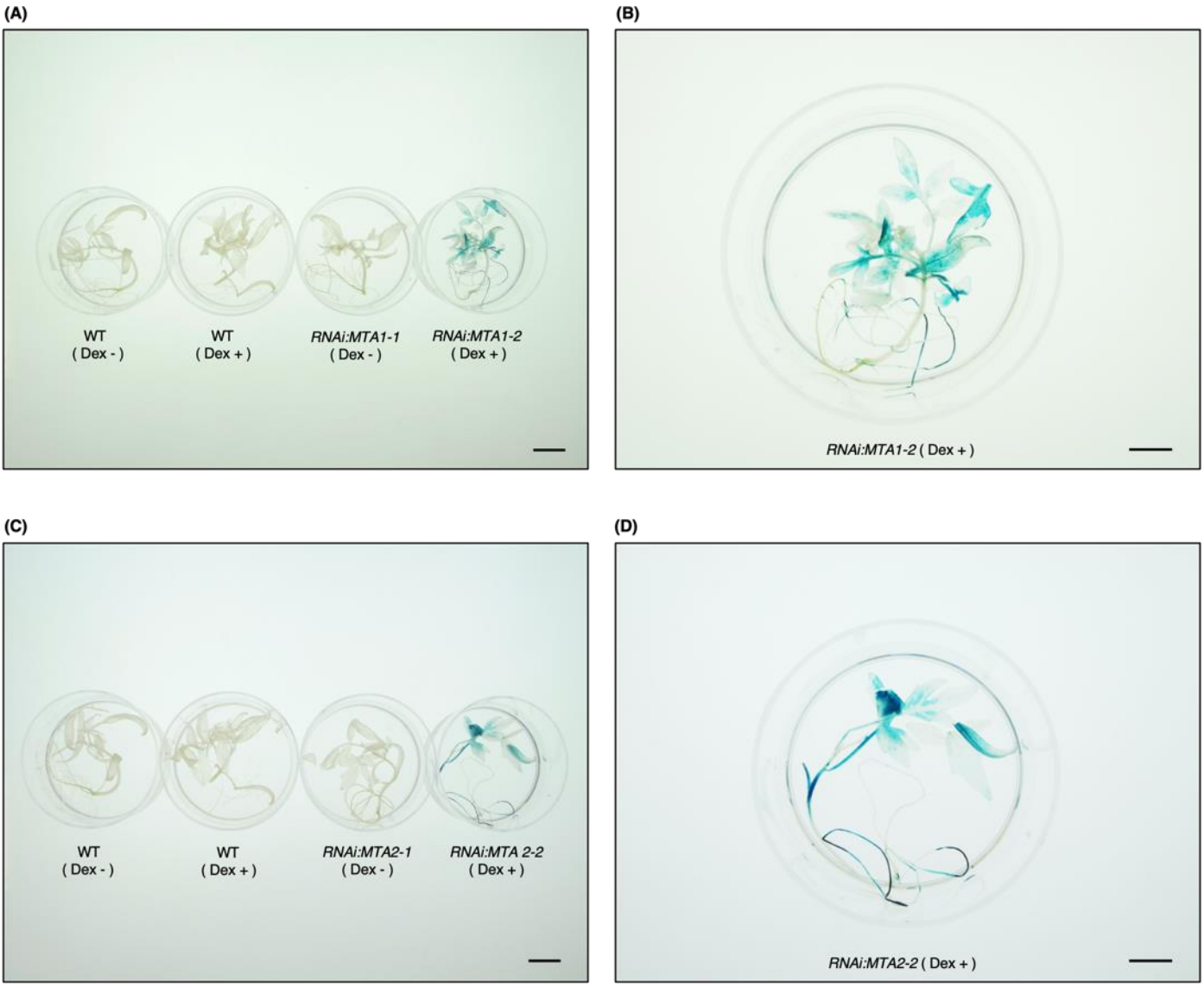


Figure 5.3 Histochemical GUS staining of *RNAi-MTA1* and *RNAi-MTA2* DEX inducible lines.

Histochemical GUS staining was performed on two independent *RNAi-MTA* lines to investigate the expression of the plasmid after dexamethasone DEX induction. (A) Seedlings of WT and *RNAi-MTA1* were subjected to staining, which showed blue GUS staining in the DEX-induced plants. (B) A close-up of *RNAi-MTA1-2*. (C) Seedlings of WT and *RNAi-MTA2* were also stained, showing blue GUS staining in the DEX-induced plants. (D) A close-up of *RNAi-MTA2-2*. Seedlings were initially germinated on ½ MS medium for 10 days and then transferred to ½ MS medium with DEX (30μM). GUS staining was performed 3 weeks after transfer to DEX (30μM) medium. The plants were grown in a Conviron growth room at 25°C with a 16-hour day length. The scale bars for (A,C) are 1.25cm, and for (B,D) they are 0.62cm.

5.3.5 Physiological characterisation of RNAi-*MTA* lines in growth media

The physiological characterisation of *RNAi-MTA* lines was conducted on tomato plant seedlings that were germinated on ½ MS media for 8 days and transferred to ½ MS medium with the addition of DEX (30µM). This ensured that plants were able to germinate without the KD effects of *MTA*, as previous studies have shown that KO of *MTA* is embryo lethal (Zhong *et al.*, 2008). DEX-induced plants displayed a range of physiological differences compared to the WT for both independent *RNAi-MTA1* and *RNAi-MTA2* lines (Figure 5.4). Plant heights for both lines did not show any statistically significant change for DEX-induced plants compared to normal medium.

The length of roots was measured and the *RNAi-MTA1* DEX-induced plants exhibited a statistically significant reduction of 21% in root length compared to plants grown under normal conditions. Similarly, *RNAi-MTA2* DEX-induced plants showed a significant reduction in root length with a decrease of 20.5% compared to plants grown under normal conditions.

Phenotypic observation involving leaf growth was also investigated in this study. *RNAi-MTA1* DEX-induced plants displayed a statistically significant decrease of 45% in true leaf length compared to untreated plants grown on normal medium (Figure 5.6). Similarly, *RNAi-MTA2* DEX-induced plants exhibited a 35% reduction in true leaf length compared to normal medium, although the reduction was not as severe as that observed in *RNAi-MTA1* plants. In terms of leaf width, both *RNAi-MTA1* and *RNAi-MTA2* DEX-induced plants showed significant reductions of 35% and 30%, respectively, compared to plants grown on ½ MS media. Overall, DEX-induced plants displayed a significant reduction in leaf growth and physiology (Figure 5.6).

In order to investigate changes in lateral branch formation, the number of lateral branches formed from auxiliary meristems was quantified. Both *RNAi-MTA1* and *RNAi-MTA2* DEX induced plants resulted in a significant increase of 2 lateral branches on average compared

to untreated plants grown on normal medium. This was accompanied by an increase in nodes, with both *RNAi-MTA1* and *RNAi-MTA2* showing a significant average increase of 2 nodes compared to untreated plants.

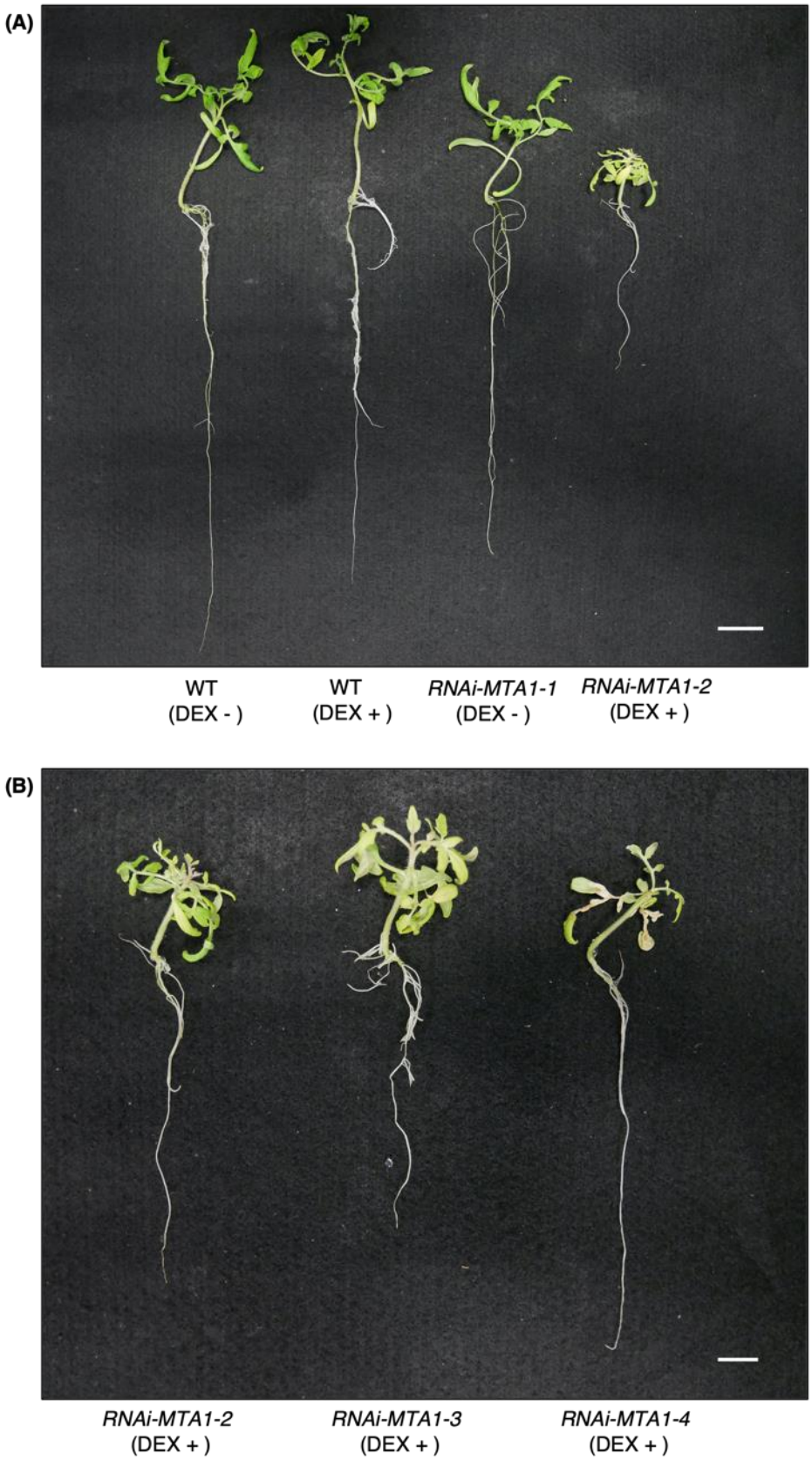


Figure 5.4 Phenotype observation of DEX induced *RNAi-MTA1* on growth media.

Phenotype observations of *RNAi-MTA1* seedlings. (A) Representative images of WT and *RNAi-MTA1* line showing no DEX induction (DEX -) and DEX induction (DEX +) of seedlings growth on ½ MS + DEX (30μM) medium. (B) Close-up images of 3 *RNAi-MTA1* DEX induced seedlings showing reduced growth phenotype. Seedlings were photographed at 5-weeks old. Scale bars: (A) 1.5cm; (B) 1.2cm.

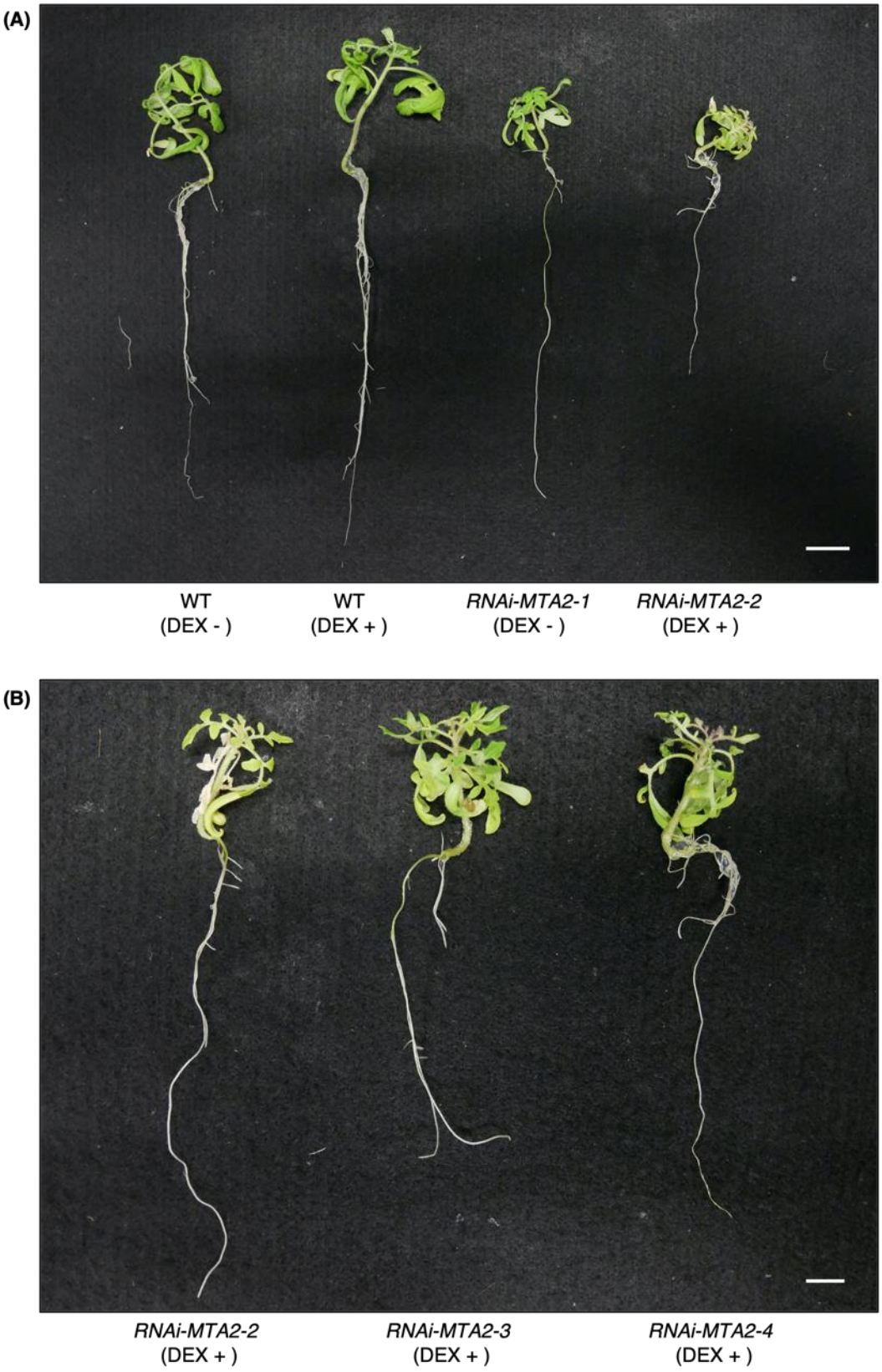


Figure 5.5 Phenotype observation of DEX induced *RNAi-MTA2* on growth media.

Phenotype observations of *RNAi-MTA2* seedlings with their respective lines. (A) Representative images of WT and *RNAi-MTA2* line showing no DEX induction (DEX -) and DEX induction (DEX +) of seedlings grown on ½ MS + DEX (30μM) medium. (B) Close-up images of 3 *RNAi-MTA2* DEX induced seedlings showing reduced growth phenotype. Seedlings were photographed at 5-weeks old. Scale bars: (A) 1.5cm; (B) 1.2cm.

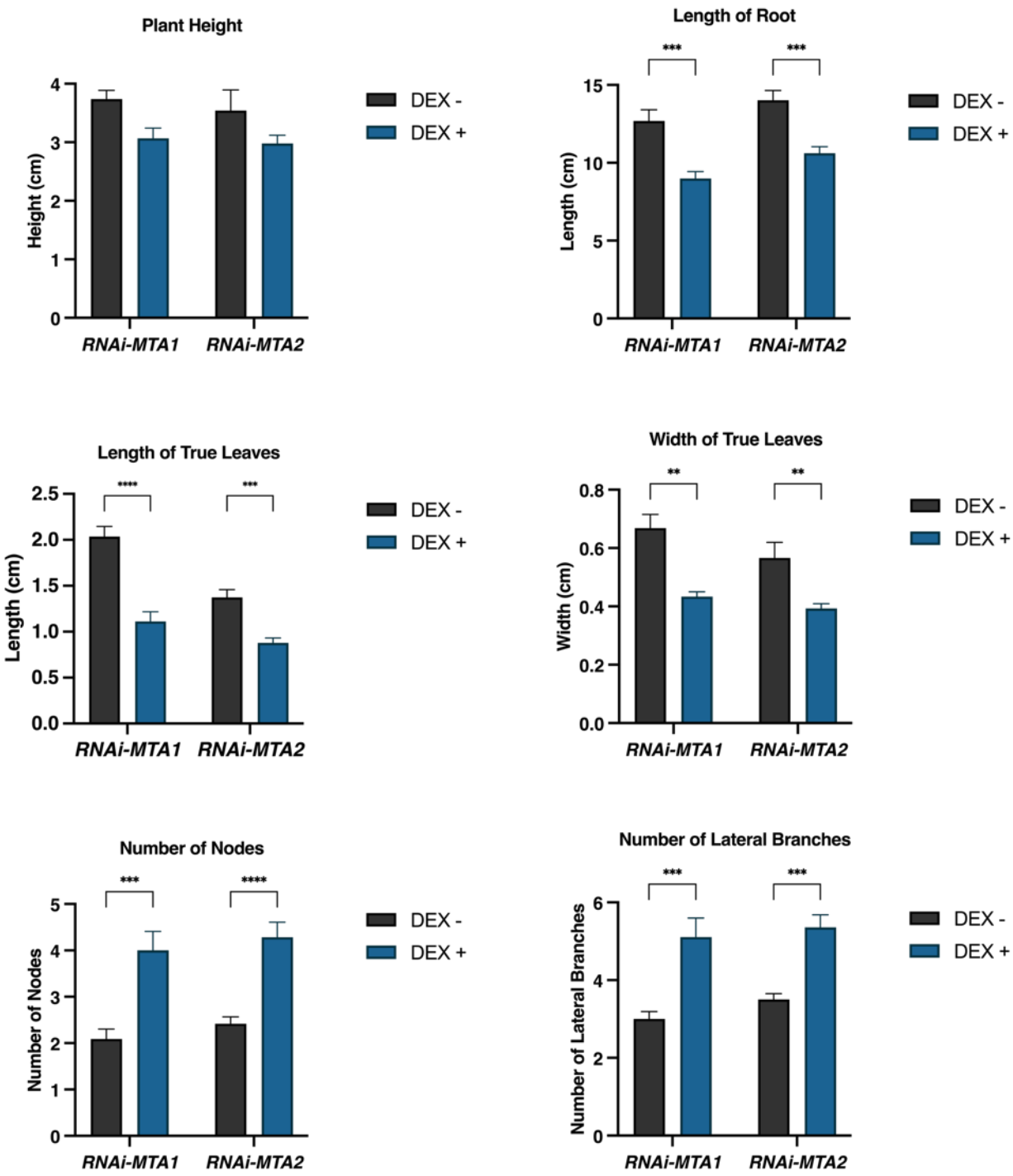


Figure 5.6 Characterisation of DEX induced *RNAi-MTA1* and *RNAi-MTA2* mutant lines grown on media. Histogram representation of key plant growth measurements.

Key plant growth measurements of *RNAi-MTA1* and *RNAi-MTA2* seedlings grown on ½ MS and ½ MS + DEX (30μM) medium. Measurements taken at 5 weeks in cm. Plants were grown on plastic pots containing respective medium in Conviron growth room at 25°C with 16-hour day length. Error bars represent SEM, n = 15-20 with * representing significant difference: ** $p < 0.005$, *** $p < 0.0005$ and **** $p < 0.0001$ respectively by two-way ANOVA.

5.3.6 Physiological characterisation of *RNAi-MTA* lines in Levington compost

In order to observe the effects of DEX induction on larger plants, *RNAi-MTA* induced plants were grown for seven weeks on Levington compost in a glasshouse environment. The method chosen for DEX induction was performed by watering the plants in a water solution containing DEX (30 µM) every four days, ensuring constant induction is achieved. Key growth parameters were measured at the end of the seven weeks. Similar to the results obtained from plants grown on media, plants grown on compost did not show a significant decrease in plant height for both *RNAi-MTA1* and *RNAi-MTA2*.

The length of leaves 3 and 4 was measured from the first true leaf to ensure consistency in plant measurements. *RNAi-MTA1* and *RNAi-MTA2* leaves exhibited a significant reduction of 19% and 28%, respectively, in length when compared to plants without DEX induction (Figure 5.9). Similarly, leaf widths of *RNAi-MTA1* and *RNAi-MTA2* showed a significant reduction of 24% and 16%, respectively, when compared to plants without DEX induction (Figure 5.9).

DEX induction resulted in a significant increase in the average number of lateral branches and nodes for both *RNAi-MTA1* and *RNAi-MTA2* mutants. Specifically, an average of four lateral branches and four nodes were observed in DEX-induced plants compared to untreated plants. Additionally, both *RNAi-MTA1* and *RNAi-MTA2* lines showed an average increase of four nodes. Overall, the phenotypic observations of larger tomato plants under DEX induction have shown a reduction in leaf morphology but an increase in the number of nodes and lateral branches (Figure 5.7+5.8).

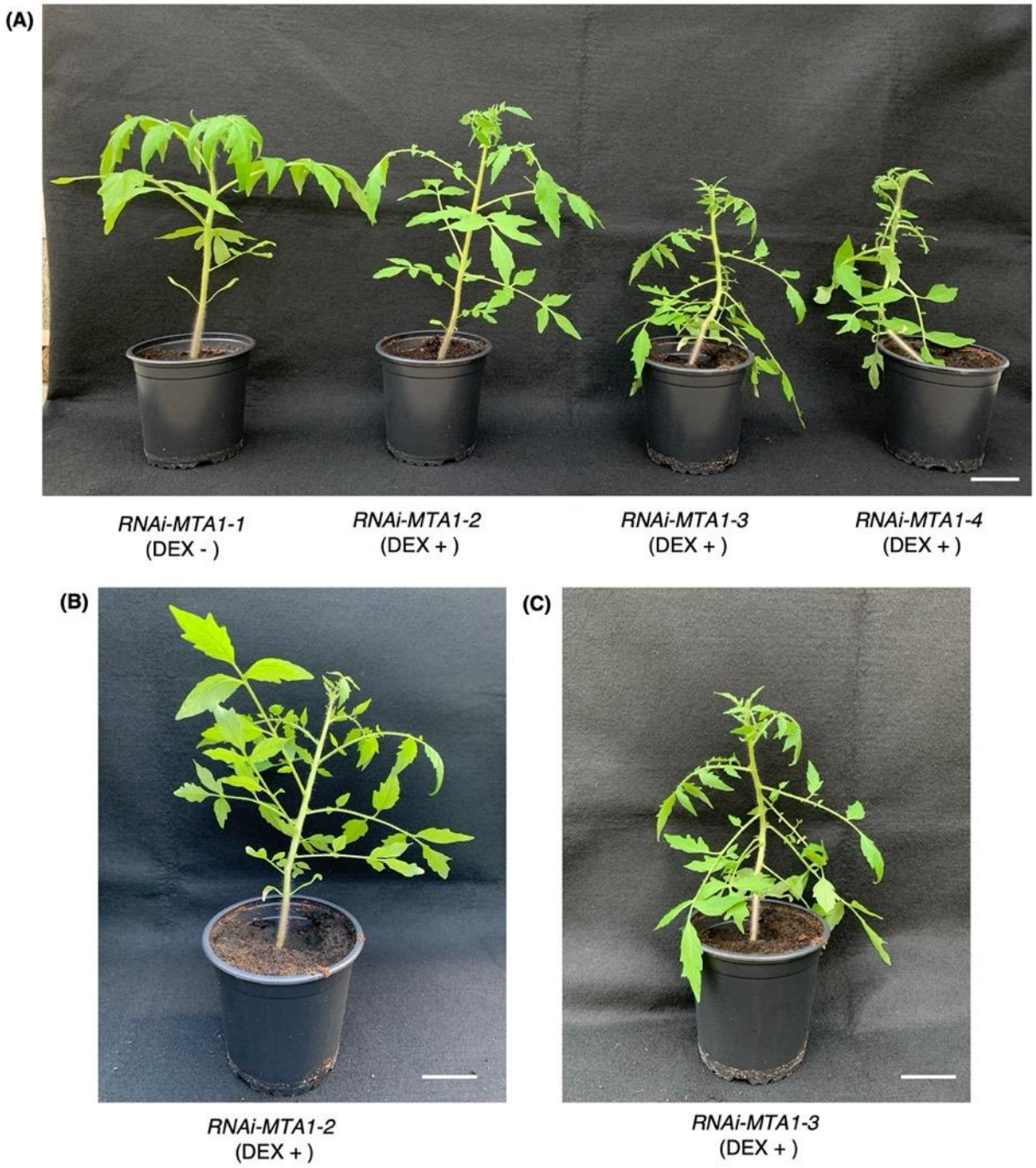


Figure 5.7 Phenotype observation of *RNAi-MTA1* on Levington compost.

Phenotype observations of tomato plants grown on Levington compost with their respective lines. (A) Representative images of *RNAi-MTA1* line showing no DEX induction (DEX -) and DEX induction (DEX +). DEX induction was done through watering of soil with DEX (30μM) every 4 days. (B) Close-up image of *RNAi-MTA1-2* DEX induced plant. (C) Close-up images of *RNAi-MTA1-3* DEX induced plant. Plants were photographed at 7-weeks old. Scale bars = 6cm.

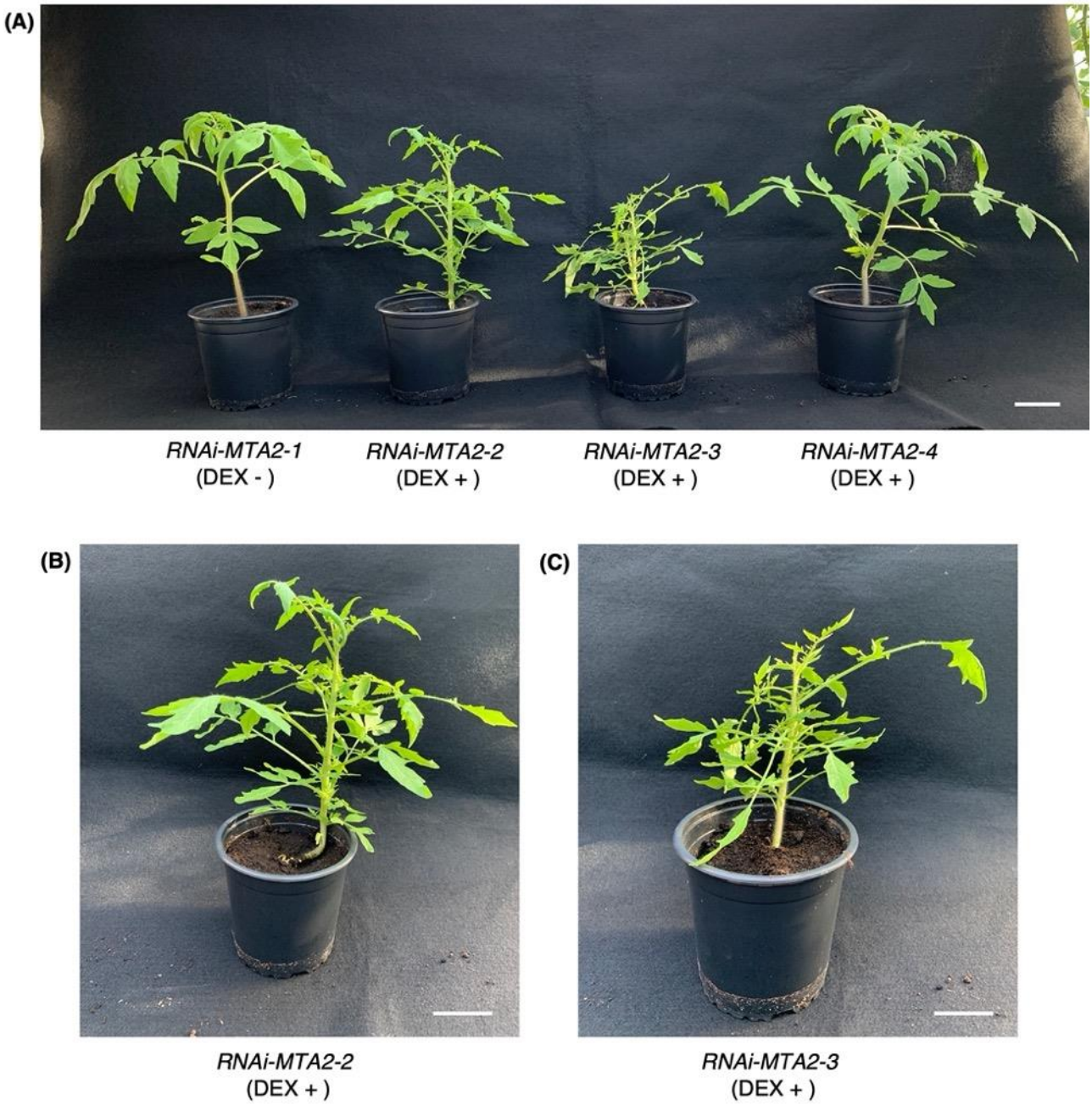


Figure 5.8 Phenotype observation of *RNAi-MTA2* on compost.

Phenotype observations of tomato plants grown on Levington compost with their respective lines. (A) Representative images of *RNAi-MTA2* line showing no DEX induction (DEX -) and DEX induction (DEX +). DEX induction was done through watering of soil with DEX (30μM) every 4 days. (B) Close-up image of *RNAi-MTA2-2* DEX induced plant. (C) Close-up images of *RNAi-MTA2-3* DEX induced plant. Plants were photographed at 7-weeks old. Scale bars = 6cm.

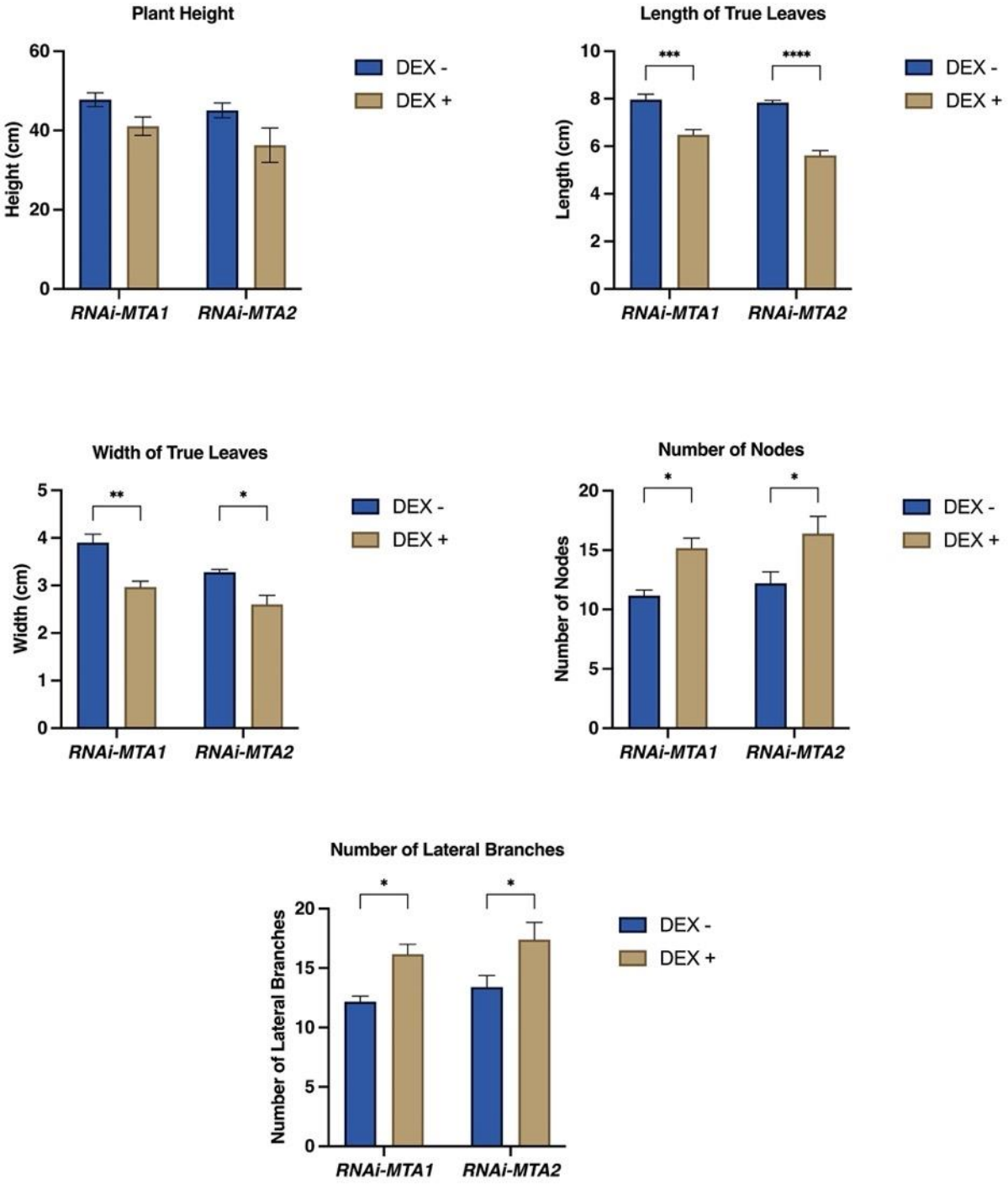


Figure 5.9 Characterisation of DEX induced *RNAi-MTA1* and *RNAi-MTA2* mutant lines grown on compost. Histogram representation of key plant growth measurements.

Key plant growth measurements of *RNAi-MTA1* and *RNAi-MTA2* plants grown on Levington compost and watered with DEX (30μM) solution for DEX induction. Measurements taken at 7 weeks in cm. Plants were grown on plastic pots containing Levington compost in a glasshouse. Error bars represent SEM, n = 9-14 with * representing significant difference: ** $p < 0.005$, *** $p < 0.0005$ and **** $p < 0.0001$ respectively by two-way ANOVA.

5.3.7 RT-qPCR measurements on *RNAi-MTA* lines show increase in *MTA* transcripts

The pOpOff2(kan) vector system was used to achieve efficient KD of the targeted gene *MTA* through hpRNAi silencing pathways in tomato. To confirm the successful KD of *MTA* and to quantify the transcript expression of the two DEX-inducible lines, we performed Reverse Transcription quantitative Polymerase Chain Reaction (RT-qPCR). Theoretically, a KD of *MTA* would be associated with a decrease in transcript expression. However, an increase in transcript expression was observed with one of the primer pairs during RT-qPCR. Therefore, we repeated the experiment using a different housekeeping gene (*SITIP41*) and two additional *MTA*-specific primer pairs. In total, we tested two housekeeping genes and three different *MTA*-specific primer pairs to rule out chances of false results that could arise from specific *MTA* primer pair and the housekeeping genes.

To further minimise the potential for false results, total RNA extracted using the Ambion RNAqueous RNA extraction kit was subjected to DNase treatment and then purified further using an RNA purification kit (Thermo Fisher), as described in the methods section. To ensure the purity of the total RNA, routine PCR of cDNA was also performed and a single band specific to the primer pair was present, while no bands were shown for gDNA contamination.

To ensure the validity of the primers used in RT-qPCR, a primer efficiency test was conducted on all primers, including the 2 housekeeping and 3 *MTA*-specific primer pairs outlined in the Chapter 2. The amplification efficiency of primers can affect the accuracy of the results in gene expression and give incorrect fold changes. To perform the primer efficiency test, cDNA from WT samples was serially diluted over four points in quadruplicate, starting with a 1:5, 1:10, 1:15, 1:20 dilution from the original cDNA. The average technical replicate Ct values for each dilution were plotted on a standard curve, and the slope of regression was calculated. Primer efficiency was calculated using the following equation shown below.

$$Efficiency\ (\%) = \left(10^{-\frac{1}{x}} - 1\right) \text{x } 100$$

The slopes of regression for the *MTA*-specific primer pairs and the housekeeping primer pair were found to range from $-3.1221x$ to $-3.8544x$ (Figure 5.10). By calculating the PCR efficiencies of all primer pairs using their respective regression values for x it was determined that all primer pairs had efficiencies between 90%-110% for RT-qPCR, which is considered optimal for gene expression quantification.

The $^{2\Delta\Delta}$Ct method was used to determine the relative expression of the target gene compared to the housekeeping gene. The relative transcript abundance of *RNAi-MTA1* and *RNAi-MTA2* lines was quantified using RT-qPCR with *MTA* primer pair 1 and housekeeping gene *SAND*, and an increase of approximately 1 and 2.9 times was observed in the DEX-induced lines compared to non-DEX-induced lines (Figure 5.11). This increase in expression was consistent when using the same primer pair with a new housekeeping gene *SITIP41*, with a relative transcript abundance increase of 0.7 and 1.2 times in DEX-induced *RNAi-MTA1* and *RNAi-MTA2* lines, respectively, compared to non-DEX-induced lines (Figure 5.11).

RT-qPCR analysis using *MTA*-specific primer pair 2 and housekeeping *SAND* demonstrated a lower increase in transcript abundance compared to *MTA*-specific primer pair 1. Specifically, *RNAi-MTA1* showed only a 0.2-fold increase in transcript abundance, while *RNAi-MTA2* showed a 1.3-fold increase compared to the non-DEX induced line respectively. Similar results were obtained when using housekeeping *SITIP41* as the housekeeping gene (Figure 5.11). The *RNAi-MTA1* and *RNAi-MTA2* showed consistent increase in relative transcript abundance by 1.3 and 1.7 times, respectively, on DEX-induced lines when using *MTA* primer pair 3 and housekeeping gene *SAND*. This trend was also observed when compared to housekeeping gene *SITIP41*, with *RNAi-MTA1* and *RNAi-MTA2* showing an increase in transcript abundance of 0.8 and 1.5 times, respectively, on both DEX-induced lines compared to non-DEX induced

lines (Figure 5.11). These findings suggest that knock-down of *MTA* results in increased abundance of *MTA* transcripts.

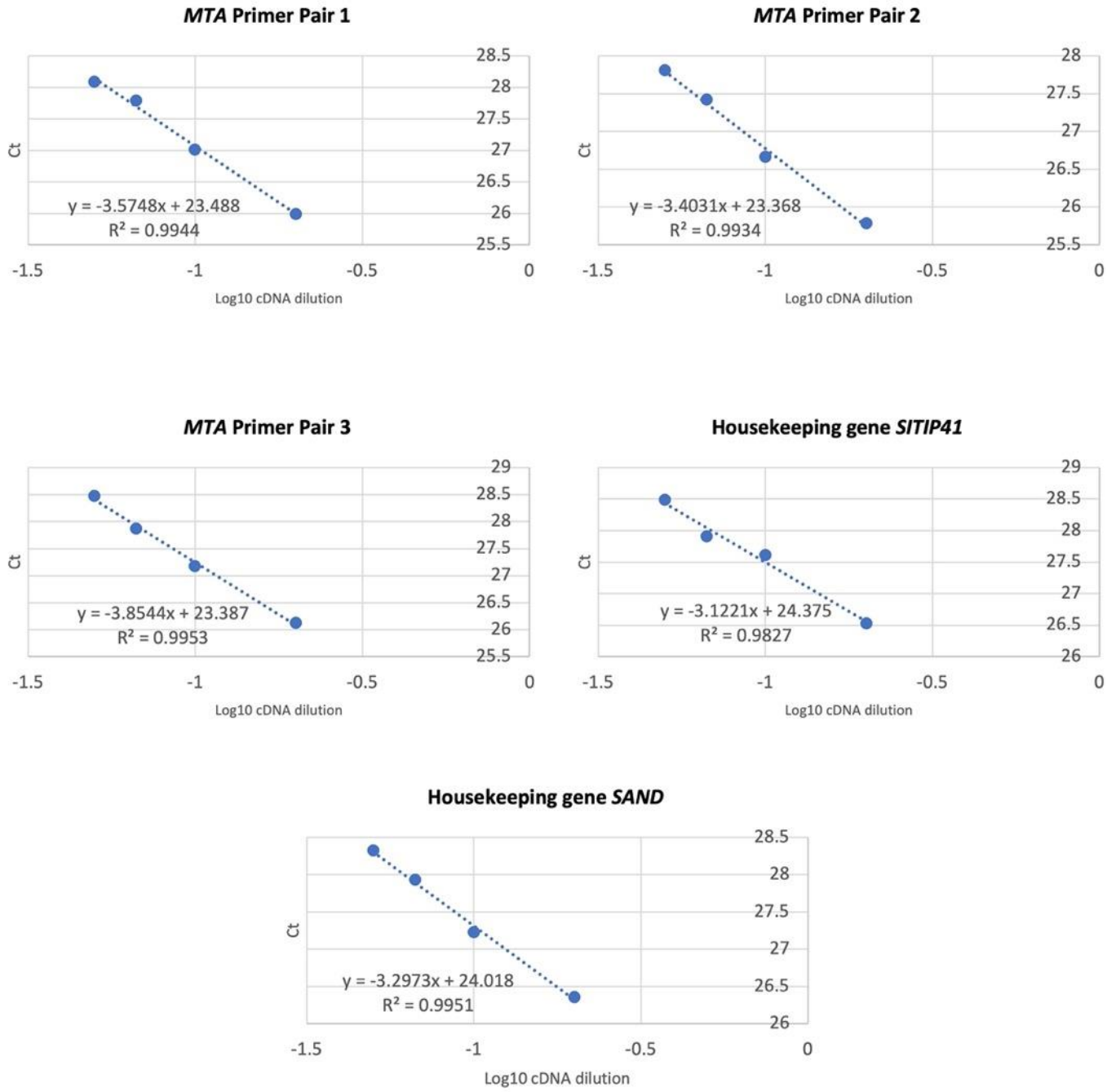


Figure 5.10 Primer check efficiency for all *MTA* specific primers and two housekeeping genes.

Standard curve plot from qPCR performed on three different *MTA* specific primer pairs on a series of four dilutions of WT cDNA starting from 1:5, 1:10, 1:15, 1:20, dilution of original cDNA. The cDNA used were generated from 5-week-old WT seedlings grown on ½ MS medium. Three technical repeats were performed.

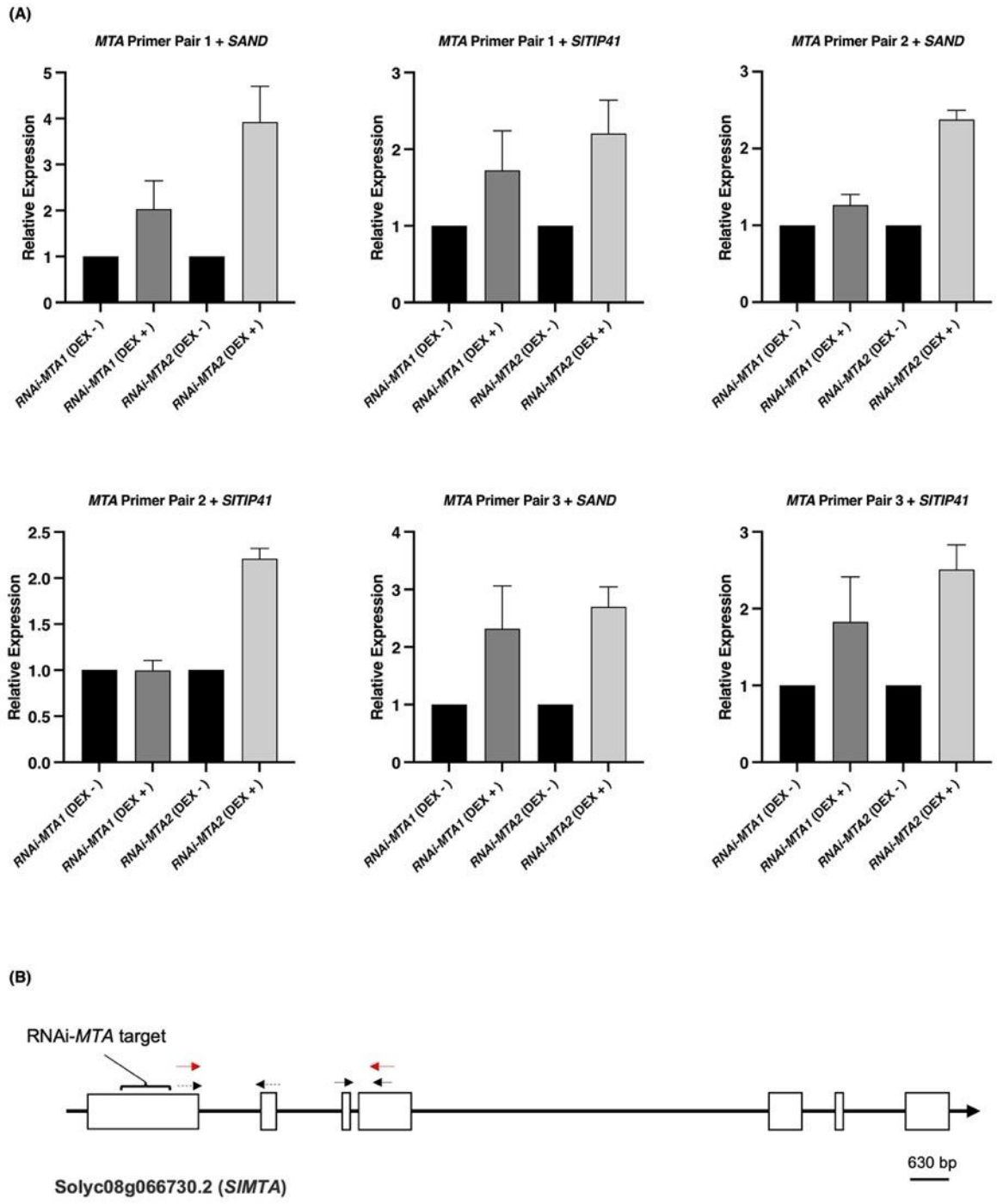


Figure 5.11 *MTA* transcript repression by DEX inducible *RNAi-MTA* show an increase in transcript abundance on DEX induced lines.

(A) Relative transcript abundance of *RNAi-MTA1* and *RNAi-MTA*2 of three different primer pairs and two different housekeeping gene during normal and DEX induced media showing changes in expression levels. Values are given relative to the respective *MTA* primer pair and housekeeping gene. The bars show mean values and error bars represent SEM from four technical repeats. (B) Schematic diagram illustrating Solyc08g066730.2 (*SlMTA*) gene showing three different primer pair location used for qPCR. Primer pair used for qPCR shown by three different arrows: Primer pair 1 = black dotted arrow, Primer pair 2 = red arrow, Primer pair 3 = solid black arrow.

5.3.8 Two-dimensional Thin Layer Chromatography showed minimal changes in m^{6}A level after DEX induction

In order to assess m^{6}A levels in a DEX-induced *RNAi-MTA* line, TLC was performed. Previous experiments confirmed positive expression of the *RNAi-MTA* vector by GUS staining and RT-qPCR expression analysis. The measurement of m^{6}A levels was conducted to determine if the direct increase in expression is reflected in an increase in m^{6}A level. The results of the TLC measurements indicate that m^{6}A levels did not change between non-DEX induction and DEX induction (Figure 5.12). However, it should be noted that this TLC analysis was performed on only one *RNAi-MTA* line (*RNAi-MTA2*) and only two biological replicates were conducted. Additionally, low level contamination by rRNA was present in the purified mRNA from total RNA used in the experiment that is seen on the TLC plates (although rRNA will not contribute to the detected m^{6}A, as it does not contain m^{6}A in a GA context).

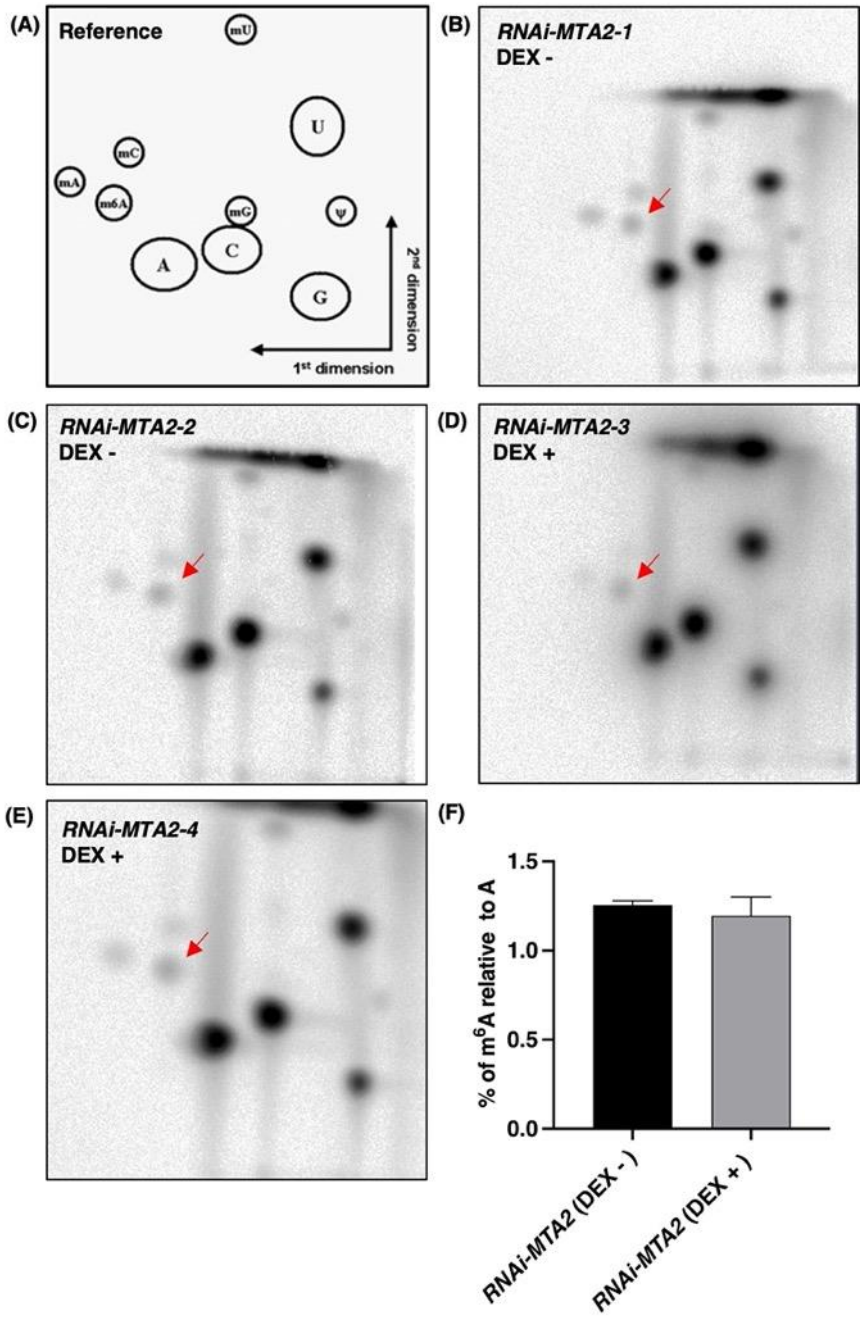


Figure 5.12 Two-dimensional Thin Layer Chromatography on *RNAi-MTA2* shows minimal changes in m^{6}A levels on DEX induction.

(A) Schematic illustrating the location of m^{6}A and non-modified nucleotides on a 2-D TLC plate. (B), (C), (D), (E) shows image of TLC plates indicating reduction in m^{6}A levels in WT, *RNAi-MTA2-1* (DEX -), *RNAi-MTA2-2* (DEX -), *RNAi-MTA2-3* (DEX +), *RNAi-MTA2-4* (DEX +) respectively going from left to right. (F) Bar graph displaying quantification of m^{6}A percentage level relative to A. Two biological replicates was done from 5-week-old seedlings grown on ½ MS plates and ½ MS + DEX (30μM). Error bars represent SEM and no significant differences was present. The red arrow indicates location of m^{6}A labelled nucleotide.

5.4 Discussion

A pivotal component of the MTC is MTA, which has been extensively characterised in *A. thaliana* and in mammalian systems. The presence of m^6A methylation in plants specifically *A. thaliana* was first reported in 2008, and it was demonstrated to be dependent on MTA (also known as the homolog of human METTL3), which acts as the catalytic subunit of the MTC (Zhong *et al.*, 2008). Subsequently, several other crucial proteins that constitute the writer complex were identified, which confirmed m^6A modification as a vital factor in various developmental processes, such as embryo and seed development, plant growth, flowering, leaf morphology, and root development (Zhong *et al.*, 2008; Bodi *et al.*, 2012; Růžička *et al.*, 2017; Arribas-Hernández *et al.*, 2018).

In contrast to mammals, the knowledge and research on m^6A modifications in plants is still lacking, with most of the focus on mammalian systems due to the connection with cancer. To date, no studies have viable KO or KD of writer complex proteins in tomato. Tomato represents an ideal dicot model system that can be utilised to investigate various physiological and morphological aspects that are less likely to be present in *A. thaliana*, such as fleshy fruit development and ripening. Moreover, the development of single and double gene disruption lines for all four *YTHDF* genes in tomato (see Chapter 6) provides a model for simultaneously knocking out/knocking down writer and reader proteins, thus creating an interesting model for identifying novel components or gene pathways related to m^6A methylation in plants.

5.4.1 Development of CRISPR/Cas9 and hpRNAi pOpOff2(kan) system

In this study, a CRISPR/Cas9 system was designed to efficiently disrupt independent lines of *MTA* and *MTB* genes in tomato. sgRNA targets were cloned and subsequently used for *Agrobacterium*-mediated transformation of WT cotyledons. However, the CRISPR/Cas9 KO system did not function as intended, as no callus shoots were obtained for screening. This may be due to the high efficiency of the CRISPR/Cas9 KO system, which was targeting a key

component of the MTC complex. Previous studies have reported that inactivation of MTA resulted in failure of the embryo to progress past the globular stage (Zhong *et al.*, 2008).

In this study, a new genetic manipulation approach was utilised to silence *MTA* expression through the use of a DEX-inducible hpRNAi system, which was introduced into tomato using the pOpOff2(kan) vector system (Wielopolska *et al.*, 2005). This system provided the ability to precisely control the expression levels of *MTA*, thus circumventing the issue of embryo lethality and allowing for time-specific knockdown of *MTA*. The pOpOff2(kan) vector contains a DEX-inducible bidirectional promoter that drives the expression of both the GUS reporter gene and the HELLSGATE12 cassette for hpRNAi (Figure 5.1). Previous studies have demonstrated that RNAi mediated by hpRNAs is an effective method for gene silencing (Varsha Wesley *et al.*, 2001; Waterhouse & Helliwell, 2003). In this system, double-stranded RNA produced from the hpRNA transgenes is processed into 21-nucleotide short interfering RNAs (siRNAs). The siRNA, along with the RNA induced silencing complex (RISC), recognises complementary nucleotide sequences, leading to the degradation of mRNA after transcription in the cytoplasm, thus preventing translation.

In order to induce gene silencing, whole T_1 seedlings after germination were transferred to ½ MS + DEX (30μM) medium and total RNA extraction, and GUS staining was performed at 5-weeks after transfer to DEX medium at the same time (Figure 5.3). The results indicated that *RNAi-MTA1* and *RNAi-MTA2* lines without DEX induction did not show GUS staining, suggesting that the expression system is stable and not expressed in tomato (Figure 5.3). Strong GUS staining was observed in the whole leaf, stems, and roots of seedlings induced with DEX for both *RNAi-MTA1* and *RNAi-MTA2*, indicating a strong responsiveness of the pOpOff2(kan) system to DEX treatment and subsequent KD of *MTA*. This positive induction is consistent with other studies using the pOpOff2 system in *A. thaliana* targeting *SAL1* (Phua *et al.*, 2018).

5.4.2 Analysis of *RNAi-MTA1* and *RNAi-MTA2* DEX induced mutant characterisation

To assess the physiological effects of DEX induction in *RNAi-MTA1* and *RNAi-MTA2* lines, growth experiments were conducted on both sterile media and compost. The phenotypic responses associated with DEX induction were observed, with notable developmental defects observed in the root growth of tomato seedlings grown on sterile media. The vertical growth of both lines upon DEX induction did not show a significant change in height. However, there was a significant reduction in root length observed for both lines on DEX media, with an approximately 20% reduction in root length observed (Figure 5.6). Additionally, both *RNAi-MTA1* and *RNAi-MTA2* lines displayed a significant decrease in the length and width of true leaves in DEX media. Interestingly, the number of nodes and lateral branches for both lines under DEX induction showed a significant increase, resulting in plants displaying a bushy phenotype with reduced apical dominance (Figure 5.4+5.5). This is similar to the phenotype seen in low m6A hypomorphic line by Bodi *et al*., (2012) and Růžička *et al.*, (2017). These observations may be related to increased stability WUSCHEL and SHOOT MERISTEMLESS transcripts (Shen *et al.*, 2016).

The physiological characterisation of *RNAi-MTA1* and *RNAi-MTA2* growth on compost displays a phenotype consistent with previous observations on growth media. Upon DEX induction, a significant reduction in the length and width of the 3rd and 4th compound leaves of each line was observed, indicating leaf developmental defects. The number of nodes and lateral branches was also significantly increased in DEX induction lines, which led to a bushy phenotype and a loss of apical dominance. This phenotype is similar with previous observations of reduced *MTA* and m^6A methylation in *A. thaliana* lines (Bodi *et al.*, 2012; Růžička *et al.*, 2017).

Although developmental reduction was seen in roots and leaves, the plant height was not significantly affected, as only a slight reduction was observed in both lines on DEX media.

This reduction is consistent with Růžička *et al.*, (2017), as their *ABI3prom:MTA* in *A. thaliana* also showed a slight reduction in vertical growth on media. These observations suggest that MTA plays an important role in leaf and root development in tomato.

5.4.3 RT-qPCR analysis on DEX induced *RNAi-MTA* lines

The comparative RT-qPCR analysis of DEX-treated *RNAi-MTA1* and *RNAi-MTA2* lines and untreated *RNAi-MTA1+2* line demonstrated a general increase in transcript abundance across both lines, which contradicts the expected gene silencing induced by the pOpOff2(kan) vector system utilising hpRNAi targeting the key writer complex gene *MTA* (Figure 5.1). Although no increase in m^6A methylation levels was detected in the TLC analysis, despite the increase in transcript abundance for *RNAi-MTA2* line. To confirm the expression of the pOpOff2(kan) vectors, GUS staining was performed, and positive staining was observed in seedlings of both lines. The RNAi inducible silencing system has been shown to impact the response of cells to the inducer, as previously described by Guo *et al.*, (2003) in the context of Cre excision by an oestradiol-inducible system, which can result in the formation of silenced sectors. Evidence of this can be observed in GUS staining, where some leaves may exhibit sections of strong GUS staining, while weak or no GUS staining is observed in other sections.

Additionally, physiological changes were present in DEX-induced medium compared to non-DEX medium. The observed results suggest the possibility that the bidirectional pOp6 promoter in the pOpOff2(kan) vector may be expressing only in one direction towards the GUS section and not towards the Gateway™ inserts. Additionally, the observed physiological changes in plant phenotype could have been induced by the presence of DEX in the media. Previous studies have reported successful application of the pOp6/LhGR system under both tissue culture and greenhouse conditions (Samalova *et al.*, 2005). However, it should be noted that these studies also reported that even with the highest concentrations of inducer, tobacco plants did not exhibit any inhibitory effects of DEX or LhGR on plant development.

One possible explanation for the observed increase in *MTA* transcript abundance is the timing of tissue extraction for total RNA and cDNA synthesis. Specifically, the tissues were collected four weeks after transfer from germination medium to DEX medium, which may not have been an optimal time for detecting changes in gene expression. Despite this, GUS staining performed at the same time as tissue collection showed positive staining on both lines, indicating continued expression of the pOpOff2(kan) vector system after four weeks on DEX medium, which should have silenced *MTA*. Further experiments with different time frames are necessary to confirm these findings.

As the siRNA system is present in the cytoplasm only, it is possible that *MTA* mRNA is involved in a negative feedback pathway. Once *MTA* mature mRNA is generated in the nucleus, it is exported to the cytoplasm where it can become an RNAi target. The KD effects of *RNAi-MTA* may degrade cytoplasmic *MTA* transcripts, leading to less MTA protein and lower levels of m^6A methylation of nascent transcripts in the nucleus. If the cell recognises low levels of MTA (or m^6A generally) in the nucleus, it may attempt to compensate by increasing *MTA* expression through activating transcription factors. Slow nuclear export of *MTA* transcripts, combined with high levels of *MTA* expression, could lead to an increase in *MTA* expression detected by RT-qPCR using poly-A RNA taken collectively from the whole cell. It has been suggested that autoregulation of the m^6A machinery by feedback loops regulating cellular mRNA m^6A metabolism related to alternative splicing is possible, based on findings in mouse embryonic stem cells (Wei *et al.*, 2021).

Further studies on the subcellular localisation of MTA-GFP in *mtb-1ABI3:MTB*, revealed that it is primarily localised in the nucleoplasm (Shen, 2023). In *AmiR-mta*, however, MTB-GFP was observed in both the nucleus and cytoplasm, which contrasts with its specific nucleoplasm localisation in WT. The study also showed that *MTA* expression was slightly up-regulated in *mtb-1ABI3:MTB*, while *MTB* expression was significantly up-regulated in *AmiR-mta* mutants (Shen, 2023). This suggest that expression levels between *MTA* and *MTB* does change in

effect with KD of the other and also with various other MTC complex proteins (Shen, 2023). This author also expanded and suggested that there is a distinct functional interdependence among individual components of the m^{6}A methyltransferase complex in *A. thaliana* at the post-translational level. The findings suggest that it is crucial to maintain protein homeostasis among various subunits of the complex to ensure the proper function of the m^{6}A methyltransferase complex in m^{6}A deposition in plants by maintaining the required protein stoichiometry (Shen, 2023). Linking back to our observed results an increase expression due to prolonged DEX induction and down-regulation of *MTA* could have triggered a response to try to maintain correct MTA protein levels.

Increasing evidence has also shown that m^{6}A methylation in specific reader proteins is linked to mRNA export. The study by Roundtree *et al.*, (2017) shows that the nuclear located m^{6}A-binding protein YTHDC1 mediates the export of methylated mRNA from the nucleus to the cytoplasm in HeLa cells. Their findings indicate that knockdown of YTHDC1 resulted in an extended residence time for nuclear m^{6}A-containing mRNA, with an accumulation of transcripts in the nucleus and depletion within the cytoplasm. Additionally, the study revealed that YTHDC1 interacts with the splicing factor and nuclear export adaptor protein SRSF3 and facilitates RNA binding to both *SRSF3* and *NXF1*. Overall, the physiological change observed with DEX induction lines along with an increase in expression levels for *MTA* without changing m^{6}A levels suggest that further experiments are needed to clarify the results seen in this chapter.

5.4.4 Summary and future work

The present work explored in this chapter overcame the embryonic lethality caused by the KO of writer protein in tomato. Future investigations may include KD or KO of both writer proteins and YTH proteins in tomato to gain a deeper understanding of the relationship between writer and reader complexes. Further KD of other writer complex proteins could also be useful, as it

could unravel interactions between MTC components and other proteins or pathways. Further experiments can be carried out to characterise the phenotypes of the *RNAi-MTA* lines more comprehensively, such as gravitropic responses, lateral branching in roots, floral patterning, and fruit development. As a subsequent experiment, overexpression constructs for *MTA* in tomato can be generated to confirm whether they display the same physiological changes in tomato under both media and compost growth conditions. Another potential avenue for future investigation involves confirming the production of dsRNA alongside the GUS expression. The popOff2(kan) construct is designed to generate dsRNA from the bidirectional promoter (pOp6), which subsequently triggers RNAi silencing through siRNA production. Conducting a northern blot analysis would provide confirmation of the presence of siRNA in the transformed plants, thereby validating the occurrence of RNAi silencing. The limitations of this study prevent any definitive conclusions regarding the autoregulatory function of m^6A in relation to KD of *MTA* and the increase in transcript abundance. Therefore, further experiments are required to enhance our understanding of the role of m^6A in tomato.

Chapter 6

Characterisation of YT521-B Double Knock-out Protein Mutants in *Solanum lycopersicum*

6 Characterisation of YT521-B double knock-out protein mutants in *Solanum lycopersicum*

6.1 Chapter Aims

The objective of work described in this chapter is to characterise YT521-B mutant lines in Tomato. The main aims are to:

- Characterise YT521-B protein single and double loss-of-function mutants.
- Identify double knock-out mutatns of YTHDF proteins in tomato.
- Phenotype single and different combinations of double mutant lines of YTHDF mutants.

6.2 Introduction

The 'epitranscriptomic' mark m^6A is the most prevalent internal modification of eukaryotic mRNAs and in recent years numerous studies have identified the expanding roles of this modification relating to gene regulation and RNA processing. RNA-binding proteins (RBPs) that read m^6A modifications can have downstream effects in RNA processing. Previous chapters have focused on the writer complex of proteins, which predominantly applies this modification to eukaryotic mRNAs. However, little is known about the reader proteins. The m^6A modification is gaining increased attention in plant research sectors as plants prove to be an ideal model for the study of this modification, which could expand our understanding of the complexity of RNA modifications in eukaryotes more generally. In the realm of studying the effects of m^6A methylation and how RBPs containing an YT521-B homology (YTH) domain mediate these effects, tomato presents an advantageous model organism. Previous studies have identified the presence of YTH domains in RBPs, which can be classified into two subsets based on their sequence and location within the cell: predominantly nuclear YTHDC

proteins and cytoplasmic YTHDF proteins (Hartmann *et al.*, 1999; Stoilov *et al.*, 2002; Zaccara *et al.*, 2019).

In *A. thaliana*, 13 proteins contain the YTH domain, of which 11 fall into the YTHDF classification and have been previously described as Evolutionarily Conserved C-terminus (ECT) proteins due to their conserved C-terminus. These proteins range from ECT1 to ECT11 and have been studied in various research works (Ok et al. 2005; Li et al. 2014; Scutenaire et al. 2018; Sun et al. 2020). The YTHDC subclade has AT4G11970 (*ECT12*) and a homolog of Cleavage and Polyadenylation Specificity Factor 4 (CPSF4), named Cleavage and Polyadenylation Specificity Factor 30 (CPSF30) (Arribas-Hernández *et al.*, 2018; Scutenaire *et al.*, 2018). CPSF30 is a m^6A reader and exists in two isoforms via alternative polyadenylation. The first isoform is a ~28 kDa protein (CPSF30-S) that harbours three zinc finger domains and is homologous to mammalian CPSF30 (CPSF4). The second isoform which is unique to plants is a ~70 kDa protein (CPSF30-L) that has the additional YTH domain, which is unique to plants (Delaney *et al.*, 2006; Hunt *et al.*, 2012; Song *et al.*, 2021).

Studying the function of each of the 11 ECT YTHDF proteins in *A. thaliana* and their potential redundancy presents a challenge. Concurrent knockout of all 11 genes is difficult, as such KOs could result in an embryo-lethal plant while also being labour intensive to produce and screen by crossing. Therefore, variations of each combination must be tested to determine the individual and overlapping functions of these proteins. Recent studies have focused on ECT2, ECT3, and ECT4, as these have been shown to be part of many important biological processes in *A. thaliana*. For example, ECT2/3 are required for the correct timing of leaf formation and for normal leaf morphology (Arribas-Hernández *et al.*, 2018). Mutants of *ect2/ect3* and *ect2/ect3/ect4* displayed slow root and stem growth, slow flower formation, defective root growth, and aberrant flower morphology (Arribas-Hernández *et al.*, 2020). In tomato, the number of YTHDF proteins is lower than the average monocot or dicot plants, as it encodes for only 4 proteins, namely SlYTHDF1, SlYTHDF2, SlYTHDF3, and SlYTHDF4. Previous work in classifying tomato YTHDF and YTHDC proteins has resulted in a variety of

nomenclature for the same members of each group, which can be confusing. In Table 6.1, a summary of the different naming systems is collated and compared. In this study and chapter, due to their separate grouping from each other the YTHDF proteins have been classified into 4 groups named SlTYHDF1 to SlYTHDF4. In this study, these will be referred to as YTHDF1 to YTHDF4. These groupings are compared to *A. thaliana* ECT proteins and are likely represented in the same groupings (Table 6.1). The *A. thaliana* CPSF30 is represented by two members in tomato, namely SlCPSF30A and SlCPSF30B, while *A. thaliana's* AT4G11970 (*ECT12*) is represented by three members in tomato, namely SlYTHDC1, SlYTHDC2A, and SlYTHDC2B. An overview of the different phylogenetic trees constructed by different studies and a comparison of trees is shown in Figure 6.1, depicting tomato, *A. thaliana*, and human YTHDF and YTHDC proteins. Our group's phylogenetic tree was used in this study, and it shows the similarity in grouping of each YTH protein compared to other research groups (Cameron Grundy, 2021).

This reduced number of YTHDF proteins in tomato provides an opportunity to efficiently study the reader proteins by creating single, double, triple, and potentially quadruple KO of these proteins. This approach has the potential to reveal novel insights into the roles of YTHDF proteins in tomato. The generation of KO mutants in tomato offers a promising strategy to accelerate the identification and transfer of knowledge to *A. thaliana* and mammals, thereby potentially advancing the understanding of the vast number of ECT proteins. The grouping of tomato and *A. thaliana* YTHDF and ECT proteins supports the idea that generation of individual KO mutants of each tomato YTHDF protein could be equivalent to ECT protein KO combinations in *A. thaliana*.

Further to this KO mutants created using CRISPR/Cas9 of SlYTHDF4 displayed pleiotropic phenotypes, such as a low seed germination rate, shortened seedling root, delayed plant growth and development during the vegetative stage, and elongated and longitudinally flattened fruit with a reduced locule number. KO of SlYTHDF4 resulted in reduced GA3 content

and downregulated expression of genes associated with the gibberellin biosynthesis pathway (Yin *et al.*, 2022b).

In previous chapters, we discussed the potential of generating KOs of writer and reader proteins in tomato to investigate the interaction between these mechanisms in relation to m^6A methylation. To this end, single and double knockouts of YTHDF1-YTHDF4 were generated by Cameron Grundy in the Fray lab, and seeds were collected from these lines. The identification of homozygous double mutants is necessary to examine the potential redundancy of each YTHDF protein in tomato and to better understand the functions of these proteins. The identification of double KO could indicate that potential groupings in *A. thaliana* could reflect the same KO and still produce viable plants.

Table 6.1 Nomenclature comparison between this study and other research groups for tomato and *A. thaliana* YTH proteins.

This study nomenclature	Accession ID (Loci)	(Shen *et al.*, 2022)	(Yin *et al.*, 2021)	(Scutenaire *et al.*, 2018)	*Arabidopsis thaliana*
SlYTHDF1	Solyc01g028860.2.1	*SLYTHDF1*	*SlYTH2*	*SlDF1A*	*ECT1, ECT3, ECT2, ECT4*
SlYTHDF2	Solyc05g032850.2.1	*SLYTHDF2*	*SlYTH3*	*SlDF1B*	*ECT5, ECT9, ECT10*
SlYTHDF3	Solyc12g099090.1.1	*SLYTHDF3B*	*SlYTH4*	*SlDF1C*	*ECT6, ECT7*
SlYTHDF4	Solyc01g103540.2.1	*SLYTHDF3A*	*SlYTH1*	*SlDF2C*	*ECT11, ECT8*
SlCPSF30A	Solyc02g021760.3	*SlCPSF30A*	*SlYTH8*	*SlDC2A*	*CPSF30*
SlCPSF30B	Solyc02g070240.3	*SlCPSF30B*	*SlYTH5*	*SlDC1A*	-
SlYTHDC1	Solyc08g007740.2	*SlYTHDC1*	*SlYTH6*	*SlDC1B*	AT4G11970 (*ECT12*)
SlYTHDC2A	Solyc08g007760.3	*SlYTHDC2A*	*SlYTH7*	-	-
SlYTHDC2B	Solyc08g007750.3	*SlYTHDC2B*	*SlYTH9*	-	-

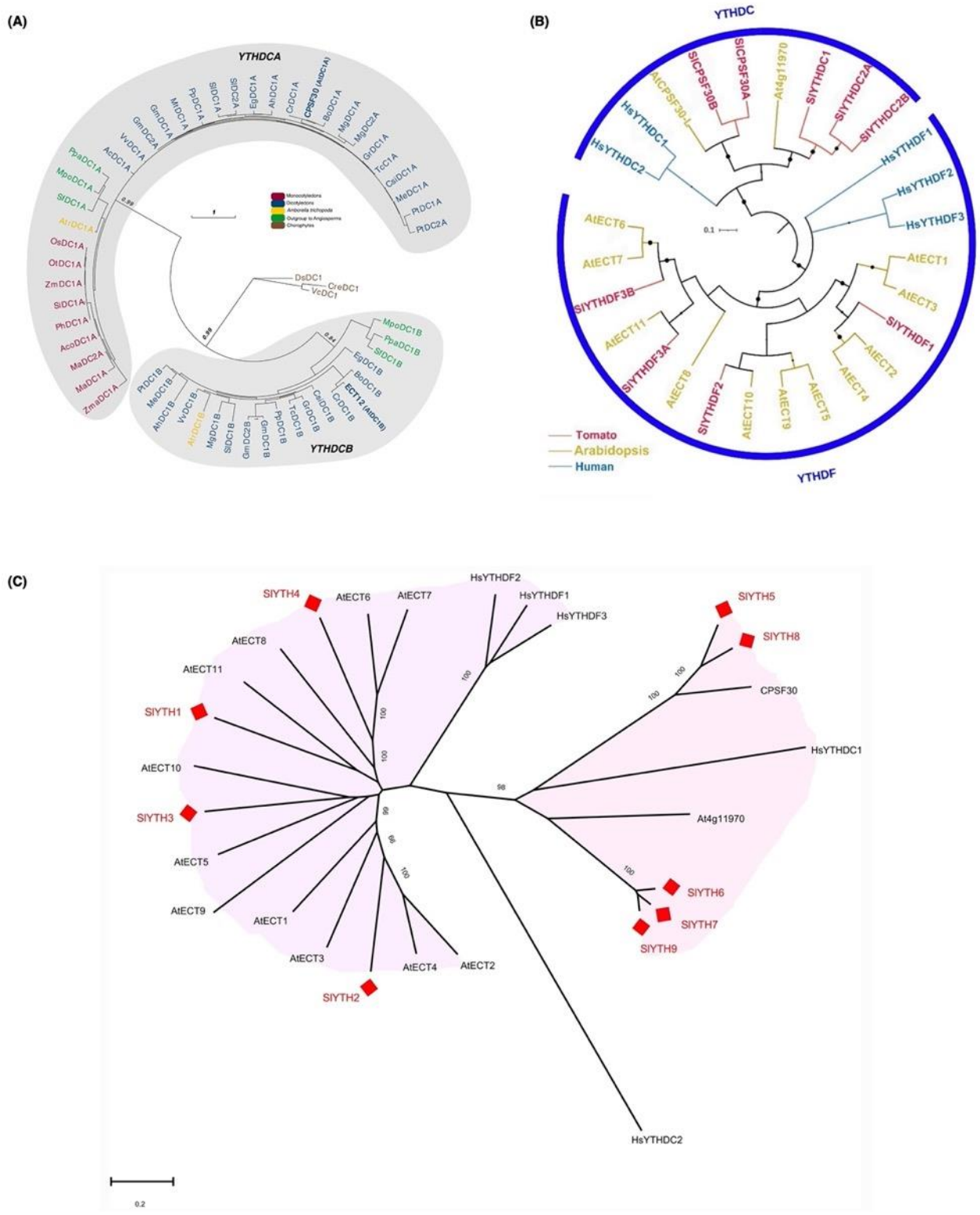


Figure 6.1 Phylogeny comparison of YTH domain-containing proteins and the different naming conventions for tomato across different research groups.

(A) Phylogeny tree taken from Scutenaire *et al.*, (2018) showing showing *YTHDCA* and *YTHDCB* subclade which has *A. thaliana* and tomato alignment. (B) Phylogeny tree taken from Shen *et al.*, (2022) illustrating YTHDF and YTHDC groups for Tomato, *A. thaliana*, and Human. (C) Phylogeny tree taken from Yin *et al.*, (2021) depicting Tomato, *A thaliana*, and Human.

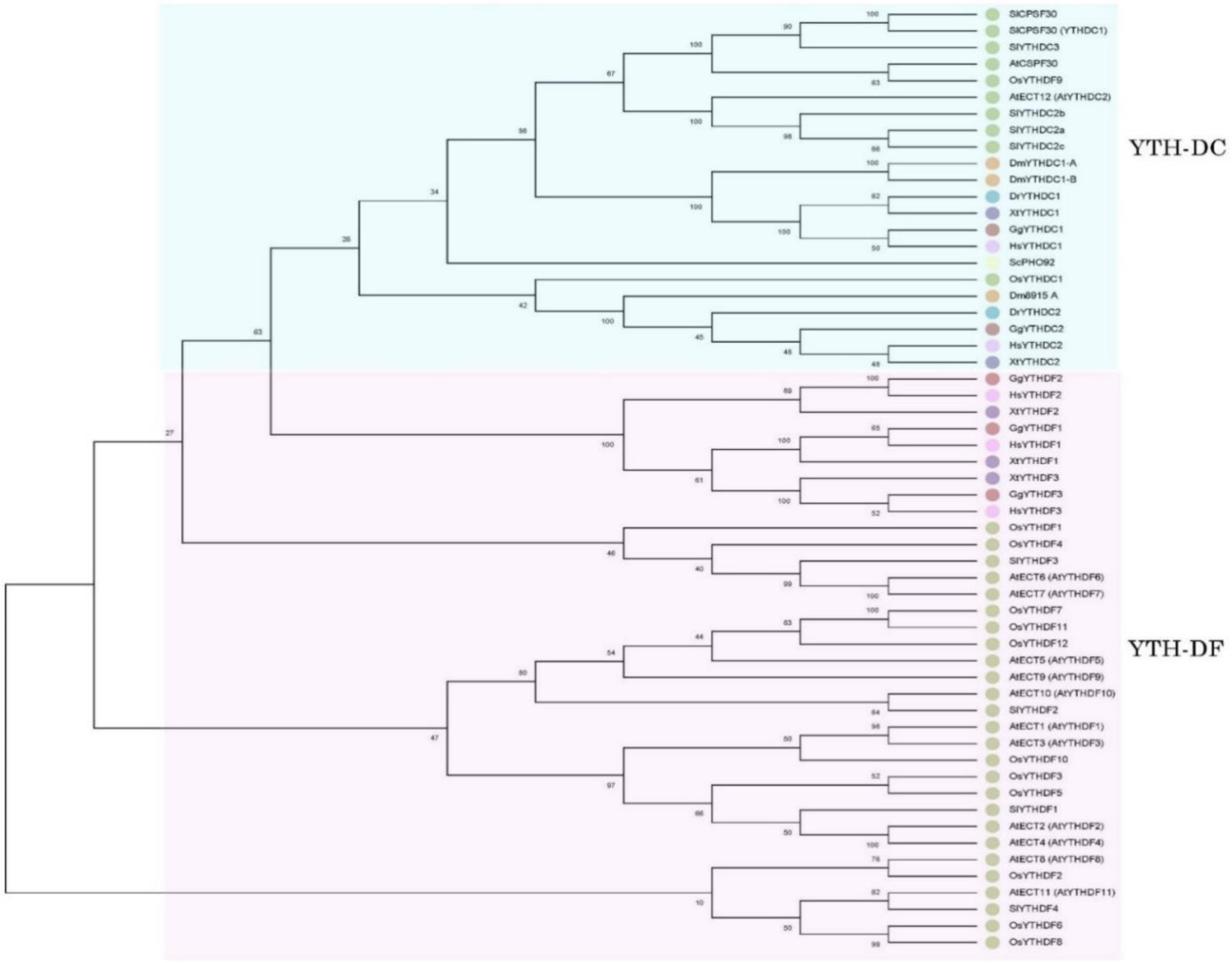


Figure 6.2 Phylogeny tree of similarity showing YTHDC and YTHDF proteins.

This phylogeny tree of similarity is the chosen tree that is referred to in this study. Comparisons between model Planta (Green), Animalia (Mixed) and Yeast (Yellow) species. At- *A. thaliana*, Dr- *Danio rario*, Dm- *Drosophila melanogaster*, Gg- *Gallus gallus*, Hs- *Homo sapiens*, Os- *Oryza sativa*, Sc- *Saccharomyces cerevisiae* Sl- *Solanum lycopersicum*, Xt- *Xenopus tropicalis*. (Cameron Grundy 2021). The protein sequence was analysed using MUSCLE 3.8.4 and to create the phylogenetic tree the alignmeents were imported to MEGA 10.1 using neighbor-joining method with standard parameters and 500 bootstrap replications.

6.3 Results

In the previous work, single KOs of *YTHDF1-YTHDF4* were generated using CRISPR/Cas9 KO with the same vector system as described in Chapter 5 by (Brooks *et al.*, 2014). *Agrobacterium*-mediated transformation was performed using cotyledons of *Solanum lycopersicum*. The T_1 lines were screened and confirmed for mutations in the single KO lines. To confirm the presence of the single KO lines in *ythdf1-ythdf4* during our physiological characterisation, PCR products from T_3 lines were sent for Sanger sequencing, and the mutations were confirmed (Figure 6.3). Specifically, the *ythdf1* single KO produced a single T insertion (Figure 6.3A), while *ythdf2* produced a 34bp deletion (Figure 6.3B), *ythdf3* produced a single T insertion (Figure 6.3C), while *ythdf4* produced a single C insertion (Figure 6.3D). The results demonstrate the presence of homozygous frameshift mutations in our plants, which were subsequently used for physiological characterisation.

6.3.1 Identification of double mutants of YTHDF gene knock-out

To identify redundant functions among the four YTHDF proteins, it is necessary to generate double KO mutants. Previous work in our group generated heterozygous double KOs for all possible combinations of *YTHDF1-YTHDF4*, resulting in six combinations that were heterozygous for both KO alleles. Further screening was performed to identify viable double homozygous KOs. Double homozygous KOs for four out of six *YTHDF* combinations was identified through PCR amplification spanning the CRISPR target sites and Sanger sequencing. The first combination was *ythdf1/ythdf3*, which resulted in a single T insertion and a C insertion, respectively (Figure 6.4A). The second combination, *ythdf2/ythdf3*, resulted in a 34bp deletion and a 5bp (CTGAT) insertion (Figure 6.4B). The third combination, *ythdf2/ythdf4*, resulted in a 34bp deletion and a single C insertion, respectively (Figure 6.4C). Finally, the fourth combination, *ythdf3/ythdf4*, resulted in a single C insertion and a C insertion, respectively (Figure 6.4D). Due to time constraints, *ythdf1/ythdf2* was screened for

heterozygous mutations for both KOs, however, further screening could not identify double homozygous mutations (Results not shown). For *ythdf1/ythdf4*, either a homozygous *ythdf1* or a homozygous *ythdf4* was identified. Further screening of subsequent progeny was carried out, but double homozygous KO could not be identified. Further experiments involving different lines and increased number of plants for screening is needed for both unidentified lines to confidently confirm if double homozygous KO mutation is not possible.

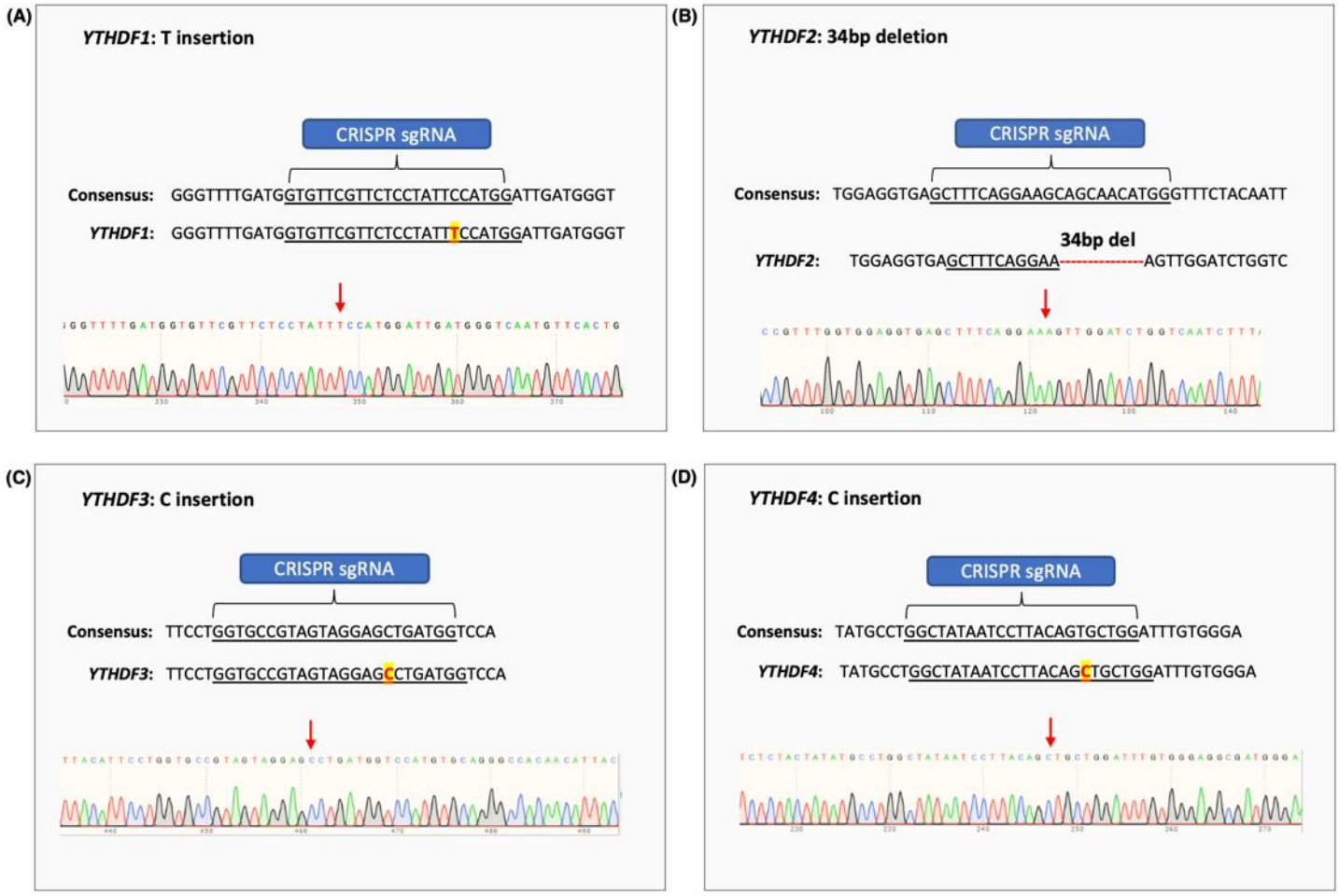


Figure 6.3 Single CRISPR/Cas9 knock-out mutation in *YTHDF1* – *YTHDF4* showing mutation detection.

YTHDF gene sequences are presented as chromatogram graphs of Sanger sequence data for (A) Homozygous *ythdf1-1* single T insertion at position 1913-1914, (B) Homozygous *ythdf2-1* with a 34bp deletion at position 5200-5234, (C) Homozygous *ythdf3-1* single C insertion at position 505-506, and (D) Homozygous *ythdf4-1* single C insertion at position 3726-3727. The sgRNA CRISPR target sites are indicated with a blue rectangle, and the position of the mutation is indicated with a red arrow and highlighted in red in the sequence.



Figure 6.4 Double knock-out of *YTHDF1* to *YTHDF4* combinations of CRISPR/Cas9 mediated mutations showing mutation detection.

Double mutation of *YTHDF* gene sequences showing chromatogram graphs of Sanger sequence data. (A) Homozygous *ythdf1-1*/*ythdf3-1* single T and C insertion respectively. (B) Homozygous *ythdf2-1*/*ythdf3-1* with 34bp and 4bp deletion respectively. (C) Homozygous *ythdf2-1*/*ythdf4-1* 34bp deletion and C insertion respectively. (D) Homozygous *ythdf3-1*/*ythdf4-1* single C and C insertion respectively. sgRNA CRISPR target sites indicated with blue rectangle. The position of mutation is indicated with red arrow and highlighted red in the sequence.

Table 6.2 Overview of mutation present in CRISPR/Cas9 mutants

Gene Name	Loci	Gene Length (bp)	Mutation	Position	Description
SlYTHDF1	Solyc01g028860.2.1	4144	T insertion	1913-1914	Single T insertion resulting in frame-shift and premature stop codon.
SlYTHDF2	Solyc05g032850.2.1	10214	34bp deletion	5200-5234	34bp deletion resulting in frame-shift and premature stop codon.
SlYTHDF3	Solyc12g099090.1.1	3784	T insertion	505-506	Single T insertion resulting in frame-shift and premature stop codon
SlYTHDF3	Solyc12g099090.1.1	3784	4bp deletion	505-509	4bp deletion resulting in frame-shift and premature stop codon
SlYTHDF4	Solyc01g103540.2.1	6008	C insertion	3726-3727	C insertion resulting in frame-shift and premature stop codon
SlYTHDF1 / SlYTHDF3	Solyc01g028860.2.1 / Solyc12g099090.1.1	4144 / 3784	T insertion / T insertion	1913-1914 / 505-506	Double homozygous knock-out
SlYTHDF2 / SlYTHDF3	Solyc05g032850.2.1 / Solyc12g099090.1.1	10214 / 3784	34bp deletion / 4bp deletion	5200-5234 / 505-509	Double homozygous knock-out
SlYTHDF2 / SlYTHDF4	Solyc05g032850.2.1 / Solyc01g103540.2.1	10214 / 6008	34bp deletion / C insertion	5200-5234 / 3726-3727	Double homozygous knock-out
SlYTHDF3 / SlYTHDF4	Solyc12g099090.1.1 / Solyc01g103540.2.1	3784 / 6008	T insertion / C insertion	505-509 / 3726-3727	Double homozygous knock-out

A table is provided to give an overview of the type, position, and description of mutations identified previously, as well as the recent double homozygous mutation. The table includes the gene name, loci, gene length, mutation, position, and description.



Figure 6.5 Phenotype displaying single homozygous knock-out of *ythdf* mutant plant.

Representative photos showing the phenotype of plant growth for single mutants of *ythdf1*, *ythdf2*, *ythdf3*, and *ythdf4* are presented. The plants were imaged at 4-weeks-old, grown on ½ MS medium in clear plastic pots, under controlled growth conditions in Conviron at 25°C with a 16-hour day length. Scale bar = 1.5 cm.



Figure 6.6 Phenotype displaying double homozygous knock-out of *ythdf* mutant plants.

Representative photos showing the phenotype of plant growth for double mutants of *ythdf1/ythdf3, ythdf2/ythdf3, ythdf2/ythdf4, and ythdf3/ythdf4* are presented. The plants were imaged at 4 weeks old, grown on ½ MS medium in clear plastic pots, under controlled growth conditions in Conviron at 25°C with a 16-hour day length. Scale bar = 1.5 cm.

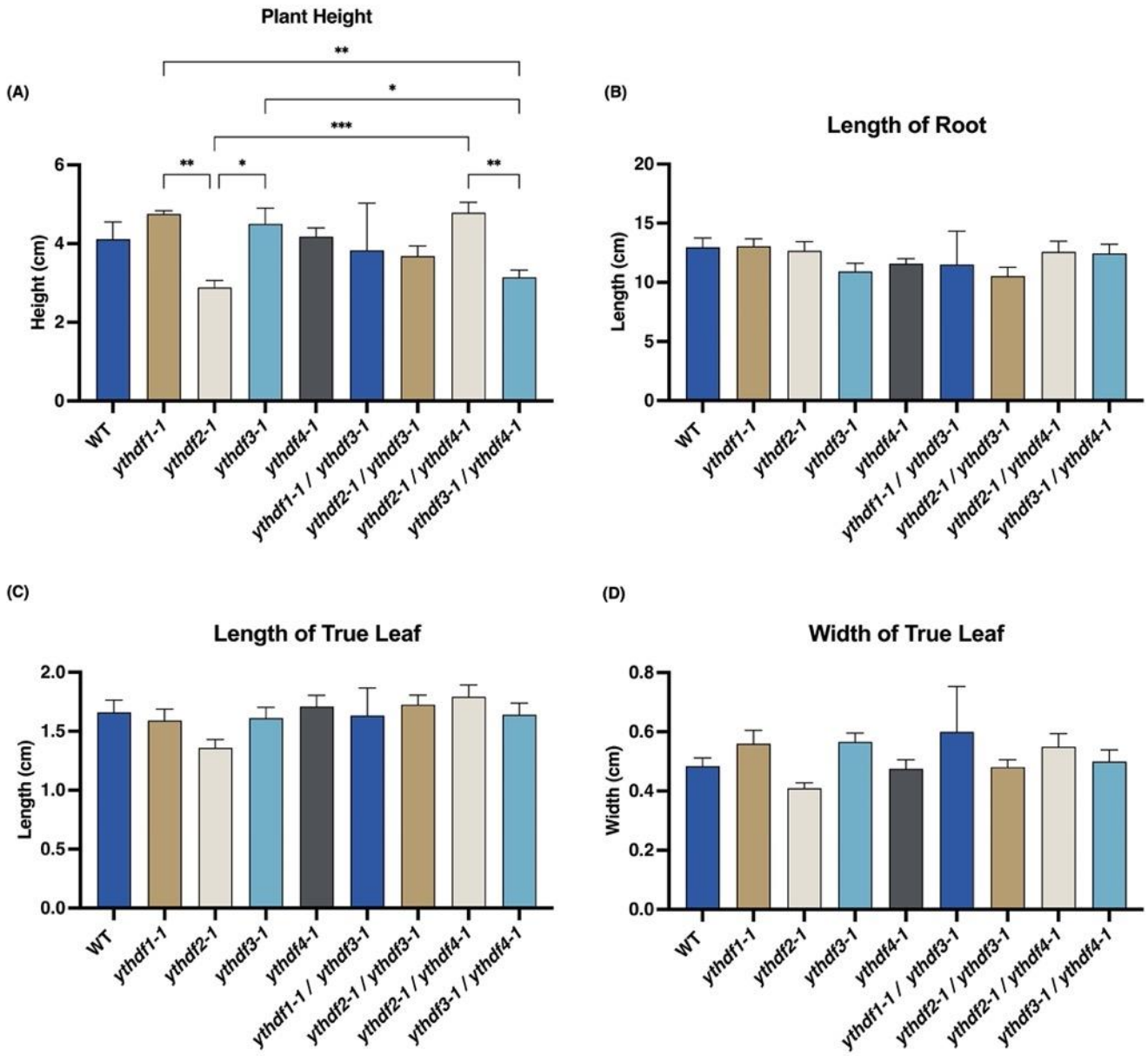


Figure 6.7 Characterisation of single and double mutant knock-out lines of *ythdf1 - ythdf4.* Histogram representation of key plant growth measurements.

Key plant growth measurements for single mutants of *ythdf1–ythdf4*, as well as for double mutants of *ythdf1/ythdf3*, *ythdf2/ythdf3*, *ythdf2/ythdf4*, and *ythdf3/ythdf4*, are presented. Plants were grown on ½ MS medium in clear plastic pots under controlled growth conditions in Conviron at 25°C with a 16-hour day length. Error bars represent SEM, n = 10-15, with * indicating a significant difference: * $p < 0.05$, ** $p < 0.005$, and *** $p < 0.0005$, respectively, as determined by ANOVA–Tukey multiple comparisons test.

6.3.2 Gene expression patterns for *YTHDF1-YTHDF4* across different tissues and developmental stages

The tomato functional genomics database is an online resource that is made available to access a wide range of data related to tomato genomics providing researchers with access to vast amounts of information on this important plant crop. Data from this database was used in this study to analyse expression data for each of the four *YTHDF* gene present in tomato. The analysis was done across different sections of the plant including leaves, roots, mature green fruits, breaker fruit, unopened flower buds, and fully opened flowers.

In leaves, the data indicate that *YTHDF2* and *YTHDF3* exhibit the lowest expression levels among the four *YTHDF* gene, while *YTHDF4* shows the highest expression level, and *YTHDF1* shows the second-highest expression level. The same trend is observed in roots, where *YTHDF4* exhibits the highest expression level and *YTHDF1* shows the second-highest expression level. Notably, the expression level of *YTHDF4* in roots is approximately 28% higher than that in leaves. Comparison of different stages of fruit ripening from mature green fruits to breaker fruits revealed a 15% reduction in expression levels for *YTHDF1*, while *YTHDF4* showed a 7% increase in expression level.

Comparison of expression levels of *YTHDF1* and *YTHDF4* was selected across different stages of flower development ranging from unopened flower buds to fully opened flowers. The data revealed that *YTHDF1* exhibited a 92.5% increase in expression levels from unopened to opened flowers, which was consistent with the 90% increase observed for *YTHDF4* expression levels. Overall, across all six parts of the plant analysed, *YTHDF2* and *YTHDF3* exhibited minimal changes in expression levels, and consistently showed lower expression levels when compared to *YTHDF1* and *YTHDF4*.

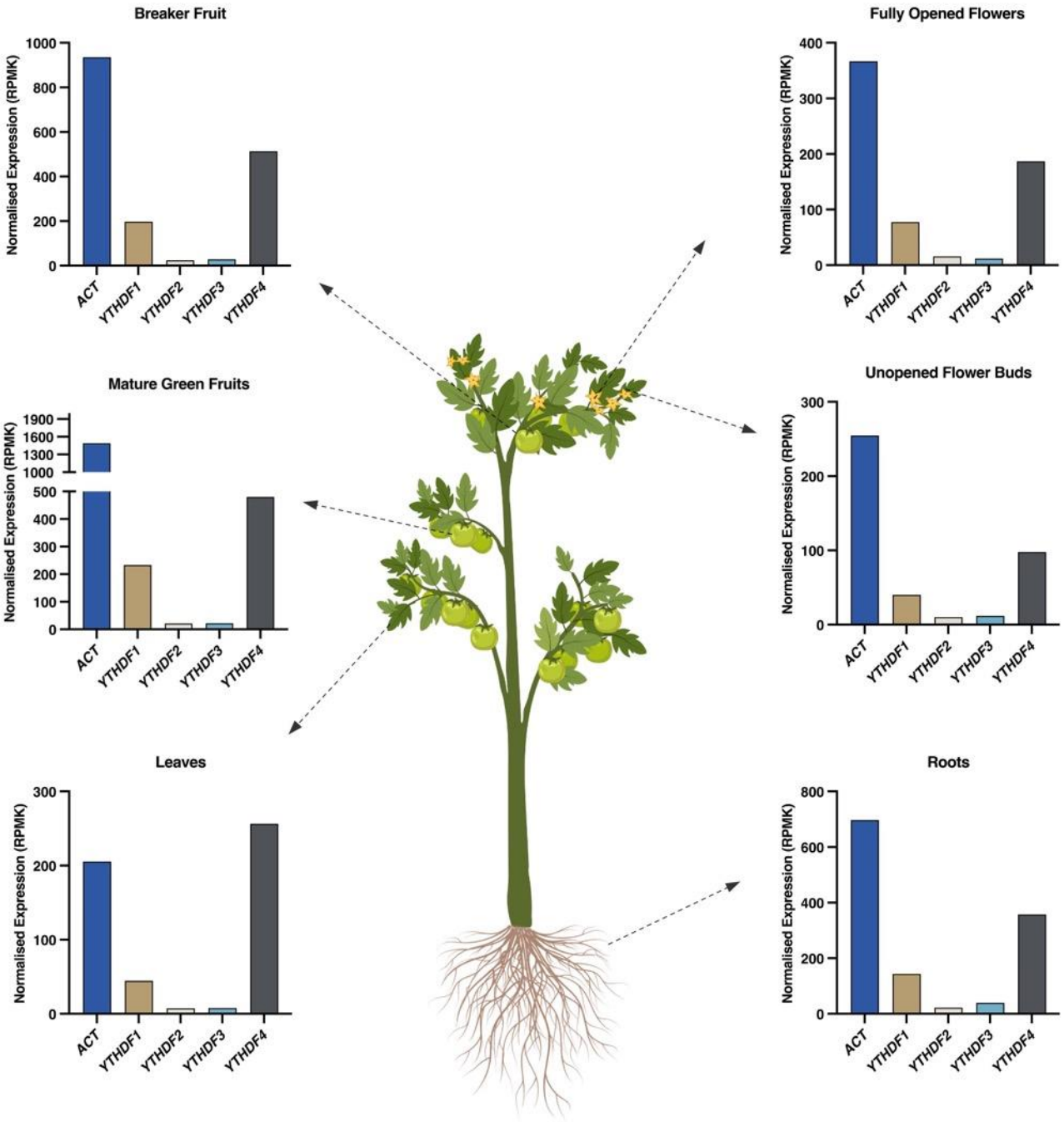


Figure 6.8 Normalised expression of *YTHDF1-YTHDF4* on different tomato physiological parts.

Transcriptome analysis of different parts of the *S. lycopersicum* seedlings and mature plant for *YTHDF1-YTHDF4*. Leaves and roots data was taken at 3 weeks old seedlings. Normalised expression was compared against *ACT* (Solyc03g078400). Data taken from, "Tomato Functional Genomics Database", (http://ted.bti.cornell.edu/).

6.3.3 Physiological characterisation of YTHDF single knock-out mutant

To enable large number of plant comparisons of the single KO mutants, the plants were phenotyped under controlled growth conditions in a growth room. The plants were grown on ½ MS medium to allow for observations of both shoots and roots. The results indicated that there were no significant differences in plant height between the single KO mutants of *ythdf1-ythdf4* when compared to the WT.

In this study, we observed no significant differences in plant height among the *YTHDF* single and double KO mutants when compared to WT. Although, *ythdf2-1* exhibited a significant decrease in plant height compared to *ythdf1-1* and *ythdf3-1*, with reductions of 39% and 36%, respectively. Conversely, a significant increase in plant height of 39% was observed in *ythdf2-1/ythdf4-1* double KO plants compared to *ythdf2-1* single KO plants. In addition, *ythdf3-1/ythdf4-1* double KO plants showed a significant decrease in plant height of 34% when compared to *ythdf1-1* single KO plants. The *ythdf3-1/ythdf4-1* double KO plants showed a 34% reduction in plant height, which was the only notable difference between them and the *ythdf2-1/ythdf4-1* double KO plants. In addition to plant height, several other plant growth measurements, including root length, true leaf count, and true leaf width, were measured for both single and double KO lines. However, there were no significant changes observed in this measurement between single and double KO lines. The *ythdf2-1* single KO did show the most significant decrease in plant height, length, and width of true leaf which could indicate the importance of this specific reader protein in tomato.

6.4 Discussion

The process of mRNA methylation is crucial for regulating gene expression, and the proteins containing YTH domains have been demonstrated to play a vital role in recognising m^6A modifications. The ability to recognise these modifications provides several levels of post-transcriptional gene control, which enhances the complexity of gene regulation associated with m^6A methylation. Several studies have investigated YTH proteins in *A. thaliana* but generating multiple KOs has been challenging due to the abundance of ECT proteins. Despite this, the functions of YTH proteins in tomato remain largely unknown. Investigating the roles of YTH proteins in tomato could offer valuable insights into RNA metabolism and gene regulation. The generation of single and double KOs for each of the YTHDF1-YTHDF4 proteins in tomato may provide insight into how these proteins function. The homozygous single KO mutants of *ythdf1-1*, *ythdf2-1*, *ythdf3-1*, and *ythdf4-1* were previously generated, and F_3 plants were subsequently checked to confirm the inheritance of the mutation. Phenotypic analysis of the single KO mutants in growth media revealed significant differences in plant height, with *ythdf2-1* showing the most pronounced reduction in height compared to *ythdf1-1* and *ythdf3-1*, with decreases of 39% and 36%, respectively.

6.4.1 Identification of double homozygous YTHDF1-YTHDF4 knock-outs

The primary objective of this chapter was to screen and identify homozygous double KOs involving all possible combinations of *YTHDF1* to *YTHDF4*. The identification of double homozygous KO lines includes *ythdf1-1/yth3-1, ythdf2-1/yth3-1, ythdf2-1/ythdf4-1*, and *ythdf3-1/ythdf4-1* was successfully achieved, indicating that these double KO proteins do not result in embryo lethality, unlike the KOs of writer complex proteins in *A. thaliana* such as MTA, MTB, VIRLIZER (Zhong *et al.*, 2008; Růžička *et al.*, 2017). Some of the double KO lines were also identified by Zuzanna Lorenz a master's student under my supervision at the time of work. The generation of stable lines for each of the double KOs may suggest redundancy among

YTHDF proteins in their respective KOs for tomato. The expansion and complexity found within the YTHDF group of proteins has proved to be a difficulty in generating simultaneous KOs of all the YTHDF proteins presented in *A. thaliana*. Current research has mainly focused on the study of *A. thaliana* ECT 2,3, and 4 proteins (Arribas-Hernández *et al.*, 2018, 2021; Scutenaire *et al.*, 2018), which together with ECT1, belong to a single cluster (homologous to tomato YTHDF1). Expanding the range of YTH mutant combinations would provide valuable insights into uncovering the roles of this protein family.

Previous work by Cameron Grundy, (2021) involved analysing full-length YTH proteins using MUSCLE and MEGA 10.1 for alignment. The resulting tree produced from the program allowed identification of the closest homologs of each tomato YTHDF protein in the groupings presented for *A. thaliana*. The phylogeny tree revealed that tomato *YTHDF1* is likely to be represented by *AtECT1*, *AtECT2*, *AtECT3*, and *AtECT4*. The tomato *YTHDF2* falls into a monophyletic subgroup with homologs containing *AtECT10* and *AtECT5*. The tomato *YTHDF3* is closely related to *AtECT6* and *AtECT7*, while the closest homolog of tomato *YTHDF4* in *A. thaliana* is *AtECT11*.

Additional research conducted by Scutenaire *et al.*, (2018) used a similar phylogenetic analysis approach, focusing solely on the YTH domains of the YTH proteins of key plant models. Their analysis identified three distinct groups. The first group, labelled DFA, included *SlYTHDF1* along with *AtECT1*, *AtECT2*, *AtECT3*, and *AtECT4*. The second group, labelled DFB, included *SlYTHDF2* along with *AtECT5*, *AtECT10*, and *AtECT9*. The third group, labelled DFC, included *SlYTHDF3* and *SlYTHDF4* along with *AtECT6*, *AtECT7*, *AtECT8*, and *AtECT11*. These findings are consistent with those of our research group phylogeny tree and provide further insights into the phylogenetic relationships among YTHDF proteins in key plant models. The grouping of YTHDF and ECT proteins based on phylogenetic analysis suggests that KOs of certain tomato SlYTHDF protein may be equivalent to simultaneous KO of several members of the *A. thaliana* family, which could mean there is functional redundancy between members in each of the four YTHDF groups. This provides valuable knowledge for designing

new experiments to investigate KO combinations of ECT proteins in *A. thaliana*. The studies on *ythdf1-1/ythdf3-1* suggest that SlYTHDF2 and SlYTHDF4 may exhibit compensatory functions in interpreting the m^6A modification. The compensatory function might imply that in the absence of SlYTHDF1 and SlYTHDF3, SlYTHDF2 and SlYTHDF4 could have recognised the m^6A mark and elucidate similar functions to allow for normal plant development, which could be a backup strategy that plants have evolved. This might explain why there are multiple reader proteins in plants. These findings raise the possibility that knockouts of *AtECT1, AtECT2, AtECT3, AtECT4, AtECT6*, and *AtECT7* in *A. thaliana* could still result in viable plants, albeit potentially with altered phenotypes or cellular characteristics in relation to m^6A methylation. Further transcriptomic analyses, such as RNA-Seq or MeRIP-Seq, could be valuable in revealing the functions of reader proteins. Similar experiments and observations may be conducted for other combinations of double KOs in tomato and *A. thaliana*.

Attempts were made to obtain double homozygous KOs of *ythdf1-1/ythdf4-1* were undertaken by screening F_2 seeds; however, only one homozygous KO of either *ythdf1-1* or *ythdf4-1*, while the other was heterozygous for the mutation, could be obtained. Despite screening a larger population and multiple lines, analysis of 12 plants derived from *ythdf1-1* heterozygous and *ythdf4-1* homozygous crosses failed to identify any double homozygous KOs. The failed results could infer that double homozygous KOs of *ythdf1-1/ythdf4-1* indicate that the proteins YTHDF1 and YTHDF4 play a crucial role in plant development, and at least one of these reader proteins is required for normal growth and development. However, further screening of a larger population and different lines is required to confirm this hypothesis. The analysis of relative expression levels of individual *YTHDF* genes in different parts of the tomato plant revealed that *YTHDF1* and *YTHDF4* had the highest expression levels among the four *YTHDF* genes. The high expression levels observed suggest the potential significance of these genes in the tomato plant, which may explain the lack of possible double KOs for both genes. Double heterozygous KOs of *ythdf1-1/ythdf2-1* were obtained; however, screening the progeny of these seeds was not feasible due to time constraints.

It is important to consider additional factors when interpreting m^{6}A modifications within the m^{6}A machinery. Intriguingly, when plants with different double KOs were cultivated to produce bulk seeds in glasshouses, they exhibited a distinctive phenotype when compared to the WT in specific KOs of *YTHDF1/YTHDF3* and *YTHDF3/YTHDF4*, including reduced apical dominance, and increased lateral branching, which resembles the phenotype observed in low methylation plants in *A. thaliana* (Figure A9.1). These findings suggest that phenotypic alterations could have occurred at different stages of the plant's life cycle, as they were not observed during early growth stages. It can be posited that different YTHDF proteins may be expressed in different tissues, such as *AtECT10* mRNA, which has been demonstrated to be highly abundant in pollen (Arribas-Hernández *et al.*, 2018).

The various combinations of single and double KOs for *YTHDF1*, *YTHDF2*, *YTHDF3*, and *YTHDF4* highlight the intricate roles of reader proteins in interpreting m^{6}A modifications and their downstream effects on cellular pathways. The complexity of reader protein functions can be attributed to one or more mechanisms in interpreting m^{6}A modifications, potentially induced by environmental stimuli, as well as differences in target-binding specificity. Further research using techniques such as transcriptomic and proteomic analyses can provide insights into the specific pathways affected by the KO of m^{6}A reader proteins.

The distribution of m^{6}A modifications on mRNA may have specific implications for the binding affinity and recognition of different YTHDF proteins. Additionally, competition for binding sites, regulation of m^{6}A methylation deposition by the writer complex, and links to histone modifications and m^{6}A enrichment are all possible factors. The findings presented in this chapter provide further evidence of the redundant functions of YTHDF proteins, however, it is necessary to gain a deeper understanding of how this redundancy is achieved in order to fully comprehend the role of YTHDF proteins in interpreting m^{6}A modifications.

6.4.2 Summary and future work

Although no significant changes were observed in plant height, length, and width of true leaves when compared to WT, further experiments on growing single and double mutants in different conditions should be performed. The limitations of ½ MS growth medium used in this study could have affected the results. Observations from plants grown in the glasshouse beyond 5 weeks have shown severe developmental phenotypes in the single and double knockout mutants. Therefore, further genetic, and biochemical studies on the YTHDF proteins are needed to understand the redundant and specific functions of each protein. Investigating the expression data of each *YTHDF* gene could provide valuable insights into their compensatory roles as described in this chapter. Conducting qRT-PCR experiments on various single and double KO lines may reveal if other sets of *YTHDFs* are upregulated in response to the KOs of certain *YTHDF* genes. Furthermore, exploring the effects of these mutant lines on fruit ripening could be an interesting area for future research. The effects of environmental factors such as temperature, light or drought stresses, or of various pathogens could also be studied using the different mutant lines.

Chapter 7

General Discussion

7 General Discussion

7.1 Suppression of *fip37-4*

The work laid out in this thesis covers two key model plant systems, *A. thaliana* and *S. lycopersicum*, exploring the epigenetic m^{6}A methylation modification on mRNA in both the writer and reader proteins. The m^{6}A modification plays a key role in gene regulation, as it has continually been highlighted that this modification is a key addition to the mRNA. In Chapter 3, the identification of two suppressor lines that exhibit a reversion of phenotype, similar to that of WT was successful. The observation of a suppressor line suggests the presence of other mechanisms that could be related to m^{6}A methylation but are currently unknown. The phenotypic observation of plant height was analysed in both suppressor lines (P3-07 + P10-23) to determine if they did show reversion of plant height back to WT. Interestingly, the *fip37-4* did show a variation in plant height, which is highly attributed to the T-DNA insertion located in the intron producing a variable KD effect of FIP37. Both suppressor lines did show a difference in height variability, as P3-07 displayed a semi-dominant mutation type of segregation growth ratio while P10-23 showed a dominant mutation segregation growth ratio.

Further phenotypic characterisation involved the observation of root architecture in both suppressor lines, P3-07 and P10-23, which both exhibited a suppression of *fip37-4* root phenotype. Previous research has also reported that hypomorphic mutants of m^{6}A writers display abnormal root phenotypes associated with root cap formation, gravity response, main and lateral root development, and auxin response defects (Zhong *et al.*, 2008; Bodi *et al.*, 2012; Růžička *et al.*, 2017; Zhang *et al.*, 2022). The PRL of P3-07 was significantly increased by 24% compared to *fip37-4*, and P10-23 displayed a complete reversion back to the WT PRL. Interestingly, under high salinity conditions, P10-23 did not exhibit any reversion back to the WT, while under drought conditions, both P3-07 and P10-23 showed full reversion back to the WT length. This observation could be an interesting avenue to explore since plants have

evolved various mechanisms to cope with drought conditions, including reducing water loss through increased diffusive resistance, enhancing water uptake through deep and prolific root systems, and minimising transpiration loss through smaller and succulent leaves. This range of physiological changes coupled with gene regulation could unravel new pathways associated with m^6A methylation in relation to plant stress response. Lateral root density was also measured and demonstrated that under normal growth conditions, both P3-07 and P10-23 reverted back to WT levels.

The trichome branching numbers was also measured and the suppressor lines exhibited a reversion back to the WT branching phenotype, as well as lateral root initiation. Specifically, P3-07 showed a reversion, with 40-50% of trichome branches ranging from three to four, whereas *fip37-4* only displayed 11% of trichome branches with three branches. P10-23 exhibited a similar reversion of 3 trichome branches more than P3-07 which indicated a stronger suppression.

To investigate the extent of the suppression of the *fip37-4* phenotype, m^6A methylation levels was measured using the TLC method, which has previously been reported to be reduced in mutants of FIP37 (Růžička *et al.*, 2017). The results revealed that while P3-07 did not show any significant changes in global m^6A levels, P10-23 exhibited a closer reversion to WT levels with only a 28% difference when compared to WT m^6A levels. It's interesting to note that P3-07 and P10-23 both showed WT-like leaf shape suppression, though P3-07 had significantly less m^6A level suppression than P10-23 both having no serrated leaf shape and larger leaf area when compared to *fip37-4*.

To analyse transcript abundance and possible splicing effects of *fip37-4* and the suppressor mutants, Northern blot was performed to determine the expression levels of FIP37. As *fip37-4* is a Salk line with the T-DNA insertion located in the intron, it produces two types of transcripts, either a WT of FIP37 or a non-functional transcript which produces a truncated protein. The results indicated that P10-23 had an increased level of correctly spliced FIP37

transcript when compared to *fip37-4* and showed similar levels to the WT. In contrast, this was different in P3-07 as it displayed only a slight increase in correctly spliced FIP37 transcripts. It is plausible to speculate that both suppressor lines are showing suppression of the phenotype through an increase in splicing efficiency, as m^6A has been linked to affect splicing. However, it should be noted that other factors may be involved in the suppression of *fip37-4* such as other mutations within the genome generated by EMS mutagenesis or other induced pathways that are generated from KO/KD of many candidate genes.

BSA, accompanied by NGS, was performed on both P3-07 and P10-23, resulting in two sets of candidate gene lists being produced. The filtering process for candidate genes associated with each line was comparably different, as P3-07 is a semi-dominant mutation, while P10-23 exhibited a dominant mutation. Firstly, P3-07 filtering of variants was performed by selecting unique homozygous variants present only in the tall DNA pool. This is reflected during BSA analysis, as only tall plants were chosen for DNA extraction and processed for sequencing. For P10-23, filtering step 1 was conducted by identifying unique homozygous variants present in the tall DNA pool. Subsequently, this was filtered by using the VEP tool, and only variants with high impact consequences were selected, resulting in a list of 55 candidate genes. A second set of filtering method was carried out to ensure that a comprehensive filtering process was performed, and other possible candidate genes were not overlooked due to bioinformatics processing.

For P3-07, filtering step 2 was done utilising NGS read counts for both mutant and WT alleles. This approach was based on the hypothesis that if a heterozygous plant (medium height) was present in the DNA pool, the resulting mutation would have been introduced, and subsequent sequencing would have shown at least 66% of the mutation allele being present. As a result, the filtering process selected a WT to mutant read count ratio of 0.66 or higher, yielding a list of 45 potential candidate genes. However, it should be noted that this filtering method is dependent on the read depth of the NGS sequencing, which in our data ranges from 2 to 60 reads, potentially giving false candidate genes in low number reads. Filtering step 2 was also

carried out on P10-23 (Dominant mutation), and both heterozygous and homozygous mutations were included in the filtering process. At least 66% of the mutation was expected to be present in the tall DNA pool. However, this filtering step resulted in a significantly higher number of candidates, with 382 potential candidate genes identified shown , compared to filtering step 1 for P10-23. Both P3-07 and P10-23 produced different sets of candidate gene lists through two different methods of filtering candidate genes. The identification of two independent suppressors, each with a different type of heritability of mutation, could indicate that different mechanisms are involved in the suppression of *fip37-4* between the two lines.

In summary, the findings in Chapter 3 demonstrate that the suppression of *fip37-4* is achievable, as evidenced by the reversion of plants to a WT phenotype, the increase in correctly spliced *FIP37* transcript, and the varying reversion of global m^6A levels. These combinations of observations suggest the presence of novel regulatory mechanisms involved in m^6A methylation. If a candidate gene was chosen, further experiments involving a double KO of the chosen candidate in a *fip37-4* background is needed. Due to *fip37-4* presence of a T-DNA insertion in the intron, it would be ideal if a CRISPR/Cas9 mutant of the candidate was made rather than crossing two Salk T-DNA lines. Subsequent, sexual crossing of the CRISPR mutant with *fip37-4* Salk line will be needed. This would also confirm that the candidate was responsible for the suppression and absent from other unrelated mutations that is present in the suppressor lines generated by EMS mutagenesis. Since this modification is a dynamic process involving reader proteins, research on both groups of proteins may uncover whether RNA-binding proteins or splicing factors are involved. Future studies could involve EMS mutagenesis on different writer or reader proteins to identify new proteins or pathways.

7.2 RNA-seq analysis on suppressor mutants

To obtain a comprehensive view of the transcriptome and uncover differential gene expression changes in the suppressor lines P3-07 and P10-23, RNA-seq was performed. The idea of an increase in splicing efficiency was proposed in Chapter 3 and this was analysed in Chapter 4

to uncover if an increase in splicing efficiency could lead to the reversion of functional FIP37. An increase in splicing efficiency could indicate the reversion of functional FIP37 transcripts. A Sashimi plot was generated, which showed a potential increase in junction reads for exon 8 exclusion in both lines. However, this exclusion of exon 8 was also present in *fip37-4*, indicating that it could be a product of the T-DNA insertion. To further analyse if splicing is causing suppression of *fip37-4*, the ratio of correctly spliced exon junctions from exon 7 to exon 8 was determined; however, this did not show any significant change from *fip37-4*. Overall, an increase in splicing efficiency could be the cause of suppression and cannot be ruled out as m^6A methylation has been linked to a variety of splicing events (Haussmann *et al.*, 2016; Li *et al.*, 2019).

Interestingly, the DEG analysis for P3-07 compared to *fip37-4* revealed the downregulation of -1.84 Log2FC for the gene *TN8* (AT1G72910) in P3-07. *TN8* is a nucleotide-binding, leucine-rich repeat, and NMD regulator (NLR) gene, which regulates autoimmunity. NMD is a pathway that removes aberrant transcripts to prevent the accumulation of truncated proteins. NLR gene activity is regulated at the transcriptional and post-transcriptional levels (Staiger *et al.*, 2013). NLR transcripts contain NMD signatures that activate the NMD machinery, resulting in transcript degradation or translational suppression. Recent studies have shown that *upf1/upf3* mutants displayed growth arrest, which was rescued by miRNA knockdown of three NLR genes, including *TN8* (Nasim *et al.*, 2020). Consistent with *TN8* activity the GO analysis for P3-07 revealed an enrichment with the term defence response. Further exploration of low methylation lines with plant defence signalling and pathogen response would be interesting, as NMD gene regulation accompanied by NLR, along with the CPSF30 machinery could unravel the interlinks between plant immunity and m^6A modification. Overall, the RNA-seq did reveal significant DEGs for both lines, and these sets of genes could contribute to the suppression of *fip37-4*. However, the list of DEGs is still extensive, and further cross referencing to the BSA candidate genes and filtering are needed on subsequent backcrosses.

7.3 Knock-down of key writer complex protein MTA in tomato

The idea of generating a combination of writer and reader KO/KD was the main reason for the efforts to produce a KO version of MTA in tomato. Initial attempts to generate KOs using CRISPR/Cas9 proved to be unreliable and a inducible hpRNAi system was utilised to generate a KD version of *MTA* in tomato. However, contrary to the expected knockdown of *MTA*, the transcript expression level measured by qRT-PCR displayed an increase in expression for *MTA* after DEX induction. It was hypothesised that accumulation of *MTA* transcripts from the presence of low MTA protein or transcript, along with slow nuclear export to the cytoplasm, contributed to this observation. Further experiments are needed to validate this hypothesis, including using different timepoints for RNA extraction and qRT-PCR after DEX induction, such as 24 hours after DEX induction, as consistent with other research papers that used the pOpOff2(kan) system.

To confirm if our generated *RNAi-MTA* construct using the pOpOff2(kan) system was efficiently expressing, GUS staining was performed on two independent lines, which showed strong staining in whole seedlings of *RNAi-MTA1* and *RNAi-MTA2*. Further phenotypic observations were carried out, which showed a significant reduction in plant height, root length, width, and height of true leaves upon DEX induction for both lines grown on media as compared to the WT. Additionally, there was an increase in the number of nodes and lateral branches compared to WT for both lines grown on media, supporting the findings. Further physiological characterisation was performed on *RNAi-MTA1* and *RNAi-MTA2*, which showed a significant decrease in the length and width of true leaves compared to WT. The number of nodes and lateral branches significantly increased as well, which is consistent with growth conditions on media.

The presence of physiological changes displayed in the *RNAi-MTA1* and *RNAi-MTA2* lines could be explained in the lack of MTA protein present; however, m^6A levels were measured and showed no changes in methylation levels in the induced line. This measurement could

explain that MTA protein levels were not decreased as a functional MTC complex is needed for m^6A methylation. MTA could have different functions in tomato as compared to *A. thaliana* which could display different outcomes of *MTA* KD.

7.4 Single and double knock-outs of YTHDF proteins in tomato

Prior to Chapter 6, the present research employed *A. thaliana* and tomato as model organisms for studying individual writer complex proteins. Based on the results of the phylogenetic analysis, tomato could be an excellent model for investigating reader proteins since it possesses only 4 YTHDF proteins, which is fewer than the model organism *A. thaliana*. It is noteworthy that plants harbour a greater number of YTH domain proteins than other eukaryotes. The creation of single KO lines for each of the YTHDF1-YTHDF4 proteins did not result in embryo lethality of the tomato plant.

In Chapter 6, double homozygous KO of four combinations of YTHDF1-YTHDF4 was positively confirmed. These four KOs also did not result in embryo lethality in the tomato plant. In contrast, two combinations of double KO in *ythdf1/ythdf2* and *ythdf1/ythdf4* resulted in no double homozygous KO. The difficulty in generating a double homozygous KO of *ythdf1/ythdf2* could indicate that these two proteins are essential in reading m^6A methylation. However, further screening is needed to confirm this claim using different mutant lines and an increase in the plant number for screening. The double KO combination of *ythdf1/ythdf4* was not identified in the screening, as either KO of YTHDF1 or YTHDF4 homozygous mutation was only possible. These observations could indicate that these two proteins may perform an essential function in plant development, as highlighted by the high expression of these two genes in different physiological parts of the plant (Figure 6.8). It is important to note that further screening of alternative lines should be performed to confidently validate that double KOs of these two combinations are not possible.

The idea of compensatory function was discussed in Chapter 6, that if a combination of 2 YTHDF protein was knocked out, the other two could potentially be compensating to play the same role as the missing proteins for reading m^6A and directing normal plant development. To further add on, each of the YTHDF protein in tomato could have independent roles apart from reading m^6A modification, which may explain the high number of reader proteins in *A .thaliana* and, to a lesser extent tomato. Comparing this idea to *A. thaliana* MTC, there is evidence that individual proteins of the MTC do have other functions apart from depositing m^6A. This is the case for MTA with miRNA as reported by Bhat *et al.*, (2020).

The vast number of reader proteins could indicate that each reader protein could be triggered in different developmental stages of the plant. This was exemplified by the severe physiological change observed in double homozygous KO of *ythdf1/ythdf3* and *ythdf3/tyhdf4* in tomato as both plants exhibited an increased in lateral branches and reduced internode length in the middle section of the plant during the later stages in the plant's life cycle Figure A9.1. As a large number of reader proteins are present, each may be expressed in different tissues, for example *AtECT10* in pollen (Arribas-Hernández *et al.*, 2018). This was taken into consideration and the approach to visualise reader proteins in tomato was undertaken. In this study, the promoter-GUS fusion constructs was made using the pGWB3 vector system by Nakagawa *et al.*, (2007) for YTHDF1 to YTHDF4. All four constructs were then transformed into tomato cotyledons via *Agrobacterium* and their successful integration was confirmed by PCR amplification. However, GUS staining experiments on two positive lines did not reveal any staining, which may be attributed to the pGWB3 vector system that was chosen.

7.5 Conclusion

In conclusion to this thesis, the main aim of Chapter 3 was to identify suppressor lines that revert *fip37-4* back to the WT phenotype. This was achieved, and further analysis of root length, m^6A levels, transcript levels, trichome branching, root initiation, and plant height was

performed to comprehensively demonstrate reversion towards the WT phenotype or levels. RNA-seq transcriptomic analysis was also conducted, revealing interesting enrichment in several key areas, such as defence response for P3-07 and DNA binding transcription factor activity for P10-23, suggesting that suppression may be linked to either of these pathways. The production of a candidate gene list for both independent lines provided an idea into the possible genes that may have caused the suppression, but further backcrossing and sequencing are necessary to carry out further experiments.

The introduction of tomato as a plant species in this thesis presented a different avenue for studying m^6A methylation. This was done in direct response to the knowledge that single homozygous KOs of reader proteins were possible, as demonstrated by previous experiments. The identification of double homozygous KOs was crucial in unravelling possible redundancy between reader proteins and presented an opportunity to create a writer and reader KO/KD in tomato, which, to date, has not been achieved.

8 References

Abas L., Kolb M., Stadlmann J., Janacek D.P., Lukic K., Schwechheimer C., … Hammes U.Z. (2020) Naphthylphthalamic acid associates with and inhibits PIN auxin transporters. *Proceedings of the National Academy of Sciences of the United States of America* **118**, 1–8.

Adams J.M. & Cory S. (1975) Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature* **255**, 28–33.

Adelman K. & Lis J.T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews. Genetics* **13**, 720–731.

Aloni Y., Dhar R. & Khoury G. (1979) Methylation of nuclear simian virus 40 RNAs. *Journal of virology* **32**, 52–60.

Amara U., Hu J., Cai J. & Kang H. (2023) FLK is an mRNA m6A reader that regulates floral transition by modulating the stability and splicing of FLC in Arabidopsis. *Molecular Plant* **16**, 919–929.

Anders S., Reyes A. & Huber W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Research* **22**, 2008–2017.

Arribas-Hernández L., Bressendorff S., Hansen M.H., Poulsen C., Erdmann S. & Brodersen P. (2018) An m6A-YTH module controls developmental timing and morphogenesis in arabidopsis. *Plant Cell* **30**, 952–967.

Arribas-Hernández L., Rennie S., Schon M., Porcelli C., Enugutti B., Andersson R., … Brodersen P. (2021) The YTHDF proteins ECT2 and ECT3 bind largely overlapping target sets and influence target mRNA abundance, not alternative polyadenylation. *eLife* **10**.

Arribas-Hernández L., Simonini S., Hansen M.H., Paredes E.B., Bressendorff S., Dong Y., … Brodersen P. (2020) Recurrent requirement for the m6A-ECT2/ECT3/ECT4 axis in the control of cell proliferation during plant organogenesis. *Development* **147**.

Beemon K. & Keith J. (1977) Localization of N6-methyladenosine in the Rous sarcoma virus genome. *Journal of molecular biology* **113**, 165–179.

Berget S.M., Moore C. & Sharp P.A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 3171–3175.

Bhat S.S., Bielewicz D., Gulanicz T., Bodi Z., Yu X., Anderson S.J., … Szweykowska-Kulinska Z. (2020) mRNA adenosine methylase (MTA) deposits m6A on pri-miRNAs to modulate miRNA biogenesis in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 21785–21795.

Bird C.R., Smith C.J.S., Ray J.A., Moureau P., Bevan M.W., Bird A.S., … Schuch W. (1988) The tomato polygalacturonase gene and ripening-specific expression in transgenic plants. *Plant Molecular Biology* **11**, 651–662.

Boccaletto P., Stefaniak F., Ray A., Cappannini A., Mukherjee S., Purta E., … Bujnicki J.M. (2022) MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Research* **50**, D231–D235.

Bodi Z., Button J.D., Grierson D. & Fray R.G. (2010) Yeast targets for mRNA methylation. *Nucleic acids research* **38**, 5327–5335.

Bodi Z., Zhong S., Mehra S., Song J., Graham N., Li H., … Fray R.G. (2012) Adenosine Methylation in Arabidopsis mRNA is Associated with the 3' End and Reduced Levels Cause Developmental Defects. *Frontiers in Plant Science* **3**, 1–10.

Bokar J.A., Shambaugh M.E., Polayes D., Matera A.G. & Rottman F.M. (1997) Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA (New York, N.Y.)* **3**, 1233–1247.

Boycheva I., Vassileva V. & Iantcheva A. (2014) Histone Acetyltransferases in Plant Development and Plasticity. *Current Genomics* **15**, 28–37.

Brooks C., Nekrasov V., Lipppman Z.B. & Van Eck J. (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiology* **166**, 1292–1297.

Bruggeman Q., Garmier M., de Bont L., Soubigou-Taconnat L., Mazubert C., Benhamed M., … 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 physiology* **165**, 732–746.

Bujnicki J.M., Feder M., Radlinska M. & Blumenthal R.M. (2002) Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m6A methyltransferase. *Journal of Molecular Evolution* **55**, 431–444.

Cameron Grundy (2021) Roles of Post-transcriptional Modifications in Solanum lycopersicum.

Chetty V.J., Ceballos N., Garcia D., Narváez-Vásquez J., Lopez W. & Orozco-Cárdenas M.L. (2013) Evaluation of four Agrobacterium tumefaciens strains for the genetic transformation of tomato (Solanum lycopersicum L.) cultivar Micro-Tom. *Plant Cell Reports* **32**, 239–247.

Chien J.C. & Sussex I.M. (1996) Differential regulation of trichome formation on the adaxial and abaxial leaf surfaces by gibberellins and photoperiod in Arabidopsis thaliana (L.) Heynh. *Plant Physiology* **111**, 1321–1328.

Chow L.T., Gelinas R.E., Broker T.R. & Roberts R.J. (1977) An amazing sequence arrangement at the 5’ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8.

Church C., Moir L., McMurray F., Girard C., Banks G.T., Teboul L., … Cox R.D. (2010) Overexpression of Fto leads to increased food intake and results in obesity. *Nature genetics* **42**, 1086–1092.

Consortium T.T.G., Sato S., Tabata S., Hirakawa H., Asamizu E., Shirasawa K., … Gianese G. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641.

Crevillén P. (2020) Histone Demethylases as Counterbalance to H3K27me3 Silencing in Plants. *iScience* **23**, 1–8.

Crick F. (1970) Central dogma of molecular biology. *Nature*.

Cuming A.C., Kennedy T.D. & Lane B.G. (1979) Wheat embryo ribonucleates. XIV. Mass isolation of mRNA from wheat germ and comparison of its translational capacity with that of mRNA from imbibing wheat embryos. *Canadian journal of biochemistry* **57**, 1170–1175.

Daffis S., Szretter K.J., Schriewer J., Li J., Youn S., Errett J., … Diamond M.S. (2010) 2’-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468**, 452–456.

Dahlberg J.E. & Blattner F.R. (1973) In vitro transcription products of lambda DNA: nucleotide sequences and regulatory sites. *Virus Research*, 533–543.

Delaney K.J., Xu R., Zhang J., Li Q.Q., Yun K.-Y.Y., Falcone D.L. & Hunt A.G. (2006) Calmodulin interacts with and regulates the RNA-binding activity of an Arabidopsis polyadenylation

factor subunit. *Plant Physiology* **140**, 1507–1521.
Desrosiers R., Friderici K. & Rottman F. (1974) Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 3971–3975.
Dominissini D., Moshitch-Moshkovitz S., Salmon-Divon M., Amariglio N. & Rechavi G. (2013) Transcriptome-wide mapping of N6-methyladenosine by m6A-seq based on immunocapturing and massively parallel sequencing. *Nature Protocols* **8**, 176–189.
Dominissini D., Moshitch-Moshkovitz S., Schwartz S., Salmon-Divon M., Ungar L., Osenberg S., … Rechavi G. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206.
Duan H.-C., Wei L.-H., Zhang C., Wang Y., Chen L., Lu Z., … Jia G. (2017) ALKBH10B Is an RNA N6-Methyladenosine Demethylase Affecting Arabidopsis Floral Transition. *The Plant cell* **29**, 2995–3011.
Egger G., Liang G., Aparicio A. & Jones P.A. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**, 457–463.
FAO (2022) *FAO Statistical Yearbook – World Food and Agriculture*. Rome, Italy.
Fernandez-Pozo N., Menda N., Edwards J.D., Saha S., Tecle I.Y., Strickler S.R., … Mueller L.A. (2015) The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. *Nucleic Acids Research* **43**, D1036–D1041.
Frayling T.M., Timpson N.J., Weedon M.N., Zeggini E., Freathy R.M., Lindgren C.M., … McCarthy M.I. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science (New York, N.Y.)* **316**, 889–894.
Fujita Y., Krause G., Scheffner M., Zechner D., Leddy H.E.M., Behrens J., … Birchmeier W. (2002) Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nature cell biology* **4**, 222–231.
Garcia D., Garcia S. & Voinnet O. (2014) Nonsense-mediated decay serves as a general viral restriction mechanism in plants. *Cell host & microbe* **16**, 391–402.
Giovannoni J.J., Wing R.A., Ganal M.W. & Tanksley S.D. (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Research* **19**, 6553–6568.
Glover-Cutter K., Kim S., Espinosa J. & Bentley D.L. (2008) RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature structural & molecular biology* **15**, 71–78.
Guo H.S., Fei J.F., Xie Q. & Chua N.H. (2003) A chemical-regulated inducible RNAi system in plants. *Plant Journal* **34**, 383–392.
Hartmann A.M., Nayler O., Schwaiger F.W., Obermeier A. & Stamm S. (1999) The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). *Molecular Biology of the Cell* **10**, 3909–3926.
Haugland R.A. & Cline M.G. (1980) Post-transcriptional modifications of oat coleoptile ribonucleic acids: 5′-terminal capping and methylation of internal nucleosides in Poly (A)-Rich RNA. *European journal of biochemistry* **104**, 271–277.
Haussmann I.U., Bodi Z., Sanchez-Moran E., Mongan N.P., Archer N., Fray R.G. & Soller M. (2016) M6 A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540**, 301–304.
Himanen K., Boucheron E., Vanneste S., de Almeida Engler J., Inzé D. & Beeckman T. (2002)

Auxin-mediated cell cycle activation during early lateral root initiation. *The Plant cell* **14**, 2339–2351.
Hou N., Li C., He J., Liu Y., Yu S., Malnoy M., … Guan Q. (2022) MdMTA-mediated m6A modification enhances drought tolerance by promoting mRNA stability and translation efficiency of genes involved in lignin deposition and oxidative stress. *New Phytologist* **234**, 1294–1314.
Hu J., Cai J., Park S.J., Lee K., Li Y., Chen Y., … Kang H. (2021) N6-Methyladenosine mRNA methylation is important for salt stress tolerance in Arabidopsis. *Plant Journal* **106**, 1759–1775.
Hu J., Cai J., Umme A., Chen Y., Xu T. & Kang H. (2022a) Unique features of mRNA m6A methylomes during expansion of tomato (Solanum lycopersicum) fruits. *Plant physiology* **188**, 2215–2227.
Hu L., Liu S., Peng Y., Ge R., Su R., Senevirathne C., … He C. (2022b) m(6)A RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nature biotechnology* **40**, 1210–1219.
Huang H., Weng H., Sun W., Qin X., Shi H., Wu H., … Chen J. (2018) Recognition of RNA N 6 - methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nature Cell Biology* **20**, 285–295.
Huang H., Weng H., Zhou K., Wu T., Zhao B.S., Sun M., … Chen J. (2019) Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. *Nature* **567**, 414–419.
Hunt A.G., Xing D. & Li Q.Q. (2012) Plant polyadenylation factors: Conservation and variety in the polyadenylation complex in plants. *BMC Genomics* **13**, 18–20.
Imai Y., Matsuo N., Ogawa S., Tohyama M. & Takagi T. (1998) Cloning of a gene, YT521, for a novel RNA splicing-related protein induced by hypoxia/reoxygenation. *Brain research. Molecular brain research* **53**, 33–40.
Jalili V., Afgan E., Gu Q., Clements D., Blankenberg D., Goecks J., … Nekrutenko A. (2021) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Research* **48**, W395–W402.
Jia G., Fu Y., Zhao X., Dai Q., Zheng G., Yang Y.Y.-G.G., … He C. (2011) N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology* **7**, 885–887.
Kassavetis G.A. & Chamberlin M.J. (1981) Pausing and termination of transcription within the early region of bacteriophage T7 DNA in vitro. *The Journal of biological chemistry* **256**, 2777–2786.
Ke S., Alemu E.A., Mertens C., Gantman E.C., Fak J.J., Mele A., … Darnell R.B. (2015) A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes and Development* **29**, 2037–2053.
Keith G. (1995) Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* **77**, 142–144.
Kim D., Langmead B. & Salzberg S.L. (2015) HISAT: A fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357–360.
Kim J., Kim Y., Yeom M., Kim J.-H. & Nam H.G. (2008) FIONA1 is essential for regulating period length in the Arabidopsis circadian clock. *The Plant cell* **20**, 307–319.
Kim Y.K.I. & Maquat L.E. (2019) UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. *RNA* **25**, 407–422.
Kingston R.E. & Chamberlin M.J. (1981) Pausing and attenuation of in vitro transcription in

the rrnB operon of E. coli. *Cell* **27**, 523–531.
Knuckles P., Lence T., Haussmann I.U., Jacob D., Kreim N., Carl S.H., … Roignant J.-Y. (2018) Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component Wtap/Fl(2)d. *Genes & development* **32**, 415–429.
Krueger F. (2021) Trim Galore. *GitHub repository*.
Kruse S., Zhong S., Bodi Z., Button J., Alcocer M.J.C., Hayes C.J. & Fray R. (2011) A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. *Scientific reports* **1**, 126.
Kurosaki T., Popp M.W. & Maquat L.E. (2019) Nonsense-mediated mRNA decay: quality and quantity control of gene expression in health and disease. *Nature Reviews Molecular Cell Biology*.
Landick R. (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society Transactions* **34**, 1062–1066.
Lence T., Akhtar J., Bayer M., Schmid K., Spindler L., Ho C.H., … Roignant J.-Y. (2016) m6A modulates neuronal functions and sex determination in Drosophila. *Nature* **540**, 242–247.
Levis R. & Penman S. (1978) 5'-terminal structures of poly(A)+ cytoplasmic messenger RNA and of poly(A)+ and poly(A)- heterogeneous nuclear RNA of cells of the dipteran Drosophila melanogaster. *Journal of molecular biology* **120**, 487–515.
Li D., Zhang H., Hong Y., Huang L., Li X., Zhang Y., … 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 molecular biology reporter* **32**, 1169–1186.
Li F., Yi Y., Miao Y., Long W., Long T., Chen S., … Zhao W. (2019) N(6)-Methyladenosine Modulates Nonsense-Mediated mRNA Decay in Human Glioblastoma. *Cancer Research* **79**, 5785–5798.
Li F., Zhao D., Wu J. & Shi Y. (2014b) Structure of the YTH domain of human YTHDF2 in complex with an m6A mononucleotide reveals an aromatic cage for m6A recognition. *Cell Research* **24**, 1490–1492.
Li X., Xiong X. & Yi C. (2017) Epitranscriptome sequencing technologies: decoding RNA modifications. *Nature Methods* **14**, 23–31.
Li Y., Wang X., Li C., Hu S., Yu J. & Song S. (2014c) Transcriptome-wide N^6-methyladenosine profiling of rice callus and leaf reveals the presence of tissue-specific competitors involved in selective mRNA modification. *RNA biology* **11**, 1180–1188.
Liao Y., Smyth G.K. & Shi W. (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.
Linder B., Grozhik A. V, Olarerin-George A.O., Meydan C., Mason C.E. & Jaffrey S.R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature Methods* **12**, 767.
Liu H., Begik O., Lucas M.C., Ramirez J.M., Mason C.E., Wiener D., … Novoa E.M. (2019) Accurate detection of m6A RNA modifications in native RNA sequences. *Nature Communications* **10**, 4079.
Liu J., Yue Y., Han D., Wang X., Fu Y., Zhang L., … He C. (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature chemical biology* **10**, 93–95.
Liu N., Dai Q., Zheng G., He C., Parisien M. & Pan T. (2015) N6 -methyladenosine-dependent

RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564.
Lloyd A.M., Schena M., Walbot V. & Davis R.W. (1994) Epidermal cell fate determination in Arabidopsis: Patterns defined by a steroid-inducible regulator. *Science* **266**, 436–439.
Luo G.-Z., MacQueen A., Zheng G., Duan H., Dore L.C., Lu Z., … He C. (2014) Unique features of the m6A methylome in Arabidopsis thaliana. *Nature Communications* **5**, 5630.
Luo J.H., Wang Y., Wang M., Zhang L.Y., Peng H.R., Zhou Y.Y., … He Y. (2020) Natural variation in RNA m6a methylation and its relationship with translational status. *Plant Physiology* **182**, 332–344.
Luo P., Di D.W., Wu L., Yang J., Lu Y. & Shi W. (2022a) MicroRNAs Are Involved in Regulating Plant Development and Stress Response through Fine-Tuning of TIR1/AFB-Dependent Auxin Signaling. *International Journal of Molecular Sciences* **23**.
Luo Q., Mo J., Chen H., Hu Z., Wang B., Wu J., … Ma J. (2022b) Structural insights into molecular mechanism for N(6)-adenosine methylation by MT-A70 family methyltransferase METTL4. *Nature communications* **13**, 5636.
Luo S. & Tong L. (2014) Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13834–13839.
Ma H., Wang X., Cai J., Dai Q., Natchiar S.K., Lv R., … He C. (2019) N6-Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. *Nature Chemical Biology* **15**, 88–94.
Mandel C.R., Bai Y. & Tong L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cellular and molecular life sciences : CMLS* **65**, 1099–1122.
Martínez-Pérez M., Aparicio F., López-Gresa M.P., Bellés J.M., Sánchez-Navarro J.A. & Pallás V. (2017) Arabidopsis m6A demethylase activity modulates viral infection of a plant virus and the m6A abundance in its genomic RNAs. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 10755–10760.
Martinez De La Cruz B., Gell C., Markus R., Macdonald I., Fray R. & Knight H.M. (2023) m6A mRNA methylation in human brain is disrupted in Lewy body disorders. *Neuropathology and Applied Neurobiology* **49**, e12885.
Mauer J., Luo X., Blanjoie A., Jiao X., Grozhik A. V, Patil D.P., … Jaffrey S.R. (2017) Reversible methylation of m(6)A(m) in the 5' cap controls mRNA stability. *Nature* **541**, 371–375.
Meyer K.D. (2019) DART-seq: an antibody-free method for global m(6)A detection. *Nature methods* **16**, 1275–1280.
Meyer K.D., Patil D.P., Zhou J., Zinoviev A., Skabkin M.A., Elemento O., … Jaffrey S.R. (2015) 5' UTR m6A Promotes Cap-Independent Translation. *Cell* **163**, 999–1010.
Meyer K.D., Saletore Y., Zumbo P., Elemento O., Mason C.E. & Jaffrey S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646.
Michelmore R.W., Paran I. & Kesseli R. V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 9828–9832.
Mielecki D., Zugaj D.Ł., Muszewska A., Piwowarski J., Chojnacka A., Mielecki M., … Grzesiuk E. (2012) Novel AlkB Dioxygenases—Alternative Models for In Silico and In Vivo Studies. *PLOS ONE* **7**, e30588.
Miesak B.H. & Coruzzi G.M. (2002) Molecular and physiological analysis of Arabidopsis mutants defective in cytosolic or chloroplastic aspartate aminotransferase. *Plant*

Physiology **129**, 650–660.
Nakagawa T., Kurose T., Hino T., Tanaka K., Kawamukai M., Niwa Y., … Kimura T. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of bioscience and bioengineering* **104**, 34–41.
Nasim Z., Fahim M., Gawarecka K., Susila H., Jin S., Youn G. & Ahn J.H. (2020) Role of AT1G72910, AT1G72940, and ADR1-LIKE 2 in Plant Immunity under Nonsense-Mediated mRNA Decay-Compromised Conditions at Low Temperatures. *International Journal of Molecular Sciences* **21**.
Nekrasov V., Staskawicz B., Weigel D., Jones J.D.G. & Kamoun S. (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nature Biotechnology* **31**, 691–693.
Nichols J.L. (1979) N6-methyladenosine in maize poly(A)-containing RNA. *Plant Science Letters* **15**, 357–361.
Palusa S.G. & Reddy A.S.N. (2010) Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *The New phytologist* **185**, 83–89.
Panneerdoss S., Eedunuri V.K., Yadav P., Timilsina S., Rajamanickam S., Viswanadhapalli S., … Rao M.K. (2023) Cross-talk among writers, readers, and erasers of m6A regulates cancer growth and progression. *Science Advances* **4**, eaar8263.
Parker M.T., Knop K., Sherwood A. V, Schurch N.J., Mackinnon K., Gould P.D., … Simpson G.G. (2019) Nanopore direct RNA sequencing maps an Arabidopsis N6 methyladenosine epitranscriptome. *BioRxiv*, 706002.
Patil D.P., Chen C.-K., Pickering B.F., Chow A., Jackson C., Guttman M. & Jaffrey S.R. (2016) m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373.
Patil D.P., Pickering B.F. & Jaffrey S.R. (2018) Reading m(6)A in the Transcriptome: m(6)A-Binding Proteins. *Trends in cell biology* **28**, 113–127.
Pendleton K.E., Chen B., Liu K., Hunter O. V, Xie Y., Tu B.P. & Conrad N.K. (2017) The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **169**, 824-835.e14.
Perry R.P. & Kelley D.E. (1974) Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42.
Perry R.P., Kelley D.E., Friderici K. & Rottman F. (1975) The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus. *Cell* **4**, 387–394.
Pesch M. & Hülskamp M. (2009) One, two, three...models for trichome patterning in Arabidopsis? *Current opinion in plant biology* **12**, 587–592.
Phua S.Y., Pornsiriwong W., Chan K.X., Estavillo G.M. & Pogson B.J. (2018) Development of strategies for genetic manipulation and fine-tuning of a chloroplast retrograde signal 3'-phosphoadenosine 5'-phosphate. *Plant Direct* **2**, e00031.
Ping X.-L., Sun B.-F., Wang L., Xiao W., Yang X., Wang W.-J., … Yang Y.-G. (2014) Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research* **24**, 177–189.
Pontier D., Picart C., El Baidouri M., Roudier F., Xu T., Lahmy S., … Lagrange T. (2019) The m(6)A pathway protects the transcriptome integrity by restricting RNA chimera formation in plants. *Life science alliance* **2**.
Proudfoot N.J. & Brownlee G.G. (1976) 3' non-coding region sequences in eukaryotic

messenger RNA. *Nature* **263**, 211–214.
Ramanathan A., Robb G.B. & Chan S.-H. (2016) mRNA capping: biological functions and applications. *Nucleic acids research* **44**, 7511–7526.
Reddy A.S.N. (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual review of plant biology* **58**, 267–294.
Rédei G.P. & Li S.L. (1969) EFFECTS OF X RAYS AND ETHYL METHANESULFONATE ON THE CHLOROPHYLL B LOCUS IN THE SOMA AND ON THE THIAMINE LOCI IN THE GERMLINE OF ARABIDOPSIS. *Genetics* **61**, 453–459.
Rodriguez C.R., Cho E.J., Keogh M.C., Moore C.L., Greenleaf A.L. & Buratowski S. (2000) Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Molecular and cellular biology* **20**, 104–112.
Roeder R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences* **21**, 327–335.
Roundtree I.A., Luo G.Z., Zhang Z., Wang X., Zhou T., Cui Y., … He C. (2017) YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *eLife* **6**, 1–28.
Růžička K., Zhang M., Campilho A., Bodi Z., Kashif M., Saleh M., … Fray R.G. (2017) Identification of factors required for m6A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytologist* **215**, 157–172.
Samalova M., Brzobohaty B. & Moore I. (2005) pOp6/LhGR: A stringently regulated and highly responsive dexamethasone-inducible gene expression system for tobacco. *Plant Journal* **41**, 919–935.
Sato H., Shibata F. & Murata M. (2005) Characterization of a Mis12 homologue in Arabidopsis thaliana. *Chromosome Research* **13**, 827–834.
Schellmann S. & Hülskamp M. (2005) Epidermal differentiation: trichomes in Arabidopsis as a model system. *The International journal of developmental biology* **49**, 579–584.
Schwartz S., Mumbach M.R., Jovanovic M., Wang T., Maciag K., Bushkin G.G., … Regev A. (2014) Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell reports* **8**, 284–296.
Scott L.J., Mohlke K.L., Bonnycastle L.L., Willer C.J., Li Y., Duren W.L., … Boehnke M. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science (New York, N.Y.)* **316**, 1341–1345.
Scutenaire J., Deragon J.M., Jean V., Benhamed M., Raynaud C., Favory J.J., … Bousquet-Antonelli C. (2018) The YTH domain protein ECT2 is an m6A reader required for normal trichome branching in arabidopsis. *Plant Cell* **30**, 986–1005.
Shen H., Luo B., Wang Y., Li J., Hu Z., Xie Q., … Chen G. (2022) Genome-Wide Identification, Classification and Expression Analysis of m6 A Gene Family in Solanum lycopersicum. *International Journal of Molecular Sciences* **23**.
Shen L. (2023) Functional interdependence of m6A methyltransferase complex subunits in Arabidopsis. *The Plant cell*, 1–16.
Shen L., Liang Z., Gu X., Chen Y., Teo Z.W.N., Hou X., … Yu H. (2016) N6-Methyladenosine RNA Modification Regulates Shoot Stem Cell Fate in Arabidopsis. *Developmental Cell* **38**, 186–200.
Shi H., Wang X., Lu Z., Zhao B.S., Ma H., Hsu P.J., … He C. (2017) YTHDF3 facilitates translation and decay of N 6-methyladenosine-modified RNA. *Cell Research* **27**, 315–328.
Shi Y., Di Giammartino D.C., Taylor D., Sarkeshik A., Rice W.J., Yates J.R. 3rd, … Manley J.L.

(2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Molecular cell* **33**, 365–376.
Shi Y. & Manley J.L. (2015) The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes & development* **29**, 889–897.
Shoaib Y., Hu J., Manduzio S. & Kang H. (2021) Alpha-ketoglutarate-dependent dioxygenase homolog 10B, an N6-methyladenosine mRNA demethylase, plays a role in salt stress and abscisic acid responses in Arabidopsis thaliana. *Physiologia Plantarum* **173**, 1078–1089.
Slobodin B., Han R., Calderone V., Vrielink J.A.F.O., Loayza-Puch F., Elkon R. & Agami R. (2017) Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* **169**, 326-337.e12.
Song P., Yang J., Wang C., Lu Q., Shi L., Tayier S. & Jia G. (2021) Arabidopsis N6-methyladenosine reader CPSF30-L recognizes FUE signals to control polyadenylation site choice in liquid-like nuclear bodies. *Molecular Plant* **14**, 571–587.
Staiger D., Korneli C., Lummer M. & Navarro L. (2013) Emerging role for RNA-based regulation in plant immunity. *New Phytologist* **197**, 394–404.
Stoilov P., Rafalska I. & Stamm S. (2002) YTH: a new domain in nuclear proteins. *Trends in biochemical sciences* **27**, 495–497.
Takenaka Y., Kato K., Ogawa-Ohnishi M., Tsuruhama K., Kajiura H., Yagyu K., … Ishimizu T. (2018) Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. *Nature Plants* **4**, 669–676.
Temple H., Phyo P., Yang W., Lyczakowski J.J., Echevarría-Poza A., Yakunin I., … Dupree P. (2022) Golgi-localized putative S-adenosyl methionine transporters required for plant cell wall polysaccharide methylation. *Nature Plants* **8**, 656–669.
Theler D., Dominguez C., Blatter M., Boudet J. & Allain F.H.-T. (2014) Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic acids research* **42**, 13911–13919.
Thomas P.E., Wu X., Liu M., Gaffney B., Ji G., Li Q.Q. & Hunt A.G. (2012) Genome-wide control of polyadenylation site choice by CPSF30 in Arabidopsis. *The Plant cell* **24**, 4376–4388.
Tian B. & Graber J.H. (2012) Signals for pre-mRNA cleavage and polyadenylation. *Wiley interdisciplinary reviews. RNA* **3**, 385–396.
van Tran N., Ernst F.G.M., Hawley B.R., Zorbas C., Ulryck N., Hackert P., … Lafontaine D.L.J. (2019) The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Research* **47**, 7719–7733.
Varsha Wesley S., Helliwell C.A., Smith N.A., Wang M., Rouse D.T., Liu Q., … Waterhouse P.M. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant Journal* **27**, 581–590.
Verelst W., Saedler H. & Münster T. (2007) MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific arabidopsis promoters. *Plant Physiology* **143**, 447–460.
Vespa L., Vachon G., Berger F., Perazza D., Faure J.D. & Herzog M. (2004) The immunophilin-interacting protein AtFIP37 from arabidopsis is essential for plant development and is involved in trichome endoreduplication. *Plant Physiology* **134**, 1283–1292.
Wahl M.C., Will C.L. & Lührmann R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718.
Wang C., Yang J., Song P., Zhang W., Lu Q., Yu Q. & Jia G. (2022) FIONA1 is an RNA N(6)-

methyladenosine methyltransferase affecting Arabidopsis photomorphogenesis and flowering. *Genome biology* **23**, 40.
Wang J., Alvin Chew B.L., Lai Y., Dong H., Xu L., Balamkundu S., … Dedon P.C. (2019) Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic Acids Research* **47**, e130–e130.
Wang P., Doxtader K.A. & Nam Y. (2016) Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Molecular Cell* **63**, 306–317.
Wang X., Lu Z., Gomez A., Hon G.C., Yue Y., Han D., … He C. (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–20.
Wang X., Wu F., Xie Q., Wang H., Wang Y., Yue Y., … Ma L. (2012) SKIP is a component of the spliceosome linking alternative splicing and the circadian clock in Arabidopsis. *The Plant cell* **24**, 3278–3295.
Wang X., Zhao B.S., Roundtree I.A., Lu Z., Han D., Ma H., … He C. (2015) N6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399.
Wang Y., Xiao Y., Dong S., Yu Q. & Jia G. (2020) Antibody-free enzyme-assisted chemical approach for detection of N(6)-methyladenosine. *Nature chemical biology* **16**, 896–903.
Watabe E., Togo-Ohno M., Ishigami Y., Wani S., Hirota K., Kimura-Asami M., … Kuroyanagi H. (2021) m6A-mediated alternative splicing coupled with nonsense-mediated mRNA decay regulates SAM synthetase homeostasis. *The EMBO Journal* **40**, e106434.
Waterhouse P.M. & Helliwell C.A. (2003) Exploring plant genomes by RNA-induced gene silencing. *Nature Reviews Genetics* **4**, 29–38.
Weber E., Engler C., Gruetzner R., Werner S. & Marillonnet S. (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE* **6**, e16765.
Wei C., Gershowitz A. & Moss B. (1975) N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs. *Nature* **257**, 251–253.
Wei G., Almeida M., Pintacuda G., Coker H., Bowness J.S., Ule J. & Brockdorff N. (2021) Acute depletion of METTL3 implicates N6-methyladenosine in alternative intron/exon inclusion in the nascent transcriptome. *Genome Research* **31**, 1395–1408.
Wei L.H., Song P., Wang Y., Lu Z., Tang Q., Yu Q., … Jia G. (2018) The m6A reader ECT2 controls trichome morphology by affecting mRNA stability in arabidopsis. *Plant Cell* **30**, 968–985.
Weng Y.-L., Wang X., An R., Cassin J., Vissers C., Liu Y., … Ming G. (2018) Epitranscriptomic m6A Regulation of Axon Regeneration in the Adult Mammalian Nervous System. *Neuron* **97**, 313-325.e6.
Wielopolska A., Townley H., Moore I., Waterhouse P. & Helliwell C. (2005) A high-throughput inducible RNAi vector for plants. *Plant Biotechnology Journal* **3**, 583–590.
Xiao W., Adhikari S., Dahal U., Chen Y.S., Hao Y.J., Sun B.F., … Yang Y.G. (2016) Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Molecular Cell* **61**, 507–519.
Xu C., Wang X., Liu K., Roundtree I.A., Tempel W., Li Y., … Min J. (2014) Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nature chemical biology* **10**, 927–929.
Xu T., Wu X., Wong C.E., Fan S., Zhang Y., Zhang S., … Shen L. (2022) FIONA1-Mediated m6A Modification Regulates the Floral Transition in Arabidopsis. *Advanced Science* **9**, 2103628.
Yin R., Chang J., Li Y., Gao Z., Qiu Q., Wang Q., … Zhang H. (2022a) Differential m(6)A RNA

landscapes across hematopoiesis reveal a role for IGF2BP2 in preserving hematopoietic stem cell function. *Cell stem cell* **29**, 149-159.e7.
Yin S., Ao Q., Qiu T., Tan C., Tu Y., Kuang T. & Yang Y. (2022b) Tomato SlYTH1 encoding a putative RNA m6A reader affects plant growth and fruit shape. *Plant Science* **323**, 111417.
Yin S., Ao Q., Tan C. & Yang Y. (2021) Genome-wide identification and characterization of YTH domain-containing genes, encoding the m6A readers, and their expression in tomato. *Plant cell reports* **40**, 1229–1245.
You Q., Wang F., Du R., Pi J., Wang H., Huo Y., … Zhu L. (2023) m6A Reader YTHDF1-Targeting Engineered Small Extracellular Vesicles for Gastric Cancer Therapy via Epigenetic and Immune Regulation. *Advanced Materials* **35**, 2204910.
Yue Y., Liu J., Cui X., Cao J., Luo G., Zhang Z., … Liu J. (2018) VIRMA mediates preferential m6A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. *Cell Discovery* **4**, 10.
Zaccara S., Ries R.J. & Jaffrey S.R. (2019) Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology* **20**, 608–624.
Zhang J., Addepalli B., Yun K.-Y., Hunt A.G., Xu R., Rao S., … Falcone D.L. (2008) A polyadenylation factor subunit implicated in regulating oxidative signaling in Arabidopsis thaliana. *PloS one* **3**, e2410.
Zhang M., Bodi Z., Mackinnon K., Zhong S., Archer N., Mongan N.P., … Fray R.G. (2022) Two zinc finger proteins with functions in m6A writing interact with HAKAI. *Nature Communications* **13**, 1–15.
Zhao T., Zhan Z. & Jiang D. (2019) Histone modifications and their regulatory roles in plant development and environmental memory. *Journal of Genetics and Genomics* **46**, 467–476.
Zhong S., Li H., Bodi Z., Button J., Vespa L., Herzog M. & Fray R.G. (2008) MTA Is an Arabidopsis Messenger RNA Adenosine Methylase and Interacts with a Homolog of a Sex-Specific Splicing Factor. *The Plant Cell Online* **20**, 1278–1288.
Zhou H., Liu Y., Liang Y., Zhou D., Li S., Lin S., … Huang L. (2020) The function of histone lysine methylation related SET domain group proteins in plants. *Protein Science* **29**, 1120–1137.
Zhu T., Roundtree I.A., Wang P., Wang X., Wang L., Sun C., … Xu Y. (2014) Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell research* **24**, 1493–1496.

9 Appendix

Table A9.1 BSA analysis identified genes from P10-23 using filtering step 2.

Gene ID	Gene Description	Consequence	AD WT	AD VAR	VAR:WT
AT1G04130	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins.	Downstream gene variant	3	2	0.67
AT1G04140	Transducin family protein / WD-40 repeat family protein	Intron variant	3	2	0.67
AT1G04150	C2 calcium/lipid-binding plant phosphoribosyltransferase family protein	Downstream gene variant	3	2	0.67
AT1G05973	Natural antisense transcript overlaps with AT1G26090	Upstream gene variant	3	2	0.67
AT1G05977	Natural antisense transcript overlaps with AT1G26100	Downstream gene variant	3	2	0.67
AT1G08910	Encodes an SP-RING domain containing protein that functions in sumolaytion and is involved in positive regulation of sulfur metabolism and stress response.	Downstream gene variant	3	2	0.67
AT1G08920	Encodes ESL1, a transporter for monosaccharides.	Upstream gene variant	3	2	0.67
AT1G09987	Natural antisense transcript overlaps with AT1G80850	Downstream gene variant	4	3	0.75
AT1G11450	nodulin MtN21-like transporter family protein	Downstream gene variant	3	2	0.67
AT1G11460	nodulin MtN21-like transporter family protein	3 prime UTR variant	3	2	0.67
AT1G11470	Pentatricopeptide repeat (PPR) superfamily protein	Downstream gene variant	3	2	0.67
AT1G11475	Non-catalytic subunit common to nuclear DNA-dependent RNA polymerases II, IV and V; homologous to budding yeast RPB10.	Upstream gene variant	3	2	0.67
AT1G11480	eukaryotic translation initiation factor-like protein	Downstream gene variant	3	2	0.67
AT1G12430	Encodes the kinesin-like protein PAK has an Armadillo motif tail and is involved in guard cell development in Arabidopsis (from Genbank record AF159052).However, no defect in stomatal complexes has been observed in loss of function mutations.	Missense variant	3	2	0.67
AT1G12440	A20/AN1-like zinc finger family protein	Downstream gene variant	3	2	0.67
AT1G26090	P-loop containing nucleoside triphosphate hydrolases superfamily protein	Downstream gene variant	3	2	0.67
AT1G26100	Cytochrome b561/ferric reductase transmembrane protein family	Intron variant	3	2	0.67
AT1G26110	Encodes Decapping 5, required for mRNA decapping, P-body formation and translational repression during postembryonic development.	Downstream gene variant	3	2	0.67
AT1G26120	alpha/beta-Hydrolases superfamily protein	Downstream gene variant	3	2	0.67
AT1G27180	disease resistance protein (TIR-NBS-LRR class)	Downstream gene variant	3	2	0.67
AT1G27190	Activated by TCP8/14/15/22, involved in modulation of GA-dependent stamen filament elongation.	Downstream gene variant	3	2	0.67
AT1G27200	glycosyltransferase family protein (DUF23)	Downstream gene variant	3	2	0.67
AT1G27960	encodes a protein whose C-terminal region is similar to ECT1 and ECT2, which bind to CIPK1.	Upstream gene variant	4	3	0.75
AT1G27970	Encodes an ortholog of yeast NTF2, a nuclear envelop transport protein that functions as the nuclear import receptor for RanGDP, an essential player in nucleocytoplasmic transport. The mRNA is cell-to-cell mobile.	Downstream gene variant	4	3	0.75
AT1G27980	Encodes an ER-localized sphingoid long-chain base-1-phosphate lyase involved in the dehydration stress response.	Intron variant	4	3	0.75
AT1G27990	transmembrane protein	Downstream gene variant	4	3	0.75

AT1G29090	Cysteine proteinases superfamily protein	Upstream gene variant	3	2	0.67
AT1G29100	Heavy metal transport/detoxification superfamily protein	Upstream gene variant	3	2	0.67
AT1G29110	Cysteine proteinases superfamily protein	Upstream gene variant	3	2	0.67
AT1G32415	Encodes a PPR protein involved in mitochondrial functioning. Mutants suppress cell wall defects caused by C17 chemical inhibitor. Mutants are defective in cytochrome c maturation and activation of mitochondrial retrograde signalling.	Downstream gene variant	3	2	0.67
AT1G32420	F-box and associated interaction domains-containing protein	Upstream gene variant	3	2	0.67
AT1G32430	F-box and associated interaction domains-containing protein	Downstream gene variant	3	2	0.67
AT1G33710	RNA-directed DNA polymerase (reverse transcriptase)-related family protein	Downstream gene variant	3	2	0.67
AT1G33720	cytochrome P450, family 76, subfamily C, polypeptide 6	3 prime UTR variant	3	2	0.67
AT1G33770	Protein kinase superfamily protein	Downstream gene variant	3	2	0.67
AT1G33780	Interacts with chloroplast chaperone CLPC1, which unfolds and delivers substrates to the stromal CLPPRT protease complex for degradation.	Upstream gene variant	3	2	0.67
AT1G40104	hypothetical protein	Missense variant	58	40	0.69
AT1G40390	DNAse I-like superfamily protein	Upstream gene variant	18	17	0.94
AT1G43760	DNAse I-like superfamily protein	Downstream gene variant	12	9	0.75
AT1G48110	evolutionarily conserved C-terminal region 7	Upstream gene variant	3	2	0.67
AT1G48120	Encodes a nuclear localized aminotransferase-like protein containing a plant mobile domain.	Missense variant	3	2	0.67
AT1G48130	encodes a protein similar to the 1-cysteine (1-Cys) peroxiredoxin family of antioxidants. Expression is limited to seed (aleurone and embryo) and is not induced by ABA or drought.	Upstream gene variant	3	2	0.67
AT1G48140	Encodes a subunit of the dolichol phosphate mannase synthase (DPMS) complex that may serve as membrane anchors for the catalytic core, DPMS1, or provide catalytic assistance. It is localized in the ER and mediates isoprenyl-linked glycan biogenesis.	Upstream gene variant	3	2	0.67
AT1G54260	winged-helix DNA-binding transcription factor family protein	Downstream gene variant	3	2	0.67
AT1G54270	member of eIF4A - eukaryotic initiation factor 4A	Downstream gene variant	3	2	0.67
AT1G54280	Encodes a member of the P4 subfamily of P-type ATPases expressed in the pollen plasma membrane. Double mutants with ALA7 display pollen and pollen tube defects.	Intron variant	3	2	0.67
AT1G54290	Translation initiation factor SUI1 family protein	Downstream gene variant	3	2	0.67
AT1G54490	Involved in the ethylene response. XRN4 does not appear to regulate ethylene signaling via an RNA-INDUCED SILENCING COMPLEX-based RNA silencing mechanism but acts by independent means. Endogenous suppressor of posttranscriptional gene silencing.	3 prime UTR variant	3	2	0.67
AT1G54500	RBD1 is a thylakoid membrane-bound iron-binding protein that is required for the proper assembly of photosystem II in Arabidopsis. It is found in all oxygenic photoautotrophic organisms (plants, algae and cyanobacteria).	Downstream gene variant	3	2	0.67
AT1G54510	Encodes AtNEK1, a member of the NIMA-related serine/threonine kinases (Neks) that have been linked to cell-cycle regulation in fungi and mammals. Plant Neks might be involved in plant development processes.	Downstream gene variant	3	2	0.67
AT1G54870	ChlADR is an aldehyde reductase that catalyzes the reduction of the aldehyde carbonyl groups on saturated and alpha,beta-unsaturated aldehydes with more than 5 carbons in vitro	Downstream gene variant	3	2	0.67
AT1G54890	Late embryogenesis abundant (LEA) protein-like protein	Downstream gene variant	3	2	0.67
AT1G54953	transmembrane protein	Downstream gene variant	3	2	0.67

AT1G54957	transmembrane protein	Downstream gene variant	3	2	0.67
AT1G54960	member of MEKK subfamily	Upstream gene variant	3	2	0.67
AT1G61190	Putative CC-NB-LRR resistance gene. It's ortholog in ecotype Wei-0 RPP39 (R_190-Wei-0, accession number JQ045573, PMID:22359513) has been shown experimentally to recognize the effector ATR39-1 produced by Hyaloperonospora arabidopsidis (Hpa).	Downstream gene variant	6	5	0.83
AT1G61200	homeobox-leucine zipper protein-like protein	Downstream gene variant	6	5	0.83
AT1G61210	DWA3 encodes a DWD(DDB1 binding WD40) protein. Invitro analyses suggest its involvement in the negative regulation of ABA responses.One of four katanin p80 subunits. Involved in targeting of katanin complex to crossover and branch points to properly sever microtubules.	Upstream gene variant	6	5	0.83
AT1G61930	senescence regulator (Protein of unknown function, DUF584)	Downstream gene variant	3	2	0.67
AT1G61940	Member of TLP family	Downstream gene variant	3	2	0.67
AT1G61950	Member of Calcium Dependent Protein Kinase, regulated by Bacillus amyloliquefaciens under high calcium stress.	Upstream gene variant	3	2	0.67
AT1G62300	Encodes a transcription factor WRKY6. Regulates Phosphate1 (Pho1) expression in response to low phosphate (Pi) stress.	Intron variant	4	3	0.75
AT1G62810	Encodes COPPER AMINE OXIDASE1 (CuAO1). Contributes to abscisic acid- and polyamine-induced nitric oxide biosynthesis and abscisic acid signal transduction.	Upstream gene variant	3	2	0.67
AT1G62820	Calcium-binding EF-hand family protein	3 prime UTR variant	3	2	0.67
AT1G62830	Encodes a homolog of human Lysine-Specific Demethylase1. Involved in H3K4 methylation of target genes including the flowering time loci FLC and FWA. Located in nucleus. Negatively regulates root elongation. Involved in repression of LRP1 via histone deacetylation.	Downstream gene variant	3	2	0.67
AT1G63540	hydroxyproline-rich glycoprotein family protein	Downstream gene variant	4	3	0.75
AT1G63550	Receptor-like protein kinase-related family protein	Downstream gene variant	4	3	0.75
AT1G63570	Receptor-like protein kinase-related family protein	Upstream gene variant	4	3	0.75
AT1G63580	Encodes a plasma membrane-localized protein with two DUF26 domains and a GPI anchor domain.	Upstream gene variant	4	3	0.75
AT1G66050	Encodes a protein that is similar to VIM1 but is not involved in cytosine methylation. This protein has an N-terminal PHD domain and two RING domains surrounding an SRA domain. ORTH5/VIM2 has E3 ubiquitin ligase activity in vitro.	Downstream gene variant	3	2	0.67
AT1G66060	hypothetical protein (DUF577)	3 prime UTR variant	3	2	0.67
AT1G66070	Translation initiation factor eIF3 subunit	Downstream gene variant	3	2	0.67
AT1G67660	Restriction endonuclease, type II-like superfamily protein	Downstream gene variant	3	2	0.67
AT1G67670	hypothetical protein	Downstream gene variant	3	2	0.67
AT1G67680	SRP72 RNA-binding domain-containing protein	5 prime UTR variant	3	2	0.67
AT1G67690	Zincin-like metalloproteases family protein	Upstream gene variant	3	2	0.67
AT1G67970	member of Heat Stress Transcription Factor (Hsf) family	Upstream gene variant	4	3	0.75
AT1G67980	Encodes S-adenosyl-L-methionine: transcaffeoyl Coenzyme A 3-O-methyltransferase. Methyltransferase in the lignin biosynthetic pathway.	Downstream gene variant	4	3	0.75
AT1G67990	Encodes a tapetum-specific O-methyltransferase. In vitro enzyme assay indicated activity with caffeoyl-CoA, caffeoyl glucose, chlorogenic acid and polyamine conjugates. RNAi mutants had impaired silique development and seed setting.	Downstream gene variant	4	3	0.75
AT1G68000	phosphatidylinositol synthase 1	Intron variant	4	3	0.75
AT1G68010	Encodes hydroxypyruvate reductase.	Upstream gene variant	4	3	0.75

AT1G70505	transmembrane protein	Downstream gene variant	4	3	0.75
AT1G70510	A member of class I knotted1-like homeobox gene family (together with KNAT1). Similar to the knotted1 (kn1) homeobox gene of maize.	Upstream gene variant	4	3	0.75
AT1G73670	member of MAP Kinase The mRNA is cell-to-cell mobile.	Downstream gene variant	3	2	0.67
AT1G73680	Encodes an alpha dioxygenase. Recombinant protein catalyzes the conversion of a wide range of fatty acids into 2(R)-hydroperoxy derivatives.	Missense variant	3	2	0.67
AT1G78950	Terpenoid cyclases family protein	Upstream gene variant	3	2	0.67
AT1G78955	Encodes a cyclase that generates predominantly a monocyclic triterpene alcohol. The product is 97% camelliol, 2% achilleol A and 0.2% beta-amyrin. Achilleol is an isomer of camelliol C with a 4-methylenecyclohexanol ring system.	Synonymous variant	3	2	0.67
AT1G80850	DNA glycosylase superfamily protein	Upstream gene variant	4	3	0.75
AT1G80860	Encodes a single-copy phospholipid N-methyltransferase, involved in phosphatidylcholine biosynthesis. Has specific activity towards phosphatidylmonomethylethanolamine and phosphatidyldimethylethanolamine, but not phosphatidylethanolamine.	Upstream gene variant	4	3	0.75
AT1G80865	hypothetical protein	Downstream gene variant	4	3	0.75
AT1G80870	Protein kinase superfamily protein	Upstream gene variant	4	3	0.75
AT1G80880	Tetratricopeptide repeat (TPR)-like superfamily protein	Downstream gene variant	4	3	0.75
AT2G07565	Ycf1 domain protein	Upstream gene variant	5	4	0.80
AT2G07981	hypothetical protein	Upstream gene variant	3	2	0.67
AT2G08035	microRNA ath-MIR5019 precursor	Downstream gene variant	4	3	0.75
AT2G08845	Unknown	Upstream gene variant	3	2	0.67
AT2G08986	hypothetical protein	Missense variant	54	48	0.89
AT2G10625	Ulp1 protease family, carboxy-terminal catalytic domain protein	Downstream gene variant	3	2	0.67
AT2G11520	high overall homology to CRCK1	Upstream gene variant	3	2	0.67
AT2G19270	mitotic checkpoint protein PRCC-carboxy-term protein	Upstream gene variant	3	2	0.67
AT2G19280	Pentatricopeptide repeat (PPR) superfamily protein	3 prime UTR variant	3	2	0.67
AT2G19290	hypothetical protein	Downstream gene variant	3	2	0.67
AT2G19300	hypothetical protein	Downstream gene variant	3	2	0.67
AT2G19310	HSP20-like chaperones superfamily protein	Downstream gene variant	3	2	0.67
AT2G21070	This gene is predicted to an encode a nuclear-localized protein that is involved in regulating the period of circadian rhythms without affecting their amplitude or robustness.	Downstream gene variant	3	2	0.67
AT2G21080	Ras guanine nucleotide exchange factor K	Missense variant	3	2	0.67
AT2G21090	Pentatricopeptide repeat (PPR-like) superfamily protein	Downstream gene variant	3	2	0.67
AT2G21100	Disease resistance-responsive (dirigent-like protein) family protein	Downstream gene variant	3	2	0.67
AT2G21105	hypothetical protein	Upstream gene variant	3	2	0.67
AT2G24500	Encodes a C2H2 zinc finger protein FZF. Involved in cold regulated rRNA processing and the accumulation of CBF protein. Cytosolic ribosomal 60S-biogenesis factor.	Upstream gene variant	4	3	0.75
AT2G24510	F-box associated ubiquitination effector family protein	Downstream gene variant	4	3	0.75
AT2G25660	Translocon at the inner-envelope membrane of chloroplasts which binds to the outer-membrane channel TOC75.	Upstream gene variant	3	2	0.67

AT2G25670	hypothetical protein	Upstream gene variant	3	2	0.67
AT2G25680	Encodes a high-affinity molybdate transporter. Mutant has reduced concentrations of molybdate in roots and shoots, and reduced shoot and root length when growing on Mo-limited medium.	Downstream gene variant	3	2	0.67
AT2G27770	DUF868 family protein (DUF868)	Upstream gene variant	3	2	0.67
AT2G30530	zinc finger CCCH domain protein	Downstream gene variant	3	2	0.67
AT2G30540	Encodes a member of the CC-type glutaredoxin (ROXY) family that has been shown to interact with the transcription factor TGA2.	5 prime UTR variant	3	2	0.67
AT2G30550	Encodes a lipase that hydrolyzes phosphatidylcholine, glycolipids as well as triacylglycerols.	Upstream gene variant	3	2	0.67
AT2G32000	DNA topoisomerase, type IA, core	Upstream gene variant	3	2	0.67
AT2G32010	Encodes an inositol polyphosphate 5?-phosphatase (5PTase). Mediating phosphoinositide signaling. Involved in establishment of foliar vein patterns.	Intron variant	3	2	0.67
AT2G32020	Acyl-CoA N-acyltransferases (NAT) superfamily protein	Downstream gene variant	3	2	0.67
AT2G33010	Ubiquitin-associated (UBA) protein	Downstream gene variant	3	2	0.67
AT2G33020	receptor like protein 24	Synonymous variant	3	2	0.67
AT2G33030	receptor like protein 25	Downstream gene variant	3	2	0.67
AT2G33040	gamma subunit of Mt ATP synthase	Downstream gene variant	3	2	0.67
AT2G36960	Arabidopsis thaliana myb/SANT domain protein	Frameshift variant	3	2	0.67
AT2G36970	UDP-Glycosyltransferase superfamily protein	Upstream gene variant	3	2	0.67
AT2G40320	Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain.	Downstream gene variant	3	2	0.67
AT2G40330	Encodes a member of the PYR (pyrabactin resistance)/PYL(PYR1-like)/RCAR (regulatory components of ABA receptor) family proteins with 14 members. PYR/PYL/RCAR family proteins function as abscisic acid sensors.	3 prime UTR variant	3	2	0.67
AT2G40340	Encodes a member of the DREB subfamily A-2 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are eight members in this subfamily including DREB2A AND DREB2B that are involved in response to drought.	Downstream gene variant	3	2	0.67
AT2G41230	Encodes an ER-localized plant hormone-responsive gene and appears to act redundantly with ARGOS and ARL during organ growth. Over-expression modifies plant sensitivity to ethylene, leading to improved drought tolerance.	Upstream gene variant	3	2	0.67
AT2G41231	transmembrane protein	Downstream gene variant	3	2	0.67
AT2G41240	Encodes a member of the basic helix-loop-helix transcription factor family protein. Functions as a key regulator of iron-deficiency responses independent of the master regulator FIT. Likely regulates genes involved in the distribution of iron within the plant.	Upstream gene variant	3	2	0.67
AT2G41250	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein	Downstream gene variant	3	2	0.67
AT2G41900	AtOXS2 specifcally entered the nuclear under salt stress. Te specifc nuclear localization of AtOXS2 could play a role in salt tolerance at the molecular level.	3 prime UTR variant	3	2	0.67
AT2G41905	transmembrane protein	Upstream gene variant	3	2	0.67
AT2G41910	Protein kinase superfamily protein	Upstream gene variant	3	2	0.67
AT3G04195	Natural antisense transcript overlaps with AT3G21820	Downstream gene variant	3	2	0.67
AT3G08220	Natural antisense transcript overlaps with AT3G52030	Downstream gene variant	3	2	0.67
AT3G08225	Natural antisense transcript overlaps with AT3G52040	Upstream gene variant	3	2	0.67

AT3G15680	Ran BP2/NZF zinc finger-like superfamily protein	Downstream gene variant	3	2	0.67
AT3G15690	Single hybrid motif superfamily protein	Frameshift variant	3	2	0.67
AT3G15700	P-loop containing nucleoside triphosphate hydrolases superfamily protein	Downstream gene variant	3	2	0.67
AT3G15710	Peptidase S24/S26A/S26B/S26C family protein	Downstream gene variant	3	2	0.67
AT3G17000	Group XIV ubiquitin-conjugating enzyme that functions negative regulation of drought stress.	Downstream gene variant	3	2	0.67
AT3G17010	transcriptional factor B3 family protein, contains Pfam profile PF02362: B3 DNA binding domain. Activated by AGAMOUS ina a cal-1, ap1-1 background. Expressed in stamen primordia, the placental region of developing carpels and the ovary.	Downstream gene variant	3	2	0.67
AT3G17020	Adenine nucleotide alpha hydrolases-like superfamily protein	Upstream gene variant	3	2	0.67
AT3G17030	Nucleic acid-binding proteins superfamily	3 prime UTR variant	3	2	0.67
AT3G17040	It is a RNA tetratricopeptide repeat-containing protein required for normal processing of transcripts from the polycistronic chloroplast psbB-psbT-psbH-petB-petD operon coding for proteins of the photosystem II and cytochrome b6/f complexes.	Downstream gene variant	3	2	0.67
AT3G17570	F-box and associated interaction domains-containing protein	Downstream gene variant	3	2	0.67
AT3G17580	SsrA-binding protein	Upstream gene variant	3	2	0.67
AT3G17590	Encodes the Arabidopsis homologue of yeast SNF5 and represents a conserved subunit of plant SWI/SNF complexes.Component of MINU-associated SWI/SNF complex, aka MAS complex (PMID: 36189880, PMID:36471048)	Downstream gene variant	3	2	0.67
AT3G17600	Encodes a member of the Aux/IAA family of proteins implicated in auxin signaling. IAA31 shares several residues with the conserved domain II region, believed to act as a degron in many of the rapidly degraded Aux/IAA family members.	Downstream gene variant	3	2	0.67
AT3G20840	Encodes a member of the AINTEGUMENTA-like (AIL) subclass of the AP2/EREBP family of transcription factors and is essential for quiescent center (QC) specification and stem cell activity.	Intron variant	3	2	0.67
AT3G20850	proline-rich family protein	Upstream gene variant	3	2	0.67
AT3G20860	Encodes a member of the NIMA-related serine/threonine kinases (Neks) that have been linked to cell-cycle regulation in fungi and mammals. Plant Neks might be involved in plant development processes.	Upstream gene variant	3	2	0.67
AT3G21810	Zinc finger C-x8-C-x5-C-x3-H type family protein	Downstream gene variant	3	2	0.67
AT3G21820	Histone lysine methyltransferase that promotes the accumulation of histone H3 proteins that are trimethylated on lysine 36 (H3K36me3) during callus formation. It is recruited to LBD promoters via interaction with ARFs.	Upstream gene variant	3	2	0.67
AT3G21825	hypothetical protein	Upstream gene variant	3	2	0.67
AT3G21830	SKP1-like 8	Downstream gene variant	3	2	0.67
AT3G21840	SKP1-like 7	Downstream gene variant	3	2	0.67
AT3G21850	one of Arabidopsis SKP1 homologues	Downstream gene variant	3	2	0.67
AT3G25410	Sodium Bile acid symporter family	Upstream gene variant	4	3	0.75
AT3G25420	serine carboxypeptidase-like 21	Downstream gene variant	4	3	0.75
AT3G25430	Polynucleotidyl transferase, ribonuclease H-like superfamily protein	Intron variant	4	3	0.75
AT3G25440	RNA-binding CRS1 / YhbY (CRM) domain protein	Upstream gene variant	4	3	0.75
AT3G25710	Encodes a basic helix-loop-helix transcription factor that is expressed in the hypophysis-adjacent embryo cells, and is	Downstream gene variant	3	2	0.67

	required and partially sufficient for MP-dependent root initiation. content.				
AT3G25715	pre-tRNA tRNA-Ile (anticodon: AAT	Upstream gene variant	3	2	0.67
AT3G25716	transmembrane protein	Frameshift variant	3	2	0.67
AT3G25717	ROTUNDIFOLIA like 16	Downstream gene variant	3	2	0.67
AT3G28810	mediator of RNA polymerase II transcription subunit-like protein, putative (DUF1216)	Intron variant	6	5	0.83
AT3G28820	mediator of RNA polymerase II transcription subunit-like protein, putative (DUF1216)	Upstream gene variant	6	5	0.83
AT3G31402	General transcription factor 2-related zinc finger protein	Downstream gene variant	3	2	0.67
AT3G43148	myosin heavy chain-like protein	Missense variant	3	2	0.67
AT3G43150	hypothetical protein	Upstream gene variant	3	2	0.67
AT3G44080	F-box family protein	Frameshift variant	4	3	0.75
AT3G44090	F-box family protein	Upstream gene variant	4	3	0.75
AT3G45920	Protein kinase superfamily protein	Downstream gene variant	6	4	0.67
AT3G45930	Histone superfamily protein	Downstream gene variant	6	4	0.67
AT3G45935	pre-tRNA tRNA-Val (anticodon: AAC);(source:Araport11, TAIR10)	Upstream gene variant	6	4	0.67
AT3G45940	Glycosyl hydrolases family 31 protein	Downstream gene variant	6	4	0.67
AT3G45950	Pre-mRNA splicing Prp18-interacting factor	Downstream gene variant	6	4	0.67
AT3G46420	Leucine-rich repeat protein kinase family protein	Intron variant	3	2	0.67
AT3G46430	Mitochondrial F1F0-ATP synthase.	Upstream gene variant	3	2	0.67
AT3G46440	Encodes a cytosolic isoform of UDP-glucuronic acid decarboxylase. UDP-glucuronic acid decarboxylase produces UDP-xylose, which is a substrate for many cell wall carbohydrates including hemicellulose and pectin.	Downstream gene variant	3	2	0.67
AT3G47040	Glycosyl hydrolase family protein	Upstream gene variant	3	2	0.67
AT3G47050	Glycosyl hydrolase family protein	Intron variant	3	2	0.67
AT3G47060	encodes an FtsH protease that is localized to the chloroplast; Plays role in abscisic acid (ABA)-dependent pathway in response to dehydration.	Upstream gene variant	3	2	0.67
AT3G48240	Octicosapeptide/Phox/Bem1p family protein	Upstream gene variant	3	2	0.67
AT3G48250	Encodes a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial nad7 transcripts.	Downstream gene variant	3	2	0.67
AT3G50770	calmodulin-like 41	Downstream gene variant	3	2	0.67
AT3G50780	BTB/POZ domain protein	5 prime UTR variant	3	2	0.67
AT3G50790	esterase/lipase/thioesterase family protein	Downstream gene variant	3	2	0.67
AT3G50800	PADRE protein.	Downstream gene variant	3	2	0.67
AT3G52010	serine carboxypeptidase-like 37	Downstream gene variant	3	2	0.67
AT3G52020	serine carboxypeptidase-like 39	Missense variant	3	2	0.67
AT3G52030	F-box family protein with WD40/YVTN repeat doamin	Upstream gene variant	3	2	0.67
AT3G52040	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	Downstream gene variant	3	2	0.67
AT3G52050	5-3 exonuclease family protein	Upstream gene variant	3	2	0.67

AT3G53820	C2H2 and C2HC zinc fingers superfamily protein	Downstream gene variant	4	3	0.75
AT3G53830	Regulator of chromosome condensation (RCC1) family protein	3 prime UTR variant	4	3	0.75
AT3G53840	Protein kinase superfamily protein	Upstream gene variant	4	3	0.75
AT3G53850	Uncharacterized protein family (UPF0497)	Downstream gene variant	4	3	0.75
AT3G55130	Encodes a vacuole localized protein of the ABC transporter White-Brown Complex (WBC) family. When overexpressed in planta, confers resistance to kanamycin.	Upstream gene variant	3	2	0.67
AT3G55140	Pectin lyase-like superfamily protein	Synonymous variant	3	2	0.67
AT3G55150	A member of EXO70 gene family, putative exocyst subunits, conserved in land plants. Arabidopsis thaliana contains 23 putative EXO70 genes, which can be classified into eight clusters on the phylogenetic tree.	Downstream gene variant	3	2	0.67
AT3G55160	THADA is an orphan gene in Arabidopsis thaliana. It is the only DUF2428 domain containing protein in the genome.	Downstream gene variant	3	2	0.67
AT3G58070	Putative transcription factor, contains C2H2 domain, regulates aspects of shoot maturation in Arabidopsis thaliana.	Upstream gene variant	3	2	0.67
AT3G58090	Disease resistance-responsive (dirigent-like protein) family protein	Downstream gene variant	3	2	0.67
AT3G58890	RNI-like superfamily protein	Downstream gene variant	3	2	0.67
AT3G58900	F-box/RNI-like superfamily protein	Downstream gene variant	3	2	0.67
AT3G58910	F-box family protein	Upstream gene variant	3	2	0.67
AT3G58920	F-box/RNI-like superfamily protein	Missense variant	3	2	0.67
AT3G58930	F-box/RNI-like superfamily protein	Downstream gene variant	3	2	0.67
AT3G58940	F-box/RNI-like superfamily protein	Downstream gene variant	3	2	0.67
AT3G61080	Protein kinase superfamily protein	Downstream gene variant	3	2	0.67
AT3G61090	Putative endonuclease or glycosyl hydrolase	Upstream gene variant	3	2	0.67
AT3G61100	Putative endonuclease or glycosyl hydrolase	Upstream gene variant	3	2	0.67
AT3G61105	pre-tRNA tRNA-Leu (anticodon: AAG);(source:Araport11, TAIR10)	Downstream gene variant	3	2	0.67
AT3G61110	Arabidopsis ribosomal protein	Intron variant	3	2	0.67
AT3G61111	Zinc-binding ribosomal protein family protein	Downstream gene variant	3	2	0.67
AT3G61113	Ubiquitin related modifier 1	Downstream gene variant	3	2	0.67
AT3G61117	This gene encodes a small protein and has either evidence of transcription or purifying selection.	Downstream gene variant	3	2	0.67
AT3G61118	other_RNA	Downstream gene variant	3	2	0.67
AT4G06210	Natural antisense transcript overlaps with AT4G14700	Downstream gene variant	3	2	0.67
AT4G06598	basic-leucine zipper (bZIP) transcription factor family protein	Upstream gene variant	4	3	0.75
AT4G06599	ubiquitin family protein	Downstream gene variant	4	3	0.75
AT4G07740	hypothetical protein (DUF3287)	Downstream gene variant	4	3	0.75
AT4G08790	NIT1 amidase involved in the breakdown of deaminated glutathione (dGSH). It is active towards dGSH and dOA with a preference for dGSH.	Upstream gene variant	3	2	0.67
AT4G08800	Protein kinase superfamily protein	Upstream gene variant	3	2	0.67
AT4G08810	Calcium binding protein involved in cryptochrome and phytochrome coaction	Downstream gene variant	3	2	0.67

AT4G12770	Negative regulation of growth and endocytosis, most likely as a result of inhibition of the recruitment of clathrin to endocytic pits.	Upstream gene variant	7	5	0.71
AT4G12780	Negative regulation of growth and endocytosis, most likely as a result of inhibition of the recruitment of clathrin to endocytic pits. Overexpression inhbits recruitment of clathrin resulting in negative regulation of endocytosis and developmental arrest.	Upstream gene variant	7	5	0.71
AT4G13400	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	Upstream gene variant	3	2	0.67
AT4G13410	encodes a gene similar to cellulose synthase	Missense variant	3	2	0.67
AT4G13420	Encodes a protein of the KUP/HAK/KT potassium channel class that is upregulated in the roots by K levels.	Downstream gene variant	3	2	0.67
AT4G14680	Encodes one of three A. thaliana ATP-sulfurylases. APS is the first enzyme of sulfate assimilation that catalyzes the formation of adenosine-5'-phosphosulfate from ATP and sulfate.	Upstream gene variant	3	2	0.67
AT4G14690	Encodes an early light-induced protein. ELIPs are thought not to be directly involved in the synthesis and assembly of specific photosynthetic complexes, but rather affect the biogenesis of all chlorophyll-binding complexes.	Downstream gene variant	3	2	0.67
AT4G14695	Uncharacterized protein family (UPF0041)	Intron variant	3	2	0.67
AT4G14700	Encodes origin of replication complex 1a subunit.The protein contains a PHD domain,binds methylated DNA and appears to function as a transcriptional activator.	Upstream gene variant	3	2	0.67
AT4G14710	RmlC-like cupins superfamily protein	Downstream gene variant	3	2	0.67
AT4G16280	Involved in the promotion of the transition of the vegetative meristem to reproductive development. Four forms of the protein (alpha, beta, delta and gamma) are produced by alternative splicing. Involved in RNA-mediated chromatin silencing.	Upstream gene variant	3	2	0.67
AT4G16295	self-incompatibility (S) protein homolog, expressed at very low levels in floral buds	Upstream gene variant	3	2	0.67
AT4G16310	FAD-dependent lysine-specific histone demethylase involved in the control of flowering time.	5 prime UTR variant	3	2	0.67
AT4G18990	xyloglucan endotransglucosylase/hydrolase 29	Upstream gene variant	3	2	0.67
AT4G19000	The C-terminal portion of this protein has homology to the C-termini of the IWS1 (Interacts With Spt6) proteins found in yeast and humans.	Downstream gene variant	3	2	0.67
AT4G19003	E2F/DP family winged-helix DNA-binding domain-containing protein	Intron variant	3	2	0.67
AT4G19006	Proteasome component (PCI) domain protein	Downstream gene variant	3	2	0.67
AT4G19010	Encodes for a 4-coumarate-CoA ligase involved in the biosynthesis of the benzenoid ring of ubiquiUnknown from phenylalanine.	Downstream gene variant	3	2	0.67
AT4G19770	Glycosyl hydrolase family protein with chitinase insertion domain-containing protein	Upstream gene variant	3	2	0.67
AT4G19800	Glycosyl hydrolase family protein with chitinase insertion domain-containing protein	Downstream gene variant	3	2	0.67
AT4G24080	aldolase like	Downstream gene variant	3	2	0.67
AT4G24090	Membrane protein located in chloroplast.	Downstream gene variant	3	2	0.67
AT4G24100	Protein kinase superfamily protein	Intron variant	3	2	0.67
AT4G24110	NADP-specific glutamate dehydrogenase	Downstream gene variant	3	2	0.67
AT4G25310	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	Downstream gene variant	3	2	0.67
AT4G25315	Expressed protein	Intron variant	3	2	0.67
AT4G25320	AT hook motif DNA-binding family protein	Upstream gene variant	3	2	0.67
AT4G31080	Encodes one of two LUNAPARK proteins in Arabidopsis. Both LNPA and LNPB are predominantly distributed throughout the ER, but not preferentially localized at the three-way junctions.	Downstream gene variant	3	2	0.67

AT4G31100	wall-associated kinase	Downstream gene variant	3	2	0.67
AT4G31110	Wall-associated kinase family protein	5 prime UTR variant	3	2	0.67
AT4G31115	DUF1997 family protein, putative (DUF1997)	Upstream gene variant	3	2	0.67
AT4G31120	Involved in vernalization. Required for epigenetic silencing of FLC, and for vernalization-mediated histone modification.	Downstream gene variant	3	2	0.67
AT4G38590	putative beta-galactosidase (BGAL14 gene)	Downstream gene variant	3	2	0.67
AT4G38600	encodes a member of HECT ubiquitin protein ligase family that is involved in trichome cell morphogenesis. Mutants in this gene exhibit supernumerary trichome branches and increased DNA content.	3 prime UTR variant	3	2	0.67
AT4G39470	Tetratricopeptide repeat (TPR)-like superfamily protein	Upstream gene variant	6	5	0.83
AT4G39480	member of CYP96A	Downstream gene variant	6	5	0.83
AT4G39490	member of CYP96A	Synonymous variant	6	5	0.83
AT4G39500	cytochrome P450, family 96, subfamily A, polypeptide 11	Downstream gene variant	6	5	0.83
AT4G39510	member of CYP96A	Downstream gene variant	6	5	0.83
AT4G39520	Encodes a member of the DRG (developmentally regulated G-protein) family. Has GTPase activity.	Downstream gene variant	6	5	0.83
AT5G03760	encodes a beta-mannan synthase that is required for agrobacterium-mediated plant genetic transformation involves a complex interaction between the bacterium and the host plant. 3' UTR is involved in transcriptional regulation and the gene is expressed in the elongation zone of the root.	5 prime UTR variant	3	2	0.67
AT5G03770	Encodes a putative KDO (3-deoxy-D-manno-octulosonate) transferase	Upstream gene variant	3	2	0.67
AT5G06600	Encodes a ubiquitin-specific protease which together with UBP13 deubiquitinates DA1, DAR1 and DAR2, hence reducing their peptidase activity. Works upstream of DA1, DAR1 and DAR2 to restrict their protease activity and hence fine-tune plant growth and development.	Intron variant	3	2	0.67
AT5G06610	DUF620 domain protein. In xylem cells BRD1 is expressed in the secondary wall pit boundaries. BRD1 interacts with and appears to be necessary to recruit WALLIN to the plasma membrane.	Upstream gene variant	3	2	0.67
AT5G07060	Encodes MAC5C, homologous to MAC5A. MAC5A is a component of the MOS4-associated complex (MAC) that contributes to snc1- mediated autoimmunity.	Upstream gene variant	4	3	0.75
AT5G07070	Encodes CBL-interacting protein kinase 2 (CIPK2).	Upstream gene variant	4	3	0.75
AT5G07080	Encodes enzymes that can efficiently convert putrescine and caffeoyl-CoA to di-caffeoyl putrescine. Can convert spermidine/spermine and feruloyl CoA to mono-feruloyl spermidine/spermine. Has a preference for feruloyl-CoA binding, but little acyl-acceptor specificity.	5 prime UTR variant	4	3	0.75
AT5G07090	Ribosomal protein S4 (RPS4A) family protein	Upstream gene variant	4	3	0.75
AT5G07100	Encodes WRKY DNA-binding protein 26 (WRKY26).	Upstream gene variant	4	3	0.75
AT5G08680	Encodes the mitochondrial ATP synthase beta-subunit. This subunit is encoded by a multigene family of three members (At5g08670, At5g08680, At5g08690) that shared 98% sequence identity at the amino acid level. The mRNA is cell-to-cell mobile.	Downstream gene variant	4	3	0.75
AT5G08690	Encodes the mitochondrial ATP synthase beta-subunit. This subunit is encoded by a multigene family of three members (At5g08670, At5g08680, At5g08690) that shared 98% sequence identity at the amino acid level. The mRNA is cell-to-cell mobile.	Intron variant	4	3	0.75
AT5G08695	RNA-binding (RRM/RBD/RNP motifs) family protein	Upstream gene variant	4	3	0.75
AT5G14860	UDP-Glycosyltransferase superfamily protein	Downstream gene variant	4	3	0.75
AT5G14870	Encodes a member of the cyclic nucleotide gated channel family that is asymmetrically localized to the plasma membrane at the growing tip of the pollen tube and is	Upstream gene variant	4	3	0.75

	involved in pollen tube growth and pollen tube guidance to ovules.				
AT5G14880	Potassium transporter family protein	Upstream gene variant	4	3	0.75
AT5G19330	Encodes an armadillo repeat protein involved in the abscisic acid response. The protein interacts with a transcription factor, ABF2, which controls ABA-dependent gene expression via the G-box-type ABA-responsive elements.	Upstream gene variant	3	2	0.67
AT5G19340	hypothetical protein	Upstream gene variant	3	2	0.67
AT5G19350	RBP45d is a component of the U1 small nuclear ribonucleoprotein and regulates alternative splicing.	Upstream gene variant	3	2	0.67
AT5G19360	member of Calcium Dependent Protein Kinase; Phosphorylates the Ser860 residue of GLR3.7.	Downstream gene variant	3	2	0.67
AT5G19600	Encodes sulfate transporter Sultr3;5.	Upstream gene variant	5	4	0.80
AT5G19610	GNOM-like 2	splice_region_variant,Synonymous variant	5	4	0.80
AT5G19620	AtOEP80 is paralog to the chloroplastic protein translocation channel Toc75. Mutations in this locus result in embryo lethality.	Upstream gene variant	5	4	0.80
AT5G19780	Encodes an isoform of alpha tubulin. Closely related to adjacent gene TUA3 suggesting recent duplication. The mRNA is cell-to-cell mobile.	Downstream gene variant	3	2	0.67
AT5G19790	encodes a member of the ERF (ethylene response factor) subfamily B-6 of ERF/AP2 transcription factor family (RAP2.11). The protein contains one AP2 domain. There are 12 members in this subfamily including RAP2.11.	Upstream gene variant	3	2	0.67
AT5G19800	hydroxyproline-rich glycoprotein family protein	Downstream gene variant	3	2	0.67
AT5G19810	Proline-rich extensin-like family protein	Upstream gene variant	3	2	0.67
AT5G19820	Encodes an importin that transports HYL1, a component of the microprocessor, from the cytoplasm to the nucleus to constitute functional microprocessor, thereby affecting miRNA processing.	Downstream gene variant	3	2	0.67
AT5G24330	Encodes a SET-domain protein, a H3K27 monomethyltransferases required for chromatin structure and gene silencing. Regulates heterochromatic DNA replication. Contains a PCNA-binding domain.	Upstream gene variant	4	3	0.75
AT5G24340	3-5 exonuclease domain-containing protein	3 prime UTR variant	4	3	0.75
AT5G24350	Member of MAG2 complex on the ER that is responsible for efficient transport of seed storage proteins, functions in protein transport between the ER and Golgi apparatus, contain a Zeste?White 10 (ZW10) domain and a Sec39 domain.	Upstream gene variant	4	3	0.75
AT5G25820	Exostosin family protein	Downstream gene variant	3	2	0.67
AT5G28080	Encodes a member of the WNK family (9 members in all) of protein kinases, the structural design of which is clearly distinct from those of other known protein kinases, such as receptor-like kinases and mitogen-activated protein kinases.	Downstream gene variant	8	6	0.75
AT5G28610	LOW protein: ATP-dependent RNA helicase DRS1-like protein	Upstream gene variant	3	2	0.67
AT5G28615	RNA-directed DNA polymerase (reverse transcriptase)-related family protein	Upstream gene variant	3	2	0.67
AT5G28640	Encodes a protein with similarity to mammalian transcriptional coactivator that is involved in cell proliferation during leaf and flower development. Loss of function mutations have narrow, pointed leaves and narrow floral organs.	Downstream gene variant	3	2	0.67
AT5G28770	BASIC LEUCINE ZIPPER protein which regulates the circadian oscillator gene PSEUDO RESPONSE REGULATOR7 (PRR7) to change the circadian phase in response to sugars. It upregulates PRR7 in response to low energy.	Upstream gene variant	3	2	0.67
AT5G34829	transmembrane protein	Upstream gene variant	5	4	0.80
AT5G34830	hypothetical protein	Downstream gene variant	5	4	0.80
AT5G35160	Endomembrane protein 70 protein family	Downstream gene variant	3	2	0.67

AT5G35170	adenylate kinase family protein	Upstream gene variant	3	2	0.67
AT5G35195	Encodes a defensin-like (DEFL) family protein.	Upstream gene variant	3	2	0.67
AT5G36980	U3 small nucleolar RNA-associated protein	Upstream gene variant	7	5	0.71
AT5G44200	Encodes a nuclear cap-binding protein that forms a heterodimeric complex with ABH1 (ATCBP80) and is likely to participate in RNA metabolism. Its mRNA is ubiquitously expressed.Loss of function mutations suggest a role in processing of pri-miRNA and mRNA splicing.	Upstream gene variant	4	3	0.75
AT5G44210	encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-9). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole.	Upstream gene variant	4	3	0.75
AT5G44220	F-box family protein	Upstream gene variant	4	3	0.75
AT5G45510	Leucine-rich repeat (LRR) family protein	Downstream gene variant	3	2	0.67
AT5G45520	Leucine-rich repeat (LRR) family protein	Intron variant	3	2	0.67
AT5G45530	transmembrane protein, putative (DUF594)	Downstream gene variant	3	2	0.67
AT5G45540	transmembrane protein, putative (DUF594)	Downstream gene variant	3	2	0.67
AT5G49560	Putative methyltransferase family protein	Upstream gene variant	4	3	0.75
AT5G49570	Encodes a protein that has peptide:N-glycanase activity in enzymatic assay in heterologous systems (although the activity was not detected in wild-type plants).	Upstream gene variant	4	3	0.75
AT5G52520	Encodes a chloroplast and mitochondria localized prolyl-tRNA synthetase.	Downstream gene variant	3	2	0.67
AT5G52530	dentin sialophosphoprotein-like protein	Upstream gene variant	3	2	0.67
AT5G52540	keratin-associated protein, putative (DUF819)	Downstream gene variant	3	2	0.67
AT5G52545	RNA-binding protein-like RNA recognition motif protein	Downstream gene variant	3	2	0.67
AT5G52547	hypothetical protein	Downstream gene variant	3	2	0.67
AT5G54800	Encodes glucose6-Phosphate/phosphate transporter 1. Essential for pollen maturation and embryo sac development. The mRNA is cell-to-cell mobile.	Downstream gene variant	3	2	0.67
AT5G54810	A.thaliana tryptophan synthase beta subunit (trpB)	Upstream gene variant	3	2	0.67
AT5G54820	F-box/RNI-like superfamily protein	Upstream gene variant	3	2	0.67
AT5G54830	DOMON domain-containing protein / dopamine beta-monooxygenase N-terminal domain-containing protein	Upstream gene variant	3	2	0.67
AT5G56000	HEAT SHOCK PROTEIN 81.4	Upstream gene variant	6	4	0.67
AT5G56010	A member of heat shock protein 90 (HSP90) gene family. Expressed in all tissues and abundant in root apical meristem, pollen and tapetum.	Intron variant	6	4	0.67
AT5G56020	Got1/Sft2-like vescicle transport protein family	Upstream gene variant	6	4	0.67
AT5G56030	A member of heat shock protein 90 (HSP90) gene family. Expressed in all tissues and abundant in root apical meristem, pollen and tapetum.	Upstream gene variant	6	4	0.67
AT5G57130	Encodes a member of an eight-gene family (SMAX1 and SMAX1-like) that has weak similarity to AtHSP101, a ClpB chaperonin required for thermotolerance.	Downstream gene variant	3	2	0.67
AT5G57140	purple acid phosphatase 28	Missense variant	3	2	0.67
AT5G57150	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	Upstream gene variant	3	2	0.67
AT5G57160	Encodes the Arabidopsis orthologue of the yeast and mammalian DNA ligase IV. Involved in the repair of DNA damage but, unlike in yeast, not required for T-DNA integration. Interacts with the Arabidopsis homologue of XRCC4.	Downstream gene variant	3	2	0.67

AT5G57220	member of CYP81F, involved in glucosinolate metabolism. Mutants had impaired resistance to fungi. The mRNA is cell-to-cell mobile.	Downstream gene variant	3	2	0.67
AT5G57230	Thioredoxin superfamily protein	Downstream gene variant	3	2	0.67
AT5G57240	OSBP(oxysterol binding protein)-related protein 4C	Upstream gene variant	3	2	0.67
AT5G57250	Pentatricopeptide repeat (PPR) superfamily protein	Downstream gene variant	3	2	0.67
AT5G65005	Polynucleotidyl transferase, ribonuclease H-like superfamily protein	Downstream gene variant	3	2	0.67
AT5G65010	Encodes asparagine synthetase (ASN2).	Downstream gene variant	3	2	0.67
AT5G65015	pre-tRNA tRNA-Arg (anticodon: ACG)	Downstream gene variant	3	2	0.67
AT5G65020	Annexins are calcium binding proteins that are localized in the cytoplasm. When cytosolic Ca2+ increases, they relocate to the plasma membrane. They may be involved in the Golgi-mediated secretion of polysaccharides.	Upstream gene variant	3	2	0.67
AT5G65030	nitric oxide synthase-interacting protein	Downstream gene variant	3	2	0.67
AT5G65040	senescence-associated family protein (DUF581)	Downstream gene variant	3	2	0.67

Table A9.2 P10-23 x *fip37-4* down-regulated genes.

Gene ID	log2FC	Adj p value	Gene Description
AT1G55430	-2.11480	0.00019	Cysteine/Histidine-rich C1 domain family protein
AT4G04565	-3.67540	0.00003	Long non-coding RNA
AT5G56700	-2.07834	0.02930	FBD / Leucine Rich Repeat domains containing protein
AT1G58037	-1.82742	0.00341	Cysteine/Histidine-rich C1 domain family protein
AT2G25460	-1.89281	0.00000	EEIG1/EHBP1 protein amino-terminal domain protein
AT2G05825	-1.50311	0.04604	Long non-coding RNA
AT1G76955	-2.02499	0.00811	Expressed protein
AT3G59930	-2.01469	0.00000	Encodes a defensin-like (DEFL) family protein that is expressed in roots and downregulated in syncytium formed during nematode infection.
AT4G05380	-3.79526	0.00001	P-loop nucleoside triphosphate hydrolase superfamily protein
AT2G01010	-1.91257	0.00000	rRNA ribosomal_rna
AT1G22470	-1.81133	0.00002	Hypothetical protein. Target of SR45.
AT3G26235	-2.00107	0.02833	Hypothetical protein
AT1G76470	-2.77435	0.00000	NAD(P)-binding Rossmann-fold superfamily protein
AT1G80100	-2.39906	0.03919	AHP6 lacks the conserved histidine residue (Asn83 in AHP6b), which is required for phosphotransfer, present in the other AHPs. AHP6 does not
AT3G18530	-1.55628	0.00000	ARM repeat superfamily protein
AT2G42660	-1.68175	0.02555	Homeodomain-like superfamily protein
AT1G24145	-2.30199	0.00000	Long non-coding RNA
AT3G46614	-4.45670	0.02627	Unknown gene other_rna
AT4G24380	-1.65936	0.00000	dihydrofolate reductase
AT5G45570	-3.02067	0.02864	Ulp1 protease family protein
AT2G11270	-2.65126	0.00002	Citrate synthase-like protein
AT1G31400	-4.90851	0.00654	TRAF-like family protein
AT2G38255	-2.43832	0.00000	Domain of unknown function 239 (DUF239) gene; conserved Neprosin-like catalytic domain; possibly involved in seed development.
AT3G04640	-1.50681	0.00000	Glycine-rich protein
AT3G04220	-1.50639	0.00025	Target of miR825/825. Mutants have decreased resistance to fungal pathogens.
AT2G01422	-2.41648	0.00055	Unknown gene
AT1G01580	-1.92024	0.00000	Encodes the low-iron-inducible ferric chelate reductase responsible for reduction of iron at the root surface. It is likely to be the major F
AT1G24147	-2.27551	0.00000	Long non-coding RNA
AT5G52020	-3.28667	0.00000	Encodes a member of the DREB subfamily A-4 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 17 members i
AT5G10130	-1.62820	0.00000	Pollen Ole e 1 allergen and extensin family protein
AT1G11785	-1.69411	0.02542	Long non-coding RNA
AT1G72240	-1.62782	0.00021	Hypothetical protein
AT4G01985	-1.55297	0.01900	Hypothetical protein
AT1G72060	-2.48913	0.02031	Serine-type endopeptidase inhibitor
AT2G37870	-1.50958	0.00075	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT3G01345	-4.32609	0.00000	Expressed protein
AT1G67148	-1.66643	0.00097	Hypothetical protein
AT1G21550	-1.61518	0.02350	Calcium-binding EF-hand family protein
AT1G60750	-2.00536	0.00559	NAD(P)-linked oxidoreductase superfamily protein
AT4G08545	-5.61964	0.00068	LRR receptor-like protein
AT3G13784	-1.82150	0.00186	Cell wall invertase 5
AT4G30420	-2.02946	0.00154	Nodulin MtN21-like transporter family protein
AT1G49570	-1.85980	0.00463	Peroxidase superfamily protein
AT5G66150	-4.87033	0.00265	Glycosyl hydrolase family 38 protein
AT5G60350	-3.56595	0.00155	Hypothetical protein
AT1G48405	-3.30542	0.00008	Kinase interacting (KIP1-like) family protein
AT1G52060	-2.31150	0.00000	Mannose-binding lectin superfamily protein
AT1G18960	-3.54767	0.00000	Myb-like HTH transcriptional regulator family protein
AT4G13620	-1.67157	0.00000	encodes a member of the DREB subfamily A-6 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 8 members in
AT3G63320	-4.01739	0.00103	Protein phosphatase 2C family protein
AT5G41550	-1.85736	0.00000	Disease resistance protein (TIR-NBS-LRR class) family
AT4G09430	-3.35404	0.00173	Disease resistance protein (TIR-NBS-LRR class)
AT4G29610	-1.72143	0.04911	Cytidine/deoxycytidylate deaminase family protein
AT1G49640	-2.00014	0.02817	Alpha/beta-Hydrolases superfamily protein
AT3G14200	-1.69831	0.00000	Chaperone DnaJ-domain superfamily protein
AT5G40000	-1.88314	0.01014	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT4G14370	-1.52917	0.00001	Disease resistance protein (TIR-NBS-LRR class) family
AT1G14550	-3.34073	0.03069	Peroxidase superfamily protein
AT5G61412	-3.05731	0.01937	Hypothetical protein
AT5G63087	-3.89666	0.02638	Encodes a Plant thionin family protein
AT5G04815	-4.38307	0.01352	Long non-coding RNA
AT3G19430	-1.56922	0.00103	Late embryogenesis abundant protein-related / LEA protein-like protein
AT2G47480	-2.75197	0.00894	DUF3511 domain protein, putative (DUF3511)
AT1G32375	-2.21250	0.00005	F-box/RNI-like/FBD-like domains-containing protein
AT3G50400	-1.50530	0.00000	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or hydrolase reactio
AT2G01008	-1.51283	0.02242	Maternal effect embryo arrest protein
AT2G15340	-4.68280	0.03520	Glycine-rich protein
AT5G41765	-2.54017	0.00060	DNA-binding storekeeper protein-related transcriptional regulator
AT2G20142	-2.23828	0.00000	Toll-Interleukin-Resistance (TIR) domain family protein
AT5G48175	-1.60046	0.01875	Long non-coding RNA
AT4G01770	-1.69001	0.04446	Encodes a protein with UDP-xylose-dependent xylosyltransferase activity, which transfers Xyl onto L-fucose and (albeit less efficiently) L-ara
AT2G34224	-3.59944	0.04752	Hypothetical protein

AT2G34030	-2.60216	0.03136	Calcium-binding EF-hand family protein
AT4G29370	-2.21899	0.01936	Galactose oxidase/kelch repeat superfamily protein
AT2G37530	-1.65004	0.00001	forkhead box protein G1
AT5G52750	-2.33381	0.00000	Heavy metal transport/detoxification superfamily protein
AT1G27565	-3.53847	0.00007	Hypothetical protein
AT2G31083	-1.69804	0.00106	Member of a large family of putative ligands homologous to the Clavata3 gene. Consists of a single exon.
AT5G41763	-2.82834	0.00091	Hypothetical protein
AT2G19890	-4.10171	0.00000	Hypothetical protein
AT1G73325	-2.63694	0.00012	Kunitz family trypsin and protease inhibitor protein
AT3G28580	-2.18451	0.00022	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT2G13430	-4.50908	0.01026	Hypothetical protein
AT1G35210	-2.49828	0.00000	Hypothetical protein
AT5G50443	-5.08560	0.00723	Long non-coding RNA
AT2G26692	-2.40826	0.00574	Natural antisense transcript overlaps with AT2G26690 antisense_long_noncoding_rna
AT2G43660	-2.18266	0.04778	Carbohydrate-binding X8 domain superfamily protein
AT4G15953	-2.13358	0.00027	Encodes a Maternally expressed gene (MEG) family protein
AT1G61750	-2.19820	0.04240	Receptor-like protein kinase-related family protein
AT5G06585	-2.16838	0.00711	Long non-coding RNA
AT2G32020	-1.51312	0.00000	Acyl-CoA N-acyltransferases (NAT) superfamily protein
AT5G03885	-1.51353	0.03484	Hypothetical protein
AT5G47130	-3.73538	0.04539	Bax inhibitor-1 family protein
AT5G22530	-2.00857	0.03816	Unknown protein, knockout shows increased sensitivity to Al stress.
AT5G58910	-1.53746	0.00172	Putative laccase, a member of laccase family of genes (17 members in Arabidopsis).
AT4G08555	-2.89977	0.00000	Hypothetical protein
AT1G14480	-1.80092	0.00000	Ankyrin repeat family protein
AT1G03710	-1.59996	0.03150	Cystatin/monellin superfamily protein
AT1G17960	-1.53506	0.00000	Threonyl-tRNA synthetase
AT5G41750	-2.49045	0.00000	Disease resistance protein (TIR-NBS-LRR class) family
AT1G18980	-1.53897	0.00717	RmlC-like cupins superfamily protein
AT3G23470	-1.95522	0.00000	Cyclopropane-fatty-acyl-phospholipid synthase
AT1G29560	-1.99644	0.00000	Zinc finger C-x8-C-x5-C-x3-H type family protein
AT1G18290	-2.28809	0.00000	PADRE protein up-regulated after infection by S. sclerotiorum.
AT1G05894	-3.01088	0.00000	Hypothetical protein
AT3G01830	-2.36939	0.00077	Calcium-binding EF-hand family protein
AT5G54165	-2.17181	0.00001	Avr9/Cf-9 rapidly elicited protein
AT5G18661	-2.27154	0.03816	This gene encodes a small protein and has either evidence of transcription or purifying selection.
AT5G03210	-3.27324	0.00034	Encodes a small polypeptide contributing to resistance to potyvirus.
AT5G24880	-1.51518	0.02533	chromo domain cec-like protein
AT2G03130	-5.60996	0.00118	Ribosomal protein L12/ ATP-dependent Clp protease adaptor protein ClpS family protein
AT1G24580	-2.01811	0.00055	RING/U-box superfamily protein
AT4G05091	-4.88056	0.00453	This gene encodes a small protein and has either evidence of transcription or purifying selection.
AT1G71280	-2.14895	0.00456	DEA(D/H)-box RNA helicase family protein
AT1G07167	-1.56858	0.01201	Long non-coding RNA
AT1G50745	-1.51547	0.02254	Plant mobile domain protein
AT2G39240	-1.63710	0.00059	RNA polymerase I specific transcription initiation factor RRN3 protein
AT4G38560	-1.94314	0.00009	Phospholipase-like protein (PEARLI 4) family protein
AT1G50060	-2.14467	0.00230	CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein) superfamily protein
AT1G04090	-1.78230	0.01565	Vacuolar sorting-associated protein (DUF946)
AT2G39920	-2.21083	0.04869	HAD superfamily, subfamily IIIB acid phosphatase
AT2G25169	-4.88809	0.02079	Long non-coding RNA
AT2G35950	-1.73307	0.00001	Embryo sac development arrest 12
AT5G09976	-2.10317	0.00429	Hypothetical protein
AT5G15940	-1.83473	0.00192	NAD(P)-binding Rossmann-fold superfamily protein
AT5G10340	-2.50136	0.04357	F-box family protein
AT1G64795	-5.18699	0.00155	Hypothetical protein
AT5G52760	-2.40890	0.00000	Copper transport protein family
AT1G17285	-1.61657	0.00593	Long non-coding RNA
AT1G78410	-1.74713	0.02567	VQ motif-containing protein
AT5G03285	-1.88339	0.00000	Unknown gene
AT4G27654	-2.29661	0.00003	Long non-coding RNA
AT5G16900	-2.45506	0.00000	Leucine-rich repeat protein kinase family protein (LFR)
AT5G19240	-1.86584	0.00000	Glycoprotein membrane precursor GPI-anchored
AT2G40450	-2.90208	0.04094	BTB/POZ domain-containing protein
AT2G37700	-2.28717	0.03966	Fatty acid hydroxylase superfamily
AT3G30460	-2.00921	0.01459	RING/U-box superfamily protein
AT3G51642	-2.73566	0.00126	This gene encodes a small protein and has either evidence of transcription or purifying selection.
AT1G12030	-2.14212	0.00001	DUF506 gene family member
AT2G29110	-2.40343	0.02251	Member of Putative ligand-gated ion channel subunit family
AT1G78030	-3.00034	0.00000	Hypothetical protein
AT1G08440	-3.89723	0.00001	Aluminum activated malate transporter family protein
AT4G03292	-4.23004	0.02995	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
AT5G24600	-1.81824	0.00029	TRP-like ion channel protein (Protein of unknown function, DUF599)
AT1G13310	-2.73688	0.00100	Endosomal targeting BRO1-like domain-containing protein
AT5G22430	-3.08743	0.04495	Pollen Ole e 1 allergen and extensin family protein
AT4G10410	-4.00900	0.00002	Leucine-rich repeat (LRR) family protein
AT1G76960	-1.85611	0.00010	Unknown protein, contains WRKY40 binding motifs.
AT1G23205	-1.55162	0.00000	Plant invertase/pectin methylesterase inhibitor superfamily protein
AT3G51360	-1.59937	0.04034	Eukaryotic aspartyl protease family protein
AT5G19100	-1.58960	0.00000	Eukaryotic aspartyl protease family protein
AT1G20310	-2.70475	0.00001	Syringolide-induced protein
AT3G49055	-1.58476	0.00000	ATP-binding protein
AT3G28840	-1.72183	0.00455	Hypothetical protein (DUF1216)
AT1G48060	-2.43582	0.00219	F-box/associated interaction domain protein
AT2G47950	-1.56187	0.02063	Myelin transcription factor-like protein
AT5G02575	-2.08351	0.00002	None small_nucleolar_rna

AT4G25980	-1.91645	0.00100	peroxidase superfamily protein
AT1G08507	-1.73252	0.02124	Long non-coding RNA
AT2G23630	-2.23529	0.00003	SKU5 similar 16
AT1G16705	-3.52527	0.00000	p300/CBP acetyltransferase-related protein-like protein
AT2G03850	-5.10133	0.00882	Late embryogenesis abundant protein (LEA) family protein
AT3G47050	-2.70948	0.01766	Glycosyl hydrolase family protein
AT3G14140	-2.16754	0.00005	2-oxoglutarate-dependent dioxygenase family protein
AT1G58225	-2.78609	0.00157	Hypothetical protein
AT1G14250	-1.73334	0.00000	GDA1/CD39 nucleoside phosphatase family protein
AT2G36295	-2.05370	0.00536	Hypothetical protein
AT4G36770	-2.45516	0.04286	UDP-Glycosyltransferase superfamily protein
AT1G67105	-2.05979	0.00000	Unknown gene
AT1G28370	-2.77202	0.00000	Encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family
AT1G32970	-2.09110	0.02441	Subtilisin-like serine endopeptidase family protein
AT5G56200	-4.94685	0.00723	Encodes a transcription factor expressed in the female gametophyte.
AT4G02170	-2.18209	0.00217	Cotton fiber protein
AT5G45095	-1.71557	0.00012	Hypothetical protein
AT1G69480	-1.77115	0.00000	EXS (ERD1/XPR1/SYG1) family protein
AT3G24542	-5.53661	0.00000	Beta-galactosidase related protein
AT2G32030	-2.23236	0.00000	Acyl-CoA N-acyltransferases (NAT) superfamily protein
AT3G03290	-3.07287	0.03546	Adenine nucleotide alpha hydrolases-like superfamily protein
AT5G07640	-2.02695	0.00000	RING/U-box superfamily protein
AT4G00695	-2.00112	0.00150	Spc97/Spc98 family of spindle pole body (SBP) component
AT5G60610	-2.61310	0.00000	F-box/RNI-like superfamily protein
AT4G27580	-3.13349	0.00258	Phosphatidylinositol transfer SFH5-like protein
AT4G27400	-1.80663	0.01584	Late embryogenesis abundant (LEA) protein-like protein
AT2G02630	-1.53401	0.00024	Cysteine/Histidine-rich C1 domain family protein
AT2G46600	-1.50654	0.00000	Calcium-binding EF-hand family protein
AT5G23115	-3.22835	0.00000	Hypothetical protein
AT1G32950	-2.36700	0.00000	Subtilase family protein
AT4G32510	-2.42768	0.00000	HCO3- transporter family
AT3G10150	-1.60441	0.00001	Purple acid phosphatase 16
AT3G29000	-3.02568	0.00000	Calcium-binding EF-hand family protein
AT1G43910	-1.54577	0.00000	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT5G02565	-1.95784	0.00000	None small_nucleolar_rna
AT3G29633	-3.72235	0.00000	Hypothetical protein
AT5G54370	-1.83655	0.00098	Late embryogenesis abundant (LEA) protein-like protein
AT2G32140	-2.30892	0.00018	Transmembrane receptor
AT1G07135	-3.02513	0.00000	Glycine-rich protein
AT2G29370	-2.97039	0.00459	NAD(P)-binding Rossmann-fold superfamily protein
AT1G24140	-2.80719	0.00000	Expression induced by fungal and bacterial pathogens (AT3-MMP)
AT5G48400	-1.72025	0.00000	Member of Putative ligand-gated ion channel subunit family (ATGLR1.2)
AT3G62760	-1.55543	0.01124	Encodes glutathione transferase belonging to the phi class of GSTs
AT3G06345	-1.60710	0.00000	MicroRNA ath-MIR5642a precursor (ATH-MIR5642A)
AT5G45630	-2.89624	0.00000	Senescence regulator (Protein of unknown function, DUF584) (ATS40-2)
AT4G12510	-1.94995	0.00002	Encodes a member of the AZI family of lipid transfer proteins (AZI5)
AT1G72210	-1.56092	0.00000	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein (BHLH096)
AT3G44006	-3.10589	0.00581	This gene encodes a small protein and has either evidence of transcription or purifying selection
AT4G29200	-3.57032	0.04938	Over-expressed by salt stress (DEG3)
AT5G54030	-1.69932	0.00572	Cysteine/Histidine-rich C1 domain family protein (DT 1053)
AT1G52343	-2.39622	0.03416	Similar to GET2, Long non-coding RNA that interacts with GET1 (G1IP-LIKE)
AT3G10185	-2.36417	0.01495	Encodes a Gibberellin-regulated GASA/GAST/Snakin family protein (GASA13)
AT1G54050	-1.64671	0.00000	HSP20-like chaperones superfamily protein (HSP17.4B)
AT1G59860	-2.56582	0.00000	HSP20-like chaperones superfamily protein (HSP17.6A)
AT1G07400	-2.63065	0.00000	HSP20-like chaperones superfamily protein (HSP17.8)
AT5G51440	-2.08175	0.00000	HSP20-like chaperones superfamily protein (HSP23.5)
AT1G73480	-2.11928	0.00000	Alpha/beta-Hydrolases superfamily protein (MAGL4)
AT1G54540	-1.75142	0.00000	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family (NHL4)
AT5G63660	-1.54299	0.00009	Predicted to encode a PR (pathogenesis-related) protein
AT5G51500	-2.25752	0.00002	Pectin methylesterase inhibitor; INVI-PMEI protein isoform which clustered with a PME (PMEI-PME60)
AT2G47050	-1.57097	0.01291	Encodes a pollen-expressed pectin methylesterase inhibitor that affects male fertility by regulating pollen viability and pollen tube growth
AT5G44567	-2.98824	0.04624	SCOOP (SERINE RICH ENDOGENOUS PEPTIDE) family member, involved in immunity against herbivorous insects (PROSCOOP2)
AT3G61390	-1.53548	0.02944	Plant U-box type E3 ubiquitin ligase (PUB36)
AT5G48570	-1.71511	0.00000	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins
AT1G32960	-2.60158	0.00000	Subtilase family protein (SBT3.3)
AT1G71890	-1.67466	0.00000	Encodes a sucrose transporter that is expressed in the endosperm. Mutants have delayed accumulation of fatty acids and embryo maturation
AT5G50790	-1.87719	0.01561	Encodes a member of the SWEET sucrose efflux transporter family proteins. Transcriptionally activated by long photoperiods; activation depends
AT4G23810	-2.68674	0.00000	Member of WRKY Transcription Factor; Group III; t WRKY53 and CRK5 are antagonistic regulators of chlorophyll synthesis/degradation, senescence
AT3G46090	-2.33251	0.00013	C2H2 and C2HC zinc fingers superfamily protein (ZAT7)
AT4G26200	-1.50340	0.00015	Member of a family of proteins in Arabidopsis that encode 1-Amino-cyclopropane-1-carboxylate synthase, an enzyme involved in ethylene biosynth
AT5G12020	-2.16882	0.00000	17.6 kDa class II heat shock protein
AT5G48390	-3.52541	0.00000	Defective in meiotic chromosome segregation. It is involved in crossover formation and involved in both male and female meiosis
AT2G40220	-4.34978	0.00937	Encodes a member of the DREB subfamily A-3 of ERF/AP2 transcription factor family (ABI4)
AT1G36180	-1.79394	0.00000	Acetyl-CoA carboxylase 2 (ACC2) The mRNA is cell-to-cell mobile
AT1G33560	-2.05708	0.00000	Encodes a NBS-LRR disease resistance protein that possesses N-terminal kinase subdomains
AT4G22950	-1.84240	0.00000	MADS-box protein AGL19; FYF-like gene that belongs to the SOC1 group; is complementary to FYF in regulating flower senescence
AT3G25250	-1.55715	0.00000	Arabidopsis protein kinase The mRNA is cell-to-cell mobile AGC2 KINASE 1 (AGC2-1)
AT1G04580	-2.08327	0.00486	Encodes aldehyde oxidase AAO4 preferentially expressed in developing seeds ALDEHYDE OXIDASE 4 (AO4)

AT1G01520	-2.13144	0.00000	RVE3 is one of eleven homologous MYB-like transcription factors in Arabidopsis and a member of the RVE8 clade. Plays a minor role in clock reg
AT3G54340	-2.56595	0.01561	Floral homeotic gene encoding a MADS domain protein homologous to SRF transcription factors
AT5G10760	-2.14669	0.00000	Eukaryotic aspartyl protease family protein APOPLASTIC, EDS1-DEPENDENT 1 (AED1)
AT4G21990	-1.50579	0.00000	Encodes a protein disulfide isomerase-like (PDIL) protein, a member of a multigene family within the thioredoxin (TRX) superfamily
AT4G09110	-2.91308	0.01340	RING/U-box superfamily protein ARABIDOPSIS T??XICOS EN LEVADURA 35 (ATL35)
AT5G61980	-1.51737	0.00000	A member of ARF GAP domain (AGD), A thaliana has 15 members, grouped into four classes. AGD1 belongs to the class 1, together with AGD2, AGD3
AT1G31290	-1.63612	0.00000	ARGONAUTE 3 (AGO3)
AT5G21150	-1.66269	0.00000	AGO9-dependent sRNA silencing is crucial to specify cell fate in the Arabidopsis ovule
AT5G44110	-2.02359	0.00000	Encodes a member of the NAP subfamily of ABC transporters whose expression pattern is regulated by light and sucrose
AT3G13090	-1.51754	0.00649	Member of MRP subfamily ATP-BINDING CASSETTE C6 (ABCC6) MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 8 (ATMRP8);ATP-BINDING CASSETT
AT3G52310	-1.57787	0.00000	ABCG27 is induced by ABA. Loss of function increases ABA concentration in pistils and leaves. ATP-BINDING CASSETTE G27 (ABCG
AT1G27770	-1.54355	0.00000	Encodes a chloroplast envelope Ca2+-ATPase with an N-terminal autoinhibitor
AT1G33960	-1.77242	0.03908	Identified as a gene that is induced by avirulence gene avrRpt2 and RPS2 after infection with Pseudomonas syringae pv maculicola strain ES4326
AT3G56970	-1.71185	0.01917	Encodes a member of the basic helix-loop-helix transcription factor family protein. bHLH IVc transcription factor repressed by bHLH11
AT3G56980	-1.79335	0.00173	Encodes a member of the basic helix-loop-helix transcription factor protein. bHLH IVc transcription factor repressed by bHLH11
AT2G41240	-1.51286	0.04504	Encodes a member of the basic helix-loop-helix transcription factor family protein. Functions as a key regulator of iron-deficiency responses
AT1G47600	-1.88697	0.00001	Encodes a myrosinase. Over-expression led to a glucosinolate profile change BETA GLUCOSIDASE 34 (BGLU34)
AT3G61190	-1.56503	0.00004	Encodes a protein with a C2 domain that binds to BON1 in yeast two hybrid analyses
AT5G01060	-1.58562	0.02989	Kinase with tetratricopeptide repeat domain-containing protein
AT4G25470	-1.93049	0.00003	Encodes a member of the DREB subfamily A-1 of ERF/AP2 transcription factor family (CBF2). The protein contains one AP2 domain
AT4G27280	-1.61945	0.00000	EF-hand Ca2 + -binding protein, which is a Ca2+-dependent transducer of auxin-regulated gene expression and interacts with ICR1
AT1G76640	-3.49058	0.00072	Calmodulin like gene, expressed in roots and induced by abiotic and biotic stimulus CALMODULIN LIKE 39 (CML39)
AT1G76650	-2.29296	0.00317	Calmodulin like gene. Inducible by nitrate and hypoxia and involved in nitrate induced BR signaling and response to hypoxia.
AT5G39670	-2.66925	0.00000	Calmodulin like protein involved in negative regulation of pattern triggered immunity CALMODULIN-LIKE 46 (CML46)
AT5G26920	-1.88284	0.00000	Encodes a calmodulin-binding protein CBP60g (calmodulin binding protein 60-like.g)
AT1G14160	-1.88474	0.00000	Uncharacterized protein family (UPF0497) CASP-LIKE PROTEIN 1A1 (CASPL1A1)
AT5G15290	-2.07069	0.00001	Uncharacterized protein family (UPF0497) CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN 5 (CASP5)
AT2G31910	-1.67867	0.00000	Member of Putative Na+/H+ antiporter family CATION/H+ EXCHANGER 21 (CHX21)
AT5G58460	-5.12896	0.00995	Member of Putative Na+/H+ antiporter family CATION/H+ EXCHANGER 25 (CHX25)
AT1G06970	-1.63131	0.03056	Member of Putative Na+/H+ antiporter family CATION/HYDROGEN EXCHANGER 14 (CHX14)
AT2G25240	-2.17495	0.02424	Serine protease inhibitor (SERPIN) family protein. Involved in stress response regulated cell death
AT3G17690	-2.40884	0.00000	Member of Cyclic nucleotide gated channel family CYCLIC NUCLEOTIDE GATED CHANNEL 19 (CNGC19)
AT3G60550	-1.52232	0.00071	Cyclin p3 CYCLIN P3;2 (CYCP3;2)
AT4G23310	-2.13347	0.04661	Encodes a cysteine-rich receptor-like protein kinase CYSTEINE-RICH RLK (RECEPTOR-LIKE PROTEIN KINASE) 23 (CRK23)
AT4G23140	-1.91258	0.00223	Arabidopsis thaliana receptor-like protein kinase
AT1G56060	-1.94968	0.00029	CYSTM3 is a mitochondrial protein that is induced by salt stress and is a negative regulator of salt stress
AT2G32200	-1.83435	0.00000	Cysteine-rich/transmembrane domain A-like protein CYSTEINE-RICH TRANSMEMBRANE MODULE 5 (ATHCYSTM5)
AT2G32210	-1.79927	0.00000	Cysteine-rich/transmembrane domain A-like protein CYSTEINE-RICH TRANSMEMBRANE MODULE 6 (ATHCYSTM6)
AT3G50930	-2.32043	0.00000	Encodes a protein that is present in a homo-multimeric protein complex on the outer mitochondrial membrane and plays a role in cell death
AT4G22710	-2.06024	0.01292	Member of CYP706A CYTOCHROME P450, FAMILY 706, SUBFAMILY A, POLYPEPTIDE 2 (CYP706A2)
AT4G31940	-2.39883	0.00000	The gene encodes a cytochrome P450 enzyme, CYP82C. It is involved in the early Fe deficiency response
AT1G64930	-1.51428	0.00920	Member of CYP89A CYTOCHROME P450, FAMILY 87, SUBFAMILY A, POLYPEPTIDE 7 (CYP89A7)
AT1G65340	-4.73076	0.03619	Member of CYP96A CYTOCHROME P450, FAMILY 96, SUBFAMILY A, POLYPEPTIDE 3 (CYP96A3)
AT1G74550	-4.17789	0.02674	Encodes a tricoumaroylspermidine meta-hydroxylase
AT1G48220	-5.64487	0.00244	Plays role in abscisic acid (ABA)-dependent pathway in response to dehydration; forms homo- or heterodimers in a kinase active dependent manner
AT4G24570	-1.65622	0.00018	Encodes one of the mitochondrial dicarboxylate carriers (DIC): DIC1 (AT2G22500), DIC2 (AT4G24570), DIC3 (AT5G09470)
AT5G09470	-2.11562	0.02189	Encodes one of the mitochondrial dicarboxylate carriers (DIC): DIC1 (AT2G22500), DIC2 (AT4G24570), DIC3 (AT5G09470)
AT2G20560	-1.89200	0.00000	DNAJ heat shock family protein DNAJ PROTEIN (DNAJ)
AT5G46140	-3.73957	0.00288	Hypothetical protein DUF295 ORGANELLAR A 13 (ATDOA13)
AT5G55270	-3.30778	0.00119	Hypothetical protein DUF295 ORGANELLAR B 14 (ATDOB14)
AT4G13680	-1.64953	0.00015	Hypothetical protein DUF295 ORGANELLAR B 3 (ATDOB3)
AT5G53230	-4.31036	0.00000	Hypothetical protein DUF295 ORGANELLAR B 6 (ATDOB6)
AT5G53240	-6.19481	0.00039	Hypothetical protein DUF295 ORGANELLAR B 7 (ATDOB7)
AT1G12610	-1.94304	0.04444	Encodes a member of the DREB subfamily A-1 of ERF/AP2 transcription factor family (DDF1)
AT3G63060	-1.66871	0.00485	EDL3 is an F-box protein involved that mediated the regulation of abscisic acid signalling EID1-LIKE 3 (EDL3)
AT5G09980	-1.52800	0.00000	Elicitor peptide 4 precursor ELICITOR PEPTIDE 4 PRECURSOR (PROPEP4)
AT4G13890	-2.90854	0.04239	Encodes a serine hydroxymethyltransferase maximally expressed in root; involved in S-adenosyl-L-methionine biosynthesis
AT5G61600	-1.84102	0.00000	Encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family

AT1G33760	-2.67692	0.00000	Encodes a member of the DREB subfamily A-4 of ERF/AP2 transcription factor family
AT4G17490	-2.11124	0.00000	Encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family (ATERF-6)
AT2G44840	-3.38759	0.00000	Encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family
AT4G01630	-1.67203	0.00310	Member of Alpha-Expansin Gene Family
AT5G06630	-1.70625	0.00547	Proline-rich extensin-like family protein EXTENSIN 9 (EXT9)
AT5G59390	-2.09402	0.00192	SGS3-like, XH/XS domain-containing protein involved in DNA methylation FACTOR OF DNA DEMETHYLATION 1 (FDDM1)
AT3G44870	-2.99446	0.00067	Encodes a protein with 93% identity to a farnesoic acid methyl transferase. SABATH family methyltransferase
AT1G47395	-1.75872	0.00000	Involved in regulation of iron deficiency response genes FE-UPTAKE-INDUCING PEPTIDE2 (FEP2)
AT1G47400	-1.70795	0.00001	Involved in regulation of iron deficiency response genes. Overexpression results in hyperaccumulation of Fe and Mn
AT5G64870	-1.72324	0.00000	Belongs to the group of plant flotillins, which are plasma membrane proteins. Flot3 is found in membrane nanodomains
AT1G30040	-1.70728	0.00000	Encodes a gibberellin 2-oxidase that acts on C-19 gibberellins. AtGA2OX2 expression is responsive to cytokinin and KNOX activities
AT1G80340	-1.58017	0.00814	Encodes a protein with gibberellin 3 hydroxylase activity
AT5G36870	-2.15217	0.04055	Encodes a gene similar to callose synthase GLUCAN SYNTHASE-LIKE 9 (GSL09)
AT5G25980	-1.70041	0.00000	Myrosinase (thioglucoside glucohydrolase) gene involved in glucosinoloate metabolism. The mRNA is cell-to-cell mobile
AT2G48150	-1.80105	0.03888	Encodes glutathione peroxidase GLUTATHIONE PEROXIDASE 4 (GPX4)
AT4G11050	-2.18965	0.00011	Glycosyl hydrolase 9C3 GLYCOSYL HYDROLASE 9C3 (GH9C3)
AT5G13900	-1.67837	0.00000	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT1G53130	-1.60056	0.04981	Encodes GRIM REAPER (GRI), involved in the regulation of cell death induced by extracellular ROS (reactive oxygen species)
AT3G12580	-1.96368	0.00000	Cytoplasmically localized member of the heat shock protein 70 family. Forms complex(heat shock granule) with J3
AT1G74310	-2.00362	0.00000	Encodes ClpB1, which belongs to the Casein lytic proteinase/heat shock protein 100 (Clp/Hsp100) family
AT5G12030	-2.06223	0.00000	Encodes a cytosolic small heat shock protein with chaperone activity that is induced by heat and osmotic stress
AT5G52640	-2.06916	0.00000	Encodes a cytosolic heat shock protein AtHSP90.1. AtHSP90.1 interacts with disease resistance signaling components SGT1b and RAR1
AT5G43840	-2.04533	0.01411	Member of Heat Stress Transcription Factor (Hsf) family HEAT SHOCK TRANSCRIPTION FACTOR A6A (HSFA6A)
AT2G26150	-2.79769	0.00000	Member of Heat Stress Transcription Factor (Hsf) family. Involved in response to misfolded protein accumulation in the cytosol
AT4G29770	-1.96038	0.00146	Target of trans acting-siR480/255. Testing HEAT-INDUCED TAS1 TARGET 1 (HTT1)
AT5G59220	-1.51651	0.00416	Encodes a member of the PP2C family (Clade A protein phosphatases type 2C). Functions as a negative regulator of osmotic stress and ABA signalling
AT4G12400	-1.53499	0.00000	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins
AT1G47510	-1.69789	0.00000	Encodes a phosphatidylinositol polyphosphate 5-phosphatase. It can dephosphorylate PI(4,5)P2, PI(3,5)P2, and PI(3,4,5)
AT3G49380	-1.85098	0.00181	Member of IQ67 (CaM binding) domain containing family IQ-DOMAIN 15 (iqd15)
AT4G19690	-2.44080	0.00000	The gene encodes Fe2+ transporter protein. It is a member of the Zrt/Irt-like protein (ZIP) family of transporters
AT1G07367	-1.91780	0.01292	Member of conserved family of peptides involved in Fe uptake and homeostasis IRONMAN 4 (IMA4)
AT1G67360	-1.71006	0.00000	Encodes a small rubber particle protein homolog. Plays dual roles as positive factors for tissue growth and development and in drought stress
AT5G59310	-1.75756	0.00530	Encodes a member of the lipid transfer protein family
AT1G72520	-2.10243	0.00003	13-lipoxygenase that is induced by or abiotic stresses and triggers defense respones LIPOXYGENASE 4 (LOX4)
AT2G28500	-2.38086	0.00000	LOB domain-containing protein 11 LOB DOMAIN-CONTAINING PROTEIN 11 (LBD11)
AT3G47870	-3.52786	0.00056	Required for normal cell division during pollen development. Mutant has extra cell in pollen of vegetative cell identity
AT5G58310	-1.52803	0.00112	Encodes a protein shown to have methyl IAA esterase activity in vitro. This protein does not act on methyl JA, MeSA, MeGA4, or MEGA9 in vitro
AT4G27765	-2.88549	0.00712	Encodes a microRNA that targets several MYB family members and the tasiRNA-generating transcript TAS4
AT4G25200	-1.79248	0.00465	AtHSP23.6-mito mRNA, nuclear gene encoding mitochondrial MITOCHONDRION-LOCALIZED SMALL HEAT SHOCK PROTEIN 23.6
AT2G47460	-2.41503	0.00000	MYB12 belongs to subgroup 7 of the R2R3-MYB family. It strongly activates the promoters of chalcone synthase (CHS)
AT3G23250	-1.79239	0.00053	Member of the R2R3 factor gene family. Key regulator of lignin biosynthesis in effector-triggered immunity
AT4G13480	-3.24306	0.04520	Member of the R2R3 factor gene family MYB DOMAIN PROTEIN 79 (MYB79)
AT4G26930	-4.06843	0.01501	Encodes a putative transcription factor (MYB97)
AT1G52040	-2.08737	0.00000	Encodes myrosinase-binding protein expressed in flowers MYROSINASE-BINDING PROTEIN 1 (MBP1)
AT2G17040	-2.90656	0.00000	Member of the NAC transcription factor family and more specifically, the ONAC022 subfamily
AT3G01600	-1.70470	0.00000	NAC domain containing protein 44 (NAC044)
AT4G28530	-1.58468	0.00004	Member of NAC family of transcription factors. Along with NAC2, KIR1 positively regulates programmed cell death of stigmatic tissue
AT2G20800	-4.37535	0.00000	NAD(P)H dehydrogenase B4 NAD(P)H DEHYDROGENASE B4 (NDB4)
AT3G52640	-2.01130	0.00000	Encodes a gamma-secretase subunit. Associates with other subunits in intracellular membrane compartments
AT5G56080	-1.95658	0.00032	Encodes a protein with nicotianamine synthase activity. Its transcript levels rise in roots in response to zinc deficiency and rise in leaves
AT2G20470	-2.19628	0.00113	AGC (cAMP-dependent, cGMP-dependent and protein kinase C) kinase family protein NUCLEAR DBF2-RELATED 5
AT5G19470	-4.18392	0.01838	Nudix hydrolase homolog 24 (NUDT24)
AT3G57260	-2.53469	0.00001	Beta 1,3-glucanase PATHOGENESIS-RELATED PROTEIN 2 (PR2)
AT4G17660	-1.57210	0.01025	Protein kinase superfamily protein PBS1-LIKE 20 (PBL20)
AT2G39110	-1.53757	0.00000	Protein kinase superfamily protein PBS1-LIKE 38 (PBL38)
AT1G26970	-1.84211	0.00522	Protein kinase superfamily protein PBS1-LIKE 4 (PBL4)
AT1G57590	-1.68713	0.00002	Pectinacetylesterase family protein (PAE2)
AT3G09405	-1.54184	0.00000	Pectinacetylesterase family protein (PAE4)

AT2G26820	-2.93231	0.04780	phloem protein 2-A3 (PP2-A3)
AT1G56250	-1.76335	0.00932	Encodes an F-box protein that can functionally replace VirF, regulating levels of the VirE2 and VIP1 proteins via a VBF-containing SCF complex
AT1G73010	-2.61854	0.00000	Encodes PPsPase1, a pyrophosphate-specific phosphatase catalyzing the specific cleavage of pyrophosphate
AT1G76530	-2.07880	0.00009	Auxin efflux carrier family protein PIN-LIKES 4 (PILS4)
AT5G54490	-1.72006	0.00000	Encodes a PINOID (PID)-binding protein containing putative EF-hand calcium-binding motifs
AT1G14880	-2.41405	0.01806	PLANT CADMIUM RESISTANCE 1 (PCR1)
AT4G28460	-3.22685	0.01667	Activates immune responses through RECEPTOR-LIKE KINASE7 (RLK7). Induces stomatal closure is dependent on RLK7 and the transcription of genes
AT4G37290	-3.20921	0.00027	Encoding a precursor protein of a secreted peptide that is responsive to flg22 stimulus
AT1G08590	-3.11270	0.00486	Encodes one of the two putative eLRR kinase closely related to PXY (At1g08590/PXL1 and At4g28650/PXL2)
AT3G30720	-2.84034	0.00000	QQS is an orphan gene that arose recently in the Arabidopsis thaliana lineage
AT5G64990	-2.47375	0.00000	RAB GTPase homolog H1A RAB GTPASE HOMOLOG H1A (RABH1a)
AT2G32680	-1.91516	0.00087	NLP20 LRR receptor protein involved in PAMP mediated immunity
AT3G28890	-3.79161	0.02188	Receptor like protein 43 (RLP43)
AT5G60900	-2.81933	0.00001	Encodes a receptor-like protein kinase RECEPTOR-LIKE PROTEIN KINASE 1 (RLK1)
AT3G49580	-1.69802	0.00000	Response to low sulfur 1 (LSU1)
AT5G24660	-2.38938	0.00000	Response to low sulfur 2 (LSU2)
AT3G49570	-2.09680	0.00000	Response to low sulfur 3 (LSU3)
AT1G65570	-2.63846	0.01610	Encodes a glycosyl hydrolase 28 (GH28) family polygalacturonase (PG) protein. Involved in root cap development
AT4G22080	-2.52520	0.00730	Root hair specific 14 (RHS14)
AT3G30350	-1.98472	0.00146	Encodes a root meristem growth factor (RGF). Belongs to a family of functionally redundant homologous peptides that are secreted
AT4G13395	-1.71217	0.00161	ROTUNDIFOLIA like 12 (RTFL12)
AT1G68825	-1.57668	0.04781	ROTUNDIFOLIA like 15 (RTFL15)
AT2G36985	-1.63012	0.02027	Encodes ROTUNDIFOLIA4, a member of the seed plant-specific family of small peptides
AT1G27730	-1.81518	0.00005	Related to Cys2/His2-type zinc-finger proteins found in higher plants. Compensated for a subset of calcineurin deficiency in yeast
AT1G73805	-3.23609	0.00000	Encodes SAR Deficient 1 (SARD1), a key regulator for ICS1 (Isochorismate Synthase 1) induction and salicylic acid (SA) synthesis
AT3G12900	-2.77167	0.00018	S8H hydroxylates scopoletin to generate fraxetin (8-hydroxyscopoletin)
AT4G36490	-2.30722	0.02909	SEC14-like 12 (SFH12)
AT3G23800	-1.72814	0.00000	Selenium-binding protein 3 (SBP3)
AT1G24260	-1.97108	0.00000	Member of the MADs box transcription factor family. SEP3 is redundant with SEP1 and 2
AT1G73300	-1.78728	0.00000	Serine carboxypeptidase-like 2 (scpl2)
AT5G48850	-1.59710	0.00024	Homologous to the wheat sulphate deficiency-induced gene sdi1. Expression in root and leaf is induced by sulfur starvation
AT5G11410	-2.42602	0.00000	Similar to receptor like kinase but does not appear to have kinase activity (psuedokinase)
AT5G05490	-1.69305	0.00000	Encodes a RAD21-like gene essential for meiosis. Encodes a 627 a.a. protein that is slightly longer in the N-terminus than SYN1 BP5
AT3G55980	-1.94221	0.00001	CCCH-type zinc finger protein involved in salt stress and immune responses TANDEM ZINC FINGER 11 (TZF11)
AT2G41800	-1.73116	0.00000	Encodes a DUF642 cell wall protein that is highly induced during the M/G1 phases of the cell cycle and is involved in hypocotyl cell elongation
AT5G57810	-2.18450	0.00113	Member of TETRASPANIN family (TET15)
AT4G30430	-3.05474	0.00000	Member of TETRASPANIN family (TET9)
AT1G72940	-1.53262	0.00007	Nucleotide-binding, leucine-rich repeat (NLR) gene regulated by nonsense-mediated mRNA decay (NMD) genes UPF1 and UPF3.
AT1G66090	-1.99155	0.00003	Disease resistance protein TIR-NBS3 (TN3)
AT1G72900	-2.01604	0.00000	Toll-Interleukin-Resistance (TIR) domain-containing protein TIR-NBS7 (TN7)
AT1G72910	-1.72136	0.00007	Nucleotide-binding, leucine-rich repeat (NLR) gene regulated by nonsense-mediated mRNA decay (NMD) genes UPF1 and UPF3.
AT1G72920	-2.80598	0.00000	Toll-Interleukin-Resistance (TIR) domain family protein TIR-NBS9 (TN9)
AT1G20350	-1.59704	0.00034	Mitochondrial inner membrane translocase TRANSLOCASE INNER MEMBRANE SUBUNIT 17-1 (TIM17-1)
AT2G36750	-1.53067	0.00080	UDP-glucosyl transferase 73C1 (UGT73C1)
AT1G05530	-1.77387	0.00206	Encodes a protein with glucosyltransferase activity with high sequence homology to UGT1 (AT1G05560)



Figure A9.1 Double KO plants showing increased branching.

Representative photos showing the phenotype of plant growth for double mutants of *ythdf3/ythdf4, and ythdf1/ythdf3.* (A) Photo showing ythdf3/ythdf4 displaying hyperbranched phenotype. (B) Close-up of ythdf3/ythdf4 fruits. (C) Photo showing ythdf1/ythdf3 displaying hyperbranched phenotype. Plants were grown in glasshouse and photographed at 12 weeks.