WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/164344

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Manipulation of meristem initiation and maintenance in wheat and rice

Cathal Meehan

This thesis is submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

University of Warwick, Department of Life Sciences

August 2021

In loving memory of

Tommy Meehan

and

Kathleen 'Baby' Kelly

Table of Contents

List of Figures xii

List of Tables xxii

Abbreviations xv

Acknowledgments xxviii

Declarations xxiv

Abstract xxx

Chapter 1: Introduction

1.1 The state of food security in the 21st century 1

1.2 Role of the grasses in food security 3

1.3 Phylogeny and Characterisation of the Grasses 5

1.4 Wheat domestication 8

1.5 Rice domestication 9

1.6 Tiller number – a grass domestication trait 11

1.7 Genetic architecture of tillering 15

1.8 Wheat and rice inflorescence morphology 19

1.9 RKDs – evolutionary conserved morphogenic regulators 25

1.10 Project aims 29

Chapter 2: Materials and methods

2.1 Materials 30

2.1.1 Plant materials 30

2.1.1.1. Fielder wheat material and growth conditions 30

2.1.1.2. Black rice material and growth conditions 30

2.1.1.3. Nipponbare rice material and growth conditions 31

2.1.2. Plant growth media 32

2.1.2.1. Hydroponics solution medium (Broughton and Dilworth, 1970) 32

2.1.2.2. Somatic embryogenesis media 32

2.1.3. Generation of *indTaRKD1-7D* Fielder transgenic lines 33

2.1.4. Genotyping of *indTaRKD1-7D* Fielder calli and plants 33

2.1.5. Generation of *indOsRKD3 Cempo Ireng* transgenic lines 34

2.1.6. Genotyping for *indOsRKD3 Cempo Ireng* calli and plants 35

2.1.7. Generation of inducible-AtRKD4 transgenic lines 36

2.1.8. Genotyping of *indTaRKD4* Nipponbare calli and plants 36

2.1.9. Somatic embryogenesis induction of immature wheat embryos 36

2.1.10. Hydroponics experiment 37

2.2. Molecular biological methods 37

2.2.1. gDNA extraction 37

2.2.2. Polymerase chain reaction 38

2.2.3. Agarose gel electrophoresis 38

2.2.4. Estradiol treatment and collection of the treated *indTaRKD1-7D* samples 39

2.2.5. DEX treatment and collection of the treated *indOsRKD3* samples 39

2.2.6. RNA extraction of wheat leaf and immature embryo tissue 40

2.2.7. RNA extraction of rice leaf tissue 40

2.2.8. cDNA synthesis and RT-PCR 41

2.3. Histology 41

2.3.1. Preparation of RNA probes 41

2.3.2. RNA *in situ* hybridization in developing axils 42

2.3.3. Toluidine Blue staining of developing axis 44

2.3.4. Antisera preparation and immunolocalization 45

2.4. Next generation sequencing library preparation 46

2.4.1. *indOsRKD3* and *indTaRKD1-7D* RNA extraction 46

2.4.2. *indOsRKD3* and *indTaRKD1-7D* RNA library preparation and sequencing 46

2.4.3. DAPseq analyses of *Triticum aestivum* genomic DNA with TaRKD1-7D 46

2.4.4. Wheat TaRKD1-7D ChIPseq library preparation and sequencing 48

2.5 Next generation sequencing analysis 48

2.5.1 Downloading raw reads from publicly available datasets 48

2.5.2 Transcriptomic analysis of wheat RNAseq datasets 48

2.5.2.1 Pre-processing and pseudoalignment 48

2.5.2.2 Sample selection and calculating differentially expressed genes 49

2.5.2.3 Visualisation of differentially expressed gene list 49

2.5.2.4 Gene set enrichment and GO term analysis 49

2.5.3 Transcriptomics analysis of rice RNAseq dataset 50

2.5.3.1 Pre-processing, alignment and counting 50

2.5.3.2 Sample selection and calculating differentially expressed genes 50

2.5.3.3 Visualisation of differentially expressed gene list 51

2.5.3.4 Gene set enrichment and GO term analysis 51

2.5.4 Wheat meristem coexpression network construction and analysis 51

2.5.4.1 Meristematic coexpression network table construction 51

2.5.4.2 DEG specific network table filtering 52

2.5.4.3 Cytoscape visualisation of the log2 fold change network 52

2.5.4.4 Cytoscape visualisation of the GO term network 52

2.5.5 Rice coexpression network construction and analysis 53

2.5.5.1 Coexpression network table construction 53

2.5.5.2 DEG specific network table filtering 53

2.5.5.3 Cytoscape visualisation of the network 53

2.5.6 Motif enrichment analysis 53

2.5.6.1 MEME-ChIP analysis of wheat DAPseq peaks 54

2.5.6.2 Predicted AtRKD2 binding sites in rice promoter sequences 54

2.5.7 DAPseq analysis of TARKD1-7D with wheat gDNA 54

2.5.7.1 Pre-processing and alignment 54

2.5.7.2 Peak calling 55

2.5.7.3 DAPseq functional and enrichment analysis 55

2.5.8 ChIPseq analysis of TaRKD1-7D in wheat hypocotyl tissue 56

2.5.8.1 Pre-processing and alignment 56

2.6.8.2. Peak calling 56

2.6.8.3. ChIPseq functional and enrichment analysis 56

2.5.9 Phylogenetic tree and protein alignments 57

2.6. Phenotypic analysis 57

2.6.1. Phenotypic screening for higher tiller number transgenic lines 57

2.6.2. Phenotypic trial for tiller traits 58

2.6.2.1. Design and conditions 58

2.6.2.2. Spike measurement 58

2.6.2.3. Seed threshing and measurements 59

2.6.3. Phenotypic trial for spike traits 59

2.6.3.1. Design and conditions 59

2.6.3.2. Spike movement 60

2.6.3.3. Seed threshing and automated grain counting 60

2.6.4. Phenotypic trial for nitrogen responces 60

2.6.4.1. Design and conditions 60

2.6.4.2. Spike measurement 61

2.6.5. Axillary meristem formation measurment 61

2.6.6. Scanning electron microscopy (SEM) 61

2.6.7. Spike x-ray Computerised Tomography scan 62

Chapter 3

3.1 Introduction 63

3.1.1 Vegetative and inflorescence architecture in wheat 63

3.1.2. Identification and characterization of candidate loci in wheat 64

3.1.3. Transformation and screening transgenic lines 65

3.1.4. Tiller phenotyping in wheat 66

3.1.5. Spike phenotyping in wheat 66

3.2. Results 67

3.2.1. Identification of embryonic *TaRKD* genes in IWGSC RefSeq v1.1 68

3.2.2. Transformation and screening of *indTaRKD1-7D* transgenic lines for higher tiller number and ectopic TaRKD1-7D expression 70

3.2.3. Increased tiller number and seed weight in two *indTaRKD1-7D* T3 lines 77

3.2.4. Increased spikelet number per spike in two *indTaRKD1-7D* T3 lines 83

3.2.5. Increased tiller number in two *indTaRKD1-7D* T3 lines independent of nitrogen availability 89

3.2.6. No increase in tiller number for estradiol treatment of *indTaRKD1-7* #1.4 92

3.3. Discussion 92

3.3.1. Zygotic RKD transcription factors in nature 92

3.3.2. Characterising RKD transcription factors in wheat and its wild progenitors 93

3.3.3. Ectopic *TaRKD1-7D* expression increase tiller number in wheat 93

3.3.4. *TaRKD1-7D* induced higher tiller number increases biomass and seed weight 94

3.3.5. *TaRKD1-7D* induced higher tiller number increases spikelet number per spike 95

3.3.6. *TaRKD1-7D* expression level dependent tiller number induction in wheat 95

3.3.7. Phenotypic model for *TaRKD1-7D* induced higher tiller number 96

Chapter 4

4.1 Introduction 97

4.1.1 Gene regulation in eukaryotes 97

4.1.2. Transcriptomics and sequencing 98

4.1.3. Aligning reads to reference assemblies of large complex genomes 99

4.1.4. Coexpression network analysis 100

4.1.5. Sequencing techniques for finding transcription factor binding sites 101

4.2. Results 103

4.2.1. *indTaRKD1-7D* #1.4 leaf tissue shows upregulation of photosynthetic genes 103

4.2.2. indTaRKD1-7D #2.28 immature embryo tissue shows upregulation of chromatin and nucleosome assembly genes 107

4.2.3. NIL tiller node tissue shows upregulation of functional terms found in *indTaRKD1-7D* 112

4.2.4. An upregulated *indTaRKD1-7D* #2.28 immature embryo shared DEGs coexpression network shows enrichment of histone H4 and HMG genes 118

4.2.5. TaRKD1-7D binds promoter regions of histone lysine methyltransferases and induces higher expression of plant architecture TFs 127

4.3. Discussion 139

4.3.1. Characterising transcriptional cascades using wheat bioinformatics resources 139

4.3.2. TaRKD1-7D binds histone and histone methylation genes, upregulating involved in chromatin accessibility, cell cycle and meristematic determinancy determinacy. 140

4.3.3. *indTaRKD1-7D* leaf tissue shows upregulation of photosynthesis genes but not histone methylation genes 142

Chapter 5

5.1 Introduction 144

5.1.1 Rice as a model organism of the grasses 144

5.1.2 Wild diversity in rice and domestication 145

5.2 Results 147

5.2.1 Ectopic expression of *indAtRKD4* in *Oryza sativa ssp. japonica* 'Nipponbare' increases vegetative tiller number 147

5.2.2 Ectopic expression of *OsRKD3* in *Oryza sativa Cempo Ireng* increases tiller number and seed weight in two *indOsRKD3* T3 lines 151

5.2.3 *indOsRKD3* leaf tissue shows upregulation of translational and metabolic processes. 160

5.2.4 Upregulated TD vs ND only DEGs with an AtRKD2 motif show enrichment for functional terms involved in transcriptional and metabolic processes 165

5.2.5 Upregulated TD vs ND only DEGs expressed in meristematic and reproductive tissue are involved in DNA binding and kinase signalling 167

5.2.6 An upregulated *indOsRKD3* TD vs ND only DEGs coexpression network shows functional enrichment for kinase signalling, stress response and cell cycle 172

5.2.7 *OsRKD3* induces higher tiller formation in *Cempo Ireng* in an expression level dependent manner 178

5.3 Discussion 180

5.3.1 Ectopic expression of zygotic *RKD*s induces higher tiller number in several BEP clade grass species 180

5.3.2 Transcriptomics analysis of *indOsRKD3 Cempo Ireng* leaf tissues shows upregulation of terms relevant to meristem and inflorescence development 181

Chapter 6

6.1 General discussion 182

Supplemental Information – Chapter 2 195

Supplemental Information – Chapter 3 198

Supplemental Information – Chapter 4 236

Supplemental Information – Chapter 5 247

References 255

List of Figures

Chapter 1

Figure 1.1 Global human population data and projections from years 1700-2100

Figure 1.2 Projected wheat yields to 2050

Figure 1.3 Grass grain and embryo anatomy

Figure 1.4 Phylogeny of the grasses

Figure 1.5 Bread wheat evolution and spike domestication

Figure 1.6 Two models for rice domestication

Figure 1.7 Inflorescence development in rice and wheat

Figure 1.8 RWP-PK phylogeny from Koi et al., 2016

Chapter 2

Figure 2.1 *indTaRKD1-7D* expression construct.

Figure 2.2 *indOsRKD3* expression construct.

Chapter 3

Figure 3.1 Heatmap of *TaRKD* expression

Figure 3.2 Phylogenetic tree of *RKD* genes in the grasses and Arabidopsis

Figure 3.3 RT-PCR of *TaRKD1-7D* expression across T2 transgenic lines in leaf tissue

Figure 3.4 RT-PCR of *TaRKD1-7D* expression across T2 transgenic lines in immature embryo tissue

Figure 3.5 Identification of *indTaRKD1-7D* T2 lines with higher reproductive tiller number

Figure 3.6 *TaRKD1-7D* expression in independent *indTaRKD1*-7D lines associated with tiller number

Figure 3.7 Tiller number in *indTaRKD1-7D #1.4* T3 plants

Figure 3.8 Reproductive tiller number two *indTaRKD1-7D* T3 lines

Figure 3.9 Plant height in *indTaRKD1-7D #1.4* T3 line

Figure 3.10 Seed weight per plant in two *indTaRKD1-7D* T3 lines

Figure 3.11 Spike length in two *indTaRKD1-7D* T3 lines

Figure 3.12 Spike number per spike in two *indTaRKD1-7D* T3 lines

Figure 3.13 Rachis elongation in two *indTaRKD1-7D* T3 lines

Figure 3.14 Images of *indTaRKD1-7D* spikes shows increased spikelet number per spike

Figure 3.15 X-ray CT scans of *indTaRKD1-7D* spikes

Figure 3.16 Nitrogen response trial shows two *indTaRKD1-7D* T3 lines with increased reproductive tiller number independent of nitrogen availability

Figure 3.17 Nitrogen response trial shows two *indTaRKD1-7D* T3 lines with no increase in nitrogen use efficiency for spike area

Chapter 4

Figure 4.1 PCA of samples and Venn diagram of DEGs from *indTaRKD1-7D #1.4* Leaf RNAseq experiment

Figure 4.2 Heatmap of shared DEGs expression from *indTaRKD1-7D #1.4* Leaf RNAseq experiment

Figure 4.3 GO term analysis of upregulated *indTaRKD1-7D #1.4* leaf shared DEGs

Figure 4.4 PCA of samples and Venn diagram of DEGs from *indTaRKD1-7D #2.28* Immature embryo RNAseq experiment

Figure 4.5 Heatmap of shared DEGs expression from *indTaRKD1-7D #2.28* immature embryo RNAseq experiment

Figure 4.6 GO term analysis of upregulated *indTaRKD1-7D #2.28* immature embryo shared DEGs

Figure 4.7 PCA of samples and Venn diagram of DEGs from NIL tiller node RNAseq experiment

Figure 4.8 Heatmap of shared DEGs expression from NIL tiller node RNAseq experiment

Figure 4.9 GO term analysis of upregulated NIL tiller node shared DEGs

Figure 4.10 Heatmap of TF family differential regulation enrichment

Figure 4.11 Toluidine staining and histone H4 RNA in-situ of *indTaRKD1-7D #1.4* tiller nodes show enrichment of cell division and histone H4

Figure 4.12 TaRKD1-7D protein immunolocalization in axillary meristem tissue

Figure 4.13 Upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network clustered and functionally annotated

Figure 4.14 Revigo analysis of GO terms from RNA processing cluster

Figure 4.15 Revigo analysis of GO terms from RNA processing cluster

Figure 4.16 Upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network annotated by differential expression

Figure 4.17 Enrichment of histone and HMG genes in nodes with highest indegree in the upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network

Figure 4.18 Enrichment of histone, HMG, and cyclin genes in nodes with highest outdegree in the upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network

Figure 4.19 ChIPseq and DAPseq *indTaRKD1-7D* shows significant enrichment of promoter regions

Figure 4.20 Heatmap of ChIPseq and DAPseq *indTaRKD1-7D* shows significant differential enrichment of promoter and downstream regions

Figure 4.21 ChIPseq and DAPseq *indTaRKD1-7D* shows higher coverage proximal to the TSS and TES

Figure 4.22 Motif similarity and central enrichment of TaRKD1-7D binding motif

Figure 4.23 GO term analysis of genes with a MACS2 ChIPseq peak called within 10kb of the TSS shows enrichment of photosynthetic and translation terms

Figure 4.24 GO term analysis of genes with a MACS2 DAPseq peak called within 10kb of the TSS shows enrichment of histone methylation and ubiquitination terms

Figure 4.25 Venn diagrams of *indTaRKD1-7D* ChIPseq, DAPseq and immature embryo RNAseq shared DEGs identify high confidence targets of TaRKD1-7D

Figure 4.26 indTaRKD1-7D DAPseq peak within 10kb of two upregulated *indTaRKD1-7D #2.28* immature embryo shared DEGs relevant to meristem determinancy

Figure 4.27 *indTaRKD1-7D* DAPseq peak within 10kb of two downregulated *indTaRKD1-7D #2.28* immature embryo shared DEG TFs relevant to cell death and DNA repair

Chapter 5

Figure 5.1 *indAtRKD4* Nipponbare lines show increased vegetative tiller number and delayed senescence

Figure 5.2 Vegetative tiller number in two *indAtRKD4* T3 Nipponbare lines

Figure 5.3 Plant height in *indAtRKD4* T3 Nipponbare lines

Figure 5.4 Increased reproductive tiller number in two *indOsRKD3 Cempo Ireng* T2 lines

Figure 5.5 Reproductive tiller number in an *indOsRKD3 Cempo Ireng* T2 line

Figure 5.6 Seed weight per plant in an *indOsRKD3 Cempo Ireng* T2 line

Figure 5.7 Panicle size, panicle branch number and seed number per panicle in two T2 *indOsRKD3 Cempo Ireng* lines

Figure 5.8 Panicle branch number in an *indOsRKD3 Cempo Ireng* T2 line

Figure 5.9 Increased tiller branch number in an *indOsRKD3 Cempo Ireng* T2 line

Figure 5.10 Increased inflorescence branches in two *indOsRKD3 Cempo Ireng* T2 lines

Figure 5.11 PCA of samples and Venn diagram of DEGs from *indOsRKD3 Cempo Ireng* leaf RNAseq experiment

Figure 5.12 Heatmap of TD vs ND only DEGs from *indOsRKD3 Cempo Ireng* leaf RNAseq experiment

Figure 5.13 GO term analysis of upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs

Figure 5.14 Enrichment of predicted AtRKD2 binding motif in 3kb upstream promoter regions of upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs

Figure 5.15 GO term analysis of upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs with a predicted AtRKD2 binding motif in their 3kb upstream promoter region shows enrichment of terms involved in transcription and metabolic process

Figure 5.16 Clustered expression plot shows tissue specific expression amongst *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs

Figure 5.17 Heatmap of Cluster 12 shows list of genes that show high expression in tissues like those involved in tiller initiation

Figure 5.18 GO term analysis of *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs Cluster 12 shows enrichment of terms involved in DNA binding and signalling

Figure 5.19 Upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs coexpression network clustered and functionally annotated

Figure 5.20 Revigo analysis of GO terms from protein phosphorylation cluster

Figure 5.21 Revigo analysis of GO terms from translation cluster

Figure 5.22 Upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs coexpression network annotated by differential expression

Figure 5.23 Enrichment of AP2 and ras-related genes in nodes with highest betweenness centrality in the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs coexpression network

Figure 5.24 Increased tiller initiation in *indOsRKD3 Cempo Ireng* T2 linein a dose dependent manner

Chapter 6

Figure 6.1 Model for TaRKD1 induced higher tiller and spikelet per spike number

Supplementary Information – Chapter 3

S. Figure 3.1 Confirmation of *TaRKD1* triad verified through protein sequence phylogeny

S. Figure 3.2 *TaRKD1* homeologue triad shows high expression in dividing microspore tissue

S. Figure 3.3 Heatmap of *AetRKD* expression from publicly available transcriptome data across tissues

S. Figure 3.4 Heatmap of *TdRKD* expression from publicly available transcriptome data across tissues

S. Figure 3.5 Polyembryonic structures in indTaRKD1-7D #2.28

S. Figure 3.6 Increased rachis length in high tiller number *indTaRKD1-7D* T3 lines

S. Figure 3.7 Increased rachis node number in high tiller number *indTaRKD1-7D* T3 lines

S. Figure 3.8 Increased rachis elongation in *indTaRKD1-7D* T3 lines

S. Figure 3.9 Increased unproductive tiller number of two *indTaRKD1-7D* T3 lines

S. Figure 3.10 Delayed flowering time of two *indTaRKD1-7D* T3 lines

S. Figure 3.11 No difference in flag leaf area of two *indTaRKD1-7D* T3 lines

S. Figure 3.12 Increased seed weight per 1000 of two *indTaRKD1-7D* T3 lines

S. Figure 3.13 Increased reproductive tiller number of two *indTaRKD1-7D* T3 lines in lower soil volume

S. Figure 3.14 Increased tiller number independent of soil volume of two *indTaRKD1-7D* T3 lines.

S. Figure 3.15 No change in reproductive tiller number of two *indTaRKD1-7D* T3 lines or wild type with mechanical stress treatment

S. Figure 3.16 No change in height of two *indTaRKD1-7D* T3 lines or wild type with mechanical stress treatment

S. Figure 3.17 Increased reproductive tiller number of two *indTaRKD1-7D* T3 lines at border positions of the trial

S. Figure 3.18 Higher axillary meristem formation in two *indTaRKD1-7D* T3 lines timecourse images

S. Figure 3.19 Higher axillary meristem formation in two *indTaRKD1-7D* T3 lines timecourse boxplots

S. Figure 3.20 Delayed anthesis time of two *indTaRKD1-7D* T3 lines

S. Figure 3.21 Increased spike area (cm^2) in two *indTaRKD1-7D* T3 lines

S. Figure 3.22 Increased spike length of two *indTaRKD1-7D* T3 lines

S. Figure 3.23 CT scan averaging to reduce background noise

S. Figure 3.24 Nitrogen response trial shows increase in reproductive tiller number of *indTaRKD1-7D #1.4* independent of nitrogen availability

S. Figure 3.25 Nitrogen response trial shows increase in reproductive tiller number of *indTaRKD1-7D #2.28* partially dependent on nitrogen availability

S. Figure 3.26 Nitrogen response trial shows increase in reproductive tiller number of Wild Type dependent on nitrogen availability

S. Figure 3.27 Nitrogen response trial shows no increase in nitrogen use efficiency for ear area in *indTaRKD1-7D #1.4*

S. Figure 3.28 Nitrogen response trial shows no increase in nitrogen use efficiency for ear area in *indTaRKD1-7D #2.28*

S. Figure 3.29 Nitrogen response trial shows no increase in nitrogen use efficiency for ear area in Wild type

S. Figure 3.30 No increase in vegetative tiller number in inducer treated *indTaRKD1-7D #1.4* in hydroponics trial

S. Figure 3.31 No increase in plant biomass in Estradiol treated *indTaRKD1-7D #1.4* in hydroponics trial

Supplementary Information – Chapter 4

S. Figure 4.1 *TaRKD1-7D* expression in the *indTaRKD1-7D #1.4* leaf RNAseq experiment shows high expression amongst transgenic samples

S. Figure 4.2 Volcano plot of the *indTaRKD1-7D #1.4* leaf TE vs NE DEGs shows greater upregulation

S. Figure 4.3 Volcano plot of the *indTaRKD1-7D #1.4* leaf TM vs NE DEGs and shows greater downregulation

S. Figure 4.4 *TaRKD1-7D* expression in the *indTaRKD1-7D #2.28* immature embryo RNAseq experiment shows high expression amongst transgenic samples

S. Figure 4.5 Volcano plot of the *indTaRKD1-7D #2.28* immature embryo IE_TE vs IE_WE DEGs shows greater upregulation

S. Figure 4.6 Volcano plot of the *indTaRKD1-7D #2.28* immature embryo IE_TM vs IE_WE DEGs shows greater downregulation

S. Figure 4.7 Significant up and downregulation of TF families in *indTaRKD1-7D #2.28* immature embryos

S. Figure 4.8 MACS2 ChIPseq peaks are distributed evenly across all 21 wheat chromosomes but with a high amount called in the unassembled region

S. Figure 4.9 MACS2 DAPseq peaks are distributed evenly across all 21 wheat chromosomes

S. Figure 4.10 Venn diagrams of *indTaRKD1-7D* DAPseq and immature embryo RNAseq shared DEGs identifies expression validated targets of TaRKD1-7D

Supplementary Information – Chapter 5

S. Figure 5.1 Increased days to maturity in an *indOsRKD3 Cempo Ireng* T2 line

S. Figure 5.2 Increased height in an *indOsRKD3 Cempo Ireng* T2 line

S. Figure 5.3 Decreased seed weight per 1000 seeds in an *indOsRKD3 Cempo Ireng* T2 line

S. Figure 5.4 Increased seed number per panicle in an *indOsRKD3 Cempo Ireng* T2 line

S. Figure 5.5 *indOsRKD3* expression in the *indOsRKD3* leaf RNAseq experiment shows high expression amongst transgenic samples

S. Figure 5.6 Heatmap of TD vs ND only DEGs from *indOsRKD3 Cempo Ireng* leaf RNAseq experiment across tissues

S. Figure 5.7 Volcano plot of the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs equal shows up and downregulation

List of Tables

Chapter 2

Table 2.1 Hydroponics solutions

Chapter 3

Table 3.1 List of *indTaRKD1-7D* lines

Chapter 5

Table 5.1 Statistics for phenotypic traits in *indOsRKD3 Cempo Ireng* trial.

Supplementary Information – Chapter 2

S. Table 2.1 Oligonucletoide sequences

S. Table 2.2 Pre-processing statistics for next generation sequencing analysis

Supplementary Information – Chapter 3

S. Table 3.1 List of *TaRKD* genes in IWGSC RefSeq v1.1. Assembly

S. Table 3.2 List of *AetRKD* genes in INSDC Aet v4.0. Assembly

S. Table 3.3 List of *TdRKD* genes in INSDC WEWSeq v1.0. Assembly

S. Table 3.4 Fold changes in phenotypic traits with mechanical stress treatment in tiller traits trials compared to WT

S. Table 3.5 P-values in phenotypic traits with mechanical stress treatment in tiller traits trials compared to WT

S. Table 3.6 Fold changes in phenotypic traits in tiller traits trial compared to WT

S. Table 3.7 P-values in phenotypic traits in tiller traits trial compared to WT

S. Table 3.8 Statistical significance test of anthesis time between genotypes from spike traits trial

S. Table 3.9 Fold changes spike traits trials compared to WT

S. Table 3.10 P-values in phenotypic spike traits trial compared to WT

S. Table 3.11 Fold changes in phenotypic traits in nitrogen response trial compared to WT

S. Table 3.12 P-values in phenotypic traits in nitrogen response trial compared to WT

S. Table 3.13 Fold changes in phenotypic traits in hydroponics traits tirals compared to mock solution

S. Table 3.14 P-values in phenotypic traits in hydroponics traits tirals compared to mock solution

Supplementary Information – Chapter 4

S. Table 4.1 *indTaRKD1-7D #2.28* immature embryo shared DEGs with a DAPseq peak 10kb within TSS relevant to cell cycle and meristems

Supplementary Information – Chapter 5

S. Table 5.1	Transgenic rice lines provided by NIAB Cambridge
S. Table 5.2	Statistics for phenotypic traits in *indAtRKD4* Nipponbare trial

Abbreviations

AG type-C agamous

APO1 aberrant panicle organization 1

AXM axillary meristem (intro)

BBM1 BABY BOOM1

BEP Bambusoideae, Ehrhartoideae (formerly Oryzoideae) and Pooideae

Bh4 Black hull 4

BOP blade on petiole 1

Br brittle rachis

CCD7 Carotenoid cleavage dioxygenas7

ChIPseq Chromatin immunoprecipitation sequencing

CNI Close-Neighbour-Interchange

CO_2 carbon dioxide

CT computerised tomography

Cul4 uniculm4

D14 Dwarf14

D27 Dwarf27

DAPseq DNA affinity purification sequencing

DEG differentially expressed gene

FAO Food and Agriculture Organisation

FM Floral meristem

FT1 flowering locus 1

FT1 Flowering time 1

FT2 Flowering time 2

FUL2 fruitfull2

GB Gibberellin

GO gene ontology

GT1 Grassy tillers 1

HadGEM2 Hadley centre global environment model version 2

HMG high mobility box gene

IFPRI International Food Policy Research Institute

IM inflorescence meristem

INSDC International Nucleotide Sequence

Database Collaboration

IRGS International Rice Genome Sequencing

IRGSP International Rice Genome Sequencing Project

IWGSC International Wheat Genome Sequencing Consortium

LEC1 LEAFY COTYLEDON1

LS lateral suppressor

MADS57 (MADS-box transcription factor 57)

MAS marker assisted selection

MFT-like Mother of FT and TFL1-like

MITE miniature inverted-repeat transposable element

MOC1 monoculm 1

MS Murashige and Skoog

ND Non-transgenic dexamethasone

NE non-transgenic estradiol

NGR5 Nitrogen mediated tiller growth response 5

NIL near isogenic line

Nin nodule inception

NLP Nin-like protein

NM Non-transgenic mock

ORF open reading frames

OSH1 oryza sativa homeobox1

PACMAD Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae

PBF Phytobiology Facility

PCA principal component analysis

PCR polymerase chain reaction

PPD1 photoperiod 1

PTFDB Plant Transcription Factor Database

qPCR quantitative polymerase chain reaction

QTL quantitative trait locus

RAM root apical meristem

RAP-DB rice annotation project database

RT-PCR reverse transcription polymerase chain reaction

SAM shoot apical meristem

SEM scanning electron microscopy

SL Strigalactone

SNS Spikelet number per spike

Suc Sucrose

TAB1 tillers absent 1

TD Transgenic dexamethasone

TE transgenic estradiol

TES transcriptional end site

TF transcription factor

Tin tiller inhibition

TM transgenic mock

TPM Transcript per million values

Tq tenacious glume

TSS transcriptional start site

UFO unusual flowering organs

VRN1 vernalisation protein 1

WE wild type estradiol

WGCNA Weighted Gene Coexpression Network Analysis

WMG Warwick Manufacturing Group

WUS wuschel

Acknowledgements

I would like to extend my gratitude to my supervisor Dr. Jose Gutierrez-Marcos for giving me the opportunity to work with him as a research assistant and to develop my knowledge and skills as a doctoral researcher. Your guidance has taught me to embrace the serendipity of science and to be passionate about research.

I would also like to thank my advisory panel, Dr. Vardis Ntoukakis and Dr. Steve Jackson for their generous time and advice. I must give a special thanks to Dr. Yang-Seok Lee, without his support and patience this research project and many others wouldn't have been possible. I must also give a special thanks to Dr. Jon Price, my bioinformatics mentor, and every past and present member of the Marcos group.

This academic journey wouldn't be possible without the support of the Molaise Meehans, the Kellys, and the McEllins, especially my father and best friend Charlie, my sister Orla, my stepmother Helen, and everyone else in Breaffy's #1 hostel. I've missed you all over this pandemic and look forward to seeing you again. During these uncertain times, I've been incredibly lucky to have my better half Marie Grypioti with me for every step and whose love and kindness has kept me.

To my friends Rory, Dan, Anna, Aaron, Shona, Genevieve, James K, Alonso, Jon, Grace, Quentin, Joanna, Rob, Habib, Sahan, Sophie, Arkadios, James T and Jack - my time in the UK wouldn't have been the same without you and I'll really miss you all. A special thanks to the all the Coventry Cubs who persevered with me on winter nights in the Chestnut Tree beer garden. I hope we stay in contact, and you visit me for a horribly themed Franco-Irish night in Paris that I will be calling the 'Sean's Elise'.

Lastly, I thank the BBSRC and the MIBTP program for their training and funding.

Declarations

This thesis was submitted to the University of Warwick for the degree of Doctor of Philosophy. The work presented here is original and has not been considered for any other award. All work here in was performed by me with exceptions outlined below.

Dr. Yang Seok (University of Warwick)

Assisted in many wet-lab related parts of this project such as tissue collection, RNA extraction, RT-PCR, qPCR, wheat histology, RNAseq and DAPseq library preparation, rice growth, rice phenotyping, and phylogenetic tree construction.

Mr. Tom Dean (University of Warwick)

Assisted in the harvesting of wheat samples and image analysis of wheat spikes.

Dr. Moussa Benhammed's lab (L'Institut des Sciences des Plantes - Paris-Saclay)

Assisted in growing wheat plants, ChIPseq library preparation and sequencing.

The Research Institute for Agriculture Technology, Yogyakarta, Indonesia)

Growth, phenotyping, and histology of black rice plants.

Dr. Paul Wilson (University of Warwick)

Visualisation of wheat spikes with x-ray computerised tomography scans.

Phytobiology Facility (University of Warwick)

Assisted in the growth of the samples and management of the growdom

Abstract

The grasses are one of the largest and most diverse families within the angiosperms and make up to 70% of all crops grown. Domestication of wild ancestors of the major cereals involved artificial selection of agronomically important traits like tillering. Plant architecture in the grasses is influenced by the initiation and outgrowth of axillary meristems that determine the number of reproductive organs formed. Inflorescence architecture has also been a major selection trait in human domestication of cereal crops and is determined by a complex genetic network of meristematic transcription factors. The RWP-RK transcription factors are a family of developmental regulators that are tightly expressed in reproductive tissue and are implicated in early embryogenesis across the plant kingdom. We've found that ectopic expression of zygotic the RWP-RK genes *TaRKD1-7D* and *OsRKD3* leads to increased tiller number, inflorescence size and seed weight in wheat and rice respectively. Interestingly, we also see greater increases of these traits in conditions of lower nitrogen availability. Through extensive bioinformatics analysis we hypothesise that ectopic expression of these transcription factors upregulates pluripotency factors, specifically cell cycle regulators, histones, histone methylation factors and meristem determinacy genes like trehalose phosphate synthases. This increases meristem formation and indeterminacy by creating and maintaining stem cell niches that determine plant architecture. This is a novel strategy for manipulating yield determining traits across a range of cultivars and cereal crop species.

Chapter 1

Introduction

1.1 The state of food security in the 21st century

Food security poses itself as one of the biggest challenges of the 21st century with the global human population set to grow to 9.7 billion by 2050, peaking at 11 billion people by the turn of the century (Figure 1.1) (United Nations Population division, 2019). To feed the global population, in line with the goals of the Millennium declaration, agricultural output will need to increase by 60-110% while using the same if not less arable land and maintaining nutritional quality (Ray et al., 2013). Achieving this monumental task in the wake of climate change will require the development of improved crop varieties, effective irrigation systems, increased nutrient use efficiency while conserving biodiversity and soil properties.

The Intergovernmental Panel on Climate Change's (IPCC) 2014 report remarked that the world may reach "a threshold of global warming beyond which current agricultural practices can no longer support large human civilizations" by the middle of 21st century. The IPCC's Third Assessment Report published in 2001 concluded that developing countries would be the worst-hit by climate change with reductions in crop yield predominantly happening in tropical and sub-tropical countries due to increased pest incidence and limited water availability. The threat of global warming via greenhouse gas emissions also raises the need to consider new approaches to sequester carbon to offset some of the worst impacts of climate change. Re-wilding agricultural

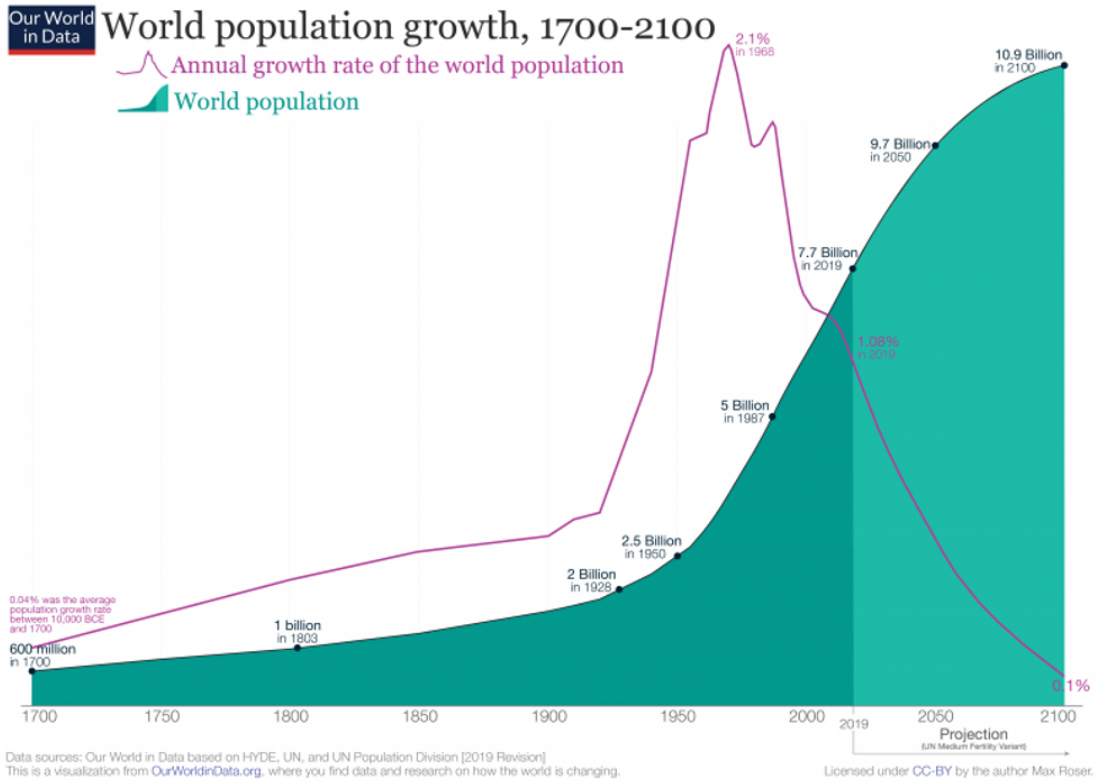


Figure 1.1. Global human population data and projections from years 1700-2100. Total population is the area of under the green curve. Recorded data are shown in dark green while projections are in light green. The pink line shows annual population growth. Visualisation taken from OurWorldinData.org (2019 Revision)

land to rainforest or woodlands has the potential to significantly counter climate change and biodiversity loss through carbon sequestration will require increased yields to free up land. Soil carbon storage has also been suggested as a land management practice to increase carbon sequestration and prevent the release of CO_2 from agricultural soil by using cover crops, reducing tillage and erosion control. Agricultural malpractice has also been linked to a 33% decline in the amount of arable land over the last 40 years due to intensive farming resulting in soil erosion and eutrophication (Benke and Tomkins, 2017). Vertical farming has been suggested as a solution to improving food production amid climate change as its estimate to use 70-95% less water and ~90% less land than conventional methods of farming. However,

vertical farming is unlikely to replace staple crops any time soon due to the prohibitive initial cost of building a vertical farm, hydroponics equipment required and to be maintained to provide the food that would be needed to feed the growing population.

Food insecurity has previously been described by the Food and Agriculture Organisation (FAO) as 'a situation that exists when people lack secure access to sufficient amounts of safe and nutritious food for normal growth and development and an active and healthy lifestyle'. Logistical and economic disruption to the food system due to the Covid-19 pandemic has significantly affected those in developing countries closest to the breadline (The World Bank, 2020) but pales in comparison to the disruption to the effects of climate change. In 2020/2021 Ethiopia has been affected by the worst locust infestation of 25 years, destroying the livelihoods of thousands of farmers while 20 million acres of Australian land has been burned with an estimated loss of 100,000 livestock and an innumerable number of crops. This trend is set to go on until we develop effective solutions to climate change and how it will affect food security.

1.2. Role of the grasses in food security

Grasses play a major economic and dietary role with cereals such as *Triticum aestivum* (Wheat), *Hordeum vulgare* (Barley), *Oryza sativa* (Rice), *Saccharum officinarum* (Sugarcane), *Sorghum bicolor* (Sorghum) and *Zea mays* (Maize) providing over 51% of our total calorific intake through direct consumption, excluding domestic animals that are raised partially or wholly on cereals, while other species like *Bambusa vulgarisas* (bamboo) and *Panicum virgatum* (switchgrass) contribute to the forestry and biofuel sectors (FAOSTAT, 2018). Grasses make up to 70% of all crops grown

and will need drastic increases in annual yields to feed a growing population while maintaining nutritional value to tackle food insecurity in the 21st century (FAOSTAT 2018). Climate change poses a particular challenge to cereal production with a 1°C increase in temperature during wheat cultivation projected to reduce yields by 3-10% (Figure 1.2) (Yu et al., 2009). Cultivating less resource intensive and drought tolerant cereal crops like millet might be a long-term solution to increasing yields in a changing climate with less land. Recently it has been shown that millet cultivation also releases fewer greenhouse gas emissions (897 kg C ha^{-1}) than conventional cereal crops such as rice (956 kg C ha^{-1}) and wheat (1042 kg C ha^{-1}) (Jain et al., 2016). The widespread adaptability, extensive germplasm, and culinary integration of the big three (maize, rice, and wheat) however, means that these crops will be a staple of human civilization for the foreseeable future, particularly amidst climate change.

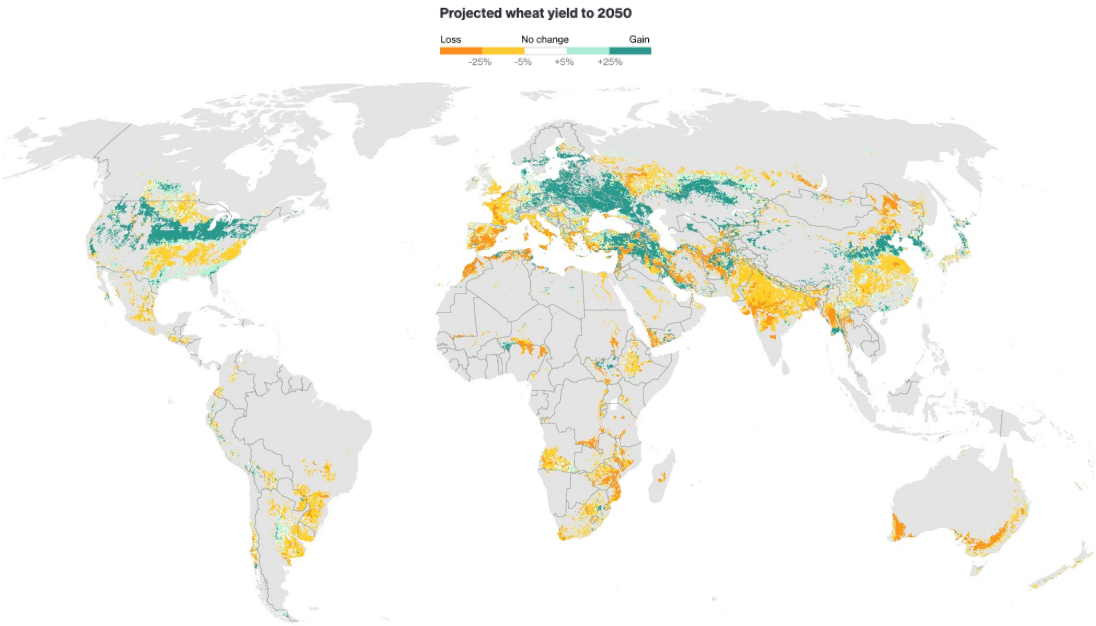


Figure 1.2. Projected wheat yields to 2050. Crop Environment Resource Synthesis wheat crop model based on past climate data and Hadley Centre Global Environment Model version 2 projections for 2050, assuming fertilizer, water management are same in both periods. Image taken from International Food Policy Research Institute (IFPRI).

1.3 Phylogeny and Characterisation of the Grasses

The grasses (Poaceae) are a family of monocotyledonous flowering plants that are distinguished by their ecological dominance with over 12,000 species covering between 20-40% of the earth's land mass (Clayton and Renvoize, 1986; Watson and Dallwitz, 1999; Shantz 1954). Major evolutionary changes in grass physiology can be deduced by comparing them to other monocotyledonous outgroups such as lilies, orchids, palms, and pineapples. Members of these outgroups show relatively undifferentiated embryos with seed maturation beginning after the formation of the shoot apical meristem (SAM). The cotyledon, leaves, root apical meristem (RAM) and vasculature typically differentiate after abscission, however in the grasses this development is accelerated relative to its outgroups (Kellogg, 2000).

Fruit structure also differs significantly between the grasses and their outgroups with most monocotyledons fusing three carpels to produce an ovary where each carpel has one locule and one ovule (Kellogg and Linder, 1995). Two of the three carpels abort during the formation of the grass ovary. This distinctive fruit of the grasses, known as the grain, is fully formed when the inner ovary wall fuses to the outer integument (Figure 1.3) (Dahlgren et al., 1985). The grasses can be divided into two major phylogenetic clades; the BEP (Bambusoideae, Ehrhartoideae (formerly Oryzoideae) and Pooideae) clade and the PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae) clade (Figure 1.4).

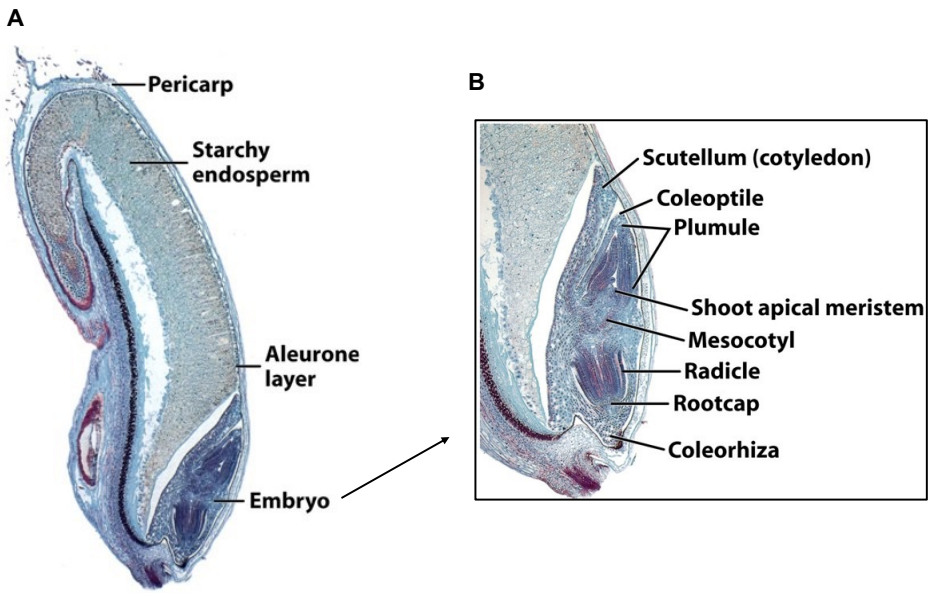


Figure 1.3. Grass grain and embryo anatomy. (A) Maize grain stained with toluidine blue to highlight areas of actively dividing cells, highlighting main anatomical structures of the grain. (**B)** Maize embryo from **A** magnified to highlight more intricate structures of the embryo. Images adapted from Fig 22-8b *Biology of Plants, Seventh Edition* 2005.

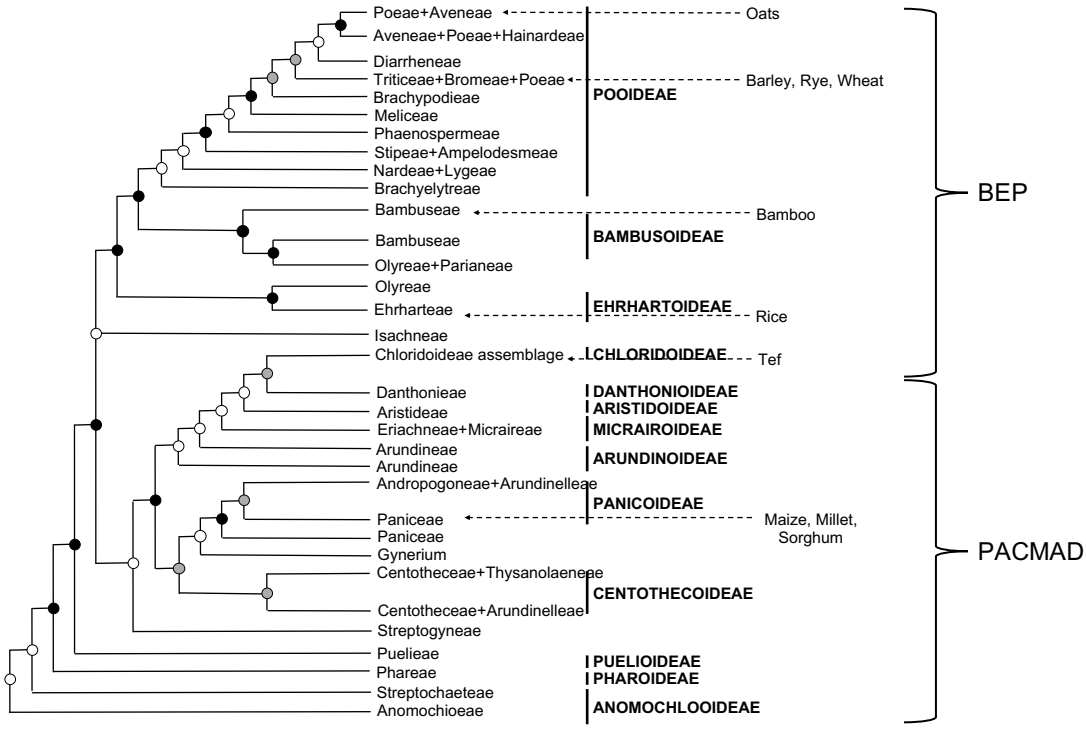


Figure 1.4. Phylogeny of the grasses. Phylogenetic tree showing sub-familial and inter-tribal relationships. White nodes indicate weak support, grey nodes indicate moderate support, and black nodes indicate strong support. Tree adapted from Bouchenak-Khelladi et al., 2008. Tribes with important cereal crops are highlighted with arrows and BEP and PACMAD clades are highlighted with brackets.

One distinct characteristic separating these two phylogenetic groups is the evolution of C4 photosynthesis. It is estimated that C4 photosynthesis has evolved between 4-10 times in the PACMAD clade, while all members in the BEP clade have been shown to use C3 photosynthesis only (Kellogg, 2000; Edwards and Smith, 2010). This carbon fixation process provides an evolutionary advantage compared to conventional C3 photosynthesis by concentrating CO_2 at the site of carboxylation in the modified bundle sheath which provides more tolerance to drought, high temperatures and carbon or nitrogen limitation (Sage and Monsor, 1999). The C4 Rice Centre has been attempting to improve photosynthetic efficiency in C3 rice (*Oryza sativa L.*) for many years by engineering a partial biochemical pathway for C4 photosynthesis pathway (Lin et al., 2020). This has shown some promising developments and may prove to be an effective breeding strategy for creating varieties of rice that are more drought and heat tolerant and require less nutrient inputs.

Distinct patterns of vegetative and inflorescence architecture aren't as clear between these clades due to the large variation and species number inherent in the grass family. Early domestication of shoot architecture in both clades involved the selection one main stem in PACMAD cereals like maize, sorghum and sugarcane or fewer basal tillers in modern bread wheat compared to its wild progenitor *Aegilops tauschii* (Doebley et al., 1997). However, during the third agricultural revolution, members of the BEP clade such as rice and wheat were selected for higher tiller number and semi-dwarf phenotypes to improve yield (Khush et al., 2001). This represents a distinct difference in modern breeding programs between these clades with increased tiller number being a favourable agronomic traits in BEP clade breeding programs with

notable exceptions like uniculm barley in boreal climates where growth seasons are shorter. (Mäkelä and Muurinen, 2012),

1.4. Wheat domestication

Plant domestication can be defined as an evolutionary process that occurs through artificial selection for beneficial agronomic traits such as vegetative or inflorescence architecture compared to their wild progenitors. *T. aestivum* is an allohexaploid with a large, complex, and repetitive genome of around 17Gbp. The first cultivation of wheat happened around the Neolithic revolution (~10,000 years ago) in the south-east of Turkey with *T. monococcum subsp. boeoticum* (einkorn - AA) and *T. turgidum subsp. dicoccoides* (wild emmer - AABB) species (Heun et al., 1997; Nesbitt, 1998). Breeding land races from wild varieties involved the selection of mutations implicated in spike morphology traits such as *brittle rachis* (*Br*), the *Q* locus and *tenacious glume*

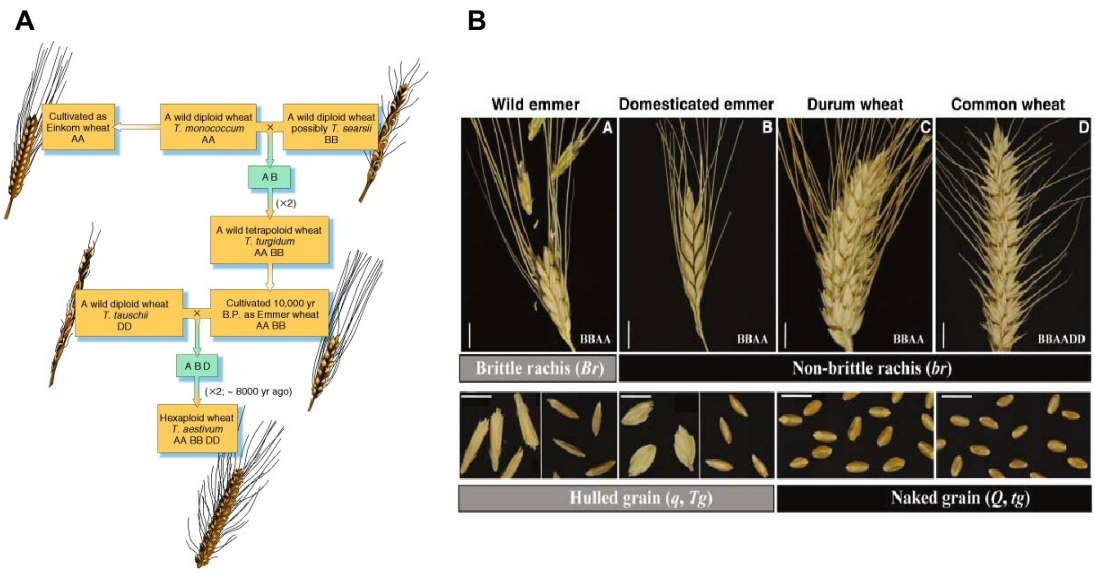


Figure 1.5. Bread wheat evolution and spike domestication. (A) Evolution of bread wheat from three wild wheat progenitors. **(B)** Phenotypes of domestication three domestication loci (*brittle rachis – br*, *Q locus* - *Q*, and *tenacious glume – tg*) in bread wheat progenitors and modern bread wheat. Image **A** taken from https://www.cerealsdb.uk.net. Image **B** taken from Dubcovsky and Dvorak 2007.

(*Tg*) that gave a naked free-threshing and non-brittle character to spikes and grains to allow for easier harvesting (Figure 1.5) (Jantasuriyarat et al., 2004; Simons et al., 2006; Nalam et al., 2006; Dubcovsky and Dvorak, 2007). Emmer wheat (AABB) is the result a hybridisation event between einkorn wheat (AA) and a non-cultivated extinct wild progenitor (BB). As cultivation spread to the near east ~9000 years ago a second hybridisation event occurred between domesticated emmer wheat (AABB) and a wild goat grass *Aegolips tauschii* (DD) to make the bread wheat *T. aestivum* (AABBDD) (Feldman et al., 2001).

Genetic diversity in land races has significantly contributed to wheat's adaptability and high yields in a multitude of environments worldwide with the International Center for Agricultural Research in the Dry Areas (ICARDA) holding around 880,000 seed samples in Svalbard despite recent disruption due to the Syrian civil war (Westengen et al., 2020). The 30,000 varieties of wheat that have been bred are grown on more land area (220.4 million hectares) than any other food crop in the world with an estimated 95% of this being bread wheat (AABBDD) and 5% being *Triticum turgidum subsp. durum* (AABB), another subspecies of tetraploid wheat primarily used in pasta production and suited to the drier climate of the Mediterranean (FAOSTAT, 2014).

1.5. Rice domestication

Oryza sativa L. is a diploid with a smaller less repetitive genome around 389Mbp that has undergone multiple selective sweeps in its domestication from its wild rice

progenitor *O. rufipogon.* Phylogenetic evidence has suggested that the two main subspecies of cultivated rice *japonica* and *indica* has arisen from a single domestication event from wild rice (Cheng et al., 2003; Londo et al., 2006). *Indica* rice tends to grow in tropical or subtropical lower latitude or altitude environments while *japonica* rice mainly grows in temperate higher latitude or altitude environments. Phylogenetic analysis of 446 *O. rufipogon* accessions and 1083 *O. sativa* varieties have shown that *japonica* was domesticated from the Or-I wild subpopulation near the middle area of the Pearl River in Southern China, while *indica* was derived from the Or-IIIa subpopulation and then introgressed with domestication alleles from japonica for key agronomic traits (Figure 1.6) (Huang et al., 2012). Highly differentiated loci between these two subspecies included *OsSOC1* and *Ghd7* (flowering time), *qCTS12* (cold tolerance) and *DPL2* (hybrid incompatibility) (Tadege et al., 2003; Andaya and Tai, 2006; Xue et al., 2008; Mizuta et al., 2010).

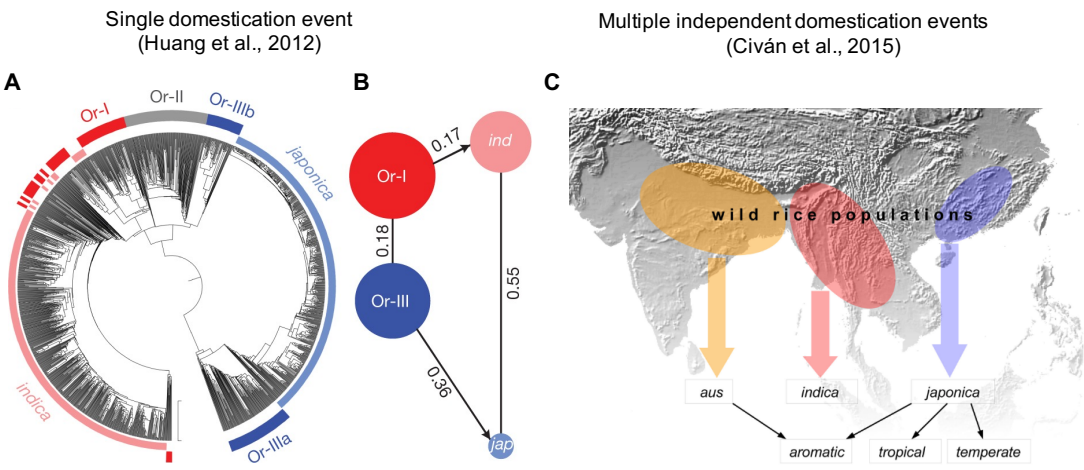


Figure 1.6. Two models for rice domestication. (A) Phylogenetic tree of 446 *O. rufipogon* accessions and 1,083 *O. sativa* varieties calculated from 8 million SNPs (*O. rufipogon* = Or-I, Or-II and Or-III and *O. sativa* = *indica* and *japonica*). (**B**) Illustration of genetic diversity and population differentiation in *O. rufipogon* and *O. sativa*. The size of the circles represents the level of genetic diversity of the groups, and the fixation index values between the groups are indicated. (**C**) Scheme of multiple independent domestication events hypothesis showing how three separate domestication events gave rise to the *O. sativa* subspecies *aus*, *indica* and *japonica*, and other modern rice varieties.

Some dispute this model of domestication, pointing out that archaeological evidence supports as many as three independent domestication origins of cultivated rice for the *japonica*, *indica* and *aus* subspecies (Civán et al., 2015). However, domestication sweeps identified by the latter study to support this model did not include well-characterised domestication genes like *An-1* (awn development number, grain size and grain number), *bh4* (grain colouration and seed shattering) and *OsC1* (anthocyanin accumulation) while the former included them suggesting technical flaws in the latter (Saitoh et al., 2004; Zhu et al., 2011; Luo et al., 2013). A summary of both theories can be found in Figure 1.6. Rice domestication events have also taken place in Africa and South America with *O. glabberima* being domesticated from *O. barthii* in West Africa around 3000 years ago and *Oryza sp.* (unnamed rice species) being domesticated from one or both endemic amazon species *O. glumaepatula* and *O. latifolia* (Wang et al., 2014; Hilbert et al., 2017). Rice today is grown predominantly in Asia with China and India contributing to 49% of global rice production, making it a staple crop for many, contributing to around 20% of the world's calorific intake (FAOSTAT, 2020). It is estimated there are 132,000 accessions stored at the International Rice Gene Bank for safeguarding biodiversity to ensure new varieties can bred to match food security challenges in a changing climate.

1.6. Tiller number – a grass domestication trait

Domestication of grasses for human agriculture involved the selection of major traits in vegetative architecture such as the pattern and timing of branching and inflorescence architecture like grain size. Evolutionarily important genes in tillering have been identified through crosses between the apically dominant maize (*Zea mays spp. mays*) and its branched wild ancestor teosinte (*Zea mays ssp. parviglumis*). Complementation

tests between these two species of maize identified the gene *TEOSINTE BRANCHED 1* (*TB1*) which restricts the outgrowth of AXMs and facilitated the domestication of teosinte through selection of locus with a distal enhancer that increases *TB1* expression in cultivated maize (Doebley et al., 1995; Doebley et al., 1997; Studer et al., 2011). Similarly, the *GRASSY TILLERS 1 (GT1)* maps to the same quantitive trait locus (QTL) as *TB1* and has been shown to promote lateral bud dormancy and suppress elongation of lateral branches through mediating the shade avoidance response (Whipple et al., 2011). This auxin-regulated process results in high apical growth to ensure the plant can rise above the surrounding canopy resulting in AXM repression.

Overexpression of *ZmTB1* in wheat has also been shown to suppress AXM outgrowth and tiller number suggesting that the genetic architecture of tillering is highly conserved amongst the grasses (Lewis et al., 2008). However, high expression alleles of *TaTB1* promote a supernumerary-spikelet trait known as paired spikelets through interaction *FLOWERING LOCUS T 1* (*FT1*), showing that *TaTB1* also has a role in spikelet formation (Dixon et al., 2018). This may suggest that while the AXM and IM share some genetic architecture, they show distinct regulatory pathways with some plant architecture transcription factors showing different roles depending on the meristem context. The evolution of apical dominance in maize through *TB1* upregulation exemplifies how members of the PACMAD clade have been domesticated for one main stem that focuses nutrients for the vigour of one inflorescence by suppressing the outgrowth of AXMs, but that these morphogenic regulators and their targets are highly evolutionarily conserved.

Domestication of the BEP clade also involved selection of less basal tillers compared to their wild ancestors. However, following the green revolution many varieties were introgressed with alleles for higher tiller number semi dwarf ideotypes to improve nitrogen response and yield which involved the selection of multiple loci controlling this complex process. Master regulators of tiller development in the pooid and erhartoid cereals have been identified and extensively studied to determine how they initiate and promote the outgrowth of AXMs. This makes tillering an important selection target for breeders to determine shoot architecture, plant biomass and grain production in the grasses (Sreenivaasulu and Schnurbusch, 2012). Branching pattern and timing is orchestrated by the actions of apical meristems and AXMs that can be characterised by three distinct developmental stages; AXM initiation, bud development and tiller outgrowth (Schmitz and Theres, 2005). Tillering has been widely studied to investigate the genes, phytohormones and environmental conditions that control the various stages of tiller development that regulate the balance between tiller number and vigour. Tillers share many characteristics with lateral branches in eudicots however, they differ in that they grow adventitious roots independently from the primary shoot. Each tiller has the potential to form a grain-bearing inflorescence but needs to be tempered with resource availability as unproductive tillers waste valuable nutrients.

High tiller number semi-dwarf varieties of wheat and rice increased crop productivity worldwide during the green revolution through introgression of the *Reduced height* mutant alleles *Rht-B1b* and *Rht-D1b*. Both *Rht-B1* and *Rht-D1* encode DELLA proteins which have been shown to inhibit GA-responsive growth by competitively binding to GA regulated genes. The *Rht-B1b* and *Rht-D1b* mutant alleles have a

truncated N-terminal domain which has been hypothesized to make them less sensitive to GA mediated ubiquitination and subsequent degradation relative to their wild type alleles (Peng et al., 1999). Introducing either allele to wheat cultivars has been shown to decrease overall plant height by ~20% and increases seed yield by 5-10% through GA insensitivity (Allan et al., 1959; Hoogendoorn et al., 1990; Flintham et al., 1997b). This phenotype is highly dependent on environmental conditions however, *Rht-B1b* show no increase in productive tiller numbers at lower plant spacings suggesting that this strategy has limitations and is not applicable to all breeding programs (Jobson et al., 2019).

Ideotype breeding for shorter plants with higher productive tiller number showed improved nitrogen response and harvest index overall in wheat (Khush, 2003). Tillers that grow late in the season may not form an inflorescence, lowering harvest index, showing that a balance is necessary between tiller number and vigour in crop breeding programs. Around 70% of modern wheat cultivars have introgressed at least one of the *Rht* mutant alleles making tiller number one of the most significant selection traits of recent agricultural history (Evans, 1999). *Rht* introgression must be done in context of wider breeding programme though, as it has been reported that *Rht* introgressed lines show reduced drought and heat tolerance due to decreased seedling emergence (Flintham et al., 1997b). Additionally, these high tiller number ideotypes can be prone to lodging, bending of stems reducing crop height uniformity for harvest, under high water and nitrogen conditions (Hedden, 2003).

1.7 Genetic architecture of tillering

Tillering is controlled by a suite of genes, hormones, and environmental conditions. The *LATERAL SUPRESSOR* (*LS*) gene, a member of the VHIID gene family, initially described in the dicot *Lycopersicon esculentum* (Tomato) is integral for the formation and maintenance of meristematic potential in axillary buds (Schumacher et al., 1999). Loss of function mutations of its rice ortholog *MONOCULM 1* (*MOC1*) results in only one main stem developing while overexpression has been shown to increase tiller number suggesting that this *MOC1* is an evolutionarily conserved AXM regulator between monocots and dicots (Li et al., 2003). The *TaMOC1-7A HapH* haplotype in hexaploid wheat, a favoured allele acquired during polyploidisation, suggests a novel role for *TaMOC1* with higher expression of *TaMOC1-7A* in this haplotype not affecting height (a correlate of higher tiller number) but increases spikelet number per spike (SNS) suggesting it controls inflorescence architecture as opposed to vegetative architecture suggesting that meristematic factors can influence both vegetative and inflorescence architecture patterns in cereals (Zhang et al., 2015).

Further evidence of this comes from the *WUSCHEL* (*WUS*) ortholog *TILLERS ABSENT1* (*TAB1*) which has been shown to localise with the indeterminacy marker *ORYZA SATIVA HOMEOBOX1* (*OSH1*) in the pre-meristem zone and precede AXM initiation in rice (Tanaka et al., 2015). *Tab1-1* shows arrested AXM formation and decreased *OSH1* expression which restricts pre-meristem zone development preventing tiller formation. While *TAB1* controls the vegetative architecture in rice *WUS* has mainly been implicated in controlling inflorescence architecture in arabidopsis through its expression in the organisation zone of the floral meristem. Characterising genes implicated in AXM formation and tillering in the *Triticeae*

family has been difficult due to their large complex genomes. Further to this, polyploid *Triticeae* genomes such as wheat introduce genetic redundancy making the mapping of causative loci and performing mutational analysis more difficult. Barely's diploid genome makes it an ideal model organism for the identification of meristematic genes in the Triticeae.

While *uniculme4* (*cul4*) was mapped in diploid barley from crosses between uniculm accessions and high tillering accessions, it's only been characterised recently with the development of advanced functional genomics in barley (Sato, 2020). *Cul4* has been shown to act at the leaf and axil boundary to control differentiation of AXMs controlling the outgrowth of axillary buds and ligule development, like its *A. thaliana* ortholog *BLADE ON PETIOLE 1* (*BOP*) that regulates AXM development and leaf morphogenesis (Tavakol et al., 2015). Further evidence of the pivotal role *CUL4* plays in AXM development with *cul4* developing leafy side shoots rather than elongated lateral culms. *MOC1* and *CUL4* are the best characterised regulators of tillering in the pooid and erhartoid cereals showing a complete absence of tillering in loss of function phenotypes (Babb and Muehlbauer, 2003). While meristematic regulators of AXM formation are necessary for tiller formation many other genes have been identified that influence the tiller number in grasses.

The rice APETALA2-domain transcription factor encoded by *NITROGEN MEDIATED TILLER GROWTH RESPONSE 5* (*NGR5*), has recently been shown to increase tiller number through interaction with the PRC2 complex to mediate H3K27me3 deposition on tiller branching inhibition genes showing how epigenetic

factors can specifically regulate processes implicated in tillering (Wu et al., 2020). Additionally, hormones such as Strigalactone (SL) and Gibberellin (GB) have also been extensively studied and identified as key regulators of lateral shoot branching. The SL pathway has yielded the most tillering related genes of any pathway with loss of function in biosynthesis and signalling genes like *DWARF14* (*D14*), *DWARF27* (*D27*), and *CAROTENOID CLEAVAGE DIOXYGENASE7* (*CCD7*) showing increased tiller number in rice (Arite et al., 2009; Lin et al., 2009; Booker et al., 2004). A complex set of interactions with the SL pathway is mediated through *MADS-box transcription factor 57* (*MADS57*) and *OsTB1* showing how this hormone interacts with well characterised regulators of apical dominance in grasses (Chen et al., 2018). GB on the other hand acts antagonistically in this pathway, repressing the expression of *OsTB1*, with overexpression of *OsGA2ox5/6* (catalyse bioactive GB) resulting in higher tiller numbers in rice (Lo et al., 2008).

Finally, the *tiller inhibition* (*tin*) gene was mapped to the short arm of chromosome 1A after showing reduced tillering, thicker stems, and large spikes. Early cessation of internode elongation of lateral culms in this genotype has been associated with hollow basal internodes, downregulation of sucrose (Suc) inducible genes, reduced Suc content in dormant *tin* buds and increased expression of *TB1-like* genes suggesting that *tin* regulates the transport of Suc to developing AXM buds (Kebrom et al., 2012). In summary, we can see that tillering is orchestrated through a complex network of interactions between genes regulating hormonal pathways, transport of essential nutrients and how morphogenic regulators involved in meristem identity and formation can affect tiller number. A summary of these tillering genes and their effects on plant architecture can be found in Figure 1.7.

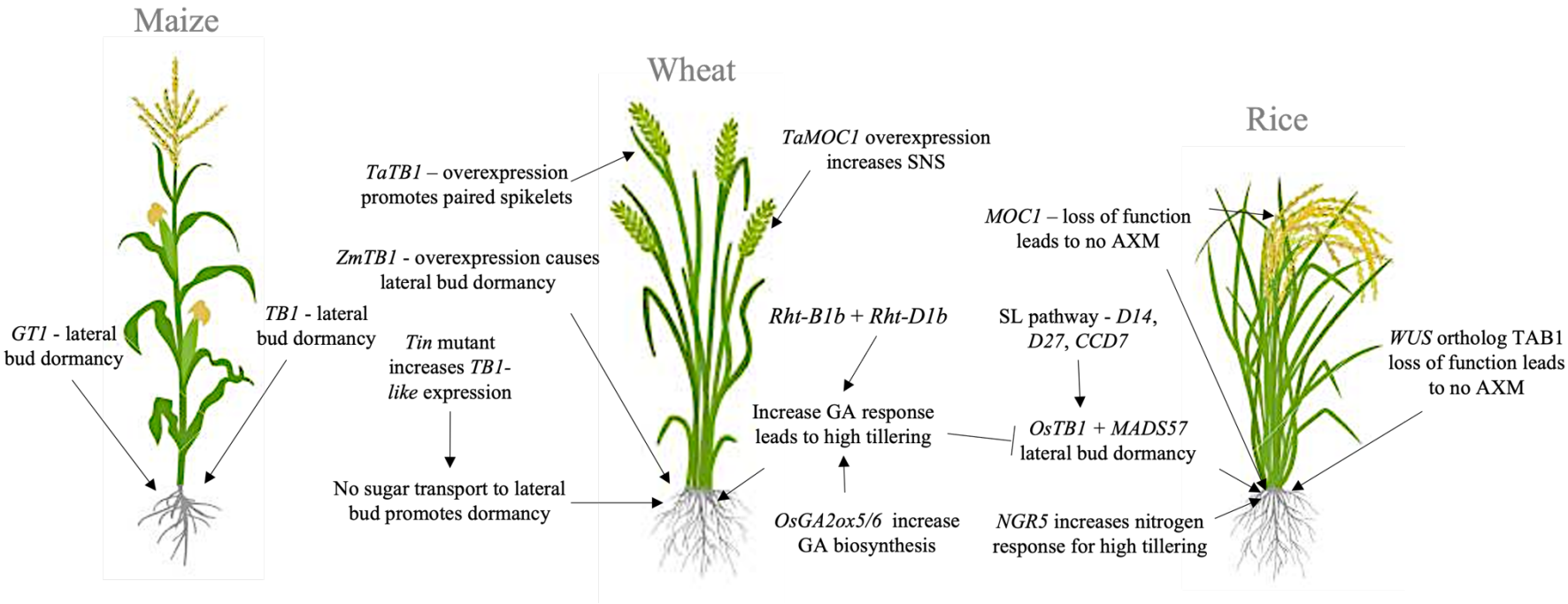


Figure 1.7. Genetics of tillering. Diagram illustrating the main genetic pathways implicated in tillering in maize, wheat and rice.

1.8. Wheat and rice inflorescence morphology

Grass inflorescence morphology (the spike) consists of smaller units called spikelets (consisting of florets and glumes) that occur in an alternating growth patten along the rachis. Inflorescence architecture has also been a major selection trait in human domestication of cereal crops and is determined by patterns of tiller architecture. Apical dominance in maize through *TB1* expression upregulates phytohormones that concentrate sugars and nutrients to the main inflorescence resulting in the formation of the characteristically large ear and large grain size (Dong et al. 2019). Inflorescence architecture in the BEP clade results in small inflorescence size due to spreading resources over a higher number of reproductive organs. Further evidence of this phenomenon comes from studying detillering where uniculm ideotypes show an increased number of grains per spikes, allowing manipulation and prediction of spike morphology in wheat (Guo et al., 2018). Meristematic determinacy in flowering plants is determined by intrinsic signals that come from resource availability and extrinsic environmental signals from photoperiod and temperature (Mouradov et al. 2002; Liu et al. 2009). This first occurs with the transformation of the shoot apical meristem (SAM) to the inflorescence meristem (IM) to signal a switch from vegetative to reproductive growth. This forms a stem cell niche of pluripotent cells capable of developing into the floral organs of the spike that are specified according to the ABCE model of flowering characterised through homeotic mutations in dicots (Krizek and Fletcher, 2005; Murai 2013).

The AP2 and MADS-box homeotic transcription factors have been implicated as master regulators of floral organ specification and development through observing

homeotic transformations that occur with deleterious mutations of these genes (Bowman et al., 1989). The *Q locus* encodes an AP2-like transcription factor that is negatively regulated by a microRNA *miR172*. Free-threshing wheat varieties have a mutation in the *Q locus*'s *miR172* binding site which confers higher AP2-like expression which is associated with a glume to floret transition consistent with the observed phenotype (Greenwood et al., 2017). It has also been shown that lower *miR172* activity is associated with increased expression of the *Q locus*'s AP2-like transcription factor and gives a free threshing character to wheat spikes further confirming this model for wheat spike domestication (Debernardi et al., 2017). Domestication of wheat spike morphology for free-threshing was one of the first major domestication events for modern wheat, initially being described by Nilsson-Ehle as a speltoid character (Nilsson-Ehle, 1917 and 1920). Wild wheat relatives such as *T. monococcum subsp. boeoticum* (einkorn), *T. turgidum subsp. dicoccoides* (emmer) and *T. aestivum subsp. spelta* (spelt) have toughened glumes which hold the grains tightly together and a semi-brittle rachis. Threshing these wheat relatives results in the spike breaking into smaller spikelets making it more laborious to collect grain and isn't amenable to agriculture. Selection of the *Q locus* in modern free-threshing wheat varieties results in fragile glumes and a non-brittle rachis allowing the spikes to break into individual grains which greatly speeds up grain collection (Dubcovsky and Dvorak, 2007).

Inflorescence development in rice is distinct from wheat as the rice panicle is indeterminate, giving rise to primary and secondary branches, with the main rachis eventually degenerates at senescence while the wheat spike is unbranched and determinate, terminating at a single spikelet (Figure 1.8). Domestication of rice

inflorescences followed similar processes through selection of loci involved in seed shattering and grain number. The *Black hull4 (Bh4)* loci encodes an amino-acid transporter and is associated with black colouration of the hull and seed shattering in wild rice, while *bh4* has a 22bp deletion in exon 3 that is associated with the straw-white hull colouration and non-shattering seeds found in cultivated rice varieties (Zhu et al., 2011). As highlighted before in wheat, seed dispersal is an important adaptive trait in wild varieties but prevents harvesting in cultivated varieties. Longer awns are also important in seed dispersal of wild varieties but are absent in domesticated rice. *An-1* encodes a basic helix-loop-helix protein that regulates cell division and has been suggested to positively regulate awn elongation (Luo et al., 2013). Interestingly, the *An-1* locus has been shown to have a significant reduction in nucleotide diversity in rice cultivars suggesting it as a domestication target as awn elongation have been shown to negative regulate grain number per panicle or spike (Luo et al., 2013; Rebetzke et al., 2016). The MADS-box identity type-C *AGAMOUS* (*AG*) has long been implicated in determining *arabidopsis* stamen and carpel development in the floral meristem through differentiating the floral meristem into a terminal state. Loss-of-function *agamous* plant are devoid of an androecium and gynoecium showing a homeotic transformation to undifferentiated petals and sepals (Bowman et al., 1989). The combined mutations of the MADS-box identity paralogs *VERNALISATION PROTEIN 1* (*VRN1*) and *FRUITFULL2* (*FUL2*) results in an elongated structure with higher spikelet number per spike (SNS) through eliminating the formation of the terminal spike, while the triple mutant *vrn1*, *ful2*, and *fruitfull3* (*ful3*) shows a conversion of lateral spikelets to vegetative tillers (Li et al., 2019).

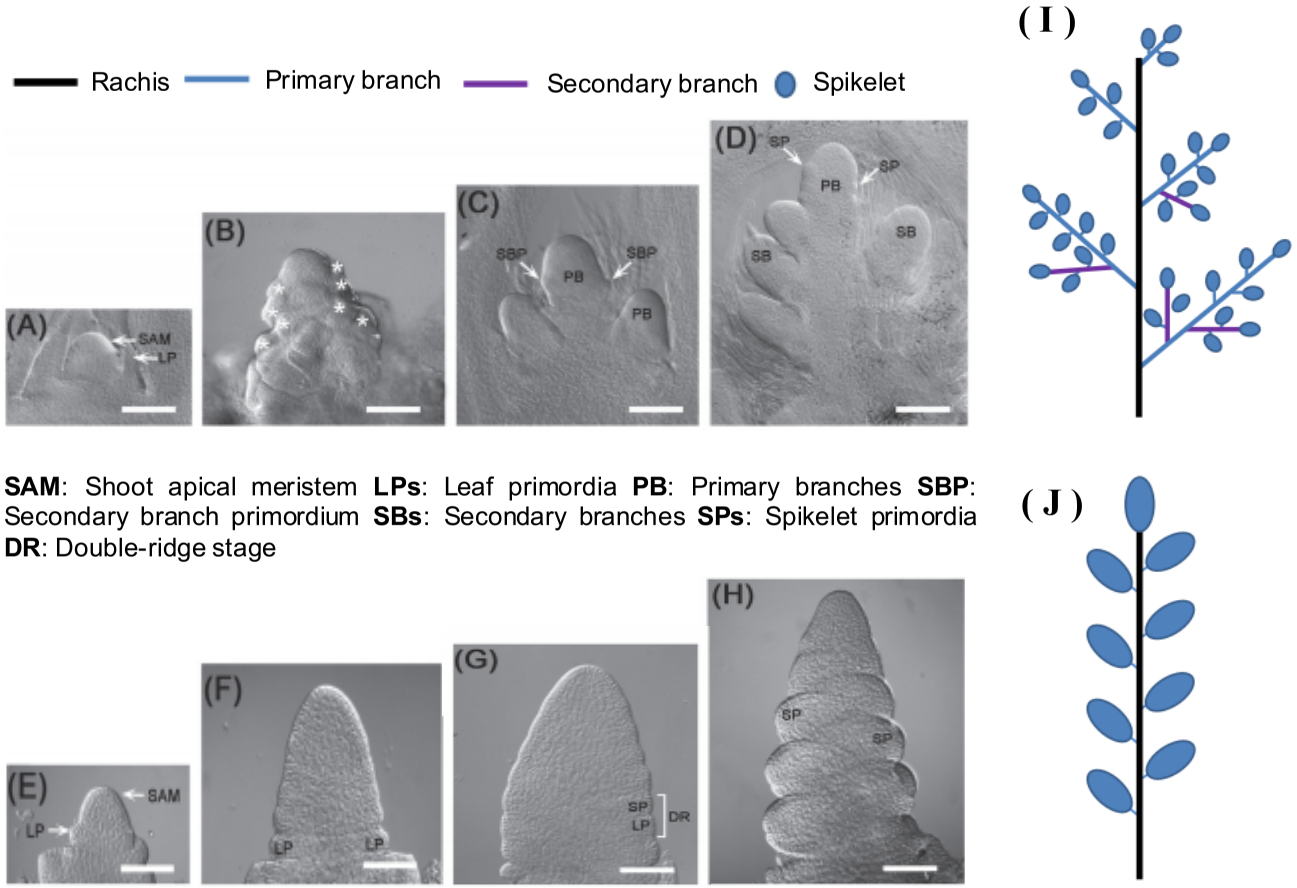


Figure 1.8. Inflorescence development in rice and wheat. Taken from Gao et al., 2019. **(A–D)** Early development of the rice panicle. **(A)** Shoot apical meristem (SAM) and leaf primordia (LPs). **(B)** Inflorescence meristem with primary branch (PB) primordia (stars). **(C)** Elongating PBs with secondary branch primordium (SBP). **(D)** Elongated PBs with secondary branches (SBs) and spikelet primordia (SPs). **(E–H)** Early development of wheat spike. **(E)** SAM and LPs. **(F)** Inflorescence meristem with LPs at single-ridge stage. **(G)** Inflorescence meristem with LPs and SPs at double-ridge (DR) stage. **(H)** Inflorescence meristem with enlarging SPs. Bars, 100 µm. **(I-J)** Branching Modes of Inflorescences in **I)** Rice , **J)**Wheat.

Consistent with this model of meristematic determinacy, overexpression of genes that upregulate these MADS-box transcription factors such as *FLOWERING TIME 1* (*FT1*) and *FLOWERING TIME 2* (*FT2*) and their upstream positive regulator *PHOTOPERIOD 1* (*PPD1*) resulting in lower SNS due to earlier flowering (Boden et al., 2015; Shaw et al., 2018; Liu et al., 2020). Flowering time and spike development have been shown to be inextricably linked through studies involving *VRN1*, *VERNALISATION PROTEIN 2* (VRN2) and *FLOWERING LOCUS T-like* (*FT*-like) genes (Dubcovsky et al., 2003; Liu et al., 2019). *FT-like* and *TERMINAL FLOWERING1-like* (TFL1-like) genes are members of the phosphatidylethanolamine binding protein (PEBP) family, both being derived from duplication and neofunctionalization from *MOTHER OF FT AND TFL1-like* (*MFT-like*) genes (Wickland and Hanzawa, 2015). *FT-like* genes encode florigens that form the florigen

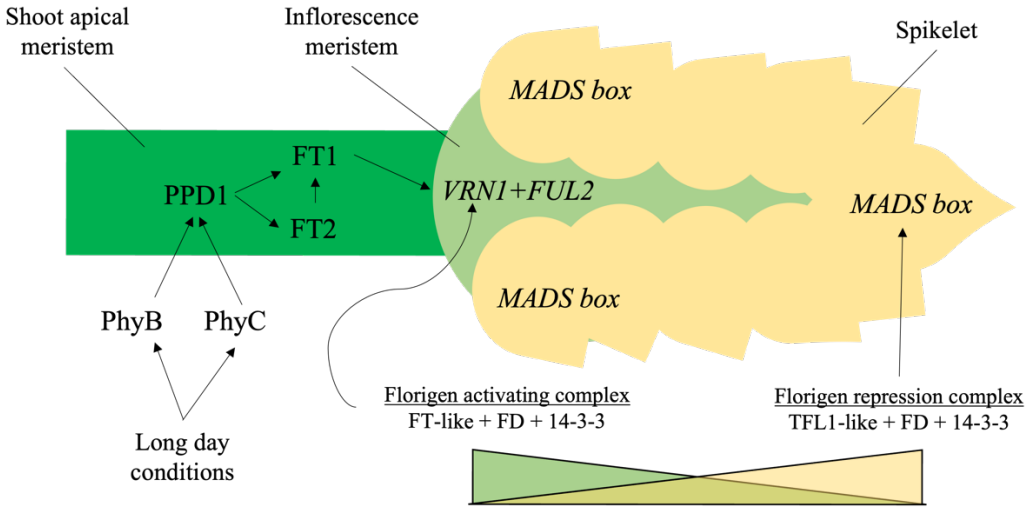


Figure 1.9. Genetics of flowering time. Diagram illustrating the main genetic pathways in flowering time and spike development in wheat.

activation complex in the SAM with bZIP transcription factor FD and 14-3-3 proteins, activating *VRN1* which initiating the transition to reproductive growth and the formation of the IM (Li and Dubcovsky, 2008). TFL1-like proteins compete for FD and 14-3-3 forming the floral repression complex which maintains IM meristematic activity, preventing the expression of MADS-box identity genes necessary for meristematic determinacy (Kaneko-Suzuki et al., 2018). Flowering time's effect on spike development highlights how environmental factors and endogenous signalling specifies plant architecture through meristematic formation and identity. A summary of these pathways can be found in Figure 1.9.

The F-box protein *UNUSUAL FLOWERING ORGANS (UFO)* has long been known to play a role in promoting floral meristem identity while suppressing development of bract-like structures in developing floral structures (Wilkinson and Haughn, 1995). Gain-of-function mutations in the *UFO* rice ortholog *ABERRANT PANICLE ORGANIZATION 1* (*APO1*) shows an increased number of spikelets by delaying the programmed shift to spikelet formation while ectopic expression showed increased meristem size by inducing higher cell proliferation (Muqaddasi et al., 2019). Loss-of-function mutations of *WAPO-A1* alleles identified in *T. monococcum* that show reduced expression have been shown to result in less spikelets per spike when introgressed into *T. dicoccoides* suggesting it may be involved in the suppression of the transition between the IM and terminal spikelet (Kuzay et al., 2019). This suggests that it may be possible to increase SNS in a range of grasses using other morphogenic regulators implicated in cell division through ectopic expression in the meristem.

1.9. RKDs - evolutionarily conserved morphogenic regulators

Germ cell formation in plants is mediated by reprogramming somatic cells in the reproductive organs of plants (Baroux et al., 2011). The RKD subfamily of the RWP-RK transcription factors has been implicated in this process across the plant kingdom, with *MpRKD* being shown to be necessary for germ cell formation in the basal plant *Marchantia polymorpha* and *RKD*s showing egg cell specific expression patterns in the angiosperms Arabidopsis and wheat (Koi et al., 2016). The RWP-RK transcription factor family was first described in *Lotus japonicus* and *Pisum sativum* (pea) in genetic studies of the *nodule inception* (*nin*) mutation and can be split into two subfamilies; the Nin-like proteins (NLPs) which are necessary for nodule formation, symbiosis, and other nitrogen related processes and the RKDs which are implicated in early embryo development (Schauser et al., 1999; Jeong et al., 2011). The *NIN* gene for which the NLPs have been named has recently been shown to regulate nodule formation in a dosage-dependent manner with downregulation and mutation of *NIN* resulting in aberrant nodule formation reducing nitrogen fixation in *Medicago truncatula* (Lin et al., 2018).

Consistent with *RKD*s playing a role in embryogenesis, *AtRKD4* has been shown to be required for zygotic elongation and the establishment of embryonic polarity in concert with the MAP kinase YDA (Waki et al., 2011). Mutations proximal to the RWP-RK domain and nuclear localisation signal of *AtRKD4* result in aberrant suspensor formation and early embryo arrest, suggesting that *RKD*s act as master regulators of early embryonic development through the activity of this DNA binding domain (Jeong et al. 2011). Induced overexpression of the evolutionarily conserved *AtRKD4* has been shown to prime root meristem and young leaf primordia cells for embryogenesis in *A.*

thaliana, showing that overexpression of zygotic factors can orchestrate reprogramming of cell identity and proliferation of callus tissue via somatic embryogenesis (Waki et al., 2011). When *AtRKD4* overexpression ceases, primed cells proliferate to form globular structures of cytoplasm rich cells and embryo-like structures that can regenerate to form new plantlets.

RKD induced somatic embryogenesis isn't a solely synthetic process however as natural occurrence of this process can be found in polyembryonic varieties of *Citrus sinensis.* Two alleles of the *CitRKD1* gene, *CitRKD1-mg1* and *CitRKD1-mg2*, in *C. sinensis* were found to control embryonic type (mono/polyembryony) due to the presence of a miniature inverted-repeat transposable element (MITE)-like insertion upstream of *CitRKD1* (Shimada et al., 2018). This upstream element facilitates ectopic expression of *CitRKD1* in flowers and young fruit, driving somatic embryogenesis and the formation of multiple embryos. Although expression patterns between female gametophytic tissue in grasses and dicots are conserved it doesn't hold that RKD induced somatic embryogenesis is maintained across these highly diverged clades. To identify RKD genes capable of driving somatic embryogenesis in cereals we can test those that show the highest homology to *AtRKD4* (Figure 1.10). This phylogeny that has been constructed from RWP-RK protein sequences across *A. thaliana* (14 proteins), the model cereal *Brachypodium distachyon* (16 proteins)*, the* unicellular algae *Chlamydomonas reinhardtii* (17 proteins), the model legume *Medicago truncatula* (8 proteins), *O. sativa ssp. japonica* (15 proteins) and the multicellular algae *Volvox carteri* (11 proteins) retrieved from the Plant Transcription Factor Database (PTFDB) shows that the grass genes *BdRKD1*, *BdRKD2* and *OsRKD3* show the closest homology to *AtRKD4* (Chardin et al., 2014). These four genes are all members of the

RKD (A) subfamily, characterised by the DNA-binding RWP-RK motif and another small motif close to the C-terminal end of the protein.

Our research group has previously found that inducible overexpression of *OsRKD3* rice leaf callus tissue is sufficient to maintain somatic embryo identity and drive callus growth suggesting that RKD's ability to influence meristematic identity and formation is conserved across dicots and grasses (Plant journal submission 2021). *TaRKD1* expression has been shown to be localised to the female gametophyte and significant efforts have been made to identify all *RWP-RK* genes in wheat to understand more about nitrogen responses and embryo development in this staple cereal crop (Kőszegi et al., 2011; Kumar et al., 2018). 18 *TaNLP* genes (nitrogen use efficiency subclade of RWP-RK) and 17 *TaRKD* genes have been identified by finding representative orthologues to the model grass species *Brachypodium distachyon* (shared >60%

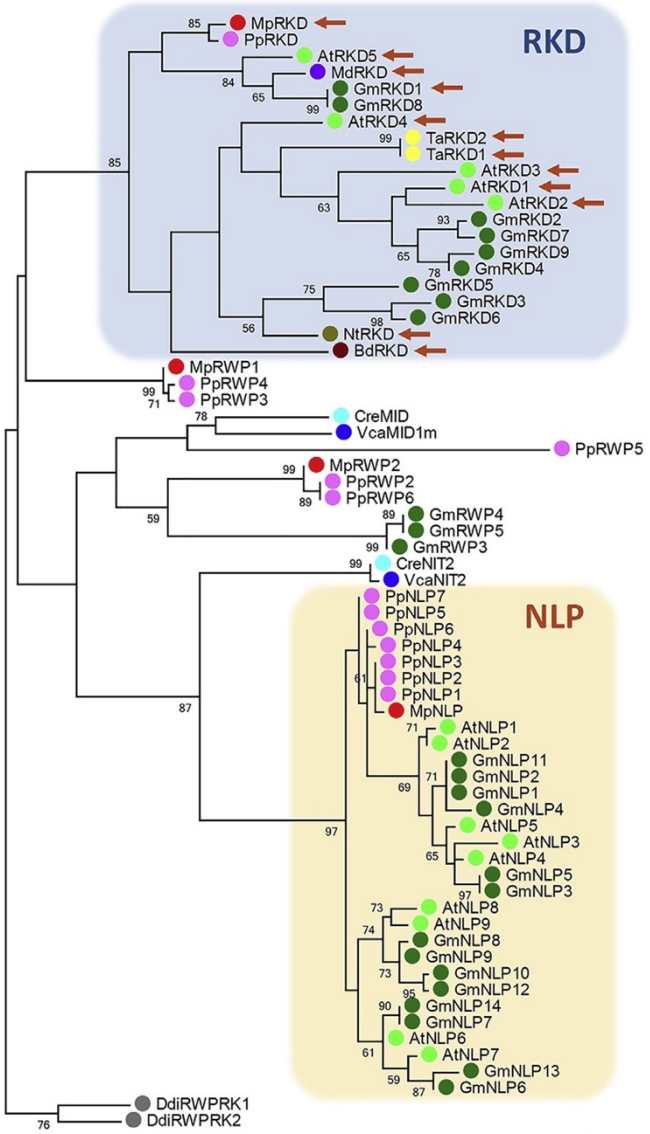


Figure 1.10. RWP-RK phylogeny from Koi et al., 2016. Maximum-likelihood tree of RWP-RK proteins constructed using the conserved RWP-RK and carboxy-terminal domains of representative plant species. Numbers at the branches indicate bo ues calculated from 1,000 replicates. Only values higher than 50 are shown. The scale bar indicates evolutionary distance expressed as the ratio of amino acid substitutions. Species are indicated by coloured circles and are abbreviated as follows: *At, Arabidopsis thaliana; Bd, Brachypodium distachyon; Cre, Chlamydomonas reinhardtii; Ddi, Dictyoste- lium discoideum; Gm, Glycine max; Md, Malus domestica; Mp, Marchantia polymorpha; Nt, Nicotiana tabacum; Pp, Physcomitrella patens; Ta, Triticum aestivum; Vca, Volvox carteri.* Arrows indicate RKD members with known expression in reproductive cells. The RKD clade is shaded in blue, whereas the NLP clade is shaded in yellow.

protein sequence identity) followed by identifying homeologues of these orthologous genes within hexaploid wheat (>90% shared protein sequence identity to the orthologous *TaRWP-RK* gene). The identification and characterisation of *TaRWP-RK* genes has focused mostly on the characterisation of *TaNLP* genes as nitrogen responses regulate the outcome to a key agricultural input. *TaRKD* genes have remained relatively unstudied for a decade and present themselves as an interesting subfamily of transcription factors for understanding embryo development and their potential roles in increasing crop yield.

1.10. Project aims

The development of advanced functional genomics and bioinformatics resources has given molecular geneticists the tools to investigate fundamental biological processes and yield determining traits in wheat. Cereal crop architecture remains one of the most important domestication traits in human history and enhancing our knowledge of how patterns of branching are determined will impact agriculture and society. The RWP-RK transcription factors have been extensively studied in model plant systems as morphogenic regulators capable of initiating the formation and maintenance of pluripotent stem cell niches. The main aim of this research project is to investigate whether RWP-RK transcription factors can be used as a novel strategy to increase tiller number by inducing the formation of axillary meristems in wheat and rice. We've also investigated whether RWP-RK transcription factors regulate meristem formation in wild wheat progenitors. By studying how plant architecture can be determined through the rate of meristem initiation and maintenance, we hope to influence future approaches to breeding in grasses.

To achieve this aim, two distinct yet complementary objectives are addressed.

1. To identify transgenic lines that show ectopic expression of RKD and display phenotypes relevant to yield determining traits in the grasses.
2. To define the molecular pathways of RKD in a variety of tissues using next generation sequencing in several grass species.

Chapter 2

Materials and methods

2.1. Materials

2.1.1. Plant materials

2.1.1.1. Fielder wheat material and growth conditions

Unless stated otherwise, all wheat plants grown were from the spring bread wheat cultivar 'Fielder'. Dried wheat seeds were stratified on moist Whatman paper at 4°C for 2 days and then transferred to damp M1 mix (85% fine peat, 15% grit, 2.7kg/m^3 osmocote). Seedling were germinated in a Growdome at the Phytobiology Facility (PBF) at University of Warwick under 16 h day/ 8 h night conditions at 22°C day/ 20°C night. Once spikes had fully senesced, they were dried at 37°C for 24 hours. Seeds were threshed and stored at 21°C.

2.1.1.2. Black rice material and growth conditions

Seeds of black rice cultivar *Cempo Ireng* were obtained from the Research Institute for Agriculture Technology (Yogyakarta, Indonesia). Mature seeds of black rice *Cempo Ireng* were dehusked and sterilized by immersion in 70% ethanol for 1 min followed by soaking for 30 min in 10% (v/v) of sodium hypochlorite (5.25 % active chlorine). Then they were washed three times with sterile water for 15 min in total. The sterilised seeds were planted on half-strength Murashige and Skoog (MS) medium containing 3% sucrose and

grown for one week at 28°C under constitutive light. The grown seedlings were transplanted into M2 soil (Levington Advance, UK) containing Osmocote slow-release fertilizer and the plants were grown at 28°C day/ 25°C night in a 16 h light/8 h dark cycle with a light intensity of 230 µE m^{-2} s^{-1}. Eight weeks after transplantation, short-day treatment (12 h light/ 12 h dark) was performed for two weeks to induce floral transition of the rice and the treated plants were grown to ripening stages under 16 h day/ 8 h night conditions.

2.1.1.3. Nipponbare rice material and growth conditions

Seeds of the cultivar Nipponbare were obtained from NIAB (Cambridge, UK). Mature seeds of Nipponbare were dehusked and sterilized by immersion in 70% ethanol for 1 min followed by soaking for 30 min in 10% (v/v) of sodium hypochlorite (5.25 % active chlorine). Then they were washed three times with sterile water for 15 min in total. The sterilised seeds were planted on half-strength Murashige and Skoog (MS) medium containing 3% sucrose and grown for one week at 28°C under constitutive light. The grown seedlings were transplanted into M2 soil (Levington Advance, UK) containing Osmocote slow-release fertilizer and the plants were grown at 28°C day/ 25°C night in a 16 h light/8 h dark cycle with a light intensity of 230 µE m-2 s-1. Eight weeks after transplantation, short-day treatment (12 h light/ 12 h dark) was performed for two weeks to induce floral transition of the rice and the treated plants were grown to ripening stages under 16 h day/ 8 h night conditions.

2.1.2. Plant growth media

2.1.2.1. Hydroponics solution medium (Broughton and Dilworth, 1970)

Hydroponics solutions were prepared by weighing and mixing anhydrous forms of the following reagents in dH_2O with a magnetic stirrer under standard conditions according to Broughton and Dilworth, 1970. To make the final hydroponics solution we mixed 500ml solution 1, 1000ml solution 2, 1000ml solution 3, 1000ml solution 4 and 100ml N solution to a final volume of 3600ml.

Table 2.1. Hydroponics solutions.

Solution	Chemical stock concentration	Final concentration
Solution 1	$CaCl_2$*$2H_2O$ 294.1g/l	1mM
Solution 2	KH_2PO_4 136.1g/l	1mM (was 0.5mM)
Solution 3	FeNAEDTA	75μM (was 11μM)
	$MgSO_4$*$7H_2O$ 123.3 g/l	1mM (was 0.25mM)
	K_2SO_4 87g/l	0.25mM
	$MnSO_4$*H_2O 0.338 g/l	6μM (was 1μM)
Solution 4	H_3BO_3 0.247g/l	20μM (was 2μM)
	$ZnSO_4$*$7H_2O$ 0.288g/l	1μM (was 0.5μM)
	$CuSO_4$*$5H_2O$ 0.10g/l	0.5μM (was 0.2μM)
	$CoSO_4$*$7H_2O$ (C046) 0.056g/l	0.05μM
	Na_2MoO_4*$2H_2O$ 0.048g/l	0.1μM
N solution	NH_3NO_3	15mM (was 30mM)

2.1.2.2. Somatic embryogenesis media

0.21% w:v Murashige and Skoog media (Merck) was supplemented with 100μM 2,4-Dichlorophenoxyacetic acid, 1% w/v Sucrose (Phyto X), 40μM estradiol and adjusted to pH 5.7 with 1 M KOH, and 0.7% w:v phytoagar (Duchefa Biochemie, Netherlands).

2.1.3. Generation of *indTaRKD1-7D* Fielder transgenic lines

For the generation of an inducible-*TaRKD1-7D* vector, we designed a codon-optimized full length CDS of *TaRKD1-7D* (*synTaRKD1-7D*) and the designed sequence was synthesized by IDT. The *synTaRKD1-7D* was introduced into a pDONR207 vector (Invitrogen, USA) via BP recombination and subcloned in pERV-BASTA or pERV-HYG vector (Park et al., 2012) by LR recombination. The pERV-BASTA or pERV-HYG vector, thereafter, named indTaRKD1-7D, was fully sequenced and the construct was introduced into Agrobacterium tumefaciens EHA105. Immature embryos from Fielder plants grown at NIAB Cambridge were collected 12-18 days after anthesis and transformed using agrobacterium transformed with the construct. Transformants were screened for the insert by plating them on media with the herbicides Bialaphos or Hygromycin. Somatic embryogenesis was induced in transformed immature embryos to form calli for regeneration. Once calli were an appropriate size they were transferred to root inducing and then shoot inducing media for regeneration of transformed plantlets. T1 regenerated plants were screened for low-copy number using DNA profiling by NIAB Cambridge and then transferred to PBF for propagation**.**

2.1.4. Genotyping for *indTaRKD1-7D* Fielder calli and plants

Genomic DNA isolation of putative transformant calli was performed using a CTAB protocol (Doyle and Doyle, 1990). PCR was used to detect the inserted gene in the genome with specific primer synTaRKD1-7D-Fwd and synTaRKD1-7D-Rev, as well as to detect HPT with a pair of primer depending on the construct, BASTA-Fwd and BASTA-Rev or the pair Hygromycin-Fwd

and Hygromycin-Rev (Table 2.2.). PCR cycling conditions included denaturing at 94°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 60s. PCR product was separated by electrophoresis on 2% agarose gels in 1x TBE buffer.

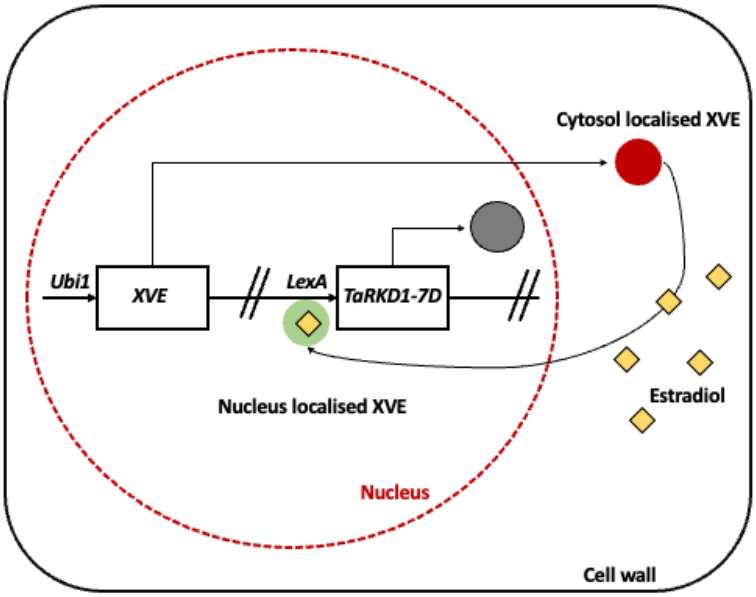


Figure 2.1. indTaRKD1-7D expression construct. Diagrammatic representation of chemical induction of *TaRKD1-7D* expression through XVE translocation into the nucleus. The chimeric *LexA* domain binding and estrogen receptor based-transactivator XVE is constitutively expressed under the *ubiquitin 1* (*Ubi1*) promoter and localised in the cytosol. Estradiol bound XVE is translocated into the nucleus where it binds to the *LexA* promoter activating expression of TaRKD1-7D.

2.1.5. Generation of *indOsRKD3 Cempo Ireng* transgenic lines

For generation of inducible-OsRKD3 vector, we designed a codon-optimized full length CDS of *OsRKD3* (*synOsRKD3*) and the designed sequence was synthesized by Integrated DNA Technology (IDT, Leuven, BE). The *synOsRKD3* was introduced into a pDONR207 vector (Invitrogen, USA) via BP recombination and subcloned in pTA7002 vector (Park et al., 2012) by LR recombination. The pTA7002-s*ynOsRKD3* vector, thereafter, named

indOsRKD3, was fully sequenced and the construct was introduced into Agrobacterium tumefaciens EHA105. Rice transformation was performed by the Agrobacterium-mediated co-cultivation method, as previously described (Lee et al., 1999, Toki et al., 2006, Susanto et al., 2020).

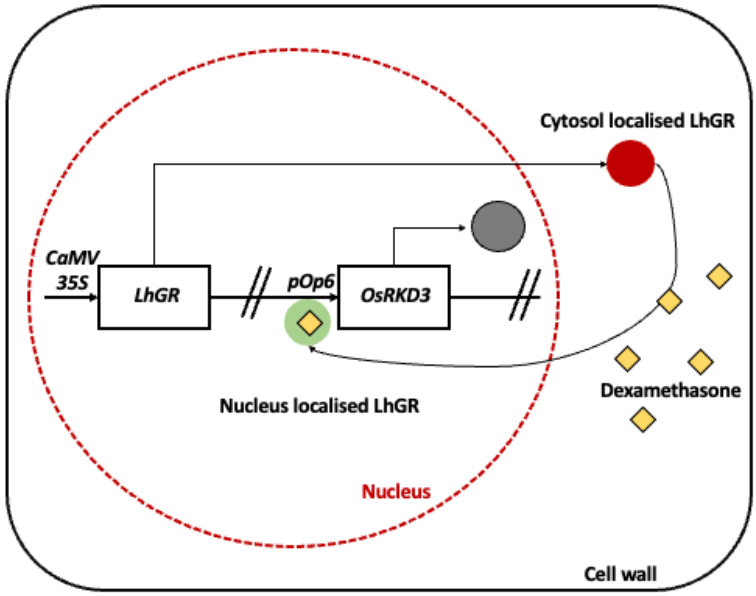


Figure 2.2. indOsRKD3 expression construct. Diagrammatic representation of chemical induction of *OsRKD3* expression through LhGR translocation into the nucleus. The chimeric *pOp6* domain binding and glucocorticoid receptor based-transactivator LhGR is constitutively expressed under the *cauliflower mosaic virus* (*CaMV 35S*) promoter and localised in the cytosol. Dexamethasone bound LhGR is translocated into the nucleus where it binds to the *pOp6* promoter activating expression of *OsRKD3*.

2.1.6. Genotyping for *indOsRKD3 Cempo Ireng* calli and plants

Genomic DNA isolation of putative transformant calli was performed using CTAB method (Doyle and Doyle, 1990). PCR was used to detect the inserted gene in the genome with specific primer *synOsRKD3*-Fwd and *synOsRKD3*- Rev, as well as to detect HPT with a pair of primer Hygromycin-Fwd and Hygromycin-Rev (Supplemental Table 2.1.). PCR cycling conditions included denaturing at 94°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 60 s. PCR product was separated by electrophoresis on 2% agarose gels in 1x TBE buffer.

2.1.7. Generation of inducible-AtRKD4 transgenic lines

For generation of inducible-AtRKD4 vector, we designed a codon-optimized full length CDS of AtRKD4 (synAtRKD4) and the designed sequence was synthesized by Integrated DNA Technology (IDT, Leuven, BE). The s synAtRKD4 was introduced into a pDONR207 vector (Invitrogen, USA) via BP recombination and subcloned in pERV1 vector (Park et al., 2012) by LR recombination. The pERV1-synAtRKD4 vector, thereafter, named indAtRKD4, was fully sequenced and the construct was introduced into Agrobacterium tumefaciens EHA105. The indAtRKD4 was introduced to Nipponbare, and rice transformation was performed by the Agrobacterium-mediated co-cultivation method, as previously described (Lee et al., 1999, Toki et al., 2006, Susanto et al., 2020).

2.1.8. Genotyping for *indAtRKD4* Nipponbare calli and plants

Genomic DNA isolation of putative transformant calli was performed using CTAB method (Doyle and Doyle, 1990). PCR was used to detect the inserted gene in the genome with specific primer *synOsRKD3*-Fwd and *synOsRKD3*-Rev, as well as to detect HPT with a pair of primer Hygromycin-Fwd and Hygromycin-Rev (Supplemental Table 2.1.). PCR cycling conditions included denaturing at 94℃ for 30 s, annealing at 55℃ for 30 s and extension at 72℃ for 60 s. PCR product was separated by electrophoresis using 2% agarose concentration in 1x TBE buffer.

2.1.9. Somatic embryogenesis induction of immature wheat embryos

Immature embryos were excised from the caryopsis 12-18 days after anthesis. Samples were sterilised in a laminar flow hood with NaOH for 5 min followed

by 70% EtOH for 15 min. Samples were washed with sterile dH_2O 5 times kept at 4°C for 60 min. Immature embryos were transferred to somatic embryogenesis medium and had the shoot apical meristem excised using a sterilised blade to prevent vegetative growth of the embryos. Immature embryos were transferred to new media weekly and incubated in the dark at 21°C.

2.1.10. Hydroponics experiment

Hoagland solution adapted from a protocol (Broughton and Dilworth, 1970) was made up to 30L using dH_2O and adjusted to a pH of 6.4 and checked using litmus paper (Broughton and Dilworth, 1970). This solution was split into two separate mixtures: one with 0.01% (w/v) Tween-20 and another with 0.01% (w/v) Tween-20 and 40μM estradiol. Plants growing in an 4:1 vermiculite:perlite mix were supplemented with the solution once per week for 8 weeks. Tiller number was measured and then vermiculite/perlite was cleaned from the roots to measure total plant biomass. Plants were then photographed using a Nikon D50 camera to measure root length and area using colour thresholding in image J.

2.2. Molecular biological methods

2.2.1. gDNA extraction

Leaf tissue from wheat or rice plants was collected in 96 well blocks with metal beads and then frozen in liquid nitrogen. Blocks were placed in a shaker at max speed for bursts of 15 seconds until all tissue was pulverised. 300 μl of extraction buffer was added followed by 40μl of 10% SDS to each well and

incubated for 15 minutes and incubated at 4°C. Blocks were centrifuged at max speed for 20 minutes and then the supernatant was discarded. Pellets were resuspended with potassium acetate by pipetting up and down until dissolved and stored on dry ice for 5 minutes. Blocks were then centrifuged at max speed for 20 minutes to pellet the mixture. Supernatant was then discarded and 400µl isopropanol was added and incubated at room temperature for 15 minutes to wash the pellet. Blocks were then centrifuged at max speed for 20 minutes and the supernatant was then quickly discarded. Blocks were washed twice by adding 400µl 70% ethanol, centrifuging at max speed for 20 minutes and discarding the supernatant. Blocks were air dried for 30 minutes to ensure no ethanol was present in the wells and the pellets were resuspended in 100µl TE buffer.

2.2.2. Polymerase chain reaction

The KAPA Taq PCR kit (KAPA Biosystems) was used to amplify DNA fragments from plant material (gDNA/cDNA template). 3µl of template DNA was added to a PCR reaction mixture (template, *1x buffer*, 100µM DNTPs, *500nM* forward primer, *500nM* reverse primer, 0.05% *KAPA Taq polymerase*) and made up to a final volume of 20µl with dH_2O. Thermocycler programs were chosen depending on the amplicon size and primer annealing temperature. PCR products were analysed by gel electrophoresis.

2.2.3. Agarose gel electrophoresis

To analyse PCR products, samples were separated according to size via agarose gel electrophoresis. Gels were made using TAE buffer (*40mM Tris base*, *20mM acetic acid* and *1mM EDTA*) supplemented with 1% w:v agarose

(Merck company) and stained with Diamond™ Nucleic Acid Dye (Promega). Samples were loaded in 6x loading dye and 1kb DNA ladder plus (New England Biolabs) was used as a molecular marker. Gels were run at 120V until sufficient separation had occurred and were imaged using a G:BOX F3 gel imager (Syngene).

2.2.4. Estradiol treatment and collection of the treated *indTaRKD1-7D* samples

Leaf segments or immature embryos of WT and pTA7002-*synTaRKD1-7D (indTaRKD1-7D*) plants were soaked in 40 μM of Estradiol solution containing 0.01% (w/v) Tween-20 and the soaked leaf segments were incubated in 28°C for 3 hours. As a mock control, leaf segments and immature embryos were also incubated in 0.02% (w/v) DMSO solution containing 0.01% Tween-20. After treatment of Estradiol and mock solutions, the rest of solution was removed from the samples as much as possible using a paper towel and the samples were put into a new Eppendorf tube. The collected samples were directly frozen in liquid nitrogen and the frozen samples were stored in -80°C before use.

2.2.5. DEX treatment and collection of the treated *indOsRKD3* samples

Leaf segments of WT and pTA7002-*synOsRKD3* (*indOsRKD3*) plants were soaked in 20 μM of DEX solution containing 0.01% (w/v) Tween-20 and the soaked leaf segments were incubated in 28°C for 3 hours. As a mock control, leaf segments were also incubated in 0.02% (w/v) DMSO solution containing 0.01% Tween-20. After treatment of DEX and mock solutions, the rest of solution was removed from the samples as much as possible using a paper

towel and the samples were put into a new Eppendorf tube. The collected samples were directly frozen in liquid nitrogen and the frozen samples were stored in -80°C before use.

2.2.6. RNA extraction of wheat leaf and immature embryo tissue

Leaf tissue from wheat seedling at Feekes development stage 1 (Feekes, 1941) was flash frozen and pulverized with metal beads in liquid N_2. Immature embryos were excised from the caryopsis 12-18 days after anthesis were flash frozen and were ground in Eppendorf tubes using an electronic drill with liquid N_2. 100mg of tissue was vortexed with 1ml TRIzol® Reagent (Invitrogen) and 200 µl of chloroform. Samples were centrifuged at maximum speed for 30 min at 4°C. The aqueous phase of the separation was incubated for 30 min with 100% isopropanol at -20°C. RNA was pelleted by centrifugation at maximum speed for 45 min at 4°C. Pellets were washed with 70% ethanol and dissolved in 20µl RNAase free H_2O.

2.2.7. RNA extraction of rice leaf tissue

Leaf tissue from rice plants early germination stage was flash frozen and pulverized with metal beads in liquid N_2. 100mg of tissue was vortexed with 1ml TRIzol® Reagent (Invitrogen) and 200 µl of chloroform. Samples were centrifuged at maximum speed for 30 min at 4°C. The aqueous phase of the separation was incubated for 30 min with 100% isopropanol at -20°C. RNA was pelleted by centrifugation at maximum speed for 45 min at 4°C. Pellets were washed with 70% ethanol and dissolved in 20µl RNAase free H_2O.

2.2.8. cDNA synthesis and RT-PCR

Total RNA was treated with Ambion® TURBO DNase kit (Life technologies, USA) as manufacturer's protocols. DNase-treated RNAs were used for cDNA synthesis using first strand cDNA synthesis kit (ThermoFisher, UK) according to manufacturer's protocols. Semi-quantitative RT-PCR was performed using templates as the synthesized cDNAs. *TaGAPDH* was used for data normalization. For Quantitative real-time RT-PCR (qRT-PCR) the optimal number of cycles was determined for each gene. PCR cycling conditions included denaturing at 95°C for 10 s, annealing at 57°C for 20 s and extension at 72°C for 30 s, using a Bio-rad qRT-PCR machine (Bio-Rad, UK). Changes in expression levels were calculated via the ΔΔCt method. To ensure primer specificity, qRT-PCR was done when the melting curve showed a single peak.

2.3. Histology

2.3.1. Preparation of RNA probes

To generate RNA probes for Histone H4, Histone H4 sequences of wheat from NCBI as previously reported (Drea et al., 2005) and selected highly conserved region of 190 bp. For in vitro transcription, T7 promoter and T3 promoter were added to each end of the selected histone H4 and we chemically synthesized the designed sequence (IDT, Leuven, BE). A histone H4 fragment was transcribed by T7 or T3 RNA polymerase (Roche) for 2 hours at 37°C, labelled by DIG-UTP (Roche) during in vitro transcription and treated by stop solution (0.1 Units·μL^{-1} DNase (Ambion, USA), 10 mM Tris-Cl, 10 mM $MgCl_2$, 50 mM NaCl, pH 7.5) for 30 minutes at 37°C. The synthesized RNA probes were precipitated in precipitation solution (0.8 M LiCl, 0.3 M NaOAc, 200 μg·mL-

[1] tRNA (Sigma, St. Louis, MO), 66.7% (v/v) ethanol, pH 5.2) overnight at -20°C, centrifuged for 10 minutes at 13000 rpm at 4°C, washed using 70% ethanol, centrifuged for 10 minutes at 13000 rpm at 4°C, discarded supernatant, air dried and dissolved in DEPC-treated distilled water. The dissolved RNA probes were hydrolysed in carbonate buffer (40 mM $NaHCO_3$, 60 mM Na_2HCO_3, pH 10.2) for 30 minutes at 60°C, precipitated in precipitation solution (10% (v/v) acetic acid, 90 mM NaOAc, 66.7% ethanol, pH 5.2) for 2 hours at -80°C, spun down for 10 minutes at 13000 rpm at 4°C, washed using 70% ethanol, spun down for 10 minutes at 13000 rpm at 4°C, discarded supernatant, air dried and dissolved in DEPC-treated distilled water.

2.3.2. RNA *in situ* hybridization in developing axils

To avoid contaminations of RNase, glassware was baked at 180ºC for 8 hours, plasticware was treated by 0.1 N NaOH (Sigma, St. Louis, MO) overnight at 37ºC and solutions were treated by 0.1% diethyl pyrocarbonate (DEPC, Sigma, St. Louis, MO). For RNA *in situ* hybridization in developing axils of WT and indTaRKD1-7D, the axils of the basal leaves of the main shoot were submerged in 4% paraformaldehyde (PFA, Agar Scientific, Essex, UK) and vacuum infiltrated three times for 10 minutes. The fixative was then replaced, and the sample incubated overnight at 4°C. After fixation, the samples were washed using phosphate buffered saline (PBS; 10 mM sodium phosphate and 150 mM NaCl, pH 7.4) with 0.1% DEPC, dehydrated via a graded ethanol series and embedded in Paraplast Plus (Sigma, St. Louis, MO). Sections were longitudinally cut at 10 µm and mounted on Superfrost Plus slides (Thermo Scientific, UK). The mounted samples were deparaffinized using Histoclear (National Diagnostics, Hull, UK), rehydrated through an ethanol series,

treated by Pronase (Merck, Darmstadt, Germany) for 10 minutes at 37ºC, fixed by 4% PFA for 10 minutes at room temperature, acetylated by acetic anhydride (Sigma, St. Louis, MO) in 0.1 M triethanolamine (Sigma, St. Louis, MO) for 10 minutes at room temperature and re-dehydrated by an ethanol series. To prepare RNA probe hybridization solution, subsequently, hybridization buffer (10 mM Tris-Cl, 0.3 M NaCl, 10 mM $NaPO_4$, 5 mM EDTA, 1.25 $mg \cdot mL^{-1}$ tRNA (Sigma, St. Louis, MO), 1.25x Denhardt's (Sigma, St. Louis, MO), 12.5% dextran sulphate (AP Biotech), 50% deionized formamide (Sigma, St. Louis, MO), pH 6.8) was mixed by vortexing, spun down, and left at room temperature. RNA probe solution was prepared with mixing between RNA probe and deionized formamide in 1:1 ratio, heat the mix solution for 2 minutes at 80ºC, spun down and cool on ice. The probe solution was mixed with the hybridization buffer in a 4:1 buffer: probe ratio, vortexed, spun down and left at room temperature. To prevent RNA probes from adhering to the coverslip, that was treated by Sigmacote (Sigma, St. Louis, MO) for 10 seconds at room temperature, washed by absolute ethanol and dried in fume hood before starting hybridization procedures. 50-100 µL of RNA probe hybridization solution was added on the slide, covered by Sigmacote-treated coverslip, incubated in humid chamber overnight at 45ºC. The slides were dipped in 2 x SSC buffer (150 mM NaCl, 15 mM NaOAc), agitated gently until the coverslips were fell off, moved to a new 2 x SSC buffer, incubated for 30 minutes at 45ºC, moved once more to a new 2 x SSC buffer, incubated for 90 minutes at 45ºC, washed twice using NTE buffer (10 mM Tric-Cl, 0.5 M NaCl, 1 mM EDTA, pH 7.5) for 5 minutes at 37ºC, treated by RNase buffer (20 $\mu g \cdot mL^{-1}$ RNase A (Sigma, St. Louis, MO) in NTE buffer) for 30 minutes

at 37ºC, washed three times using NTE buffer for 5 minutes at 37ºC, incubated in 0.5 x SSC buffer for 1 hour at 45ºC and washed using PBS for 5 minutes at room temperature. The washed slides were incubated for 5 minutes in DIG Buffer I (100 mM Tris-Cl, 150 mM NaCl, pH 7.5) at room temperature, blocked for 30 minutes in DIG Buffer II (0.5% (w/v) blocking reagent (Boehringer,) in DIG buffer I) at room temperature, transferred to DIG Buffer III (0.1% (w/v) BSA, 0.3% (v/v) Triton X-100 (Sigma, St. Louis, MO) in DIG Buffer I), incubated for 30 minutes in DIG Buffer III at room temperature, transferred to DIG Buffer IV (1:3000 anti-digoxigenin-AP (Roche) in DIG Buffer III), incubated for 90 minutes in DIG Buffer IV at room temperature, washed four times using DIG Buffer III for 20 minutes at room temperature, washed using DIG Buffer I for 5 minutes at room temperature, washed using DIG Buffer V (100 mM Tris-Cl, 100 mM NaCl, 50 mM $MgCl_2$, pH 9.5) for 5 minutes at room temperature and incubated for 6 to 36 hours in Buffer VI (75 $mg \cdot mL^{-1}$ nitroblue tetrazolium (NBT, Sigma, Pooler, UK), 50 $mg \cdot mL^{-1}$ 5-bromo-4-chloro-3-indolylphosphate (BCIP, Sigma, Pooler, UK), 24 $mg \cdot mL^{-1}$ Levamisole (Sigma, Pooler, UK) in DIG Buffer V) under dark conditions at room temperature, incubated in TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 7.5) for 10 minutes at room temperature, washed for 2 minutes in distilled water, dehydrated through an ethanol series, dried completely, mounted in Entellan (Merck, UK) and covered by coverslip.

2.3.3. Toluidine Blue staining of developing axils

For toluidine bule staining of developing axils in WT and indTaRKD1-7D, sections of developing axils were longitudinally cut at 10 µm and mounted on Superfrost Plus slides (Thermo Scientific, UK). The mounted samples were

deparaffinized using Histoclear (National Diagnostics, Hull, UK), rehydrated through an ethanol series, stained with Toluidine Blue solution for 2 minutes at room temperature, washed for 2 minutes in distilled water, dehydrated via an ethanol series, dried completely, mounted in Entellan (Merck, UK), and covered by coverslip.

2.3.4. Antisera preparation and immunolocalization

Polyclonal antisera were raised in rabbit against synthetic peptides for TaRKD1-7D (CFKENYKRRRAAASVN and ARWPHRKMKSLRSLI) (Eurogentec, Liege, BE). The raised antisera were affinity purified with peptide (CFKENYKRRRAAASVN) using a Sulpholink coupling gel system (Pierce, Rockford, IL). For immunolocalization of TaRKD1-7D in axillary meristem, sections were longitudinally cut at 10 μm and mounted on Superfrost Plus slides (Thermo Scientific, UK). The mounted samples were deparaffinized using Histoclear (National Diagnostics, Hull, UK), rehydrated through an ethanol series, and blocked in 0.1% bovine serum albumin (BSA) in phosphate buffered saline (PBS; 10 mM sodium phosphate and 150 mM NaCl, pH 7.4) for 30 mins at room temperature and incubated overnight with purified anti-TaRKD1-7D (diluted 1:20) antibody at 4℃. The immunoreactions were performed using goat anti-rabbit alkaline phosphatase-conjugated secondary antibody (Sigma, St. Louis, USA; diluted 1:100) and nitroblue tetrazolium/5-bromo-4-chloro-3-indolylphosphate (NBT/BCIP; Sigma, Pooler, UK) was used as detection substrate. The stained slides by NBT/BCIP were mounted in Entellan (Merck, UK).

2.4. Next generation sequencing library preparation

2.4.1. *indOsRKD3* and *indTaRKD1-7D* RNA extraction

Total RNAs were extracted from the leaves with Qiagen RNeasy Plant Mini kit (Qiagen, Germany) as manufacturer's protocols. RNAs for sequencing were extracted using Direct-zol RNA miniprep kit (ZYMO Research, Cambridge). The extracted RNAs were quantified using Nanodrop (ThermoFisher, UK) and qualities of those were also checked through Bioanalyzer 2100 (Agilent, UK).

2.4.2. *indOsRKD3* and *indTaRKD1-7D* RNA library preparation and sequencing

RNAs for sequencing were extracted using Direct-zol RNA miniprep kit (ZYMO Research, Cambridge). The extracted RNAs were quantified using Qubit HS kit (Invitrogen, UK) and qualities of those were also checked through Bioanalyzer 2100 (Agilent, UK). Total RNA libraries were prepared using the TruSeq RNA Sample Preparation Kit (Illumina, UK) and sequenced in single-end 150 base mode an Illumina NexSeq platform at the University of Warwick.

2.4.3. DAPseq analyses of Triticum aestivum genomic DNA with TaRKD1-7D

For construction of HaloTag fused with TaRKD1-7D, pDONR207-synTaRKD1-7D vector was cloned into pIX-HALO in vitro expression vector via LR recombination, transformed into the BL21DE3 codon plus expression strain (Stratagene, USA), protein was expressed as the manufacturer's manual.

Genomic DNA libraries were prepared according to Bartlett et al (2017), and genomic DNA was extracted from wheat leaves using Plant DNeasy kit (QIAGEN, Germany) as manufacturer's recommendations. 20 µg of genomic DNA was diluted in EB (10 mM Tris-HCl, pH 8.5) and sonicated to 200 bp fragments in a covaris S2 sonicator. DNA was purified using AmpureXP beads at a 2:1 bead:DNA ratio. Samples were then end repaired using the End-It kit (Lucigen) and cleaned using Qiaquick PCR purification (Qiagen) according to the manufacturer's recommendations. Purified samples were A-tailed using Klenow 3–5′exo- for 30 min at RT and then purified using Qiaquick PCR purification as described above. Purified samples were then ligated overnight with a truncated Illumina Y-adapter as described in Bartlett et al. Libraries were purified by bead cleaning using a 1:1 bead: DNA ratio, eluted from the beads in 30 µl of EB, and quantified with the Qubit HS fluorometric assay. A quantity of 20 µl of purified GST-ARF protein (5–20 µg) was diluted in 400 µl of 1X PBS containing 25 µl of washed MagneGST beads (Promega) for 1h at RT on a rotator to bind the protein to the beads. In addition to the GST-TaRKD1-7D samples, a negative control GST-HALO sample was performed using protein expressed in the TNT wheatgerm expression system (Promega). Beads were washed four times in 1X PBS + NP40 (0.005%) and resuspended in 100 µl of 1X PBS. One microgram of genomic DNA library generated from spikes was diluted to a final volume of 60 µl in 1X PBS and added to the protein bound beads. Protein bound beads and gDNA were rotated for 1h at RT. Beads were washed four times in 1X PBS + NP40, followed by two washes with 1X PBS. Beads were transferred to a new tube and DNA was recovered by resuspending in 25 µl

EB and boiling for 10 min at 98 °C. Eluted samples were enriched and tagged with dual indexed multiplexing barcodes by performing 20 cycles of PCR in a 50 µl reaction. Samples were pooled and sequenced on a NExtSeq500 with 75 bp single end reads.

2.4.4. Wheat indTaRKD1-7D ChIPseq library preparation and sequencing

Seeds were sent to the lab of Dr. Moussa Benhammed at L'Institut des Sciences des Plantes - Paris-Saclay (IPS2). Samples were grown at 16 h day/ 8 h night conditions at 22°C day/ 20°C night until Feekes seedling stage (Feekes, 1941). Hypocotyl tissue was dissected, RNA extracted and ChIP-seq libraries were prepared and sequenced according to Concia et al. 2020.

2.5. Next generation sequencing analysis

All scripts from NGS analysis can be found at https://github.com/meehanca .

2.5.1. Downloading raw reads from publicly available datasets

The fastq-dump program from the SRA-toolkit was used to download publicly available SRA files for subsequent analysis using the –-split-files and –-gzip parameters (Sherry et al., 2012).

2.5.2. Transcriptomic analysis of wheat RNAseq datasets

2.5.2.1. Pre-processing and pseudoalignment

Sequencing reads were pre-processed using fastp with default settings and the overrepresentation analysis parameter to remove reads with low quality score, irregular GC content, short length, and sequencing adapters present (Chen et

al., 2018). Trimmed reads were then quality checked using FastQC v0.11.5 to ensure trimming had been successful (Andrews, 2010). Trimmed reads were then pseudoaligned using Kallisto against an indexed *Triticum aestivum* IWGSC RefSeq v1.1 cDNA genome under default parameters (Bray et al., 2015; IWGSC et al., 2018).

2.5.2.2. Sample selection and calculating differentially expressed genes

Raw counts were imported into R and normalised before analysis using DESeq2 (Love et al. 2014; Team R, 2015). Replicate quality was verified using Principal Component Analysis (PCA) following a variance stabilised transformation of the samples and scatterplots of replicates were visualised using logarithmically transformed counts using DESeq2 functions and base R. Differentially expressed genes (DEGs) were calculated using DESeq2 with an FDR<0.05 and a lfcThreshold>0.5.

2.5.2.3. Visualisation of differentially expressed gene list

Heatmaps were plotted using logarithmically transformed normalised counts with the pheatmap package using genes (rows) to scale (Kolde and Kolde, 2015). MA-plots and volcano plots were made using DESeq2 functions and the EnhancedVolcano package (Blighe et al., 2020).

2.5.2.4. Gene set enrichment and GO term analysis

Gene set enrichment analysis was performed using the gage package using genes that fit the DESeq2 dispersion plot with a set size of 800, rank test set to true with a non-paired comparison (Luo et al., 2009). GO terms enrichment analysis was calculated for all DEG list using gProfiler with a Benjimini-

Hochberg correction set at a threshold of 0.05. GO term enrichment was visualised using a custom script with ggplot2 (Wickham 2011; Raudvere et al., 2019). All scripts can be found at https://github.com/meehanca/IE-AXM-RWP-RK-2021.

2.5.3. Transcriptomics analysis of rice RNAseq dataset

2.5.3.1. Pre-processing, alignment and counting

Sequencing reads were pre-processed using Trimmomatic v0.36 to remove reads with low quality score, irregular GC content, short length, and sequencing adapters present (Bolger et al., 2014). Trimmed reads were then quality checked using FastQC v0.11.5 and mapped to the *Oryza sativa ssp. japonica* genome (release MSU 7.0) using HISAT2 v2.1.0 (Ouyang et al., 2007; Andrews, 2010; Kim et al., 2019). Alignments were sorted using samtools, counted using LiBiNorm and then imported into R (Li et al., 2009; Dyer et al., 2019; Team R, 2015).

2.5.3.2. Sample selection and calculating differentially expressed genes

Raw counts were imported into R and normalised before analysis using DESeq2 (Love et al. 2014; Team R, 2015). Replicate quality was verified using principal component analysis (PCA) following a variance stabilised transformation of the samples and scatterplots of replicates were visualised using logarithmically transformed counts using DESeq2 functions and base R. differentially expressed genes (DEGs) were calculated using DESeq2 with an FDR<0.05 and a lfcThreshold>0.

2.5.3.3. Visualisation of differentially expressed gene list

Heatmaps were plotted using logarithmically transformed normalised counts with the pheatmap package using genes (rows) to scale (Kolde and Kolde, 2015). MA-plots and volcano plots were made using DESeq2 functions and the EnhancedVolcano package (Blighe et al., 2020).

2.5.3.4. Gene set enrichment and GO term analysis

GO terms enrichment analysis was calculated for all DEG list using gProfiler with a Benjimini-Hochberg correction set at a threshold of 0.05 (Raudvere et al., 2019). GO term enrichment was visualised using a custom script with ggplot2 (Wickham 2016). RKD-motif enrichment analysis was carried out on 3000bp upstream sequences using HOMER under default parameters (Heinz et al., 2010). Densigrams were plotted using ggplot2 (Wickham, 2011).

2.5.4. Wheat meristem coexpression network construction and analysis

2.5.4.1 Meristematic coexpression network table construction

The weighted gene coexpression network was constructed using transcript per million abundance (tpm) values over 0.5 from meristematic tissue samples taken from the JIC developmental timecourse (Xiang et al., 2019). WGCNA and GGM were used to calculate weighted edge values for the 53,000 genes that passed the filter (Langfelder and Horvath, 2008; Schaefer et al., 2015). Average values from both functions were taken and normalised against randomly generated edge values for the genes used to create the final coexpression table.

2.5.4.2. DEG specific network table filtering

Edge values that had at least one of the indTaRKD*1-7D* RNAseq DEGs (set at lfcThreshold = 0, p.adj = 0.05) and had a value over 0.7 were filtered from the final coexpression table due to a high number coexpressed nodes below this value.

2.5.4.3. Cytoscape visualisation of the log2 fold change network

The network was then visualised and annotated using Cytoscape 3.7.2 under an edge-weighted spring directed layout (Shannon et al., 2003). Nodes were annotated with a black border of 2 to indicate a DAPseq binding peak within 10kb of the upstream promoter, coloured according to the log_2fold change value, and the size was increased if the node was a member of a transcription factor family. This coexpression was then uploaded to the nDEX database under the name '*indTaRKD1-7D - logfoldchange'*.

2.5.4.4. Cytoscape visualisation of the GO term network

Clusters were automatically curated using the AutoAnnotate cytoscape app and GO terms enrichment analysis was calculated for all clusters list using gProfiler with a Benjimini-Hochberg correction set at a threshold of 0.05 (Kucera et al., 2016; Raudvere et al., 2019). Clusters with significant GO terms were coloured for a separate network and the most significant non-redundant term was used to name this cluster. Revigo was used with default parameters to remove redundant GO-terms and plotted using an R script provided through the site. This coexpression network was then uploaded to the nDEX database under the name '*indTaRKD1-7D-GO term'*.

2.5.5. Rice coexpression network construction and analysis

2.5.5.1 Coexpression network table construction

The weighted gene coexpression network table file was constructed using WGCNA, bc3net and GGM were used to calculate weighted edge values from several tissues of a rice transcriptomics developmental timecourse and uploaded to the Netminer github (https://github.com/czllab/NetMiner) (Yu et al., 2018).

2.5.5.2. DEG specific network table filtering

Edge values that had at least one of the *indOsRKD3* RNAseq DEGs (set at lfcThreshold = 0, p.adj = 0.05) and had a value over 0.15 were filtered from the final coexpression table due to a high number coexpressed nodes below this value, but lower than the wheat coexpression network. This cut off value allowed spatial separation of clusters for visualisations of individual relationships between nodes.

2.5.5.3. Cytoscape visualisation of the network

The network was then visualised and annotated using Cytoscape 3.7.2 under an edge-weighted spring directed layout (Shannon et al., 2003). Nodes were annotated with a black border of 2 to indicate a predicted RKD2 binding motif within 2kb of the upstream promoter, coloured according to the log2fold change value, and the size was increased if the node was a member of a transcription factor family. This coexpression was then uploaded to the nDEX database under the name '*indOsRKD3'*.

2.5.6. Motif enrichment analysis

2.5.6.1. MEME-ChIP analysis of wheat DAPseq peaks

For MEME-ChIP analysis the parameters set for any number of repetitions, minimum motif size 6, max motif size 15, palindromes only, non-randomised order, stop searching after 5 motifs had been identified by the dreme algorithm, and centrimo set to local central enrichment (McLeay and Bailey, 2010). Inputs were 50bp flanking regions the centre of overlapping peaks called by GEM and MACS2 callpeak (Zhang et al., 2008; Guo et al., 2012).

2.5.6.2. Predicted AtRKD2 binding sites in rice promoter sequences

HOMER's findMotif.pl script was used to find instances of the AtRKD2 motif promoter regions 3kb upstream of the gene TSS retrieved from the RAP-DB database (Heinz et al., 2010).

2.5.7. DAPseq analysis of TaRKD1-7D with wheat gDNA

2.5.7.1. Pre-processing and alignment

Sequencing reads were pre-processed using fastp with default settings and the overrepresentation analysis parameter to remove reads with low quality score, irregular GC content, short length, and sequencing adapters present (Chen et al., 2018). Trimmed reads were then quality checked using FastQC v0.11.5 (Andrews, 2010). Trimmed reads were mapped to the IWGSC RefSeq v1.1 reference genome (nuclear chromosomes only) using bowtie2 (Langmead, 2013; IWGSC et al., 2018).

2.5.7.2. Peak calling

Mapped reads were filtered for reads containing >MAPQ30 using samtools (samtools view –b –q 30) to restrict the number of reads mapping to multiple locations in the genome (Li et al., 2009). MAPQ filtered reads were used for all subsequent analysis. Peaks were called using GEM and MACS2 callpeak (Zhang et al., 2008; Guo et al., 2012). For MACS2 callpeak we set an effective genome size of 17e9, a minimum q value of 0.05, the no model parameter and an extsize of 200 with GST-HALO negative as a control. For GEM we used GST-HALO negative as the control sample for background subtraction and an FDR of 0.00001 (--q 5) (Guo et al., 2012).

2.5.7.3 DAPseq functional and enrichment analysis

Venn diagram of the number of overlapping peaks was visualised using the R package ggplot2 (Wickham, 2011). Peak coverage, and feature distribution and enrichment of overlapping peaks was analysed using ChIPAnnopeak (Zhu, 2013). GO term enrichment of genes within 10kb of peaks were calculated using gProfiler under standard parameters (Reimand et al., 2016). The DNA sequences of peaks were extracted using bedtools and analysed for motif enrichment using a command line installation of MEME-ChIP with the parameters set for any number of repetitions, minimum motif size 6, max motif size 15, palindromes only, non-randomised order, stop searching after 5 dreme-motifs, and for centrimo to look for local central enrichment (McLeay and Bailey, 2010; Quinlan and Hall, 2010). Normalised coverage proximal and across gene bodies was computed from bigwigs using deepTools (Ramirez et al., 2014).

2.5.8. ChIPseq analysis of TaRKD1-7D in wheat hypocotyl tissue

2.5.8.1. Pre-processing and alignment

Sequencing reads were pre-processed using fastp with default settings and the overrepresentation analysis parameter to remove reads with low quality score, irregular GC content, short length, and sequencing adapters present (Chen et al., 2018). Trimmed reads were then quality checked using FastQC v0.11.5 (Andrews, 2010). Trimmed reads were mapped to the IWGSC RefSeq v1.1 reference genome (nuclear chromosomes only) using bowtie2 (Langmead, 2013; IWGSC et al., 2018).

2.5.8.2. Peak calling

Mapped reads were filtered for reads containing >MAPQ30 using samtools (samtools view –b –q 30) to restrict the number of reads mapping to multiple locations in the genome (Li et al., 2009). MAPQ filtered reads were used for all subsequent analysis. Peaks were called using MACS2 callpeak (Zhang et al., 2008). For MACS2 callpeak we set an effective genome size of 17e9, a minimum q value of 0.05, the no model parameter and an extsize of 200.

2.5.8.3 ChIPseq functional and enrichment analysis

Venn diagram of the number of overlapping peaks was visualised using the R package ggplot2 (Wickham, 2011). Peak coverage, and feature distribution and enrichment of overlapping peaks was analysed using ChIPAnnopeak (Zhu, 2013). GO term enrichment of genes within 10kb of peaks were calculated using gProfiler under standard parameters (Reimand et al., 2016).

Normalised coverage proximal and across gene bodies was computed from bigwigs using deepTools (Ramirez et al., 2014).

2.5.9. Phylogenetic tree and protein alignments

RKD protein sequences were obtained from NCBI, TAIR, MSU TIGR and Ensemble database. The alignment of RKD proteins was computed using MAFFT in Jalview (Waterhouse et al., 2009). Phylogenetic tree was constructed using MEGA X (Kumar et al., 2016) using the minimum evolution method (Rzhetsky and Nei, 1992). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the Close-Neighbour-Interchange (CNI) algorithm (Nei and Kumar, 2000) at a search level of 1. The Neighbour-joining algorithm (Saitou and Nei, 1987) was used to generate the initial tree. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1486 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018; Stecher et al., 2020).

2.6. Phenotypic analysis

2.6.1 Phenotypic screening for higher tiller number transgenic lines

10 plants from 30 T2 lines were genotyped and grown under standard conditions, using an alpha lattice incomplete randomised design computed by the R package agricolae, to screen for T2 lines showing higher tiller number

phenotypes. Once half of the spikes for each plant had senesced in the plants of each line, productive tiller number was counted. Measurements were imported into R, and w statistically evaluated using appropriate methods based on the distribution, sample number and number of categories within the data before visualisation using ggstatsplots (Team R, 2015; Patil, 2021). The statistical defaults are designed to follow the best practices in both data visualisation and statistical reporting. Additional details can be found in the paper quoted above.

2.6.2 Phenotypic trial for tiller traits

2.6.2.1. Design and conditions

The high replication trial was arranged using a complete block design with the R package agricolae with 12cm row spacing in p40, p7, and p11 pots in PBF during Spring 2019 (de Mendiburu and de Mendiburu, 2019). Plants were grown for 100 days and the number of productive tillers, height, spikelet number and flag leaf length (from randomly selected leaves) were all counted once more than half of the spikes had fully senesced. Vegetative tiller counts were taken once every 10 days. Mechanical stress was applied by using a 0.7kg weight for 3 seconds, three times in 4-day intervals from the onset of Feekes stage 3 growth (Feekes, 1941).

2.6.2.2. Spike measurement

Spikes were photographed on a black background with a ruler for scale using a Nikon D50 under auto setting from a fixed height with a ruler. Photos were imported into ImageJ and spike length was measured manually after setting a scale based on the ruler using the line and measure tool. Rachis elongation was

estimated by measuring the length of the rachis and dividing this by the number of nodes recorded in ImageJ. Measurements were imported into R, and were both statistically evaluated using appropriate methods based on the distribution, sample number and number of categories within the data and visualised using ggstatsplots (Team R, 2015; Patil, 2021). The statistical defaults are designed to follow the best practices in both data visualisation and statistical reporting. Additional details can be found in Patil 2021.

2.6.2.3. Seed threshing and measurements

Once all plants had fully senesced, seeds were threshed, collected, counted, and weighed. Spikes were then threshed for grain and grain number was estimated by finding the volume of 100 seeds in a graduated cylinder and extrapolating this to the total grain number per plant. Measurements were imported into R, and were both statistically evaluating using appropriate methods based on the distribution, sample number and number of categories within the data and visualised using the ggstatsplots approach (Team R, 2015; Patil, 2021). The statistical defaults are designed to follow the best practices in both data visualisation and statistical reporting. Additional details can be found in Patil 2021.

2.6.3 Phenotypic trial for spike traits

2.6.3.1. Design and conditions

The trial was arranged using a complete randomised block design with the R package agricolae with 12cm row spacing in p11 pots and were surrounded by wild type plants to control for border effects in PBF during Spring 2020 (de Mendiburu and de Mendiburu, 2019). Anthesis was measured once per week.

2.6.3.2. Spike measurement

When plants were fully dried, spikes were taken and photographed using a Nikon D50 camera set with an ISO of 400 using the table method outlined in the WERecognizer manual (Genaev et al., 2019). Images were then analysed using the WERecognizer application with command line on a Windows OS. Measurements were imported into R, and were both statistically evaluating using appropriate methods based on the distribution, sample number and number of categories within the data and visualised using the ggstatsplots approach (Team R, 2015; Patil, 2021). The statistical defaults are designed to follow the best practices in both data visualisation and statistical reporting. Additional details can be found in Patil 2021.

2.6.3.3. Seed threshing and automated grain counting

Once all plants had fully senesced, seeds were threshed, collected, counted, and weighed. Spikes were threshed for grain and photographed using a Nikon D50 on auto setting against a blue background. Images were then imported into ImageJ and a macro was made to automate the counting of grain number.

2.6.4 Phenotypic trial for nitrogen responses

2.6.4.1. Design and conditions

The trial was arranged using a complete randomised block design with the R package agricolae with 12cm row spacing in 11 pots with three nitrogen soil conditions in PBF during Winter/Spring 2021; M1 soil, M1 soil with 3g/1L osmocote added, and M1 soil with 6g/1L osmocote added (de Mendiburu and de Mendiburu, 2019). Tillers were measured once per week until senescence.

2.6.4.2. Spike measurement

When plants were fully dried, spikes were taken and photographed using a Nikon D50 camera set with an ISO of 400 using the table method outlined in the WERecognizer manual (Genaev et al., 2019). Images were then analysed using the WERecognizer application with command line on a Windows OS. Measurements were imported into R, and were both statistically evaluating using appropriate methods based on the distribution, sample number and number of categories within the data and visualised using the ggstatsplots approach (Team R, 2015; Patil, 2021). The statistical defaults are designed to follow the best practices in both data visualisation and statistical reporting. Additional details can be found in Patil 2021.

2.6.5. Axillary meristem formation measurement

Plants were grown for AXM formation measurement under standard conditions. 25 plants were dissected and assessed for AXM formation once a week for 5 weeks. All plots were generated in R using the package ggplot2 (Wickham, 2011; Team R, 2015).

2.6.6. Scanning Electron Microscopy (SEM)

For scanning electron microscopy (SEM) analysis, 2-3 mm immature inflorescences were isolated, those were submerged in fixative solution (2% paraformaldehyde, 2.5% glutaraldehyde, 25 mM phosphate buffer (pH 7.0)), and incubated at 4°C for 16 hours. After fixation, samples were washed with 25 mM phosphate buffer three times for 10 minutes each with gentle agitation, were dehydrated via a graded ethanol series, and chemically dried using HMDS (Sigma, UK). Pre-treatment and SEM imaging were performed in

Warwick Electron Microscopy Suite, the sputter coater was used to perform sample coating with gold, and SEM imaging was performed.

2.6.7 Spike x-ray Computerised Tomography scan

Ten spikes from three genotypes from the spike traits trial were visualised by Dr. Paul Wilson of the Warwick Manufacturing Group (WMG). The wheat spikes were scanned on a Tescan Unitom XL Micro CT scanner. They were scanned at 40kV and 1121µA with an exposure of 70ms over 1583 with 4 averages using Acquila (Tescan). The resulting radiographs were then reconstructed 3D voxelised volume using Acquila Reconstructor (Tescan), with a resulting voxel resolution of 50µm.

Chapter 3

3.1. Introduction

3.1.1. Vegetative and inflorescence architecture in wheat

Vegetative architecture in the grasses is determined by the pattern and timing of the formation of lateral branches called tillers. Tillering is one of the most important domestication traits in wheat and its wild relatives and has had a significant impact on wheat yield through breeding high tiller number semi-dwarf varieties via introgression of the *Rht* mutant alleles (Chapter 1). Ideotype breeding for high tiller number typically happens at the expense of apical growth, resulting in shorter plants and has also been shown to improve nitrogen use efficiency (Khush, 2003). Many researchers have increased tiller number through ideotype breeding, resistance to abiotic and biotic stresses and increased nitrogen response however, increases in plant growth and biomass overall has remained an elusive goal.

Inflorescence architecture is more complex than vegetative architecture, involving an intricate genetic network of transcription factors that regulate the meristematic determinacy of inflorescence and floral meristems. Like tiller development, inflorescence formation is tempered by resource availability and extrinsic environmental signals such as photoperiod and temperature (Mouradov et al. 2002; Liu et al. 2009). Domestication of the wheat inflorescence, the spike, has involved the selection of traits more amenable to agriculture such as free threshing through the Q locus (Chapter 1). Here, we aim to identify new morphogenic regulators of vegetative and inflorescence architecture by characterising a family of transcription factors, the

RKDs, that have long been implicated in pluripotency and the formation of meristem-like structures through ectopic expression (Waki et al., 2011). We hypothesise that ectopic expression of *TaRKDs* may induce plant architecture phenotypes relevant to crop improvement in wheat. Manipulation of axillary meristem formation through the mutation of meristematic genes has been done in rice with loss of function mutations in *TILLERS ABSENT 1* (*TAB1*), an ortholog of *WUSCHEL* (*WUS*) (Chapter 1). Additionally, gain of function mutations in *ABERRANT PANICLE ORGANIZATION 1* (*APO1*), a suppressor of meristematic determinacy, has been shown to increase grain number in rice panicles by delaying spikelet formation and was further confirmed through ectopic expression which showed an increase in meristem size by inducing higher cell proliferation (Muqaddasi et al., 2019).

3.1.2. Identification and characterisation of candidate loci in wheat

Historically, identification of causative loci in crop breeding had been performed through genetic mapping allowing researchers to track quantitative trait loci (QTL) from crosses and characterise their phenotypic effects (Würschum, 2012). Advanced functional genomics has allowed further resolution of QTLs to identify individual genetic elements within or proximal to genes that are causative of the phenotype. The development of bioinformatical resources such as reference assemblies and annotations has allowed characterisation and a more systematic understanding of important traits in crop breeding in staple and orphan crops (Hu et al., 2018; Ribaut et al., 2019). Identification of transcription factors implicated in plant architecture has been greatly improved by these developments, allowing the phylogenetic and transcriptomic characterisation of less well studied members of that family, allowing us to infer roles for them in biological processes and identify those with the potential

to improve crop yield and plant development. Despite its agronomic importance, wheat research hasn't had the same bioinformatics resources available as crops that contribute similarly to global human calorific intake such as rice due to its large and complex polyploid genome. The International Wheat Genome Sequencing Consortium (IWGSC) has generated a high-quality annotated reference for wheat allowing molecular geneticists and breeders to identify and characterise genes implicated in wheat development and agronomically interesting processes to increase yield and better understand plant development (Jia et al., 2018; IWGSC, 2018).

3.1.3. Transformation and screening transgenic lines

Investigating ectopic expression of genes of interest can be performed through two-component inducible expression systems, such as the XVE two component system which induces ectopic expression in the presence of Estradiol (Brand et al., 2006). Somatic embryogenesis in wheat varies significantly with variety, elite cultivars like 'Fielder' and 'Bobwhite' are typically chosen as they show high efficiencies at producing fertile transgenic plants compared to others like 'Haruyokoi' and 'Gamtoos' being completely recalcitrant to agrobacterium transformation (Fennell et al., 1996; Zale et al., 2004; Hamada et al., 2017). Independent transformation events require screening and characterisation for ectopic gene expression as integration of transgenes into heterochromatic regions can result in low or no ectopic expression. Additionally, integration of multiple transgenes into the genome can lead to transcriptional/post-transcriptional gene silencing (Vaucheret et al., 1998). Screening ectopic expression of transgenic lines can be performed through reverse transcription polymerase chain reaction (RT-PCR) of RNA extracted from tissues of interest as chromatin accessibility differs between tissues allowing researchers make associations between

tissue specific ectopic expression and phenotypes (Becker-André and Hahlbrock, 1989).

3.1.4. Tiller phenotyping in wheat

Plant phenotyping is the primary metric that breeding programs measure crop improvement in plant genotypes and requires randomised trials to control for abiotic and biotic genetic-environmental interactions that can introduce biases into measurements. Tillering has been shown to be regulated by many environmental factors such as mechanical damage, nitrogen availability and soil volume. The Japanese farming technique 'touatsu' (translated as 'treading pressure') involves application of mechanical damage to wheat and barley seedling by stepping on them and has been used since the 17th century to prevent spindled growth by increasing tiller number and shortened plant height (Iida, 2014). Nitrogen availability has been shown to increase plant biomass and tiller number in rice and wheat due to its critical role in plant growth and has been shown to vary by cultivar (Power and Alessi, 1978; Amin et al., 2006). Soil volume also been shown to affect tiller number in wheat, being synomonous to plant spacing in field conditions, with plants sensing substrate volume through root density which proactively modulates shoot growth (Wheeldon et al., 2021). Here, we aim to characterise tillering phenotypes extensively through the introduction of these 3 factors into randomised design trials.

3.1.5. Spike phenotyping in wheat.

In recent years, there's been a dramatic increase in the number of state-of-the-art technological platforms for phenotyping a range of plant traits in both controlled and field conditions (NSF, 2011; Lawrence-Dill et al., 2019). Automated image analysis

of phenotypic images has also become standard practice through analysis software such as ImageJ and custom programs like WERecognizer allowing researchers to provide unbiased results capable of replication (Rasband, 1997; Genaev et al., 2019). WERecognizer analyses spike morphology based on 2D images of dry wheat spikes and uses a colour checker to control for slight changes in ambient lighting making it a widely available, low-cost phenotypic technique. Here, we aim to utilize state-of-the-art automated image analysis to remove bias in measuring spike phenotypes and traits.

Chapter aims and hypothesis

The overall aim of this chapter is to identify and characterise zygotic RKD transcription factor and their effect on vegetative and inflorescence architecture in wheat. This will be addressed through a phylogenetic and transcriptomic analysis of RKDs in hexaploid wheat and its wild progenitors to identify a zygotic *TaRKD* transcription factor, followed by transformation and phenotypic characterisation of the wheat cultivar 'Fielder' using an ectopic expression system for this gene. The question being addressed in this chapter is:

- Can plant architecture be manipulated through ectopic expression of zygotic RKDs?

We aim to provide a resource for others investigating the function of the RKDs in wheat and its wild progenitors by characterising unannotated members in IWGSC RefSeq v1.1. Further to this, by establishing whether zygotic transcription factors can modulate plant architecture through effecting meristem formation we aim to learn more about how yield-determining traits can be manipulated in cereal crops.

3.2. Results

3.2.1. Identification of embryonic *TaRKD* genes in IWGSC RefSeq v1.1

To identify an embryonic *TaRKD* gene that could be used as a morphogenic regulator for increasing tiller number in wheat, we took open reading frames (ORF) from 17 previously annotated *TaRKD* genes and compared them to 2 *TaRKD* cDNA sequences from a cDNA embryo specific expression library (Kőszegi, 2008; Kumar et al., 2018). *TaRKD1*-7A showed the highest homology to embryo specific *TaRKD*s but had no annotated homeologues in IWGSC RefSeq v1.1. Two homoeologous sequences that showed high sequence similarity to the 4 *TaRKD1-7A* ORF sequences in syntenic loci on chromosomes 7B and 7D were annotated. To confirm a homoeologous relationship between *TaRKD1-A*, *TaRKD1-7B* and *TaRKD1-7D,* a TaRKD amino acid sequence phylogeny was constructed using *TaRKDs* and *RKDs* from other grass species (Supplemental Table 3.1; Supplemental Figure 3.1).

To confirm expression in all *TaRKD1* homeologues, whole genome RNAseq data from microspore tissue was obtained and analysed, verifying *TaRKD1* expression in dividing microspore tissue of all three *TaRKD1* homeologues (Supplemental Figure 3.2). Expression profiles of *TaRKDs* were investigated by obtaining and analysing whole genome RNAseq data from using publicly available data from a developmental timecourse and hierarchically clustered using a heatmap (Figure 3.1). Two distinct expression profiles of *TaRKD* expression were found: a global vegetative and a restricted embryonic expression profile. To confirm that this expression profile was evolutionarily retained in wheat progenitors, publicly available transcriptome data from a developmental timecourse in the diploid progenitor *Aegolips tauschii* and the tetraploid progenitor *Triticum dicoccum* was analysed using high confidence *TaRKD*

orthologs in the respective species (Supplementary Table 3.2, Supplementary Figure 3.3 and Supplementary Table 3.3, Supplementary Figure 3.4). A phylogenetic tree of RKD amino acid sequences in grass species and *A. thaliana* confirmed that *TaRKD1*, *AetRKD1* and *TdRKD1* cluster together and show high similarity to the well characterised zygotic transcription factor *AtRKD4* (Figure 3.2).\

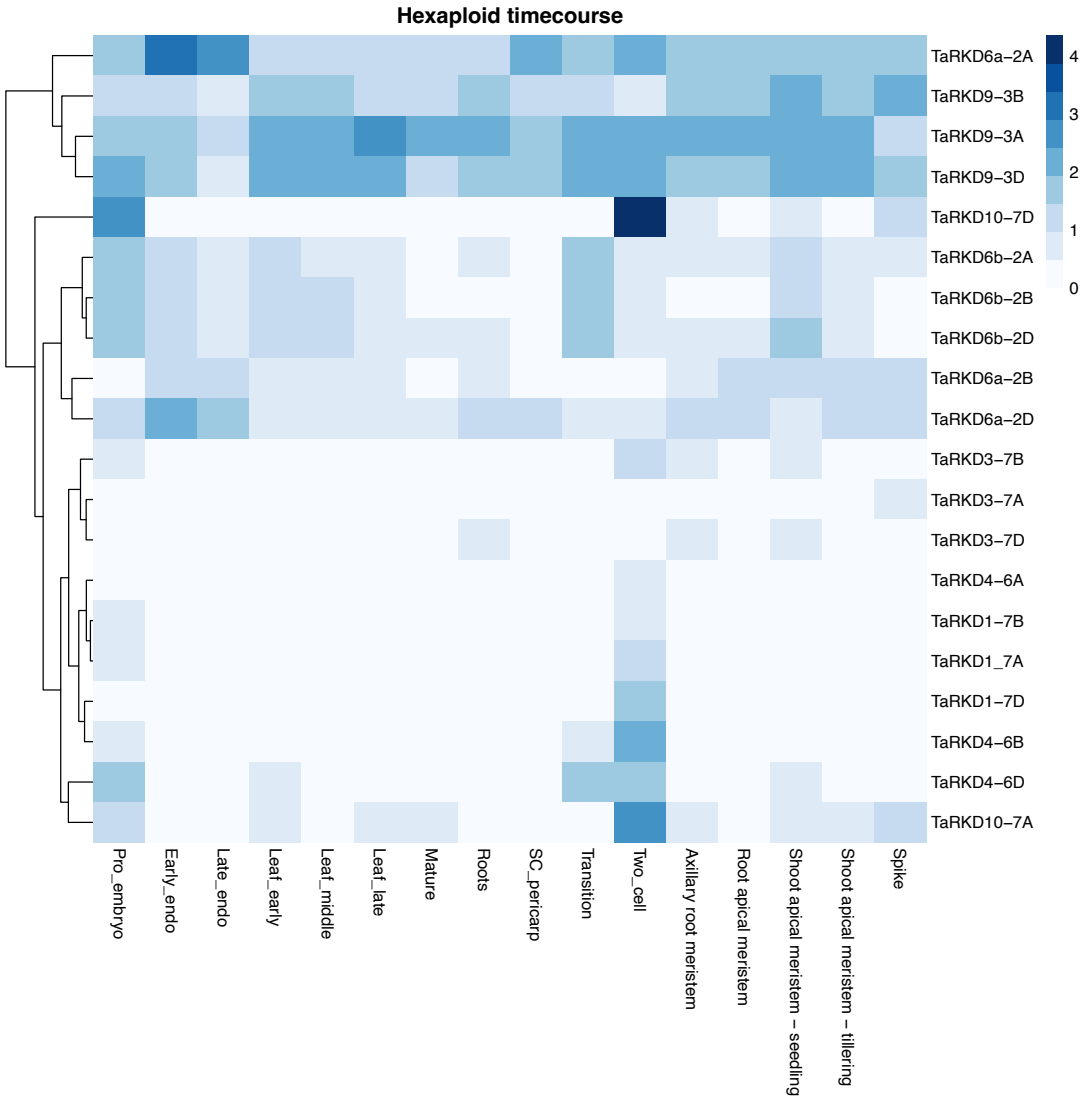


Figure 3.1. Heatmap of *TaRKD* expression. Heatmap of re-analysed publicly available developmental timecourse transcriptome samples shows two distinct expression profiles of *TaRKD* genes from Supplementary Table 3.1: a global vegetative expression profile (top) and a restricted embryonic expression profile (bottom). Scale displayed is logarithmically transformed transcript per million values from Kallisto quant. Hierarchical clustering was performed on *TaRKD* genes and is displayed as a dendrogram on the side of the heatmap.

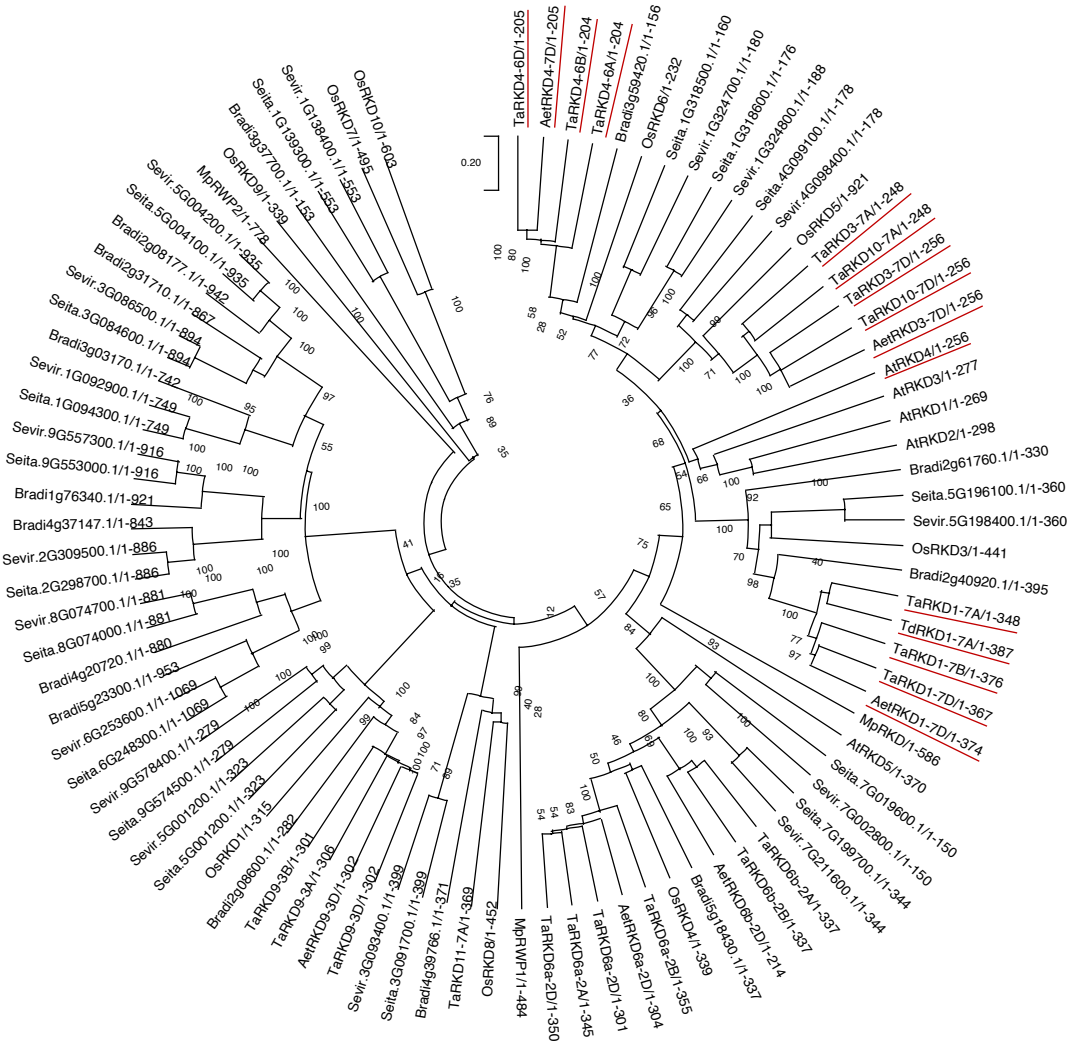


Figure 3.2. Phylogenetic tree of RKD genes in the grasses and Arabidopsis. 92 protein sequences of RKD genes from several grass species and the dicot *Arabidopsis.* Values at branches represent evolutionary distance between sequences. Species key; *Aet = Aegolips tauschii, At = Arabidopsis thaliana, Bradi = Brachypodium distachyon, Mp = Marchantia polymorpha, Os = Oryza sativa ssp. japonica, Seita = Setaria indica, Sevir = Setaria virdis, Ta = Triticum aestivum, Td = Triticum dicoccum.* Zygotic *RKD*s from Figure 3.1 are underlined in red.

3.2.2. Transformation and screening of *indTaRKD1-7D* transgenic lines for higher tiller number and ectopic TaRKD1-7D expression

A two-component XVE inducible *indTaRKD1-7D* expression construct was transformed into 'Fielder' by NIAB Cambridge. Transgenic lines were screened for high copy T-DNA insertion number using DNA profiling and 55 *indTaRKD1-7D* T1 lines were transferred and propagated at the phytobiology facility (PBF) at the University of Warwick. To confirm that all transgenic lines were single copy, 55 *indTaRKD1-7D* T2 lines were genotyped for multiple T-DNA insertions. A summary of these results can be found in Table 3.1. To construct an *indTaRKD1-7D* expression profile for transgenic lines, 40 high confidence single copy T2 lines were screened for *indTaRKD1-7D* in leaf tissue using RT-PCR with the chemical inducer Estradiol (Figure 3.3).

16 transgenic T2 lines that showed moderate to high *indTaRKD1-7D* expression were also screened for *indTaRKD1-7D* in immature embryo tissue using RT-PCR with and without the presence of the chemical inducer Estradiol (Figure 3.4). This showed that a high number of transgenic lines had ectopic *indTaRKD1-7D* expression in vegetative and reproductive tissues. Somatic embryogenesis was tested in 9 lines that showed immature embryo expression but couldn't induce callus formation in the scutellum. However, *indTaRKD1-7D* #2.28 showed a polyembryonic phenotype with a frequency of 2% (4/200) suggesting that ectopic *indTaRKD1-7D* expression can drive somatic embryogenesis in certain tissue contexts (Supplemental Figure 3.5).

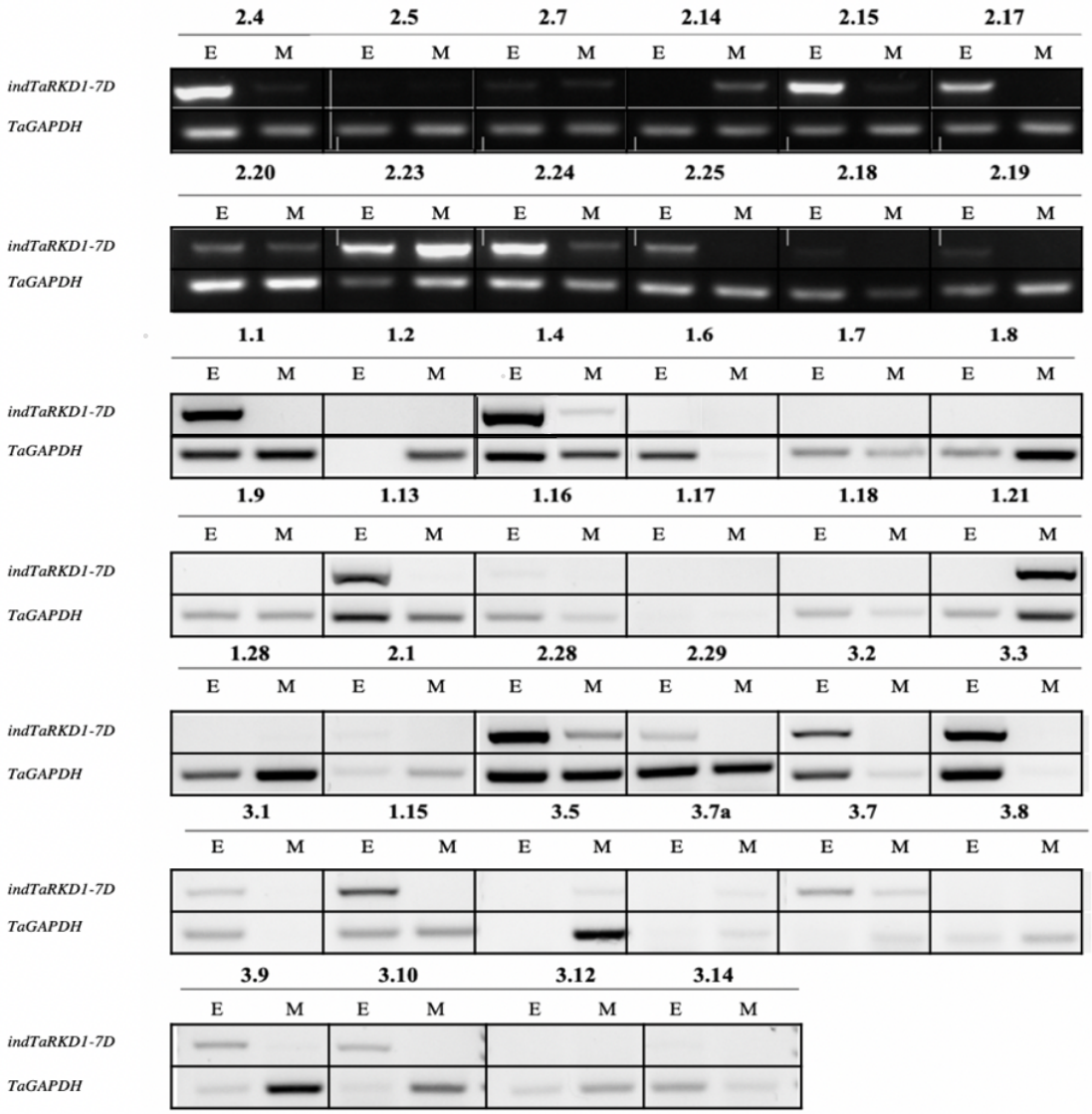


Figure 3.3. RT-PCR of *TaRKD1-7D* expression across T2 transgenic lines in leaf tissue. *TaRKD1-7D* expression was quantified through RT-PCR for 32 cycles using *TaGAPDH* as a control at 24 cycles. Leaf tissue was soaked in 40M of estradiol or a mock solution containing 0.01% (w/v) Tween-20 (E and M). *TaRKD1-7D* and *TaGAPDH* amplicons are labelled on the left of the panels and solution treatment above. Each box corresponds to the transgenic line labelled above, details for which can be found in Table 3.1.

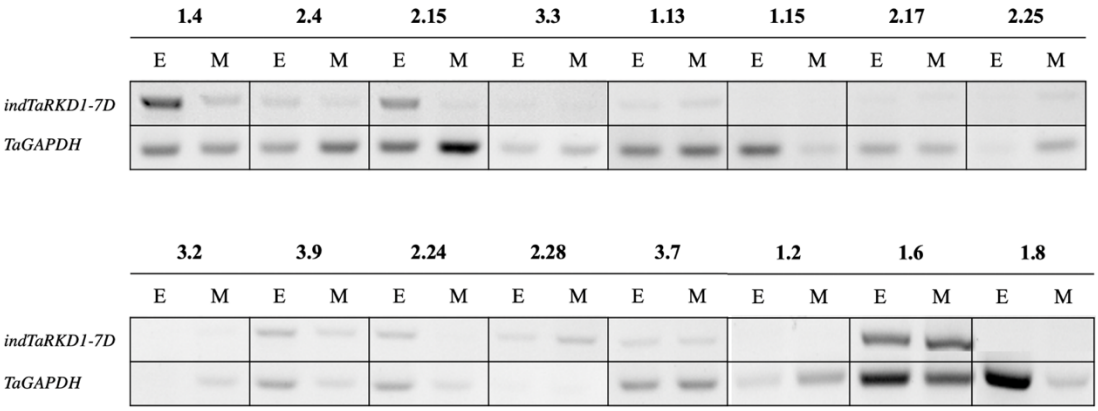


Figure 3.4. RT-PCR of *TaRKD1-7D* expression across T2 transgenic lines in immature embryo tissue. *TaRKD1-7D* expression was quantified through RT-PCR for 32 cycles using *TaGAPDH* as a control at 24 cycles. 20 Immature embryos were excised from the caryopsis 12-18 days after anthesis and were soaked in 40M of estradiol or a mock solution containing 0.01% (w/v) Tween-20 (E and M). *TaRKD1-7D* and *TaGAPDH* amplicons are labelled on the left of the panels and solution treatment above. Each box corresponds to the transgenic line labelled above, details for which can be found in Table 3.1.

Table 3.1. List of *indTaRKD1-7D* lines.

NIAB construct	Sample name	Construct	Resistance	PCR positives	Multiple copy
bar 1_	1._	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 1_1	1.1	pERV-Basta-TaRKD1-7D	basta	8	No
bar 1_12	1.12	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 1_13	1.13	pERV-Basta-TaRKD1-7D	basta	4	No
bar 1_15	1.15	pERV-Basta-TaRKD1-7D	basta	5	No
bar 1_16	1.16	pERV-Basta-TaRKD1-7D	basta	9	Yes
bar 1_17	1.17	pERV-Basta-TaRKD1-7D	basta	6	No
bar 1_18	1.18	pERV-Basta-TaRKD1-7D	basta	8	No
bar 1_2	1.2	pERV-Basta-TaRKD1-7D	basta	7	No
bar 1_20	1.20.	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 1_21	1.21	pERV-Basta-TaRKD1-7D	basta	3	No
bar 1_24	1.24	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 1_25	1.25	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 1_4	1.4	pERV-Basta-TaRKD1-7D	basta	8	No
bar 1_6	1.6	pERV-Basta-TaRKD1-7D	basta	2	No
bar 1_7	1.7	pERV-Basta-TaRKD1-7D	basta	4	No
bar 1_8	1.8	pERV-Basta-TaRKD1-7D	basta	7	No
bar 1_9	1.9	pERV-Basta-TaRKD1-7D	basta	2	No
bar 1_control 1	Bar 1.C.1	None	control	0	No
bar 1_control2	Bar 1.C.2	None	control	0	No
bar 2_1	2.1	pERV-Basta-TaRKD1-7D	basta	9	Yes
bar 2_10	2.10.	pERV-Basta-TaRKD1-7D	basta	9	Yes
bar 2_12	2.12	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 2_13	2.13	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 2_14	2.14	pERV-Basta-TaRKD1-7D	basta	5	No
bar 2_15	2.15	pERV-Basta-TaRKD1-7D	basta	4	No
bar 2_17	2.17	pERV-Basta-TaRKD1-7D	basta	7	No
bar 2_18	2.18	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 2_19	2.19	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 2_2	2.2	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 2_20	2.20.	pERV-Basta-TaRKD1-7D	basta	8	No
bar 2_23	2.23	pERV-Basta-TaRKD1-7D	basta	4	No
bar 2_24	2.24	pERV-Basta-TaRKD1-7D	basta	6	No
bar 2_25	2.25	pERV-Basta-TaRKD1-7D	basta	2	No
bar 2_26	2.26	pERV-Basta-TaRKD1-7D	basta	9	Yes
bar 2_28	2.28	pERV-Basta-TaRKD1-7D	basta	7	No
bar 2_29	2.29	pERV-Basta-TaRKD1-7D	basta	8	No
bar 2_31	2.31	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 2_34	2.34	pERV-Basta-TaRKD1-7D	Basta	9	Yes
bar 2_4	2.4	pERV-Basta-TaRKD1-7D	basta	4	No
bar 2_5	2.5	pERV-Basta-TaRKD1-7D	basta	6	No
bar 2_6	2.6	pERV-Basta-TaRKD1-7D	Basta	4	No
bar 2_7	2.7	pERV-Basta-TaRKD1-7D	basta	7	No
bar 2_8	2.8	pERV-Basta-TaRKD1-7D	basta	10	Yes
bar 2_9	2.9	pERV-Basta-TaRKD1-7D	Basta	9	Yes
bar 2_control1	Bar 2.C.1	None	control	0	No
bar 2_control2	Bar 2.C.2	None	control	0	No
hyg 3_1	3.1	pERV-Hyg-TaRKD1-7D	hygromycin	7	No
hyg 3_10	3.10.	pERV-Hyg-TaRKD1-7D	hygromycin	2	No
hyg 3_11	3.11	pERV-Hyg-TaRKD1-7D	hygromycin	10	Yes
hyg 3_12	3.12	pERV-Hyg-TaRKD1-7D	hygromycin	6	No
hyg 3_14	3.14	pERV-Hyg-TaRKD1-7D	hygromycin	5	No
hyg 3_14a	3.14a	pERV-Hyg-TaRKD1-7D	hygromycin	10	Yes
hyg 3_3	3.3	pERV-Hyg-TaRKD1-7D	Hygromycin	4	No
hyg 3_5	3.5	pERV-Hyg-TaRKD1-7D	Hygromycin	8	No
hyg 3_7	3.7	pERV-Hyg-TaRKD1-7D	hygromycin	3	No
hyg 3_7a	3.7a	pERV-Hyg-TaRKD1-7D	hygromycin	10	Yes
hyg 3_8	3.8	pERV-Hyg-TaRKD1-7D	hygromycin	4	No
hyg 3_9	3.9	pERV-Hyg-TaRKD1-7D	hygromycin	8	No
hyg 3_control	Hyg 3.C	None	control	0	No
hyg 3_control1	Hyg 3.C.1	None	control	0	No

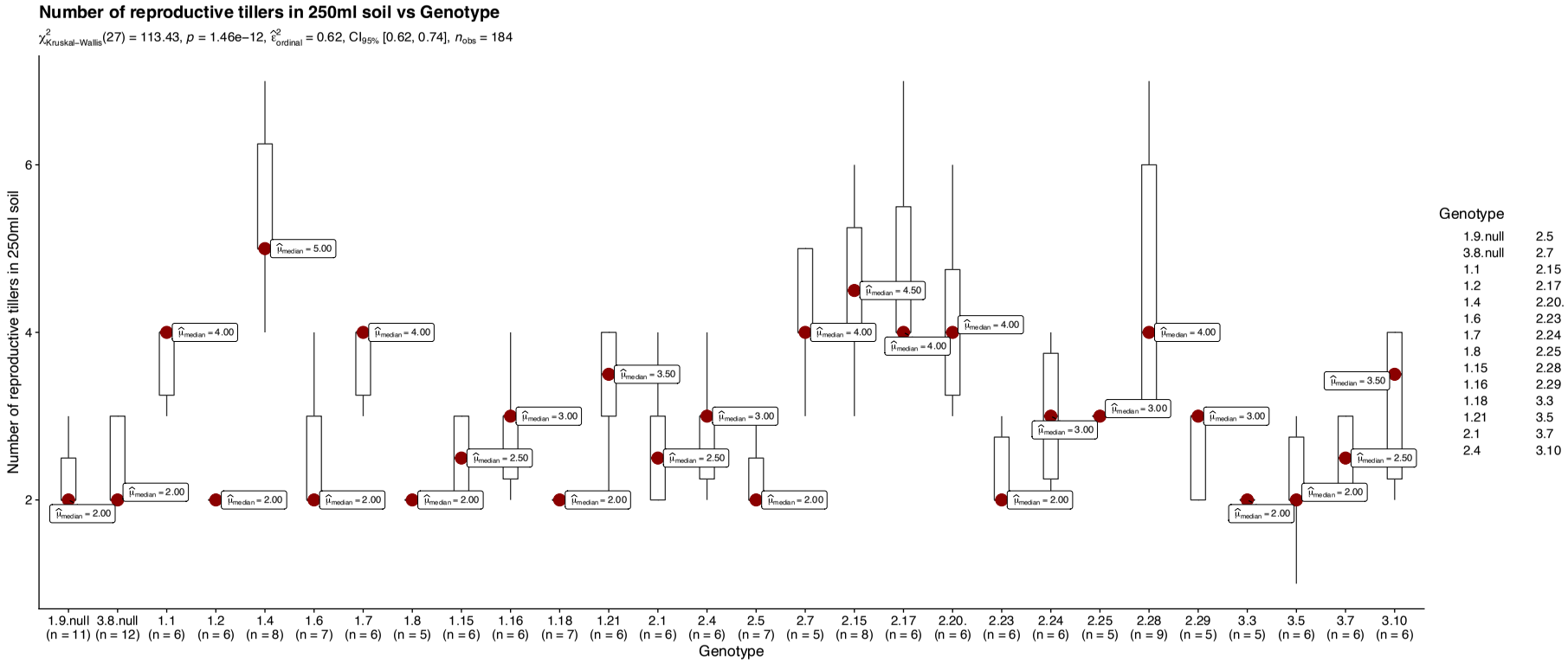


Figure 3.5. Identification of *indTaRKD1-7D* T2 lines with higher reproductive tiller number. Reproductive tiller number from genotypically positive plants measured from counting the number of fully developed spikes at senescence. Pairwise significance comparisons could not be visualised here due to high number of lines. #1.4, #2.15, #2.17, and #2.28 were all found to show significantly different reproductive tiller number to null segregants when tested with a Games Howell t-test. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot. Additional statistics can be found on the top left of the plot.

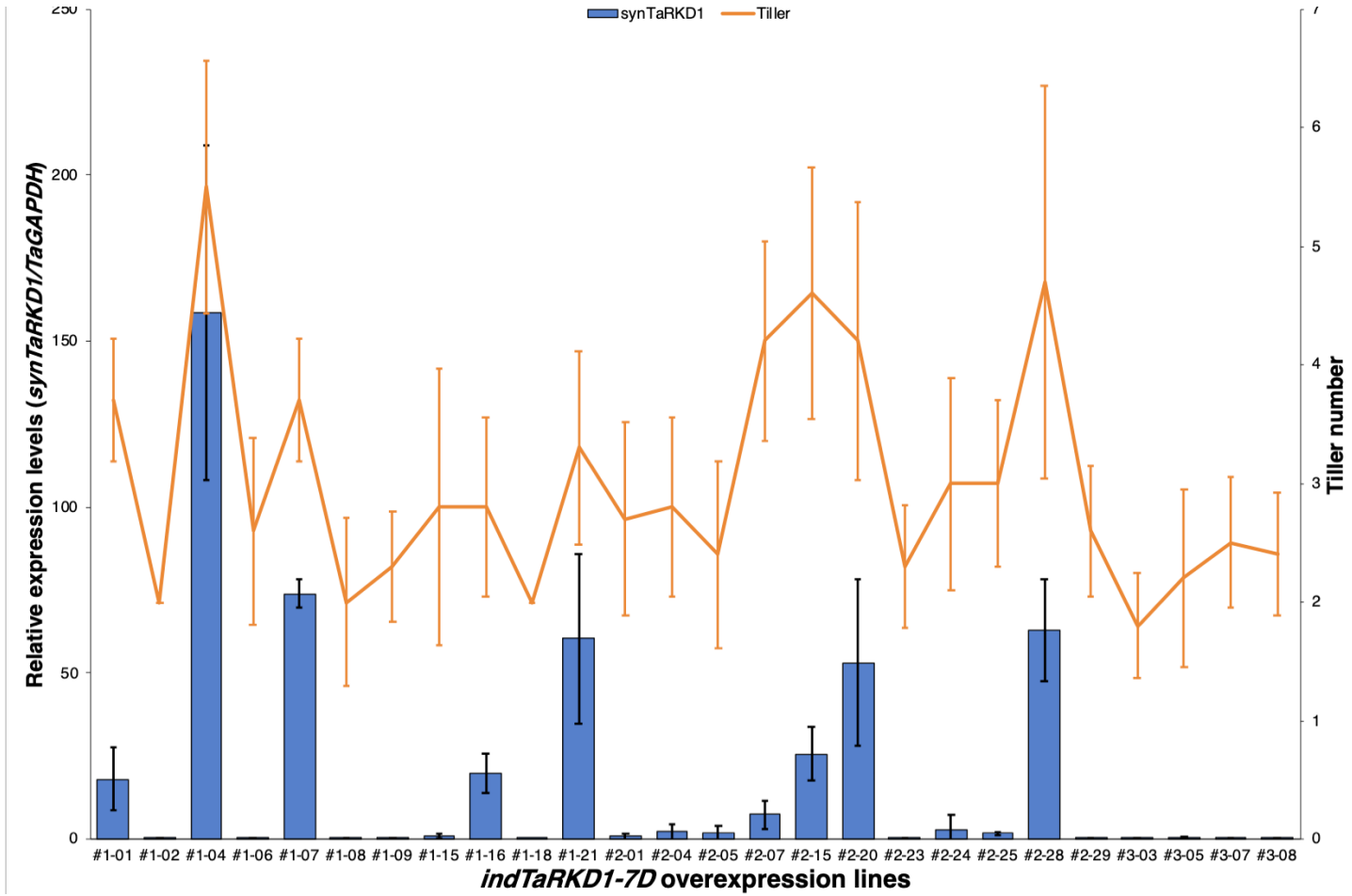


Figure 3.6. *TaRKD1-7D* expression in independent *indTaRKD1*-7D lines associated with tiller number. Relative *TaRKD1-7D/GAPDH* qPCR values are plotted in blue bars while reproductive tiller number is plotted in orange line plots.

To investigate whether ectopic expression of *indTaRKD1-7D* could increase tiller number we measured the reproductive tiller number of 26 transgenic lines (including all lines showed moderate to high *indTaRKD1-7D* leaf expression) that were genotypically positive for *indTaRKD1-7D* and 2 null segregants. *indTaRKD1-7D* qPCR was performed using *TaGAPDH* as a control to confirm the relationship between *indTaRKD1-7D* ectopic expression and reproductive tiller number showing a moderate positive Pearson correlation of 0.497 with relative *TaRKD1-7D* expression. *#1.4* and *#2.28* were chosen for a high replication trial for tiller traits as they exhibited ectopic expression and increased tiller number (Figure 3.3, Figure 3.6). Increased rachis elongation amongst high tillering lines was also observed suggesting ectopic *indTaRKD1-7D* expression had spike trait phenotypes (Supplementary Figure 3.6, Supplementary Figure 3.7, and Supplementary Figure 3.8).

3.2.3. Increased tiller number and seed weight in two *indTaRKD1-7D* T3 lines.

Homozygous T3 lines were selected for *indTaRKD1-7D* #1.4 and #2.28 to be grown for a tiller traits trial. Two *indTaRKD1-7D* T3 lines were selected for each independent transgenic event (#1.4 and #2.28) and grown with a randomised design for 100 days. Soil volume and mechanical stress treatment were used as factors in the experiments as they've been described in the literature to play a role in tiller development. Increased reproductive tiller number was observed in *indTaRKD1-7D* T3 lines at senescence (Figure 3.7, Figure 3.8).

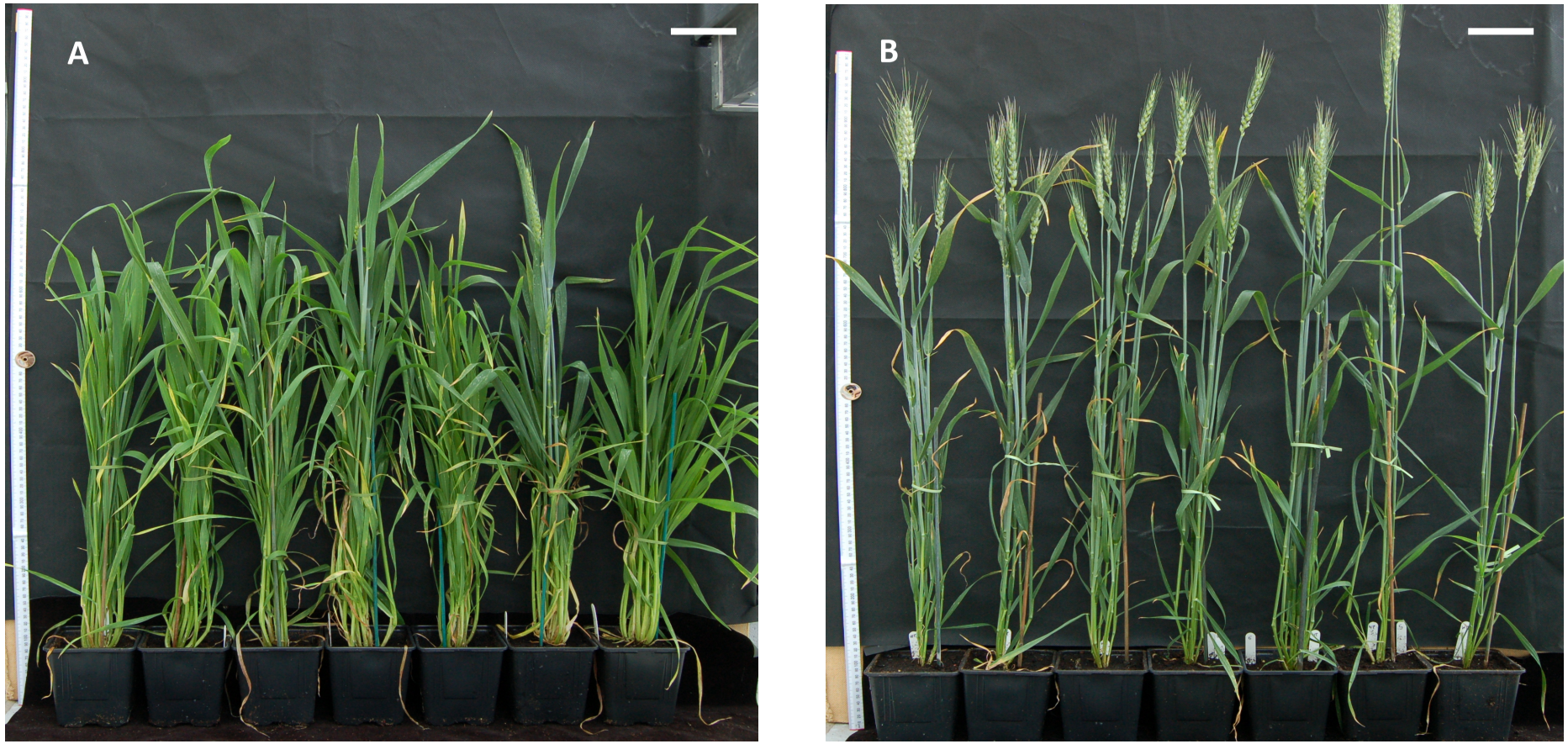


Figure 3.7. Tiller number in *indTaRKD1-7D #1.4* T3 plants. Photographs were taken 60 days after sowing with a metre stick for measurement. White scale bar in top right = 10cm. **(A)** *indTaRKD1-7D #1.4* 1000ml soil **(B)** Wild type 1000ml soil.

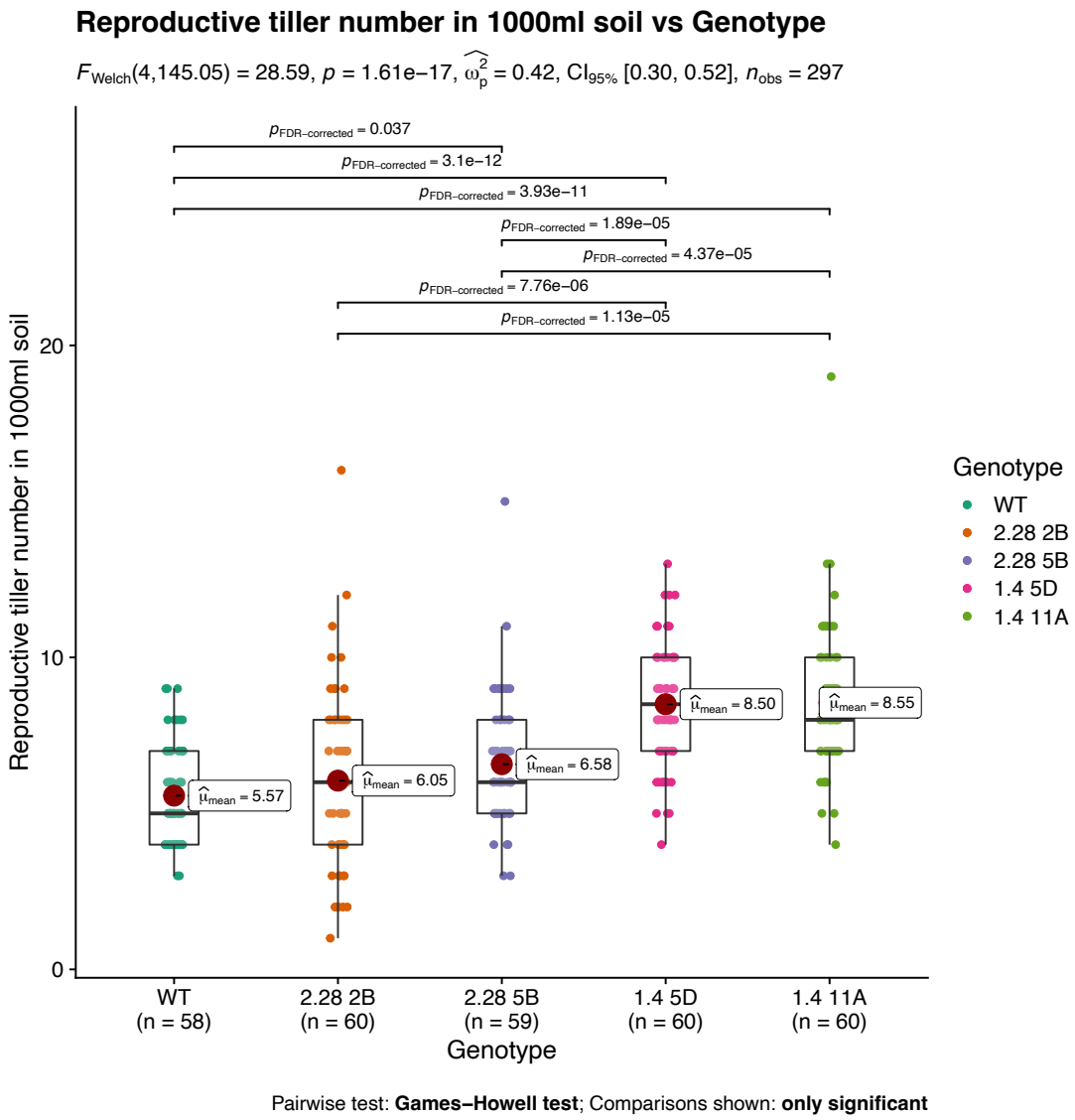


Figure 3.8. Reproductive tiller number two *indTaRKD1-7D* T3 lines. Reproductive tiller number from plants measured from counting the number of fully developed spikes at senescence. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

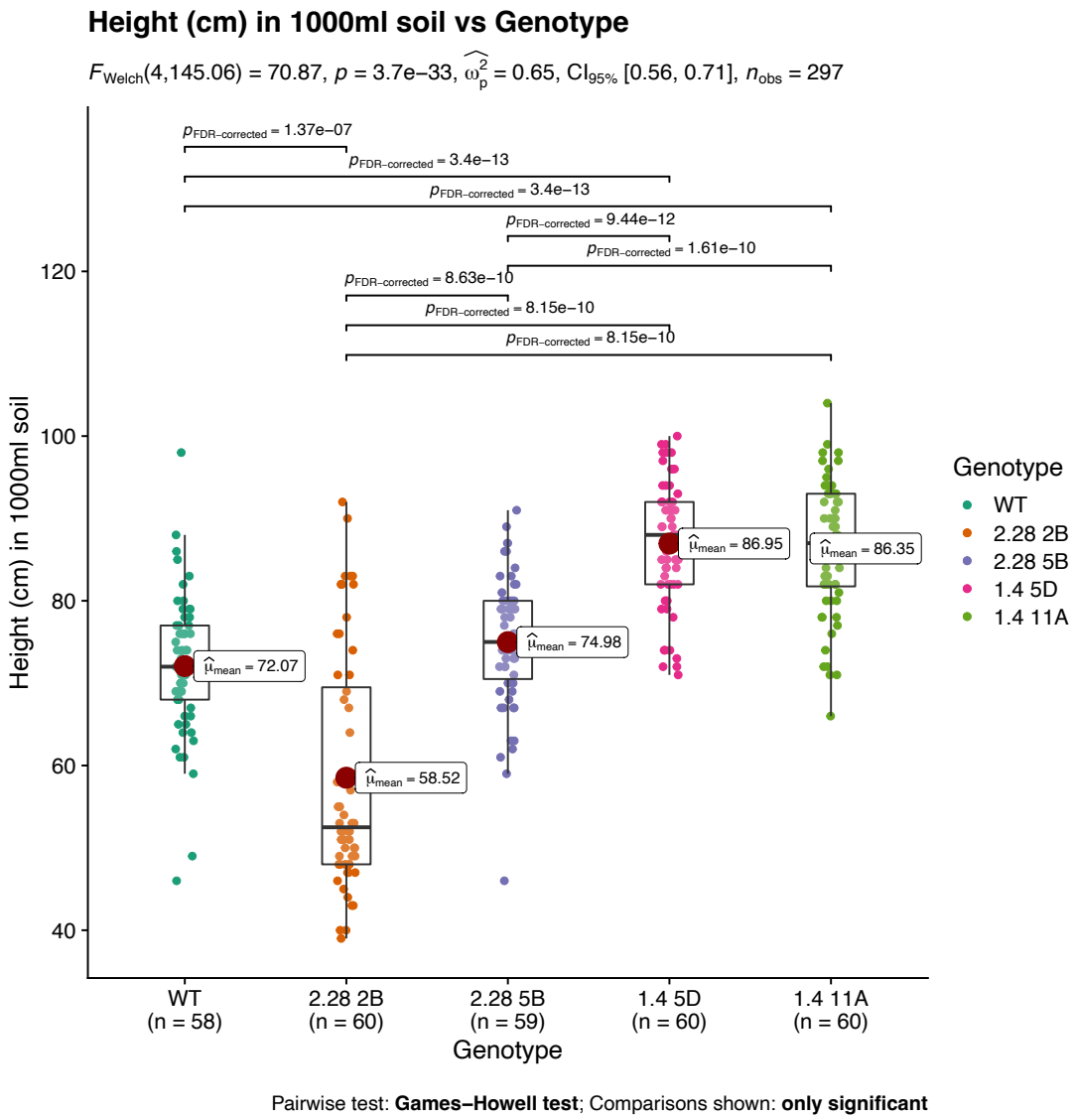


Figure 3.9. Plant height in *indTaRKD1-7D #1.4* T3 line. Height measured from the base of the plant to the tallest spike at senescence. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

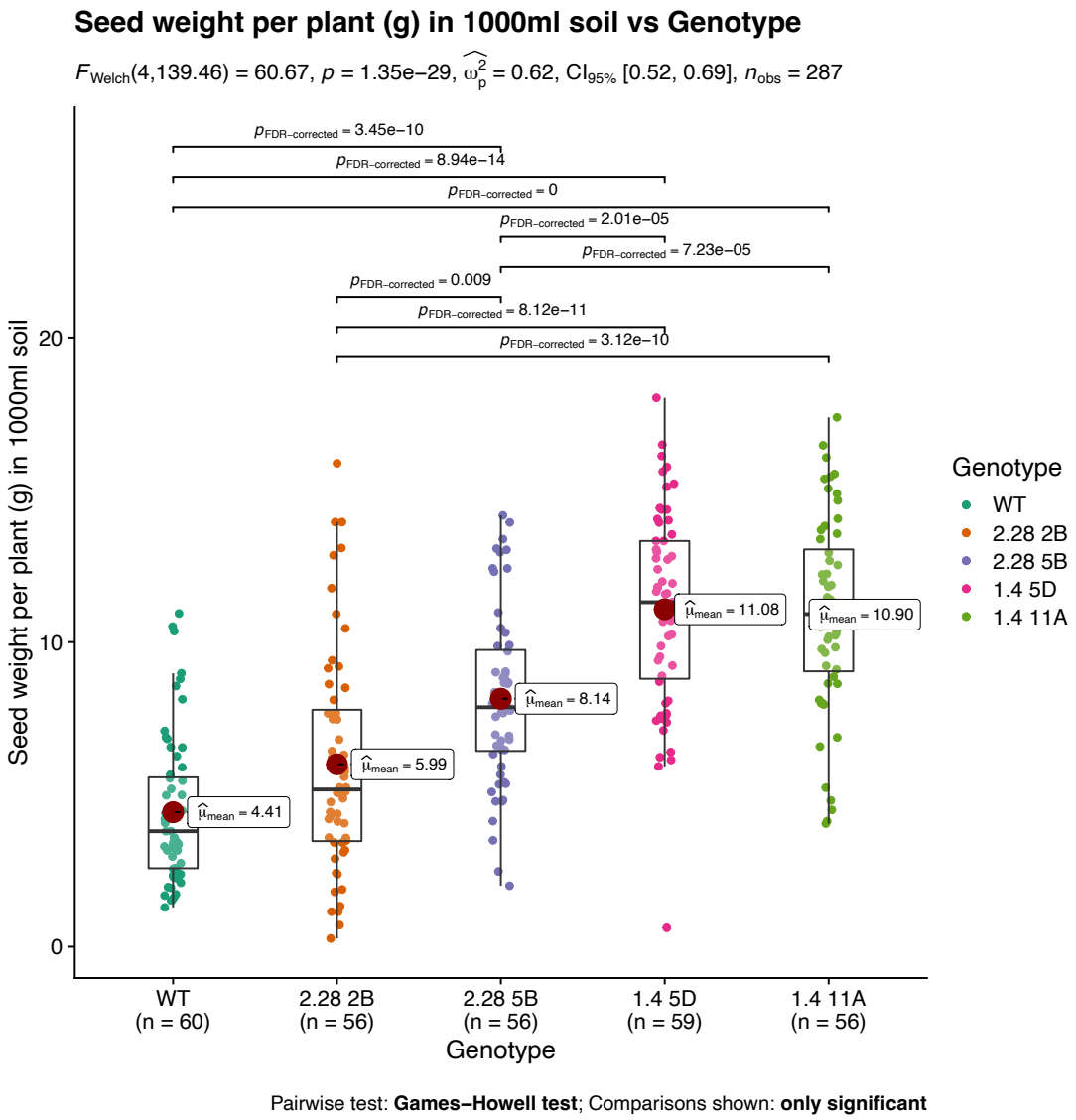


Figure 3.10. Seed weight per plant in two *indTaRKD1-7D* T3 lines. Seed weight measured after threshing fully dried spikes. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

However, an increase in unproductive tillers was also observed in *indTaRKD1-7D* #1.4 (Supplementary Figure 3.9). Days to maturity was significantly delayed in both *indTaRKD1-7D* T3 lines, a common characteristic of high tiller number phenotypes (Supplementary Figure 3.10). Additionally, increased height was also observed amongst *indTaRKD1-7D* T3 lines, in contrast to most characterised high tiller number phenotypes in the literature (Figure 3.9). No significant differences were found in flag leaf area (Supplementary Figure 3.11). Increased seed weight per plant and per 1000 seeds was observed in both *indTaRKD1-7D* T3 lines suggesting that ectopic *TaRKD1-7D* might work as a breeding strategy for increasing yield in wheat (Figure 3.10, Supplementary Figure 3.12). Increased tiller number and height was observed across 1000ml and 250ml soil volume conditions, and while *indTaRKD1-7D* lines in 50ml soil volume showed higher vegetative tiller number at 50 days after sowing, no significant difference in reproductive tiller number was observed (Supplementary Figure 3.13, Supplementary Figure 3.14).

No significant differences were found between mechanical stress treated and non-treated plants (Supplementary Figure 3.15, Supplementary Figure 3.16). Finally, Increased spike length was observed in both *indTaRKD1-7D* T3 lines to a similar extent after manually measured image analysis (Figure 3.11). Significant border effects in the tiller trial as spike phenotypes have been previously shown to be very sensitive to environmental variation (Supplemental Figure 3.17). A summary of these results can be found in Supplementary Table 3.6 and Supplementary Table 3.7. An independent time course trial measuring axillary meristem formation in two *indTaRKD1-7D* T3 lines confirmed higher tiller formation from axillary meristem formation and outgrowth (Supplementary Figure 3.18, Supplementary Figure 3.19).

3.2.4. Increased spikelet number per spike in two *indTaRKD1-7D* T3 lines

To confirm spike phenotypes in indTaRKD1-7D, a second trial with a border of wild type plant was conducted. Two homozygous *indTaRKD1-7D* T3 lines were selected from independent transgenic events (#1.4 and #2.28) and grown with wild type and an *indTaRKD1-7D* null segregant in a randomised design with a border of wild type plants for 100 days. Anthesis time was significantly delayed in both *indTaRKD1-7D* T3 lines from 60 to 80 days after sowing, consistent with the delayed maturity phenotype seen in the previous trial (Supplementary Figure 3.20, Supplementary Table 3.4). Spike traits were measured after senescence using an automated image analysis program called WERecognizer. Increased area and length were found in *indTaRKD1-7D* spikes compared to both wild type and null segregants (Supplementary Figure 3.21, Supplementary Figure 3.22). Increased spikelet number per spike was found amongst *indTaRKD1-7D* T3 lines from grain number data collected by automated image analysis but with no increase in interspikelet rachis elongation as previously suggested by the tiller traits trial (Figure 3.12, Figure 3.13). A summary of these results can be found in Supplementary Table 3.9 and Supplementary Table 3.10. Spikes from both *indTaRKD1-7D #1.4* and *#2.28* were visualised with scanning electron microscopy (SEM) and light microscopy to show increased spikelet number during spike development (Figure 3.14). X-ray computerise tomography (CT) scanning was also performed to image spikes for phenotypic differences (Figure 3.15). An increase in spikelet size could be seen, (but was not statistically evaluated) between 2 *indTaRKD1-7D* T3 lines and the wild type from 4 averaged scans (Supplementary Figure 3.23, Supplemental Figure 3.24). Following the characterisation of tiller and spike traits in *indTaRKD1-7D* lines we investigated whether these phenotypes were linked to nitrogen use efficiency.

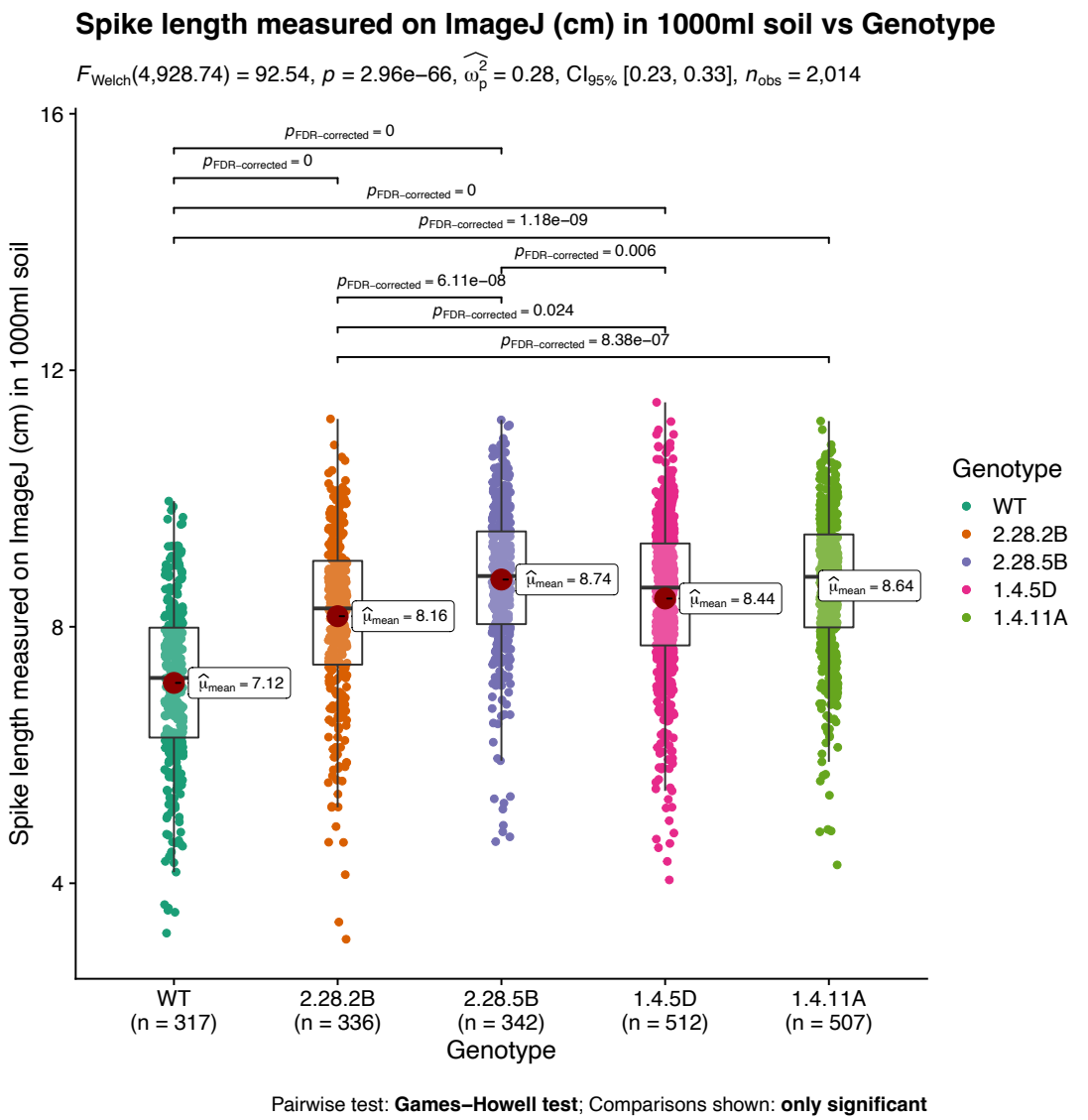


Figure 3.11. Spike length in two *indTaRKD1-7D* T3 lines. Spike length of fully dried spikes measured by ImageJ. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

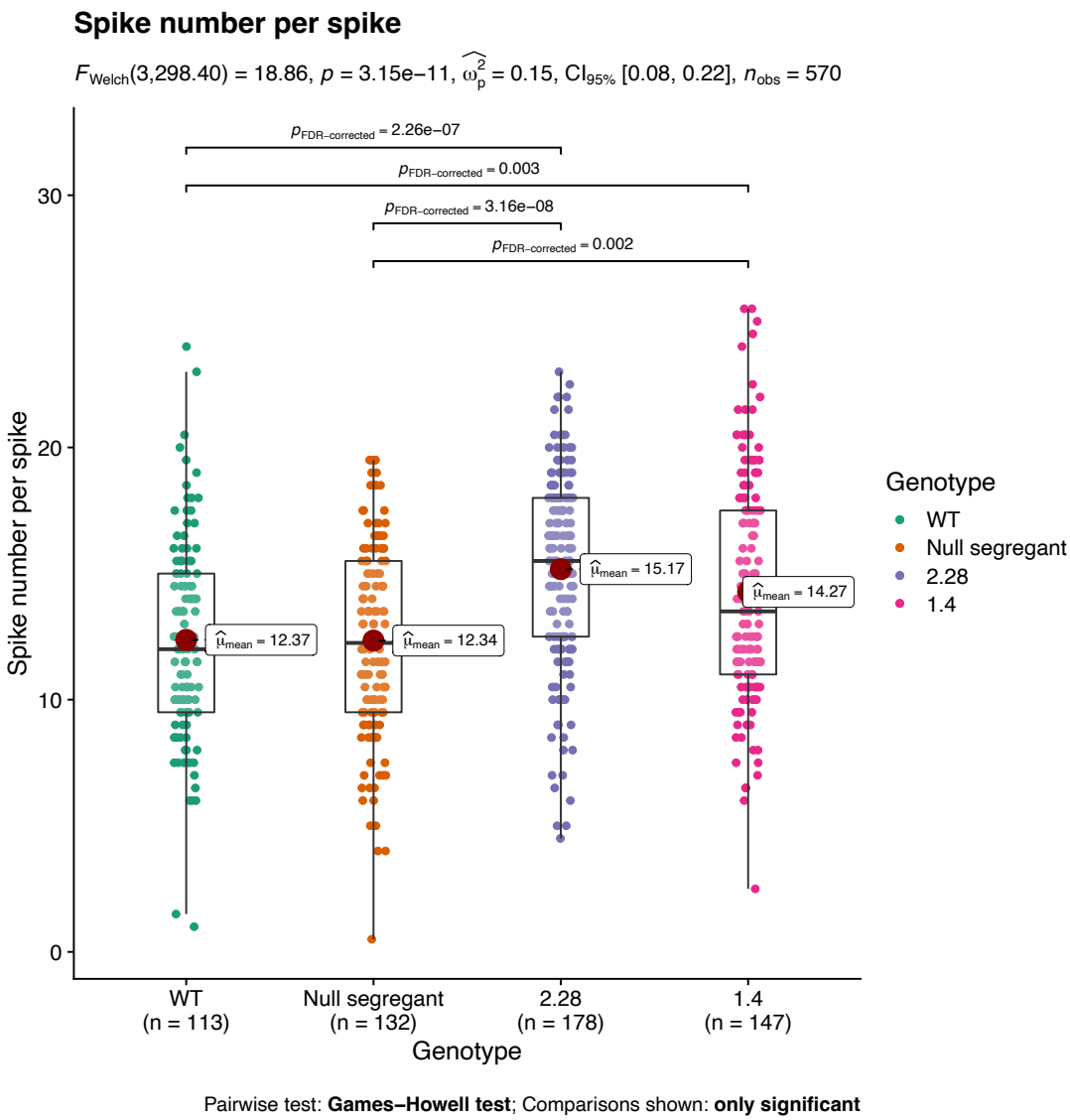


Figure 3.12. Spikelet number per spike in two *indTaRKD1-7D* T3 lines. Spikelet number per spike of fully dried spikes calculated by dividing the grain number by two and subtracting 1. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

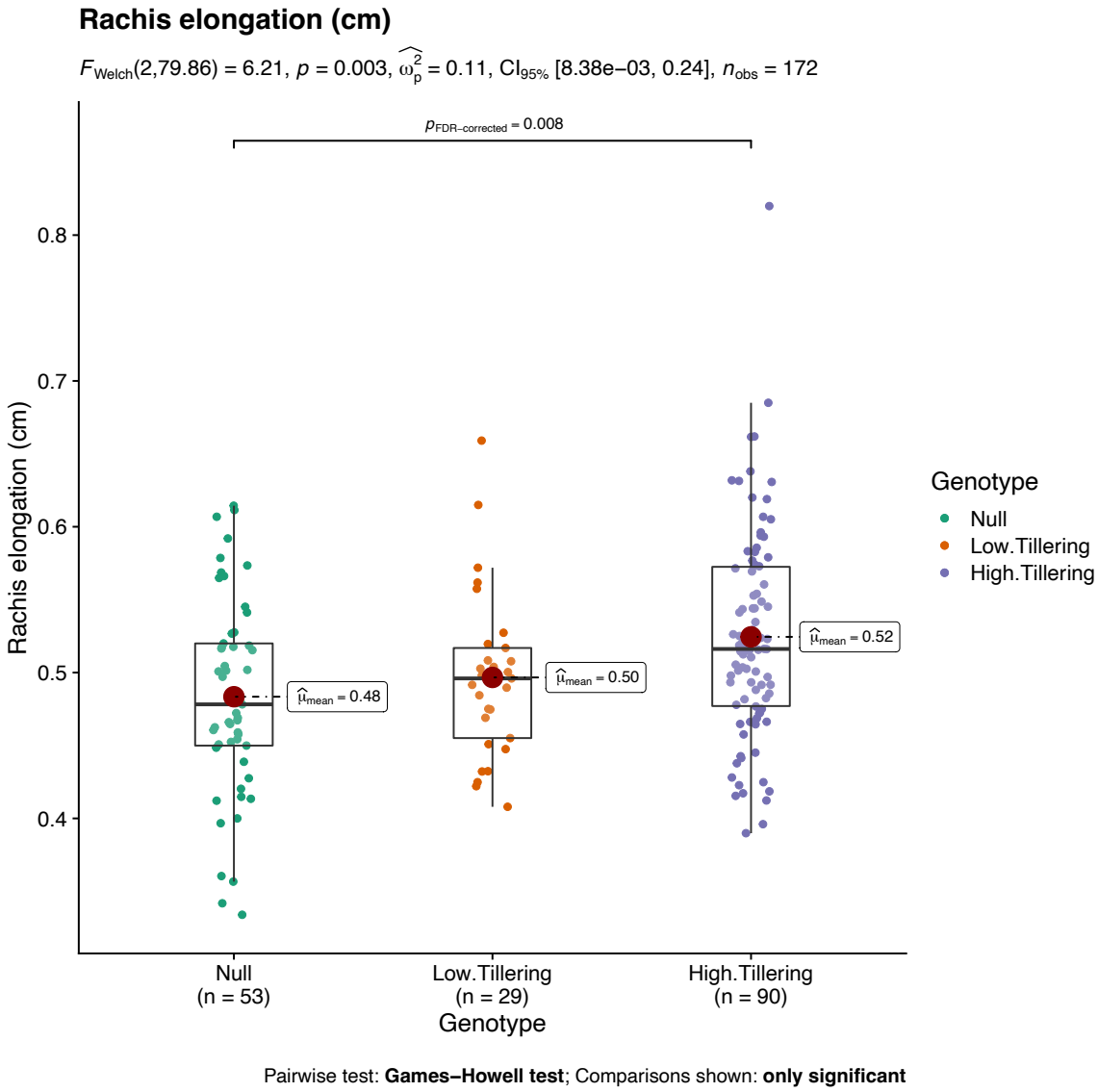


Figure 3.13. Rachis elongation in two *indTaRKD1-7D* T3 lines. Spike length measured by WERecognizer was divided by grain number measured using an imageJ macro to calculate rachis elongation. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

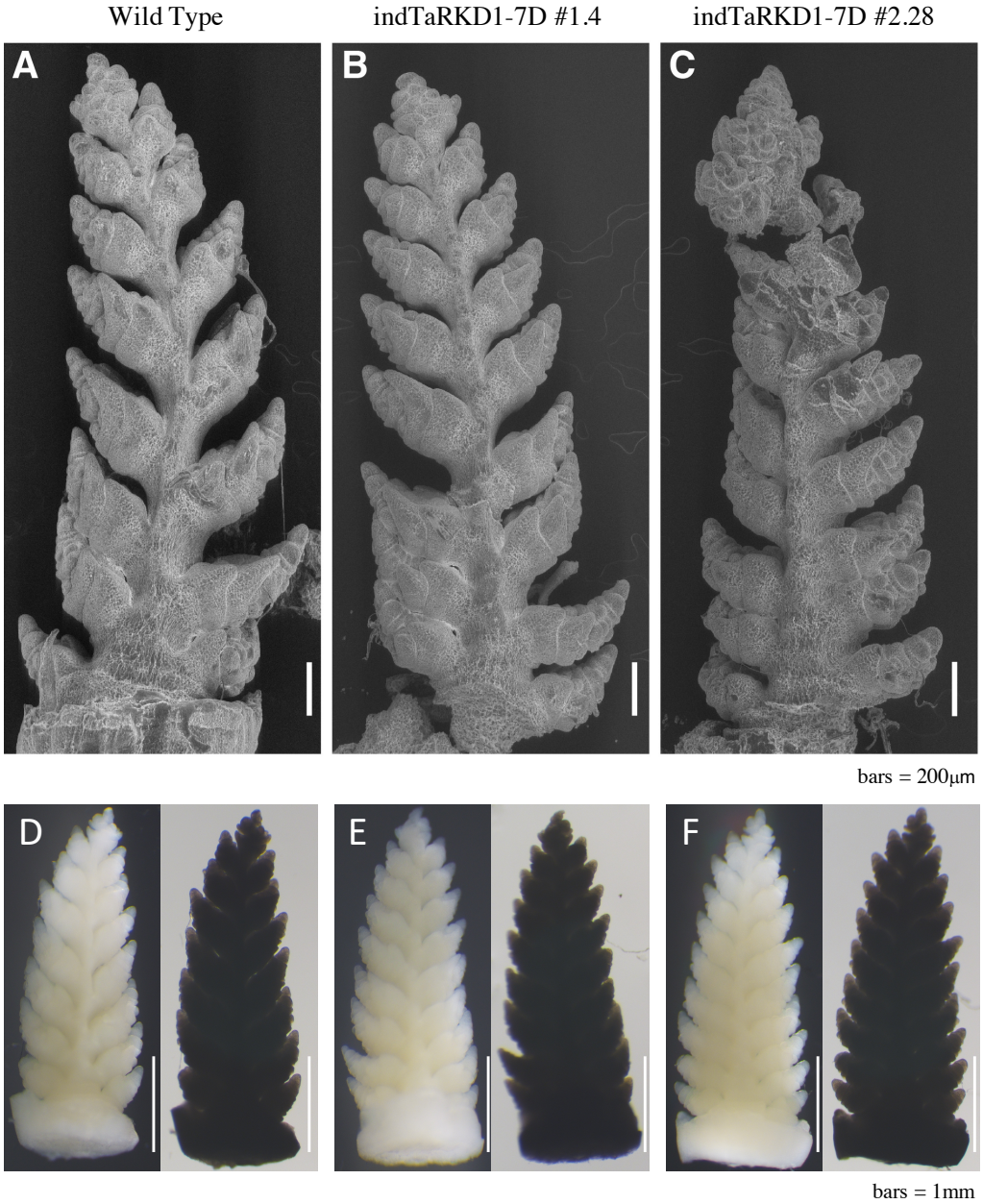


Figure 3.14. Images of *indTaRKD1-7D* spikes shows increased spikelet number per spike. Representative spikes from the spike traits trial were taken and imaged using Scanning Electron Microscopy (SEM). Bars in A, B, and C are set to 200µm while bars in D, E, and F are set to 1mm. **(A)** SEM image of Wild type spike. **(B)** SEM image of *indTaRKD1-7D #1.4* spike. **(C)** SEM image of *indTaRKD1-7D #2.28* spike. **(D)** Light microscope image of Wild type spike. **(E)** Light microscope image of *indTaRKD1-7D #1.4* spike. **(F)** Light microscope image of *indTaRKD1-7D #2.28* spike. Scale bars = 1mm

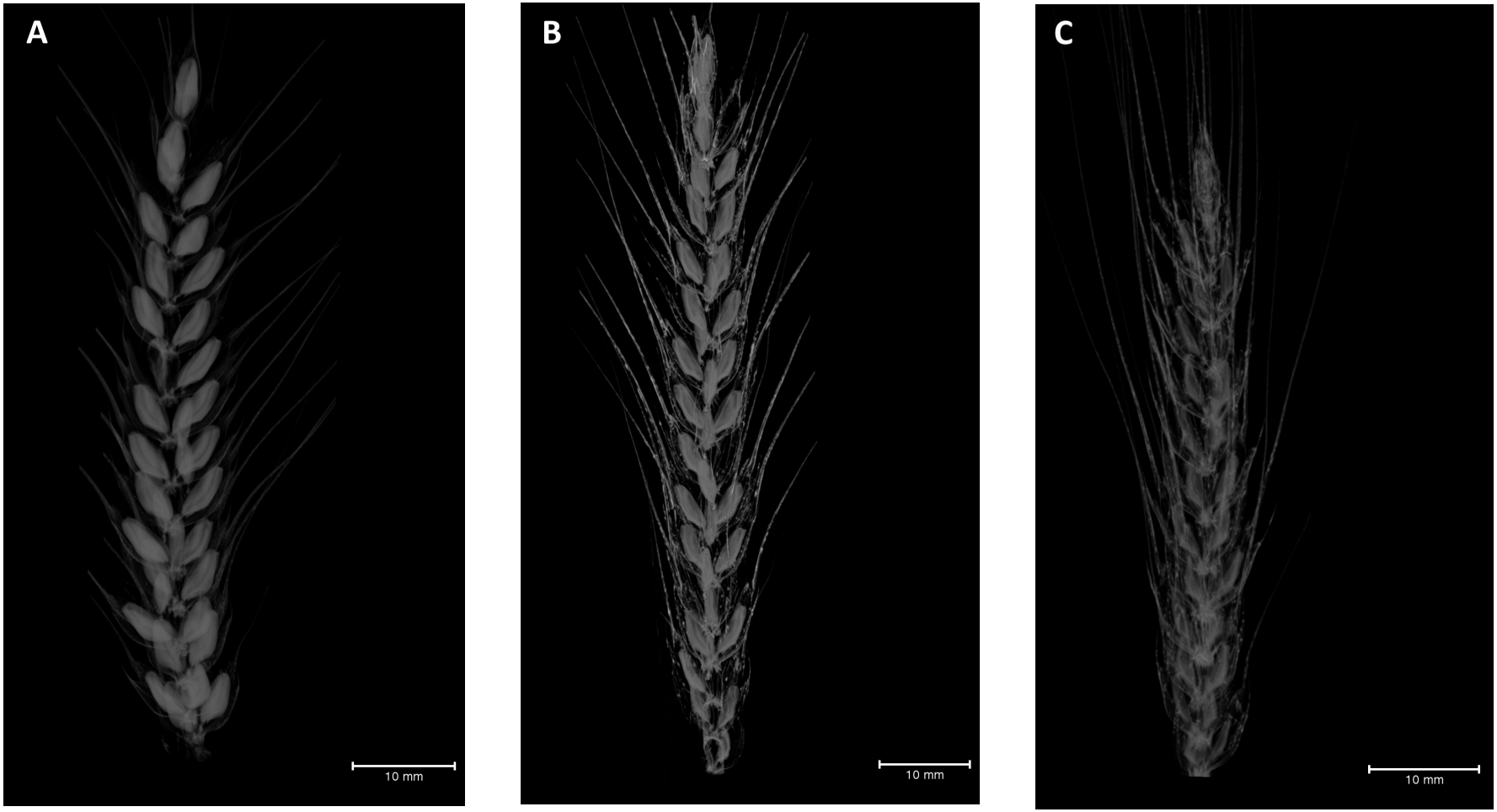


Figure 3.15. X-ray CT scans of *indTaRKD1-7D* spikes. 30 randomly chosen spikes from the spike traits trial were taken and imaged using a computerized tomography (CT) scan. Representative spikes from each genotype are shown here. **(A)** *indTaRKD1-7D* #1.4 **(B)** *indTaRKD1-7D* #2.28 **(C)** Wild type

3.2.5. Increased tiller number in two *indTaRKD1-7D* T3 lines independent of nitrogen availability

Two homozygous *indTaRKD1-7D* T3 lines were selected from independent transgenic events (#1.4 and #2.28) and grown in a randomised design with a border of wild type plants for 100 days and three nitrogen availability conditions; M2 soil with 0g added osmocote, M2 soil with 3g added osmocote and M2 soil with 6g added osmocote. Increased reproductive tiller number was observed in *indTaRKD1-7D* #1.4 independent of nitrogen availability as there was no significant difference between any of the soil conditions (Supplemental Figure 3.25). Increased reproductive tiller number was observed in *indTaRKD1-7D* #2.28 partially dependent on nitrogen availability as there was only a significant different between 0g and 6g (Supplemental Figure 3.26). Wild type showed significant differences between 0g and 3g, and 0g and 6g (Supplemental Figure 3.27, Figure 3.16). Interestingly, lower nitrogen concentrations (0g and 3g) showed greater differences in reproductive tiller number in *indTaRKD1-7D* #1.4 and #2.28 compared to wild type than at higher nitrogen concentrations (6g).

Spike traits were measured after senescence using an automated image analysis program called WERecognizer. Spike area and length showed comparable differences between the three nitrogen availability conditions across all genotypes (Supplemental Figure 3.28, Supplemental Figure 3.29, Supplemental Figure 3.30, Figure 3.17). A summary of these results can be found in Supplementary Table 3.11 and Supplementary Table 3.12. Finally, we next investigated whether increased *indTaRKD1-7D* expression through treatment of *indTaRKD1-7D #1.4* plants with the chemical inducer Estradiol in a semi-hydroponics system would lead to an increase in tiller number.

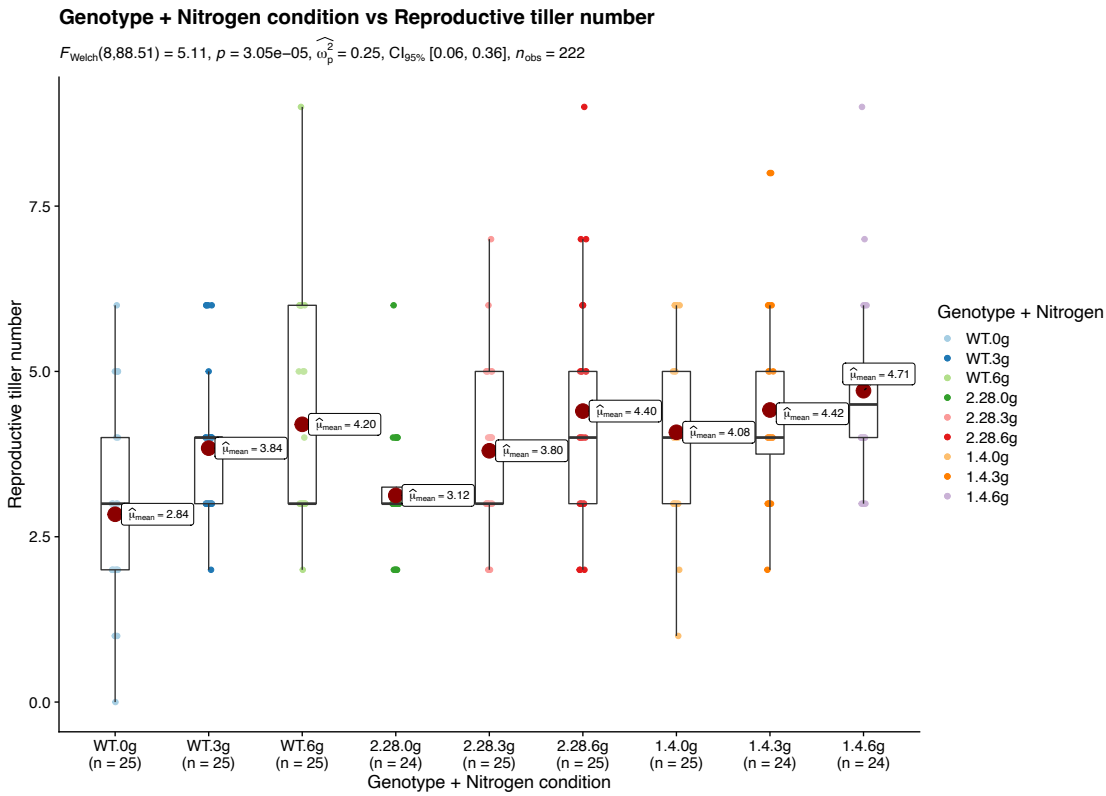


Figure 3.16. Nitrogen response trial shows two *indTaRKD1-7D* T3 lines with increased reproductive tiller number independent of nitrogen availability. Reproductive tiller number from plants measured from counting the number of fully developed spikes at senescence. 0g, 3g and 6g represent the amount of Osmocote added o M2 soil. Incremental increases can be seen across the three nitrogen conditions in WT, between 0g and 6g in 2.28 and no significant differences in 1.4. Better resolution of nitrogen response with reproductive tiller number can be seen in Supplemental Figures 3.23, 3.24 and 3.25. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Statistical significance hasn't been plotted here due to high condition number. Additional statistics can be found on the top left of the plot.

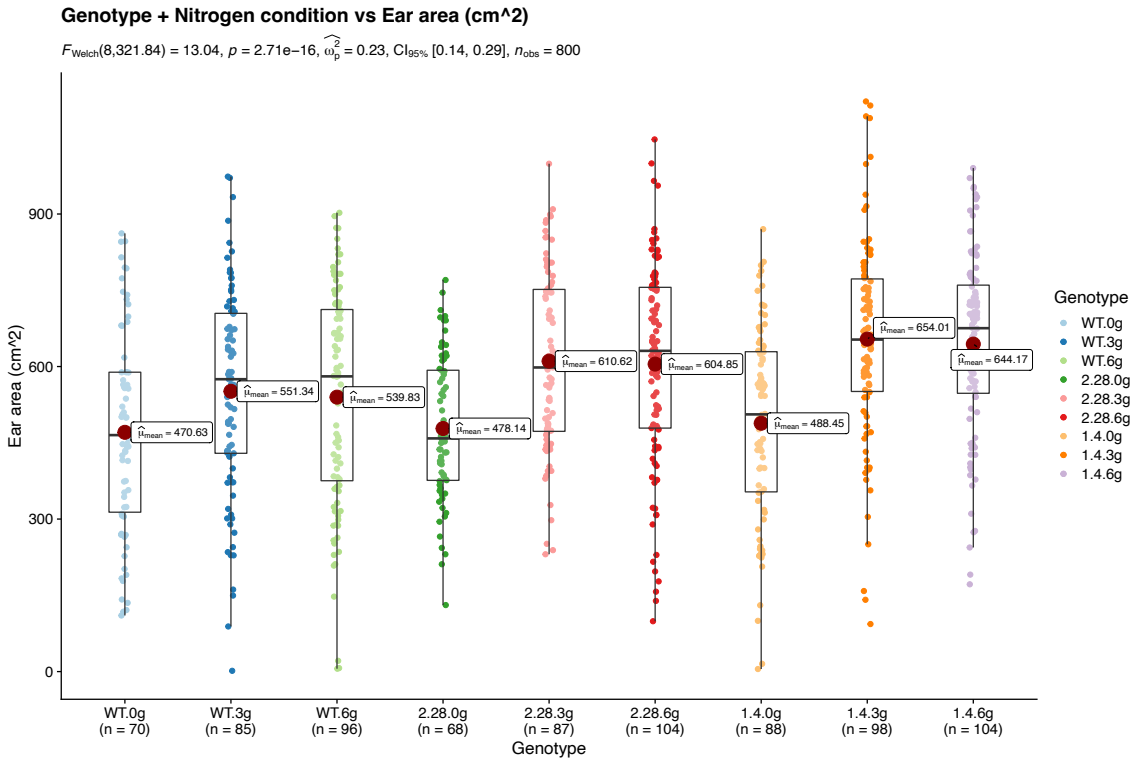


Figure 3.17. Nitrogen response trial shows two *indTaRKD1-7D* T3 lines with no increase in nitrogen use efficiency for spike area. Ear area of fully dried spikes calculated by WERecognizer. 0g, 3g and 6g describes osmocote added to the M2 soil. Comparable incremental increases can be seen across the three nitrogen conditions in all genotypes. Better resolution of nitrogen response with Ear area can be seen in Supplemental Figures 3.20, 3.21 and 3.22. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Statistical significance hasn't been plotted here due to high condition number. Additional statistics can be found on the top left of the plot.

3.2.6. No increase in tiller number for estradiol treatment of *indTaRKD1-7D #1.4*

A homozygous *indTaRKD1-7D #1.4* T3 line was grown in a semi hydroponics system for 50 days and treated with Estradiol or a mock solution. No increases in tiller number or total plant biomass were observed between *indTaRKD1-7D #1.4* treatments (Supplementary Figure 3.31, Supplementary Figure 3.32.). A summary of these results can be found in Supplementary Table 3.13 and Supplementary Table 3.14.

3.3. Discussion

3.3.1. Zygotic RKD transcription factors in nature

RWP-RK transcription factors have been extensively studied through their effects on nitrogen use efficiency and nodule developments using the NLP transcription factors and somatic embryogenesis, pluripotency, and embryo development through the zygotic RKD transcription factors (Chapter 1). Manipulation of meristem initiation and determinacy to modulate vegetative and inflorescence architecture has been previously shown through a variety of morphogenic regulators in many grass species as highlighted in the introduction of this chapter. Most of these morphogenic regulators have a role in either axillary meristem (AXM) initiation or floral meristem (FM) determinacy. Ectopic expression of zygotic RKD transcription factors in mandarins has been shown to drive somatic embryogenesis in nucellar tissue leading to a polyembryonic phenotype (Shimada et al., 2018). No evidence of a natural role in meristem formation for zygotic RKD transcription factors has been found in the grasses.

3.3.2. Characterising RKD transcription factors in wheat and its wild progenitors

Bioinformatics resources such as the wheat expression browser have provided invaluable information for investigating expression of wheat genes in a variety of tissues (Ramírez-González et al., 2018). Incomplete annotation of zygotic *TaRKDs* in IWGSC RefSeq v1.1., likely due to their low tissue-specific expression counts, means that phylogenetic and transcriptomic characterisation of this gene family had to be carried out by reanalysing publicly available developmental time course data and for a custom annotation file to be created for all subsequent analysis. In this thesis we've provided a thorough annotation of the *TaRKD*, *AetRKD* and *TdRKD* families in *Triticum aestivum*, *Aegolips tauschii*, and *Triticum dicoccum* and characterised their expression profiles in a variety of tissues from publicly available sequencing data.

3.3.3. Ectopic *TaRKD1-7D* expression increases tiller number in wheat

Here, we've shown that ectopic expression *TaRKD1-7D* in vegetative tissue increases reproductive tiller number in the wheat cultivar 'Fielder'. Additionally, it was found that vegetative tiller number could be increased even in plants with limited nutrient availability in a soil volume of 250ml. This suggests that *TaRKD1-7D* induces higher tiller number independent of intrinsic signalling of resource allocation and environmental conditions which typically temper tiller number for vigour. Nitrogen response trials showed that no significant difference in reproductive tiller number could be seen in high *TaRKD1-7D* ectopic expression regardless of nitrogen availability confirming that higher tiller number was induced independent of environmental factors. Additionally, this would allow plants to induce higher tiller number in environments with less nutrient availability (Supplementary figure 3.24 and Supplementary figure 3.26). To address the limitation of this as a potential breeding

strategy, it must be noted that increased unproductive tillers were found in high *TaRKD1-7D* ectopic expression lines and that in many field conditions it is not favourable to have phenotypes that increase tiller number regardless of environmental conditions. However, it have extensively been shown in the literature that higher tiller number phenotypes improve yield through increased nitrogen response, including ones linked to epigenomic regulation (Gooding et al., 2012; Wu et al., 2020).

3.3.4. *TaRKD1-7D* induced higher tiller number increases total plant biomass and seed weight

Ideotype breeding for higher tiller number tends to result in reduced plant height as more resources are allocated for axillary rather than apical growth (Chapter 1). We've found that high ectopic *TaRKD1-7D* expression increases plant height and total plant biomass suggesting decreased meristematic determinacy in the shoot apical meristem (SAM). So, while our main phenotype described here remains reproductive tiller number, it must be noted that this is phenotypically very different to other classic semi-dwarf high tiller number varieties in the literature such as the Rht1 mutant alleles (Allan et al., 1959). We also see high amounts of histone H4 and toluidine staining in SAM tissue like in AXM tissue suggesting ectopic *TaRKD1-7D* expression can increase the formation of AXM tissue and decrease determinacy in SAM by a similar mechanism as seen in Chapter 4. Increased seed weight per plant was observed in high and moderate ectopic *TaRKD1-7D* expression lines suggesting that despite an increase in unproductive tillers, grain yield is increased with ectopic *TaRKD1-7D* expression. Seed weight per 1000 also showed a significant increase with ectopic *TaRKD1-7D* expression however, this phenotypic trait has been shown to vary significantly in trials and may need a field trial for confirmation. Typically, it has been shown that without

increased spacing or longer growing seasons, higher tiller number leads to less seed weight per 1000 (Duggan et al., 2005; Jobson et al., 2019).

3.3.5. *TaRKD1-7D* induced higher tiller number increases spikelet number per spike

As mentioned in the introduction, inflorescence architecture is determined through a complex network of genetic and environmental factors. High and moderate ectopic expression of *TaRKD1-7D* has similar effects on spikelet number per spike with both showing a significant increase relative to wild type and a null segregant. Whether this is a downstream effect of delayed flowering or decreased meristematic determinacy which has been shown in the literature is hard to determine due to *TaRKD1-7D's* ectopic expression (Brassac et al., 2021; Glenn et al., 2021). Perhaps one way to confirm this would be to assess *TaRKD1-7D* expression in the FM, however full confirmation of *TaRKD1-7D's* as a potential meristematic determinacy regulator will require further work beyond this thesis.

3.3.6. *TaRKD1-7D* expression level dependent tiller number induction in wheat

Our failure to show an expression level dependent tiller number phenotype in wheat here is at least partially down to our choice of chemical inducer. While others have managed to use it effectively for ectopic expression it appears to be difficult to use in large and complex tissues (Chapter 4 discussion). Further confirmation of this hypothesis comes later in the thesis where we show an expression dependent tiller number phenotype in rice using the chemical inducer Dexamethasone (Chapter 5).

3.3.7. Phenotypic model for *TaRKD1-7D* induced higher tiller number

We hypothesize a model where ectopic *TaRKD1-7D* expression in pre-meristematic tissue can increase AXM initiation, formation, and maintenance through dedifferentiation of vegetative tissue, leading to an increase in vegetative tiller number. Following this, we speculate that other phenotypic traits in *TaRKD1-7D* induced higher tiller number plants such as increased plant height and increased spikelet number per spikes are caused through decreased meristematic determinacy of SAM and FM tissue respectively, through ectopic expression of *TaRKD1-7D* in these stem cell niches. This agrees with the extensively characterised role of zygotic RKD transcription factors of inducing somatic embryogenesis, a process that dedifferentiates vegetative tissues to a more pluripotent state (Waki et al., 2011). In the next chapter, we will characterise the molecular pathways of *TaRKD1-7D* in a variety of tissues to gain a greater understanding of how ectopic *TaRKD1-7D* expression increases meristem formation and determinacy in wheat.

Chapter 4

4.1. Introduction

4.1.1. Gene regulation in eukaryotes

Eukaryote development is orchestrated through a complex set of interactions between an organism's genome and environmental cues. The central dogma of molecular biology defines the two major stages of how genetic information relates to functional processes as transcription and translation. Eukaryotic transcription is regulated by three RNA polymerases and master regulators called transcription factors allowing expression profiles in tissues and cells that are spatially and temporally specific. RNA polymerases bind promoter regions along with transcription factors to transcribe complementary mRNA sequences that are translated into amino-acid sequence and folded with ribosomes rand chaperones to form proteins, the functional unit of the cell.

Eukaryotic transcription has many unique regulatory elements compared to prokaryotes. While prokaryotes have proximal promoter elements that regulate the spatial and temporal expression of genes, eukaryotes have multiple unique promoter elements such as the TATA box and downstream core elements allowing the formation of complex secondary structures not typically found in prokaryotic gene regulation (Horokoshi et al., 1990; Lee et al., 2005). Other processes such as 5' capping, polyadenylation and alternative splicing regulate which exons are translated and the half-life of an mRNA strand allowing post-transcriptional control of gene expression (Rogers and Wall, 1980; Colgan and Manley, 1997; Ramanathan et al., 2016).

Spatial and temporal expression profiles are regulated by chromatin accessibility in eukaryotes by determining whether RNA polymerases and transcription factors can bind to motifs to induce transcriptional cascades involved in biological processes allowing the complex tissue specific expression patterns that can be seen in multicellular eukaryotes [ref]. While this enables certain genes to show very restricted expression profiles limited to specific developmental stages or as responses to environmental signals, others show ubiquitous expression profiles being expressed in multiple tissues playing 'house-keeping' roles in cellular functions. Accessible regions of chromatin are called euchromatin and are typically associated with histone modifications such as acetyl groups or the well characterised H3K4me3 mark, while inaccessible regions of chromatin called heterochromatin are associated with negative marks such as H3K27me3 (Ferrari et al., 2014; Howe et al., 2017).

4.1.2. Transcriptomics and sequencing

The transcriptome is the complete set of all transcripts in a cell at a particular developmental stage or condition, and their quantity. The species of transcript isolated or excluded depends on the library preparation technique, most procedures involve poly-A tail enrichment or rRNA exclusion for a high-quality mRNA library as rRNA is estimated to make up 90% of the transcriptome (Zhao et al., 2018). Since the sequencing of the human genome in 2003 there has been extraordinary progress in genome sequencing technologies with short read sequencing providing accurate and affordable data that allows researchers to investigate molecular processes in their organism of interest.

RNAseq is commonly used for interrogating transcriptomics questions and has multiple advantages compared to previous transcriptomic profiling techniques such as microarrays and EST-libraries like lower background noise (Wang et al., 2009). While microarrays required pre-defined sequence probes to estimate quantities of transcripts, RNAseq can measure the quantity of any sequence allowing transcriptomic profiling of non-model organisms and novel isoforms of transcripts. RNAseq analysis allows investigation into multiple biological phenomena helping researchers draw conclusions based on differential transcriptome profiles between conditions such as species, genotypes, tissues, timepoints and treatments. Sample number and read coverage also play a significant role in the experimental design although most analyses are recommended to prioritise sample number over read depth unless the RNA species being investigated has low expression (Baccarella et al., 2018). Library preparation can be an important consideration too as amplification cycles of adapter ligated sequences can introduce amplification biases to transcripts with higher GC content, this has led to the development of multiple PCR-free library preparations (Aird et al., 2011).

4.1.3. Aligning reads to reference assemblies of large complex genomes

The development of advanced functional genomics in wheat to identify and characterise genes implicated in developmental and agronomically interesting processes hasn't progressed as quickly as in other cereal crops such as barley and rice due to its large and complex polyploid genome. Recent developments have resulted in the International Wheat Genome Sequencing Consortium (IWGSC) generating a high-quality annotated reference for wheat allowing molecular geneticists and breeders to identify and characterise genes implicated in wheat development to provide a greater

understanding of how we can develop novel breeding strategies to meet the increased demand of wheat yield in the 21st century (Jia et al., 2018; IWGSC, 2018). Aligning trimmed reads to a genome can be performed with splice aware alignment software such as HISAT and STAR (Dobin et al., 2013; Kim et al., 2015). With large complex genomes like wheat, we require an alternative strategy to process reads. Sailfish showed promise for this as an aligner that was less frugal with the alignment step which led to a much smaller run-time compared to other conventional aligners (Patro et al., 2014). Kallisto dropped the alignment paradigm completely by mapping trimmed reads directly to a cDNA library without processing exact alignments between the reads and the reference sequence, outputting the most probable transcript, making the process much more computationally efficient and applicable to wheat (Bray et al., 2016). This pseudoalignment strategy also has the added benefit of providing direct quantification of transcripts, significantly speeding up analysis. We've used the pseudoaligner kallisto in this chapter to quantify transcripts in a variety of wheat RNAseq experiments.

4.1.4. Coexpression network analysis

Weighted correlation network analysis has been integral to systems biology since microarray sequencing was common practice. Finding highly correlated clusters of genes (modules) allows us to construct a spatial relationship of how genes in a transcriptional cascade are regulated and annotate nodes (genes) with information like fold change from RNAseq, ChIPseq or DAPseq data and whether the node is a transcription factor or has a particular biological function from databases like Interpro. Network analysis is a strong exploratory tool for identifying potential candidates and choosing those correlated with other nodes involved in biological processes of interest.

Weighted Gene Coexpression Network Analysis (WGCNA) and GeneNet are typically used for modelling and inferring gene networks as well as less common packages such as bc3net (Kolpakov et al., 1998; Langfelder et al., 2008; de Matos Simoes and Emmert-Streib, 2016). Integrating multiple network algorithm to control for statistical biases inherent in certain approaches can be used to create high confidence correlation networks (Yu et al, 2018). Here, we've generated and analysed a high confidence gene coexpression correlation network in wheat meristematic tissue.

4.1.5. Sequencing techniques for finding transcription factor binding sites

Chromatin immunoprecipitation sequencing (ChIPseq) and DNA affinity purification sequencing (DAPseq) are two methods of interrogating the interactions between proteins and chromatin (Ren et al., 2000; Bartlett et al., 2017). ChIPseq is an in vivo method of mapping DNA-binding proteins or histone modifications across the genome in a particular chromatin accessibility context. Chromatin accessibility is a widely studied characteristic of eukaryotic genomes as its profile varies between cells, tissues, and environmental stimuli. This marks a distinct difference to DAPseq as an in vitro method of mapping of tagged DNA-binding proteins to the genome irrelevant of a tissue's chromatin context. Intersecting ChIPseq, DAPseq and RNAseq data allows high confidence identification of candidate gene implicated in transcriptional cascade induced by transcription factor binding. We've utilised all 3 of these sequencing techniques here to fully elucidate the transcriptional cascades implicated in ectopic *TaRKD1-7D* expression.

Chapter aims and hypothesis

The overall aim of this chapter is to molecularly characterise how ectopic *TaRKD1-7D* expression induces higher tiller number and potentially higher spikelet number per spike. This aim will be addressed by the analysis of several whole genome RNA-seq experiments from *indTaRKD1-7D* lines and publicly available data relevant to tillering, as well as investigating binding targets of TaRKD1-7D to elucidate its full transcriptional cascade. Additionally, RNA in situ experiments and coexpression network analysis will be performed to characterise any molecular pathways or models proposed. The question being addressed in this chapter is:

- What molecular mechanisms are involved in TaRKD1-7D induced meristem formation and indeterminacy?

By investigating and analysing bioinformatics data relevant to *indTaRKD1-7D* and tillering, molecular and phenotypic data to propose a novel strategy for how yield-determining traits are determined and can be manipulated.

4.2. Results

4.2.1 *indTaRKD1-7D #1.4* leaf tissue shows upregulation of photosynthetic genes

To investigate the molecular pathways implicated in ectopic expression of *indTaRKD1-7D* in wheat an RNAseq experiment was performed with a transgenic line that had shown high vegetative tissue expression. Whole genome RNAseq data from *indTaRKD1-7D #1.4* leaf tissue were obtained using three sample conditions: transgenic □stradiol (TE), transgenic mock ™, and non-transge□stradioldiol (NE). *TaRKD1-7D* expression was verified in all transgenic samples and replicate concordance was confirmed using principal component analysis (PCA) (Supplementary figure 4.1, Figure 4.1). Pairwise comparisons for differentially genes (DEGs) were made between *indTaRKD1-7D #1.4* and wild type samples to obtain a high confidence list of 5651 overlapping DEGs (shared DEGs) (Figure 4.1). Both TE vs NE DEGs (56% up) and TM vs NE DEGs (53% up) showed greater upregulation of DEGs (Supplemental Figure 4.2, Supplemental Figure 4.3). Shared DEGs clustered together and showed similar expression profiles across both *indTaRKD1-7D* conditions (Figure 4.2). Gene ontology (GO) term analysis was carried out on upregulated shared DEGs to test for functional enrichment of biological processes, cellular components, and molecular functions, showing significant enrichment of photosynthetic terms (Figure 4.3). Since transcriptomic differences between *indTaRKD1-7D #1.4* and wild type could have been the result of *TaRKD1-7D* upregulating photosynthetic genes or energy demands for increased vegetative tissue growth, a less differentiated pluripotent tissue, immature embryo, was chosen for further RNAseq analysis to identify molecular pathways implicated in tiller development.

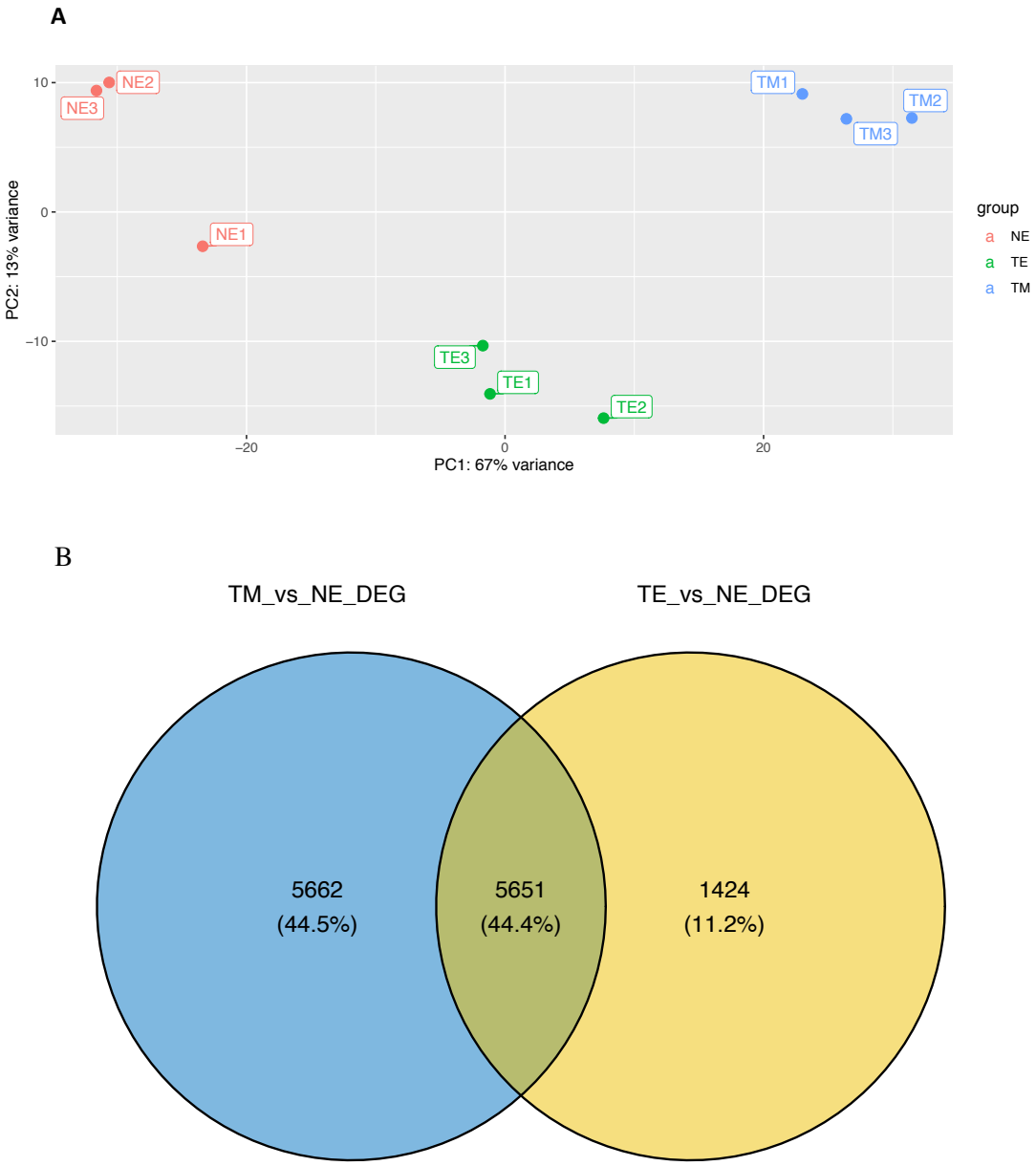


Figure 4.1. PCA of samples and Venn diagram of DEGs from *indTaRKD1-7D #1.4* Leaf RNAseq experiment. TE = Transge□stradioldiol leaf, TM = Transgenic mock leaf, and NE = non-transge□stradioldiol leaf. **(A)** Principal component analysis (PCA) of *indTaRKD1-7D #1.4* Leaf RNAseq samples coloured according to genotype and treatment condition as described in the legend on the right. Percentage of variance accounted for in each principal component is denoted beside the label. **(B)** Venn diagram of differentially expressed genes (DEGs) shared component analysis (PCA) and 2 pairwise differential comparisons TM vs NE and TE vs NE. Percentages highlight proportion of total genes between both pairwise differential comparisons.

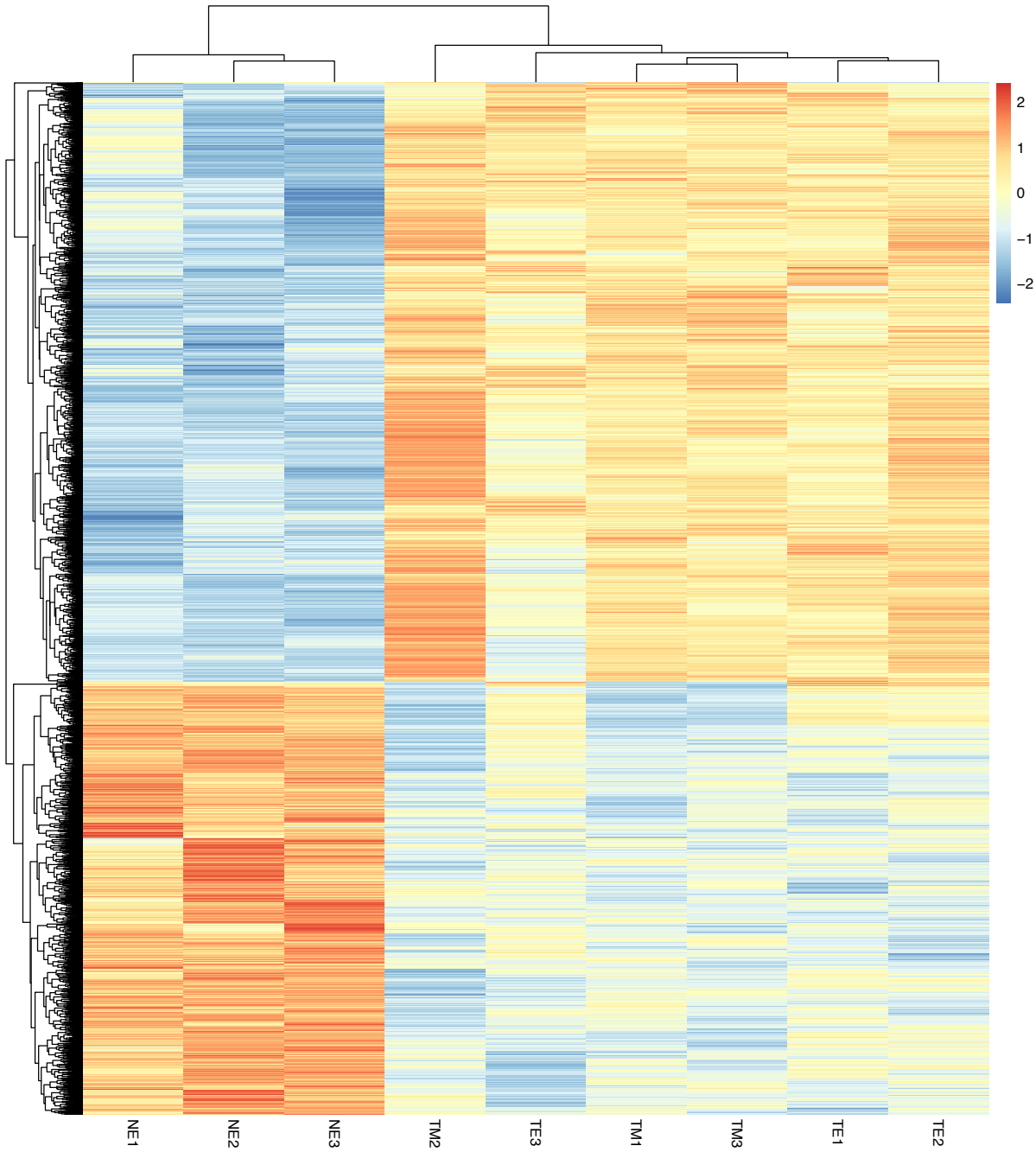


Figure 4.2. Heatmap of Shared DEGs expression from *indTaRKD1-7D #1.4* Leaf RNAseq experiment. TE = Transge□stradioldiol leaf, TM = Transgenic mock leaf, and NE = non-transge□stradioldiol leaf. Scale displayed is row scaled logarithmically transformed DESeq2 normalised count values from Kallisto quant. Hierarchical clustering was performed on *indTaRKD1-7D #1.4* leaf shared DEGs and is displayed as a dendrogram on the side of the heatmap. Hierarchical clustering was also performed on samples and is displayed as a dendrogram on the top of the heatmap.

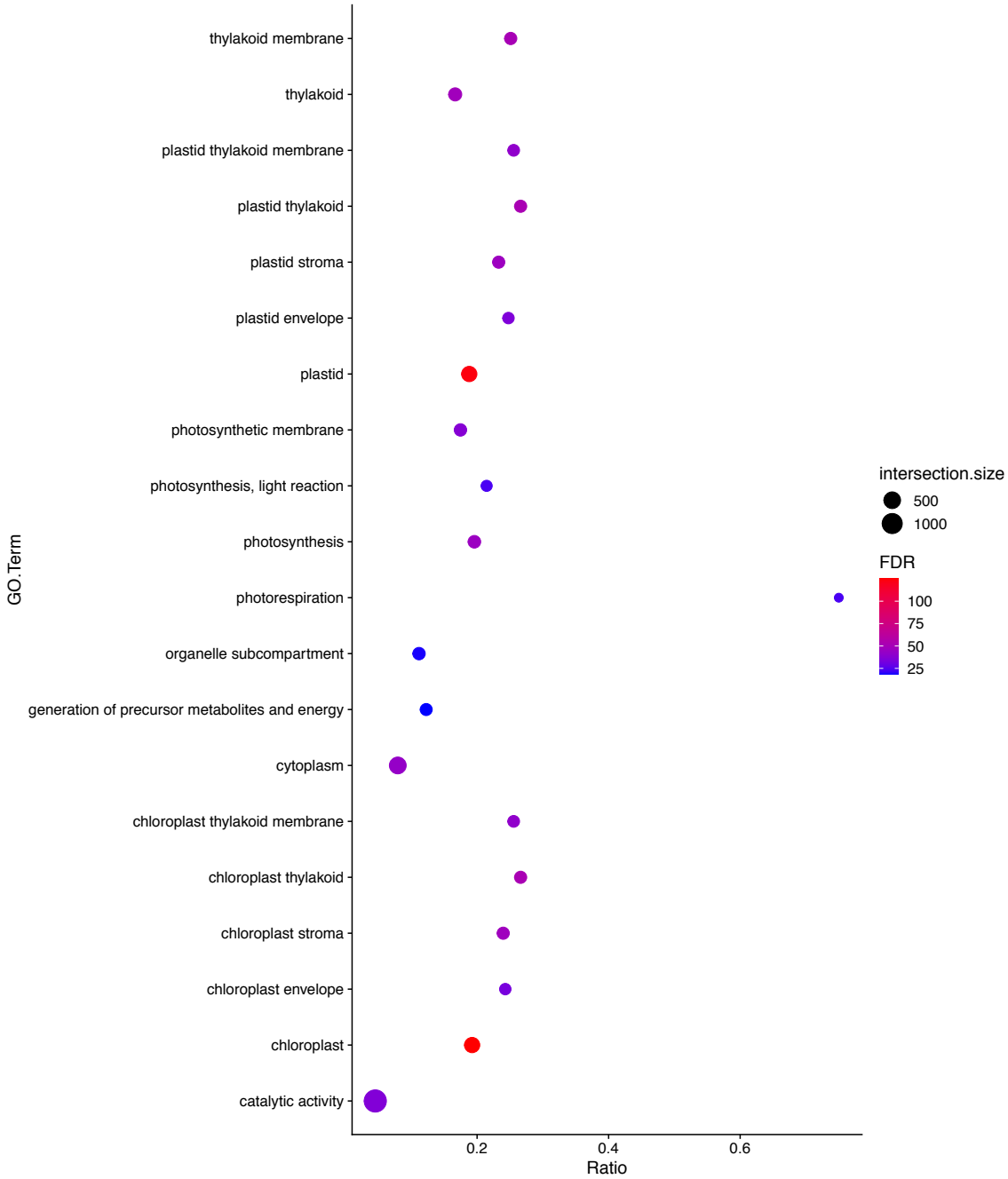
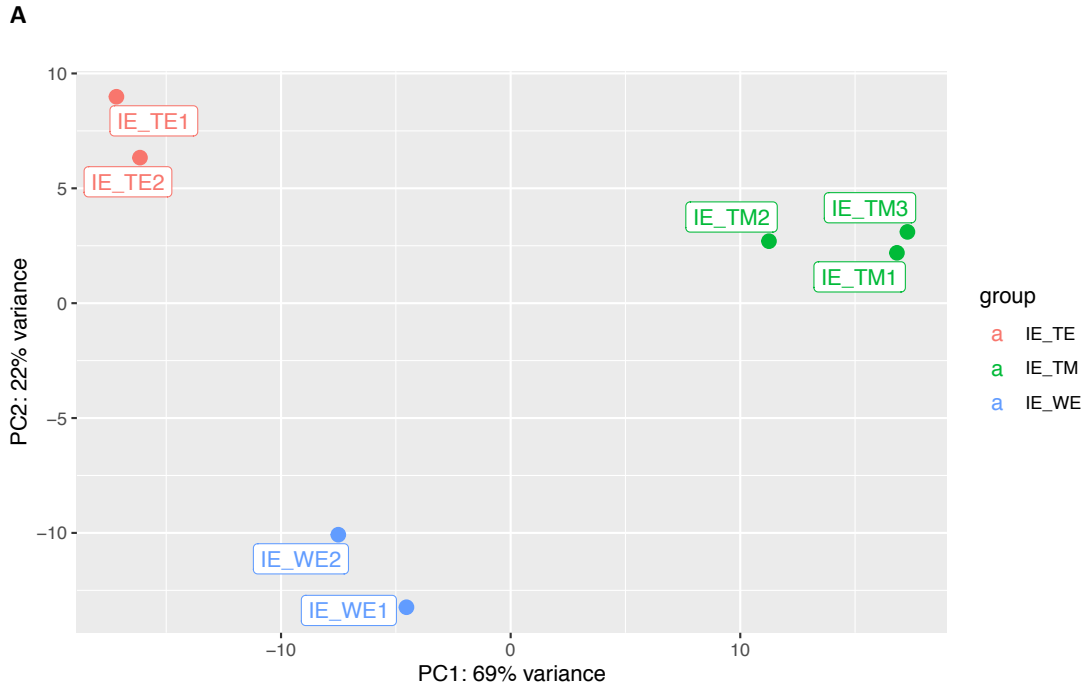


Figure 4.3. GO term analysis of upregulated *indTaRKD1-7D #1.4* leaf shared DEGs. The top 20 most significant gene ontology (GO) terms from the upregulated *indTaRKD1-7D #1.4* leaf shared DEGs are plotted here. GO terms are labelled on the y axis and the ratio of genes for a GO term in the upregulated *indTaRKD1-7D #1.4* leaf shared DEGs compared to the total number of genes corresponding to that GO term in the genome is plotted on the x axis. Point size is determined by the number of genes from that GO term in the *indTaRKD1-7D #1.4* leaf shared DEGs as denoted on the legend on the right. False discovery rate (FDR) is computed by taking the $-\log_{10}$ of the p.value determined from an enrichment test with a Benjimini-Hochberg correction and is coloured as denoted on the legend on the right.

4.2.2. *indTaRKD1-7D #2.28* immature embryo tissue shows upregulation of chromatin and nucleosome assembly genes

To investigate the molecular pathways directly implicated in *indTaRKD1-7D* induced higher meristem formation and tiller number, an RNAseq experiment was performed with immature embryo tissue with a transgenic line that had shown high immature tissue expression. Whole genome RNAseq data from *indTaRKD1-7D #2.28* immature embryo tissue were obtained using three sample conditions: transge□stradioldiol (TE), transgenic m™(TM), and wi□stradiolstradiol (WE). *TaRKD1-7D* expression was verified in all transgenic samples and replicate concordance were confirmed using principal component analysis (PCA) (Supplementary Figure 4.4, Figure 4.4). Pairwise comparisons for differentially expressed genes (DEGs) were made between *indTaRKD1-7D #2.28* and wild type to obtain a high confidence list of 2073 overlapping DEGs (shared DEGs). Both TE vs WE DEGs (75% up) and TM vs WE DEGs (61% up) showed greater upregulation of DEGs (Supplemental Figure 4.5, Supplemental Figure 4.6). Shared DEGs clustered together and showed similar expression profiles across both *indTaRKD1-7D* conditions (Figure 4.5). Gene ontology (GO) term analysis was carried out on upregulated shared DEGs to test for functional enrichment of biological processes, cellular components, and molecular functions, showing significant enrichment of chromatin and nucleosome assembly terms (Figure 4.6). We also tested for differential expression of transcription factor (TF) families in *indTaRKD1-7D* samples and found significant upregulation of the stem cell niche promoting TFs BBR/BPC and significant downregulation of abiotic stress-related TFs such as AP2/ERFs (Mizoi et al., 2012; Theune et al., 2019) (Supplementary Figure 4.7). This was consistent with a model where ectopic *indTaRKD1-7D* expression was inducing higher cell proliferation in axillary meristem

tissue or increased the formation of meristematic stem cell niches that later developed into tillers. Additionally, the well characterised role of RKDs in the literature as zygotic TFs that promote the formation of dedifferentiated tissue through activation of cell division and epigenetic regulators supported this model. To investigate whether these molecular pathways were restricted to *indTaRKD1-7D* induced higher tiller number, we decided to analyse other high tiller number phenotypes in near isogenic lines (NILs) from publicly available RNAseq data from tiller nodes.



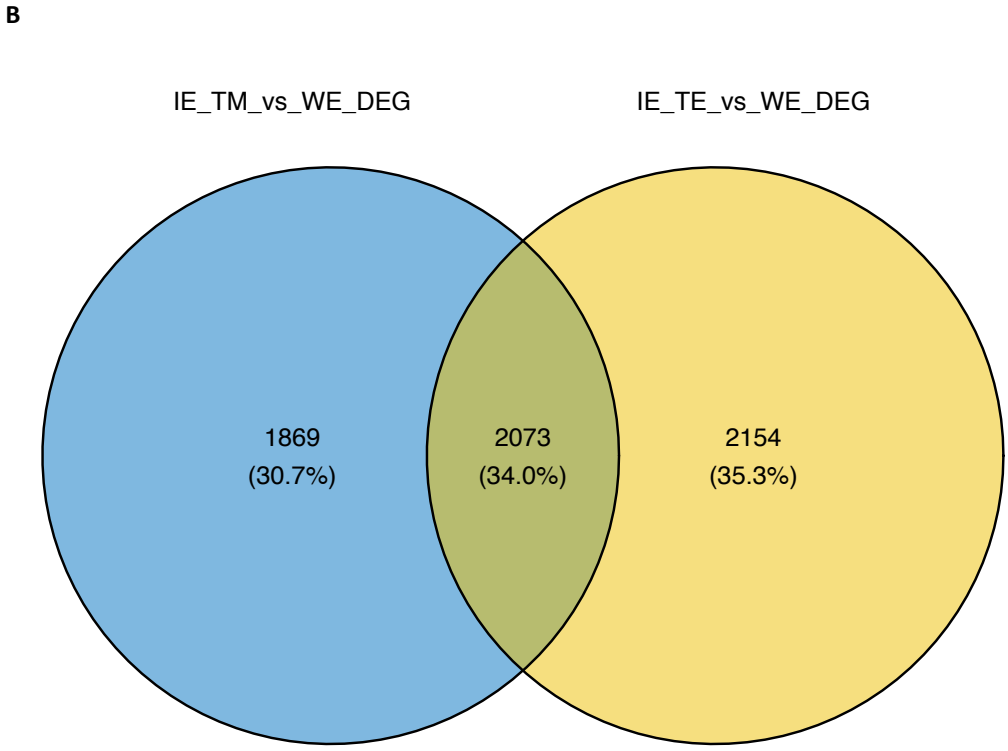


Figure 4.4. PCA of samples and Venn diagram of DEGs from *indTaRKD1-7D #2.28* Immature embryo RNAseq experiment. IE_TE = Immature embryo transgenic estradiol, IE_TM = Immature embryo transgenic mock, and IE_WE = Immature embryo wild type estradiol. **(A)** Principal component analysis (PCA) of *indTaRKD1-7D #2.28* Immature embryo RNAseq samples coloured according to genotype and treatment condition as described in the legend on the right. Percentage of variance accounted for in each principal component is denoted beside the label. **(B)** Venn diagram of differentially expressed genes (DEGs) shared between the pairwise differential comparisons IE_TM vs IE_WE and IE_TE vs IE_WE. Percentages highlights the proportion of total genes between both pairwise differential comparisons.

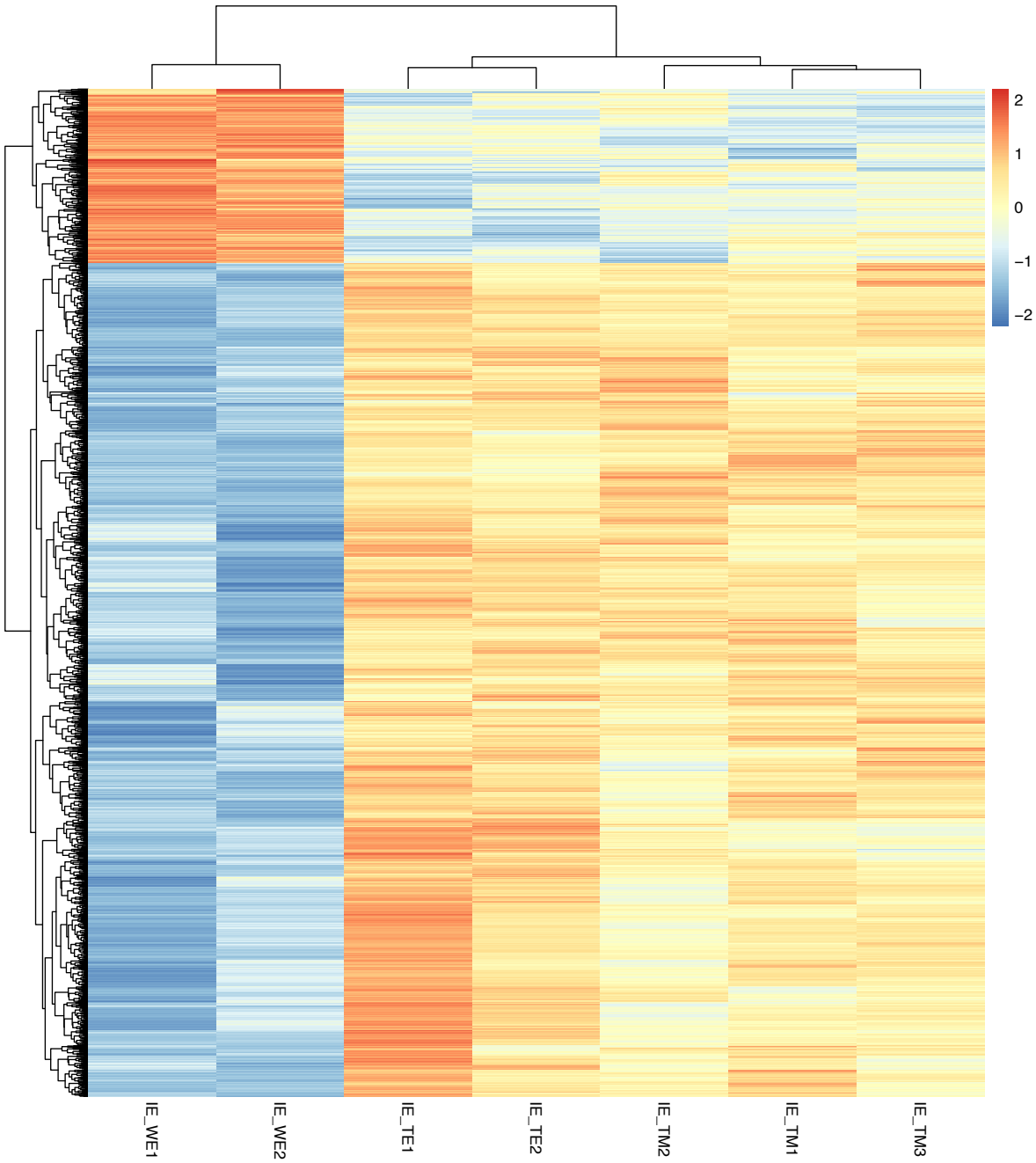


Figure 4.5. Heatmap of Shared DEGs expression from *indTaRKD1-7D #2.28* immature embryo RNAseq experiment. IE_TE = Immature embryo transgenic estradiol, IE_TM = Immature embryo transgenic mock, and IE_WE = Immature embryo wild type estradiol. Scale displayed is row scaled logarithmically transformed DESeq2 normalised count values from Kallisto quant. Hierarchical clustering was performed on *indTaRKD1-7D* immature embryo shared DEGs and is displayed as a dendrogram on the side of the heatmap. Hierarchical clustering was also performed on samples and is displayed as a dendrogram on the top of the heatmap.

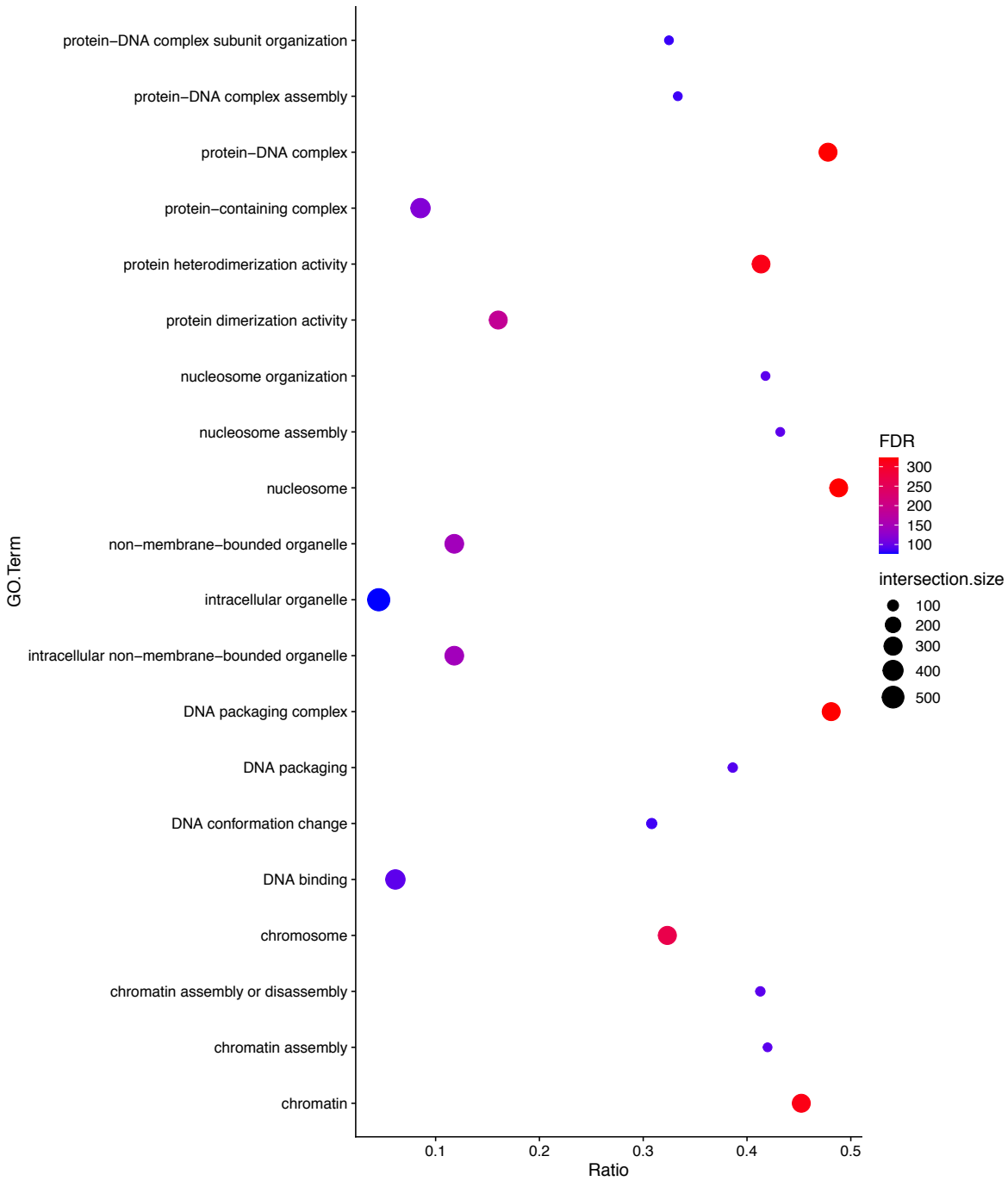


Figure 4.6. GO term analysis of upregulated *indTaRKD1-7D #2.28* immature embryo shared DEGs. The top 20 most significant gene ontology (GO) terms from the upregulated *indTaRKD1-7D #2.28* immature embryo shared DEGs are plotted here. GO terms are labelled on the y axis and the ratio of genes for a GO term in the Shared DEGs compared to the total number of genes corresponding to that GO term in the genome is plotted on the x axis. Point size is determined by the number of genes from that GO term in the upregulated *indTaRKD1-7D #2.28* immature embryo shared DEGs as denoted on the legend on the right. False discovery rate (FDR) is computed by taking the -log10 of the p.value determined from an enrichment test with a Benjimini-Hochberg correction and is coloured as denoted on the legend on the right.

4.2.3. NIL tiller node tissue shows upregulation of functional terms found in *indTaRKD1-7D* leaf and immature embryo tissue

To investigate whether indTaRKD1-7D induced higher tiller number shared molecular pathways with other high tiller number phenotypes we reanalysed publicly available RNAseq data from tiller nodes. Tiller node samples were taken 30 days after sowing from two pairs of two near isogenic lines, NIL 1A (free tillering – wild type) and NIL 1B (low tillering - *Qltn.sicau-2D*), and NIL7A (free tillering – wild type) and NIL7B (low tillering - *Qltn.sicau-2D*), that differed with the integration of the low tillering *Qltn.sicau-2D* allele into the free tillering CN16 Chinese spring cultivar (Wang et al., 2016a). This was seen as an appropriate tissue and genotypic context to investigate what genes were upregulated in other high tiller number phenotypes. Whole genome RNAseq data from NIL tiller node tissue were downloaded and analysed. Replicate concordance was confirmed using principal component analysis (PCA) (Figure 4.7). Pairwise comparisons for differentially expressed genes (DEGs) were made between high tiller number NILs and low tiller number NILs from their respective crosses to a obtain high confidence list of 4914 overlapping DEGs (shared DEGs). Shared DEGs clustered together and showed similar expression profiles across both high tiller number conditions (Figure 4.8). Gene ontology (GO) term analysis was carried out on upregulated shared DEGs to test for functional enrichment of biological processes, cellular components, and molecular functions, showing significant enrichment photosynthetic terms such as thylakoid and chromatin/nucleosome assembly terms like DNA packaging complex (Figure 4.9). We decided to test for differential expression of transcription factor (TF) families in all three analysed RNAseq experiments and found two distinct clades of TF differential expression formed; an upregulation clade with TFs involved in pluripotency and cell cycle such as E2F/DF,

RB and BBR/BPC and a downregulation clade with TFs involved in stress response such as tify, WRKY and AP2/ERF (Ren et al., 2002; Chunge et al., 2009; Mizoi et al., 2012; Phukan et al., 2016; Theune et al., 2019) (Figure 4.10). This provided further evidence that cell cycle pathways were implicated in transcriptional cascades implicated in tiller development and not restricted to indTaRKD1-7D induced higher tiller number.

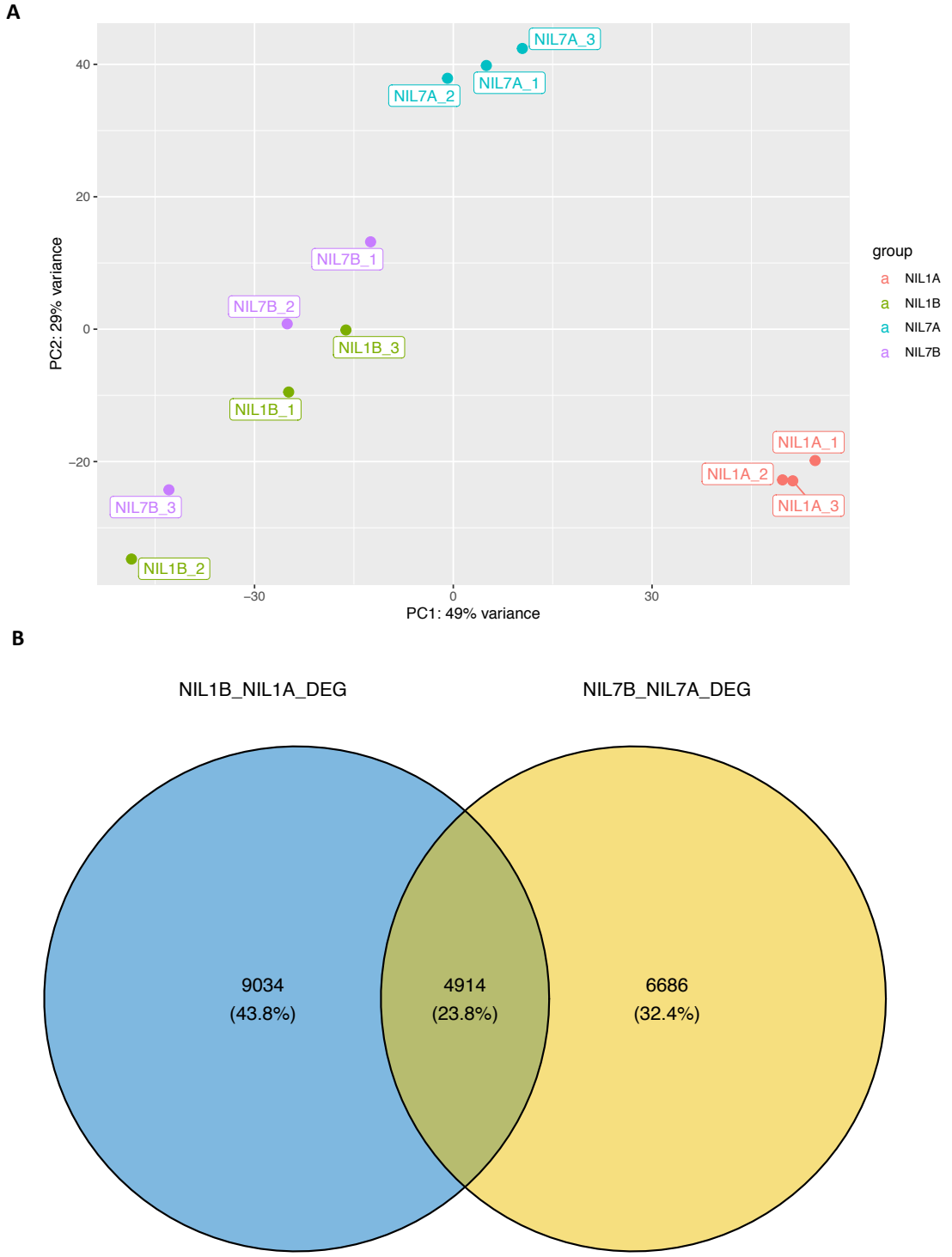


Figure 4.7. PCA of samples and Venn diagram of DEGs from NIL tiller node RNAseq experiment. NIL1B = NIL1B tiller node (low tillering - Qltn.sicau-2D), NIL7B = NIL7B tiller node (low tillering - Qltn.sicau-2D), NIL1A = NIL1A tiller node (free tillering – wild type), NIL7A tiller node (free tillering – wild type). **(A)** PCA of NIL tiller node samples coloured according to genotype and treatment condition as described in the legend on the right. Percentage of variance accounted for in each principal component is denoted beside the label. **(B)** Venn diagram of DEGs shared between the pairwise differential comparisons NIL1A vs NIL1B and NIL7A vs NIL7B. Percentages highlight proportion of total genes between both pairwise differential comparisons.

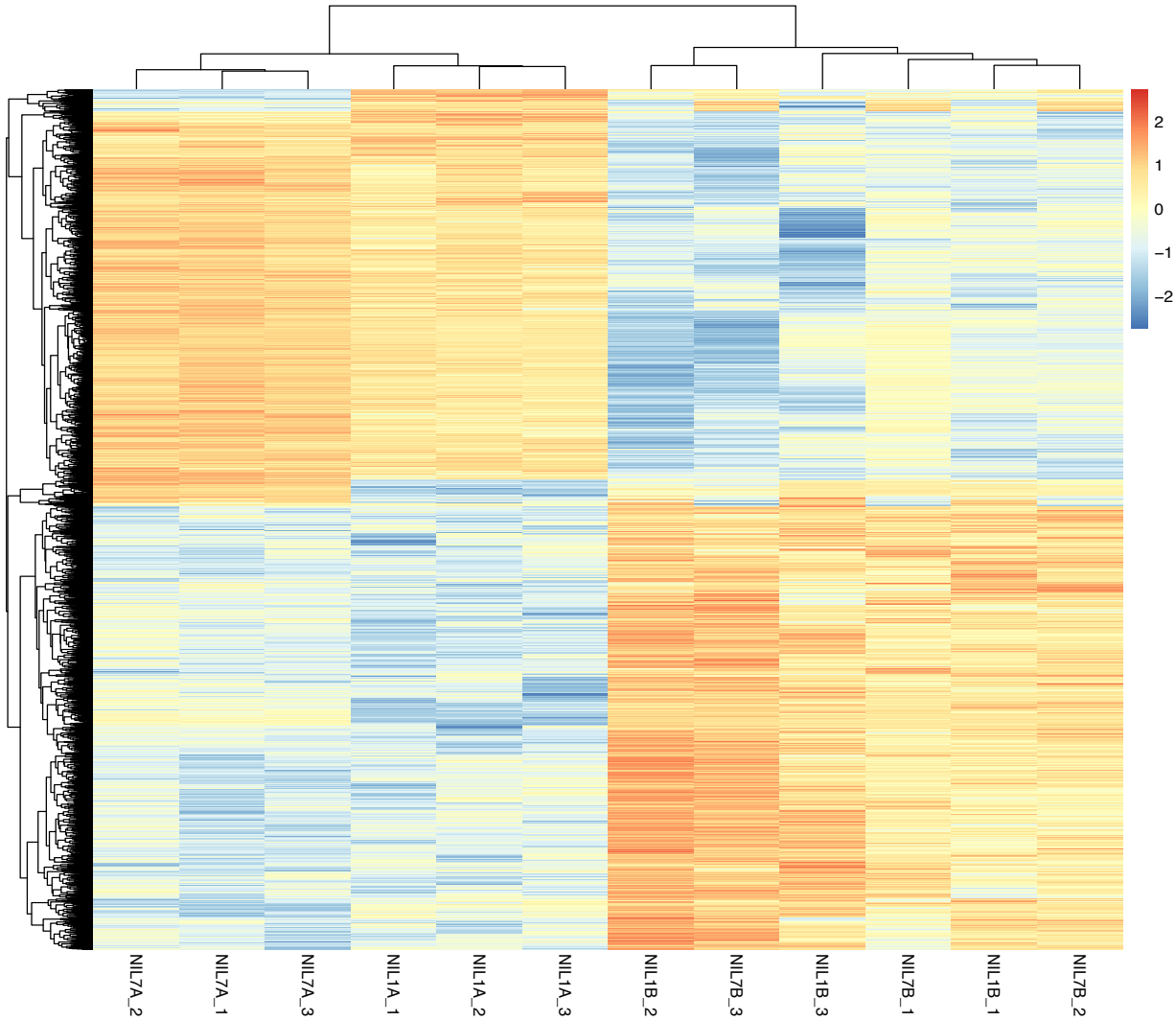


Figure 4.8. Heatmap of Shared DEGs expression from NIL tiller node RNAseq experiment. NIL1B = NIL1B tiller node (low tillering - Qltn.sicau-2D), NIL7B = NIL7B tiller node (low tillering - Qltn.sicau-2D), NIL1A = NIL1A tiller node (free tillering – wild type), NIL7A tiller node (free tillering – wild type). Scale displayed is row scaled logarithmically transformed DESeq2 normalised count values from Kallisto quant. Hierarchical clustering was performed on NIL tiller node shared DEGs and is displayed as a dendrogram on the side of the heatmap. Hierarchical clustering was also performed on samples and is displayed as a dendrogram on the top of the heatmap.

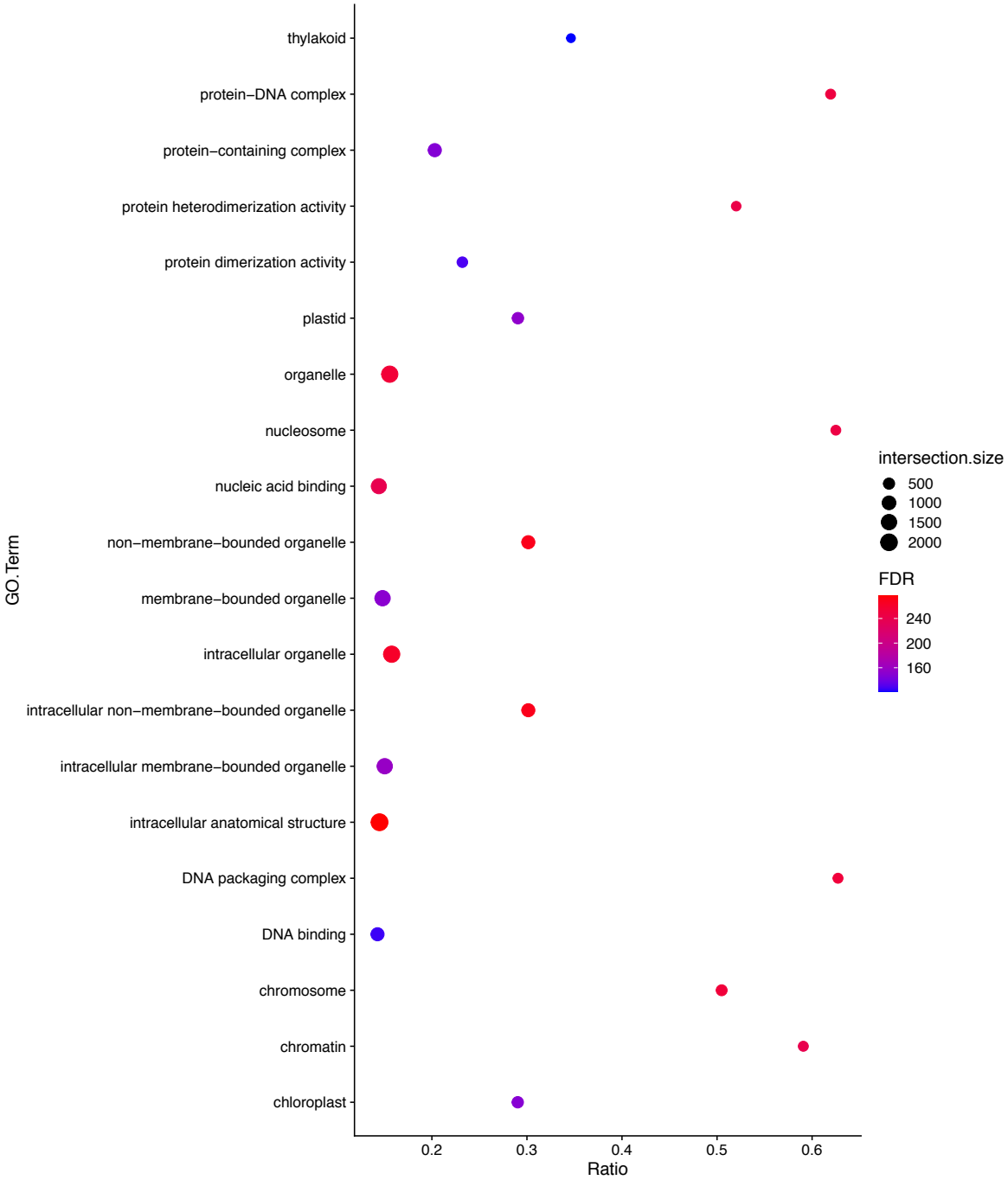


Figure 4.9. GO term analysis of upregulated NIL tiller node shared DEGs. The top 20 most significant gene ontology (GO) terms from the upregulated NIL tiller node shared DEGs are plotted here. GO terms are labelled on the y axis and the ratio of genes for a GO term in the upregulated NIL tiller node shared DEGs compared to the total number of genes corresponding to that GO term in the genome is plotted on the x axis. Point size is determined by the number of genes from that GO term in the upregulated NIL tiller node shared DEGs as denoted on the legend on the right. False discovery rate (FDR) is computed by taking the -log10 of the p.value determined from an enrichment test with a Benjimini-Hochberg correction and is coloured as denoted on the legend on the right.

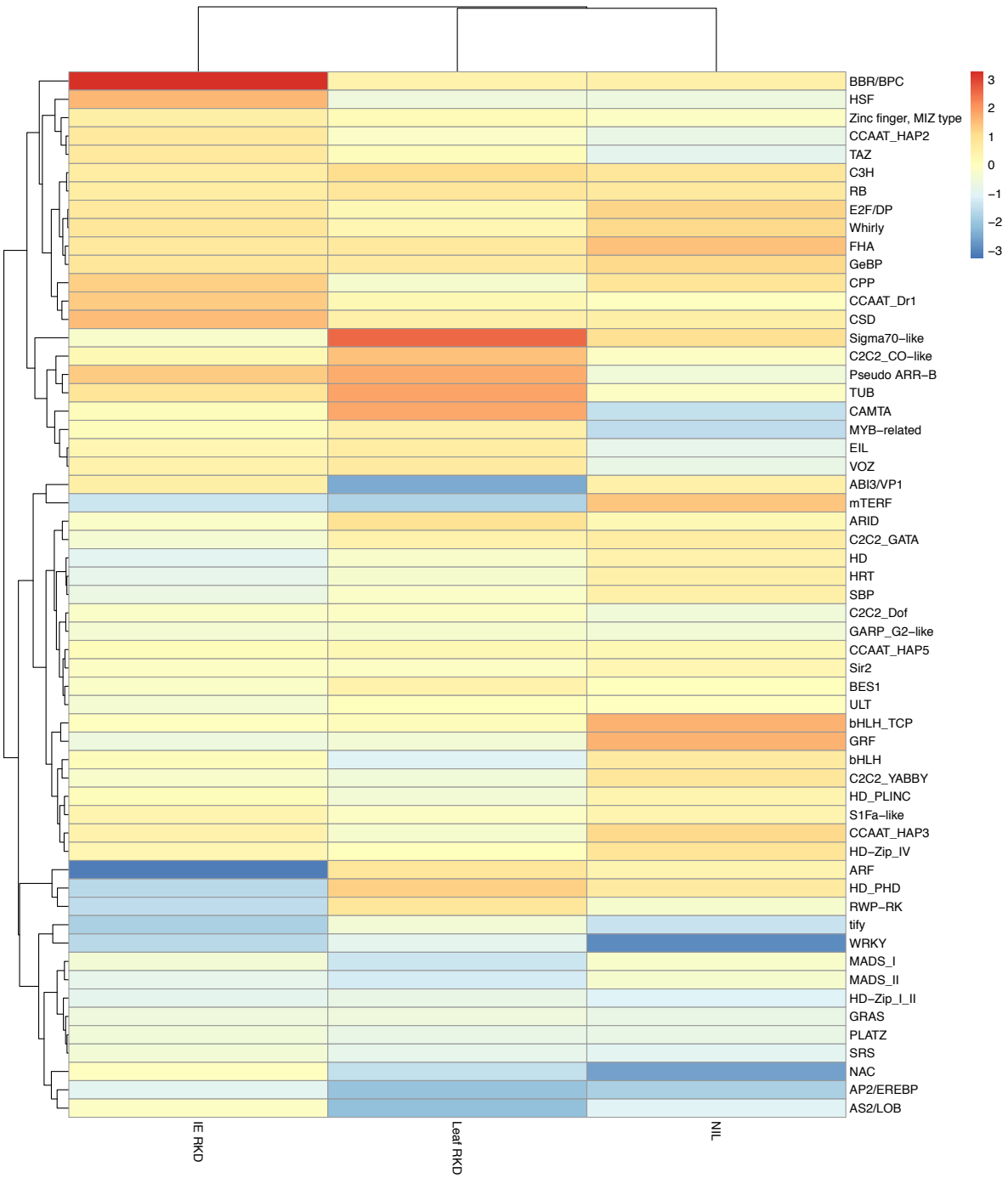


Figure 4.10. Heatmap of TF family differential regulation enrichment. Differential regulation of transcription factor (TF) families from the *indTaRKD1-7D* #2.28 immature embryo (IE RKD), indTaRKD1-7D #1.4 leaf (Leaf RKD) and NIL tiller node (NIL) RNAseq experiments. Two distinct clades of TF differential expression profiles can be seen here: an upregulated clade (top) and a downregulated clade (bottom). Scale displayed is column scaled logarithmically transformed differential values from gage. Hierarchical clustering was performed on NIL tiller number shared DEGs and is displayed as a dendrogram on the side of the heatmap. Hierarchical clustering was also performed on samples and is displayed as a dendrogram on the top of the heatmap.

4.2.4. An upregulated *indTaRKD1-7D #2.28* immature embryo shared DEGs coexpression network shows enrichment of histone H4 and HMG genes

Based on our transcriptomics analyses we hypothesized that ectopic expression of *indTaRKD1-7D* was increasing axillary meristem formation through chromatin remodelling around cell cycle genes. An RNA in-situ of tiller nodes using a histone H4 probe was performed as histone H4 is highly expressed in actively dividing cells. Increased histone H4 localisations was observed 14 days after sowing and toluidine blue staining showed a higher number of axillary meristems present in *indTaRKD1-7D* tissue (Figure 4.11). Additionally, immunolocalization of TaRKD1-7D using a TaRKD1-7D antibody was carried out in axillary meristem tissue to confirm colocalization of both TaRKD1-7D and histone H4 in developing axils (Figure 4.12). To investigate the molecular pathways induced by TaRKD1-7D implicated in cell cycle and chromatin remodelling in meristematic tissue, a coexpression network was constructed and analysed using *indTaRKD1-7D #2.28* immature embryo shared DEGs and publicly available meristematic tissue timecourse data. This network was filtered for significant coexpression correlations with at least one *indTaRKD1-7D* #2.28 immature embryo shared DEG and seperated into coexpression clusters. GO term analysis of coexpression clusters showed significant enrichment of functional terms implicated in DNA conformational change, RNA processing and cell cycle (Figure 4.13). This network is available for visualisation at nDEX under 'Immature embryo upregulated DEGs Meristem GO term'. Redundant GO terms in the two largest clusters were removed and visualised in semantic space showing terms for DNA replication and cytokinin signalling in the RNA processing cluster and terms for histone conformational change cluster (Figure 4.14, Figure 4.15).

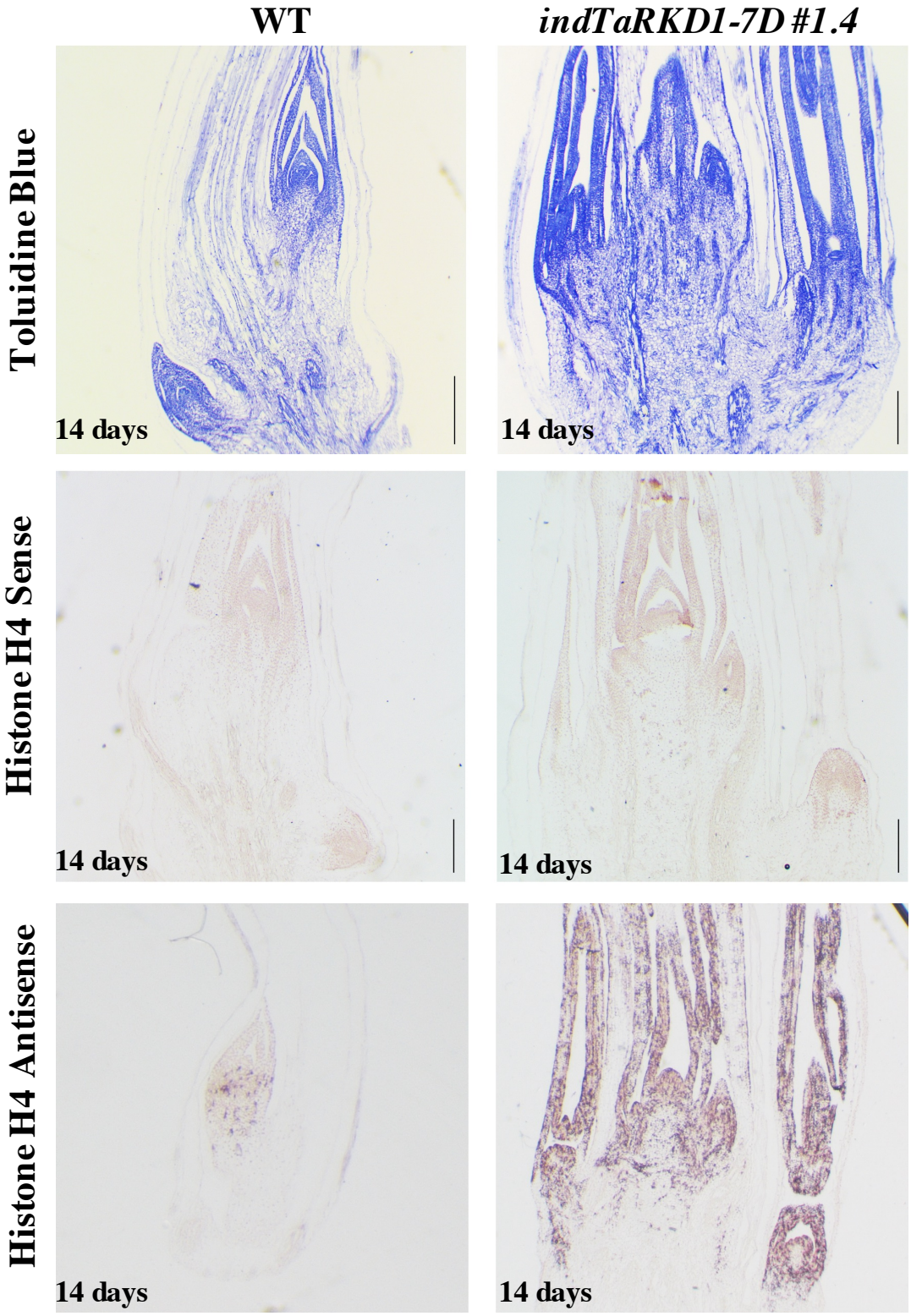


Figure 4.11. Toluidine staining and histone H4 RNA in-situ of *indTaRKD1-7D #1.4* tiller nodes shows enrichment of cell division and histone H4. Tiller nodes from *indTaRKD1-7D #1.4* and wild type (WT) plants were taken 14 days after sowing and stained with Toluidine. RNA in-situ was performed with *indTaRKD1-7D #1.4* and WT tiller nodes at 14 days and at 32 days with wild type only to correct for delayed tiller node formation.

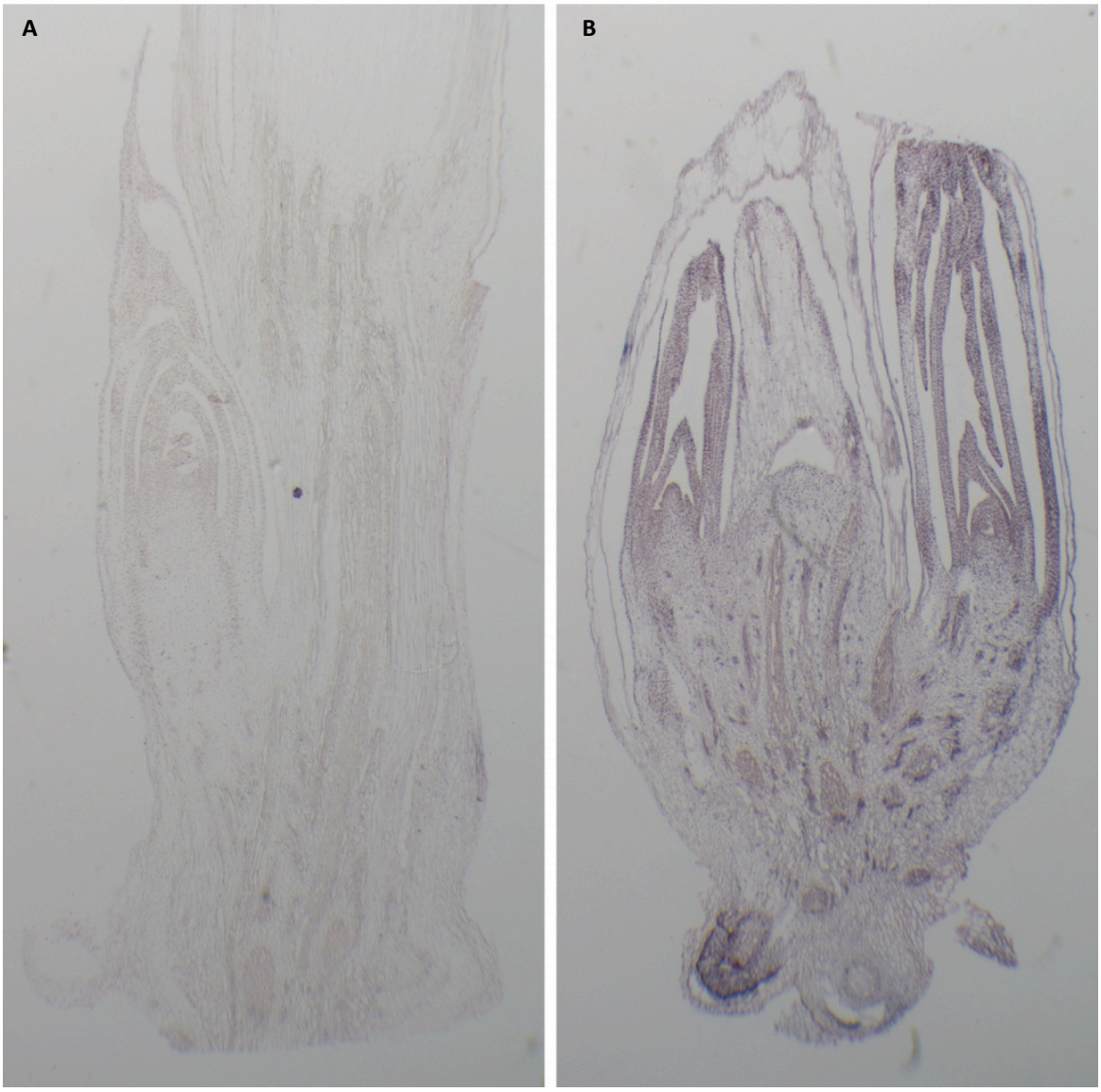


Figure 4.12. TaRKD1-7D protein immunolocalization in axillary meristem tissue. Immunodetection of TaRKD1-7D using a TaRKD1-7D antibody on longitudinal sections of developing axils. **A)** Wild type developing axil. **B)** *indTaRKD1-7D* #1.4 developing axil.

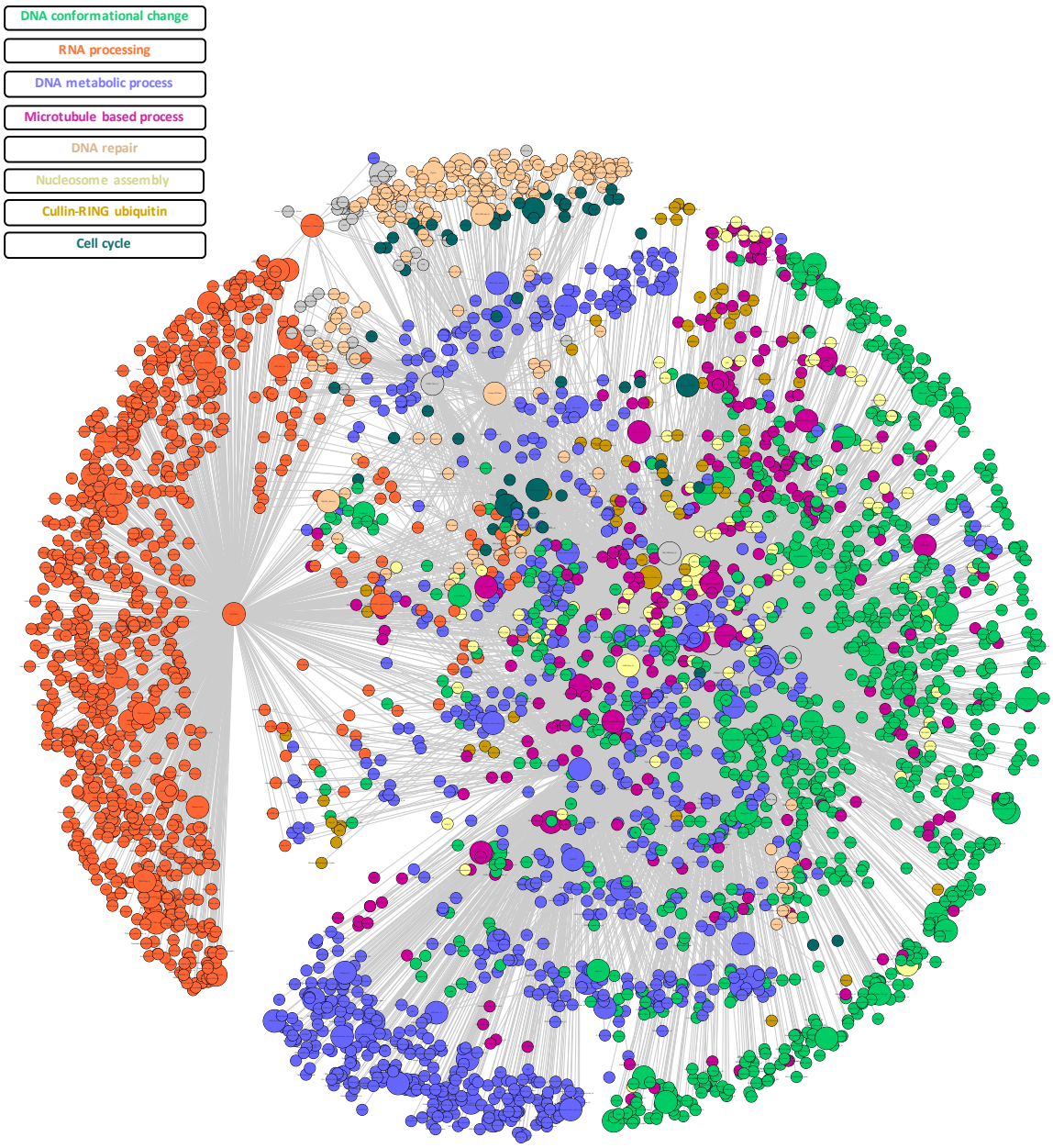


Figure 4.13. Upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network clustered and functionally annotated. Nodes (genes) and edges (coexpression) are spatially arranged with an edge-directed springboard layout in cytoscape and clustered using AutoAnnotate. Larger nodes represent transcription factors. Each colour represents a cluster with a significant GO term that can be found in its representative colour in the top left.

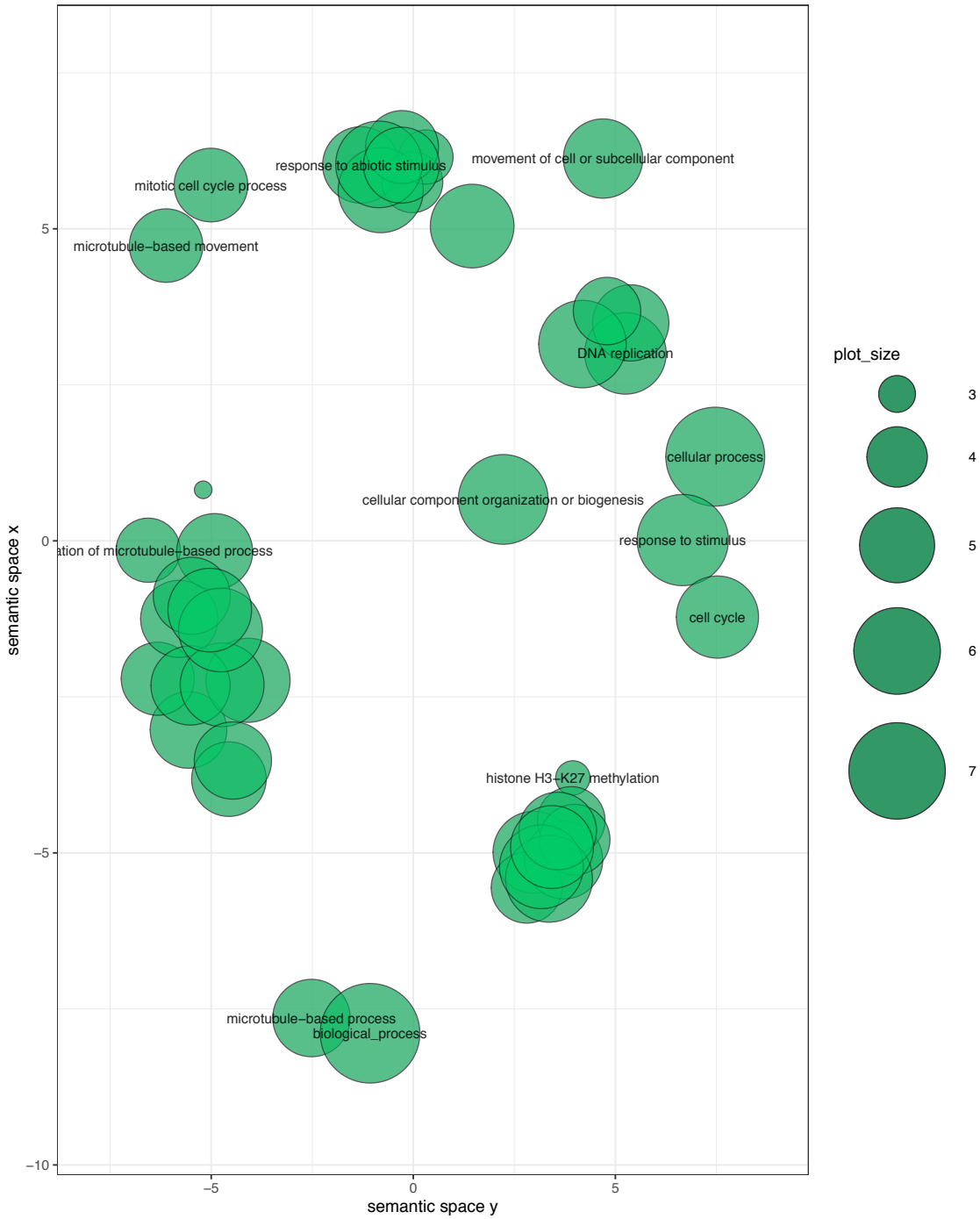


Figure 4.14. Revigo analysis of GO terms from RNA processing cluster. GO terms coloured as found in the red 'RNA processing' cluster of Figure 4.12 plotted in semantic space based on GO term similarity with the term size represented on the right of the plot.

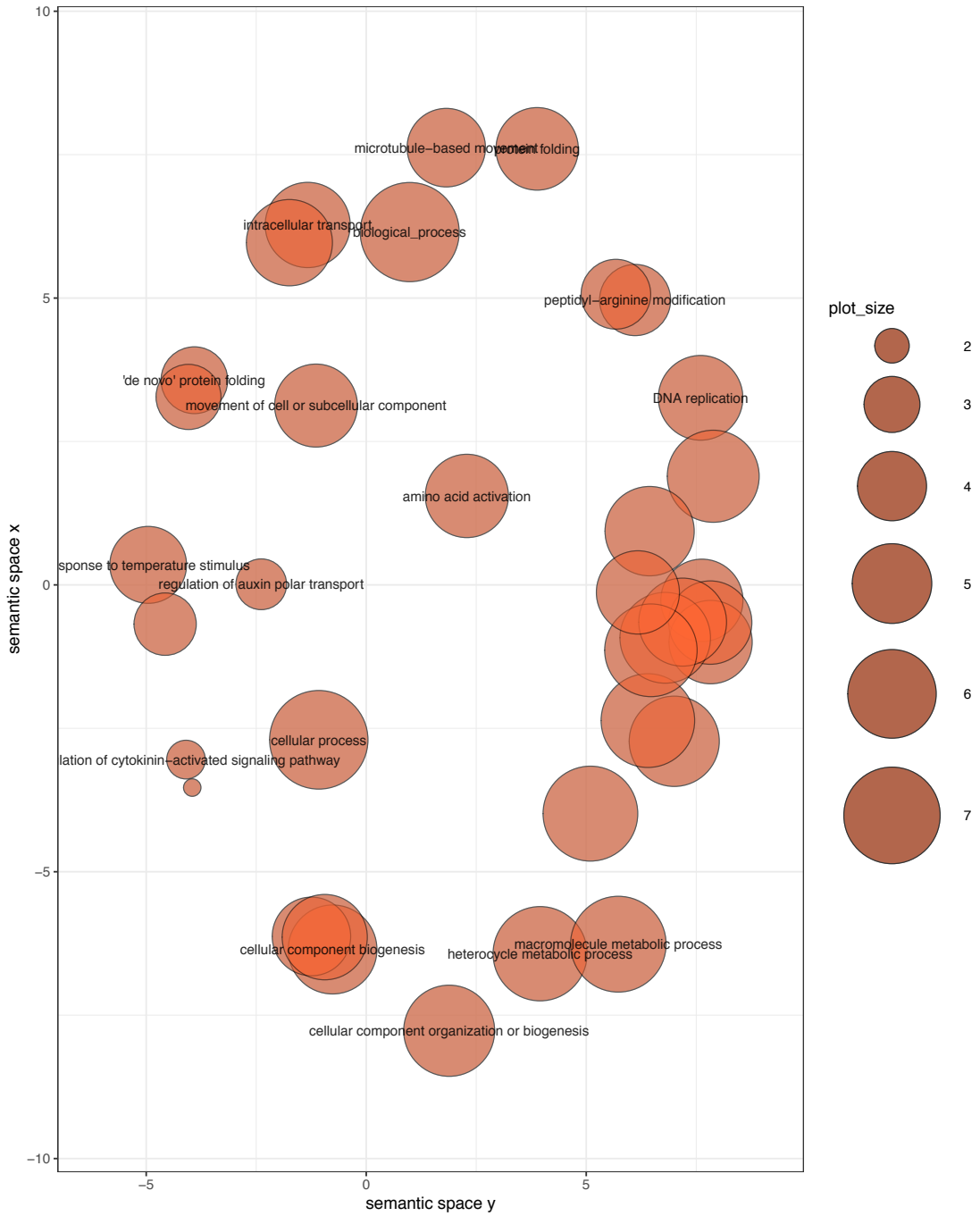


Figure 4.15. Revigo analysis of GO terms from RNA processing cluster. GO terms coloured as found in the green 'DNA conformational change' cluster of Figure 4.12 plotted in semantic space based on GO term similarity with the term size represented on the right of the plot.

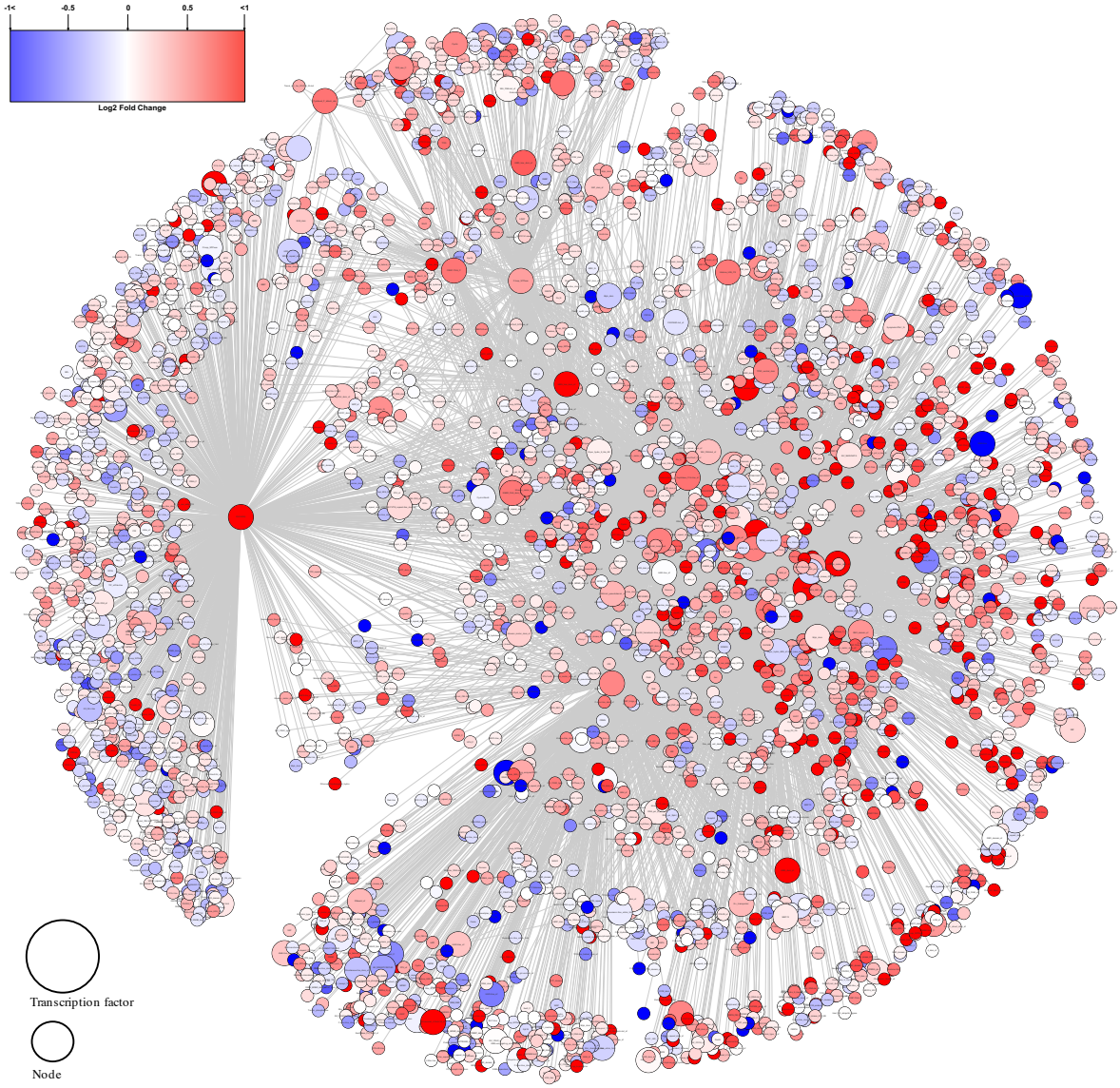


Figure 4.16. Upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network annotated by differential expression. Nodes (genes) and edges (coexpression) are spatially arranged with an edge-directed springboard layout in cytoscape. Larger nodes represent transcription factors. Nodes are coloured by log2 fold change in expression values from the *indTaRKD1-7D #2.28* IE_TM vs IE_WE comparison as denoted in the top left.

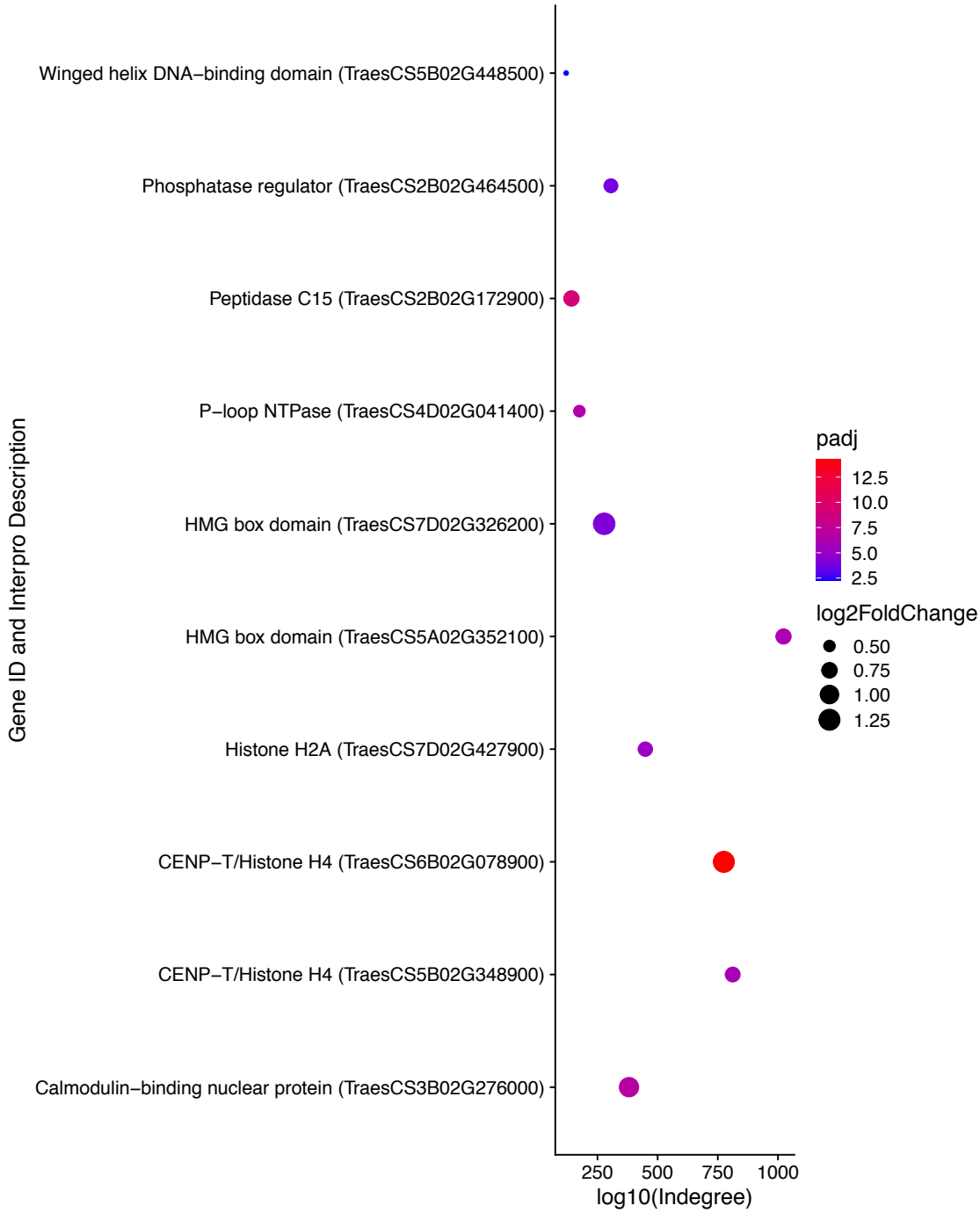


Figure 4.17. Enrichment of histone and HMG genes in nodes with highest indegree in the upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network. Top 10 nodes with the highest indegree of the upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network are plotted with a $\log_{10}$ transformation. Indegree is the number of coexpression correlation directed towards that node. Point size is determined by the differential $\log_2$ fold change of that gene in the IE_TM vs IE_WE pairwise comparison as denoted on the legend on the right. The adjusted p value is computed by taking the $-\log_{10}$ of the differential expression p value from the IE_TM vs IE_WE pairwise comparison for that gene and is coloured as denoted on the legend on the right.

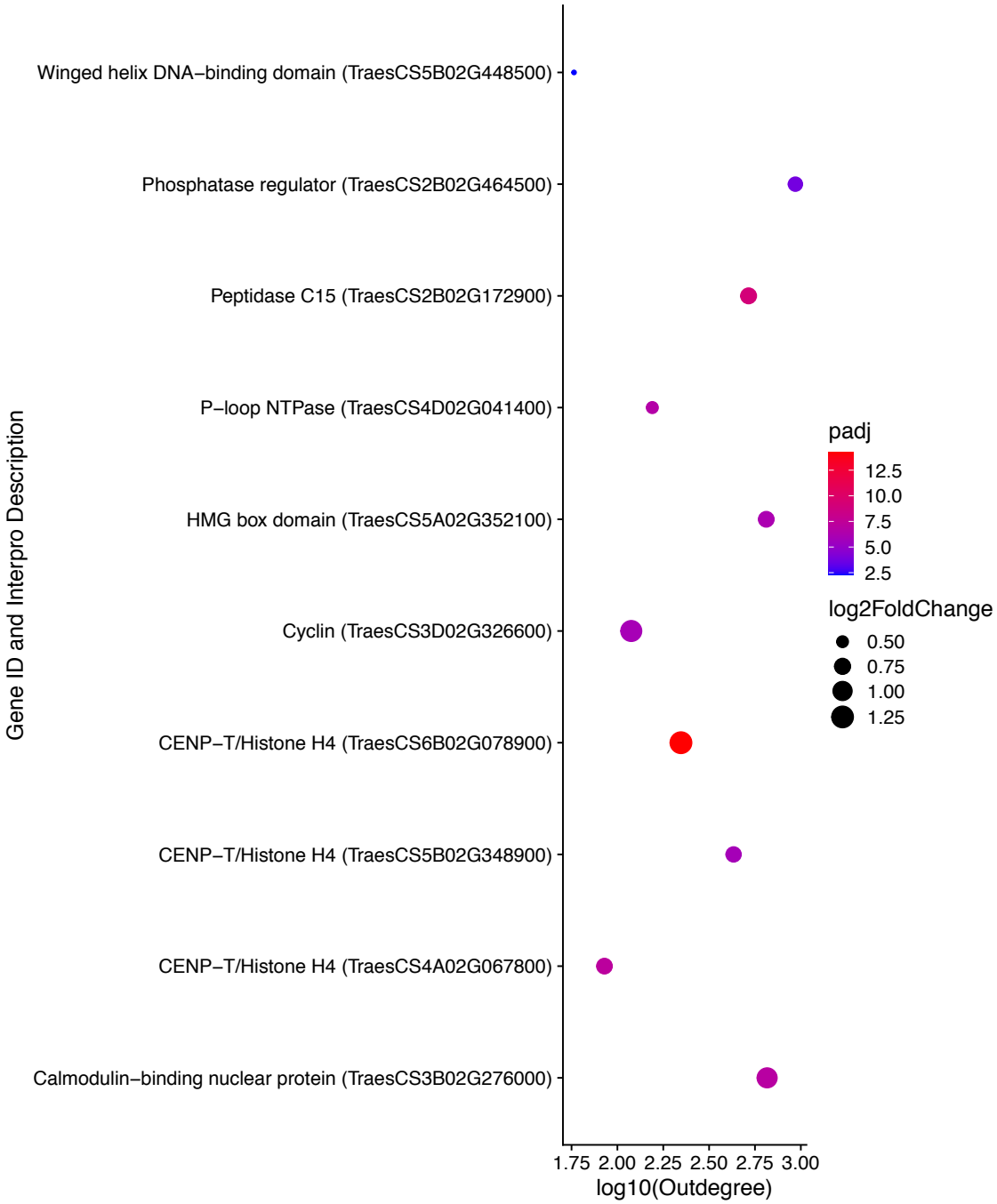


Figure 4.18. Enrichment of histone, HMG, and cyclin genes in nodes with highest outdegree in the upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network. Top 10 nodes with the highest indegree of the upregulated *indTaRKD1-7D #2.28* shared DEGs coexpression network are plotted with a $\log_{10}$ transformation. Outdegree is the number of coexpression correlation directed out of that node. Point size is determined by the differential $\log_2$ fold change of that gene in the IE_TM vs IE_WE pairwise comparison as denoted on the legend on the right. The adjusted p value is computed by taking the $-\log_{10}$ of the differential expression p value from the IE_TM vs IE_WE pairwise comparison for that gene and is coloured as denoted on the legend on the right.

The coexpression network was annotated by log2 fold change from the *indTaRKD1-7D #2.28* immature embryo TM vs WE pairwise comparison showed that most genes were upregulated (Figure 4.16). This network is available for visualisation at nDEX under 'Immature embryo upregulated DEGs Meristem Logfold'. Network analysis showed nodes with the highest indegree and outdegree were histones and high mobility box (HMG) genes (Figure 4.17, Figure 4.18). This suggested that transcriptional cascades induced by TaRKD1-7D were heavily involved in DNA conformational change and cell cycle through the upregulation of histones and HMGs as hypothesised. Following this we decided to investigate what were the direct binding targets of TaRKD1-7D to elucidate the full molecular pathway of TaRKD1-7D from binding targets to transcriptional cascades to phenotype.

4.2.5. TaRKD1-7D binds promoter regions of histone lysine methyltransferases and induces higher expression of plant architecture TFs

To investigate the binding targets of TaRKD1-7D, we obtained and analysed hypocotyl tissue ChIPseq and *in vitro* DAPseq data to elucidate the *in vitro* and tissue specific binding targets involved in TaRKD1-7D induced tiller development. Peaks called in both experiments showed equal distribution across all chromosomes with a slightly higher number of peaks called to chromosome ends and the unassembled chromosome (Supplementary Figure 4.8, Supplementary Figure 4.9). Peaks called also showed high enrichment of promoter (10kb upstream) binding relative to the expected distribution of this feature in the wheat genome (Figure 4.19, Figure 4.20). Coverage was also found to be higher proximal to the transcriptional start site (TSS) and end site (TES) of genes compared to proximal sites consistent with TaRKD1-7D's role as a TF (Figure 4.21).

Nucleotide sequences surrounding peaks called from the DAPseq experiment was obtained from the IWGSC RefSeq v1.1. assembly and were analysed for motif enrichment. The most significant enriched motif showed high similarity to the Arabidopsis RWP-RK TF AtRKD2 (Figure 4.22). Central enrichment analysis of the the TaRKD1-7D motif and another RWP-RK TF AtNLP4 were performed and showed significant central enrichment in the sequences of peaks called. Both these results confirmed that TaRKD1-7D was binding to promoter regions of target genes and that it had a highly evolutionarily conserved binding motif. To investigate this, GO term analysis was carried out on genes that had either a ChIPseq or DAPseq peak within 10kb of their TSS. ChIPseq genes showed significant enrichment of functional terms implicated in photosynthesis, DNA metabolic processes and translation (Figure 4.23). This was characteristic of vegetative tissue and was similar to the transcriptomic difference between leaf tissue and immature embryo tissue observed in our RNAseq experiments. DAPseq gene showed significant enrichment of functional terms implicated in histone lysine methylation and ubiquitination (Figure 4.24). This supported our hypothesis that downstream transcriptional cascades of TaRKD1-7D were involved in chromatin remodelling as histone lysine methylation has been shown extensively in the literature to be involved in chromatin remodelling. A high confidence list of genes with a ChIPseq and DAPseq motif within 10kb of their TSS was obtained to overlap with the IE RNAseq DEGs to investigate genes that were bound by TaRKD1-7D and upregulated in a pluripotent tissue context (Figure 4.25). Interestingly, an equal number of upregulated and downregulated DEGs had ChIPseq and DAPseq or DAPseq only binding sites,

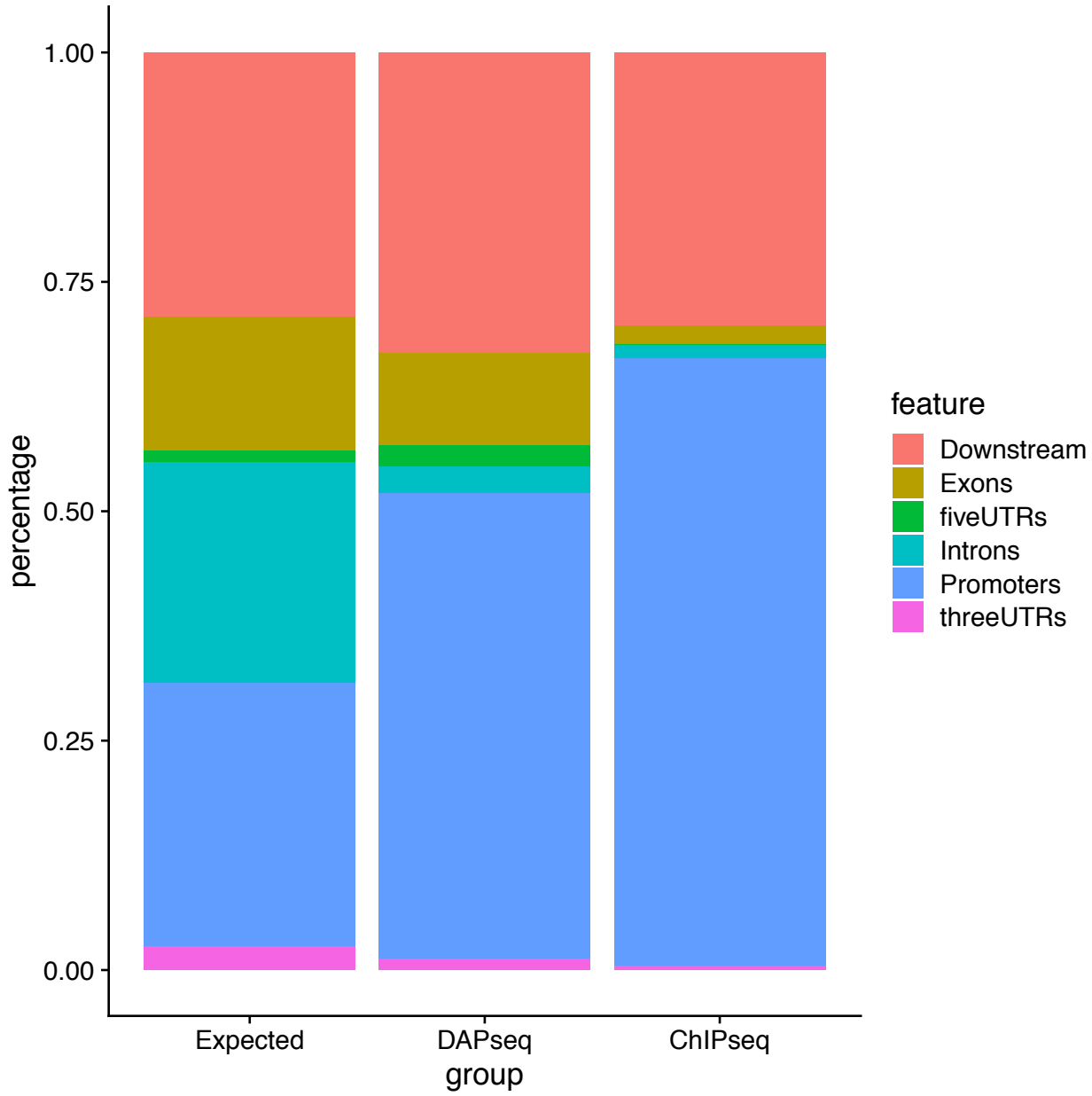


Figure 4.19. ChIPseq and DAPseq *indTaRKD1-7D* shows significant enrichment of promoter regions. Feature distribution plot from MACS2 peaks called in DAPseq and ChIPseq samples. Intergenic regions have been excluded to highlight changes in promoter regions. Features correspond to colours denoted on the legend to the right.

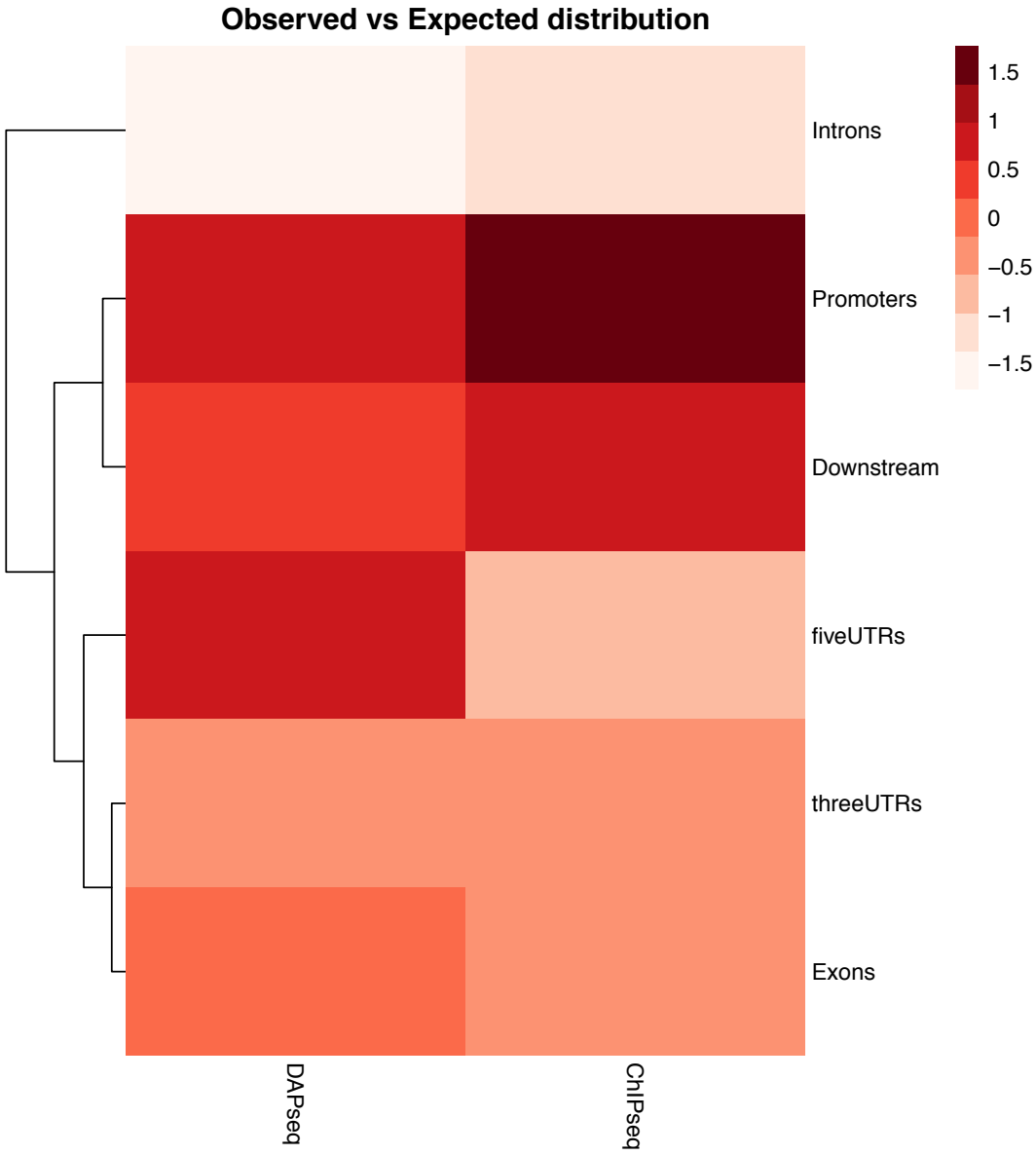


Figure 4.20. Heatmap of ChIPseq and DAPseq *indTaRKD1-7D* shows significant differential enrichment of promoter and downstream regions. Feature distribution plot from MACS2 peaks called in DAPseq and ChIPseq samples. Intergenic regions have been excluded to highlight changes in promoter regions. Features correspond to colours denoted on the legend to the right. Scale displayed is row scaled logarithmically transformed observed/expected values. Hierarchical clustering was performed on features and is displayed as a dendrogram on the side of the heatmap.

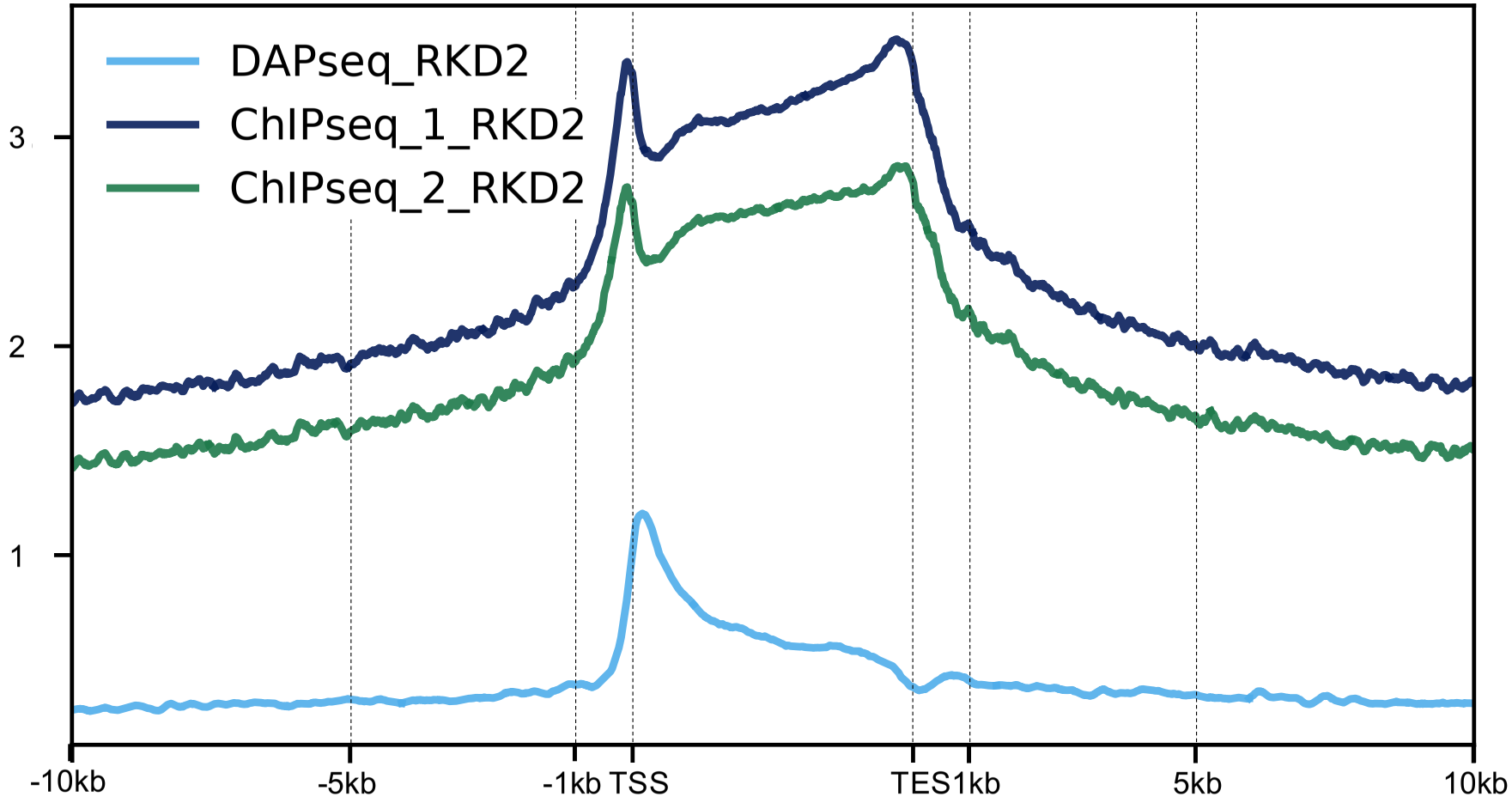


Figure 4.21. ChIPseq and DAPseq *indTaRKD1-7D* shows higher coverage proximal to the TSS and TES. All instances labelled TaRKD2 are TaRKD1-7D. Normalised coverage values were generated from bigwigs for each sample and shown on the y axis. Graduations along the x-axis shows distances from the transcriptional start site (TSS) and are denoted with -1kb, -5kb and -10kb when relative to the TSS or from the transcriptional end site (TES) and are denoted with 1kb, 5kb and 10kb when relative to the TES.

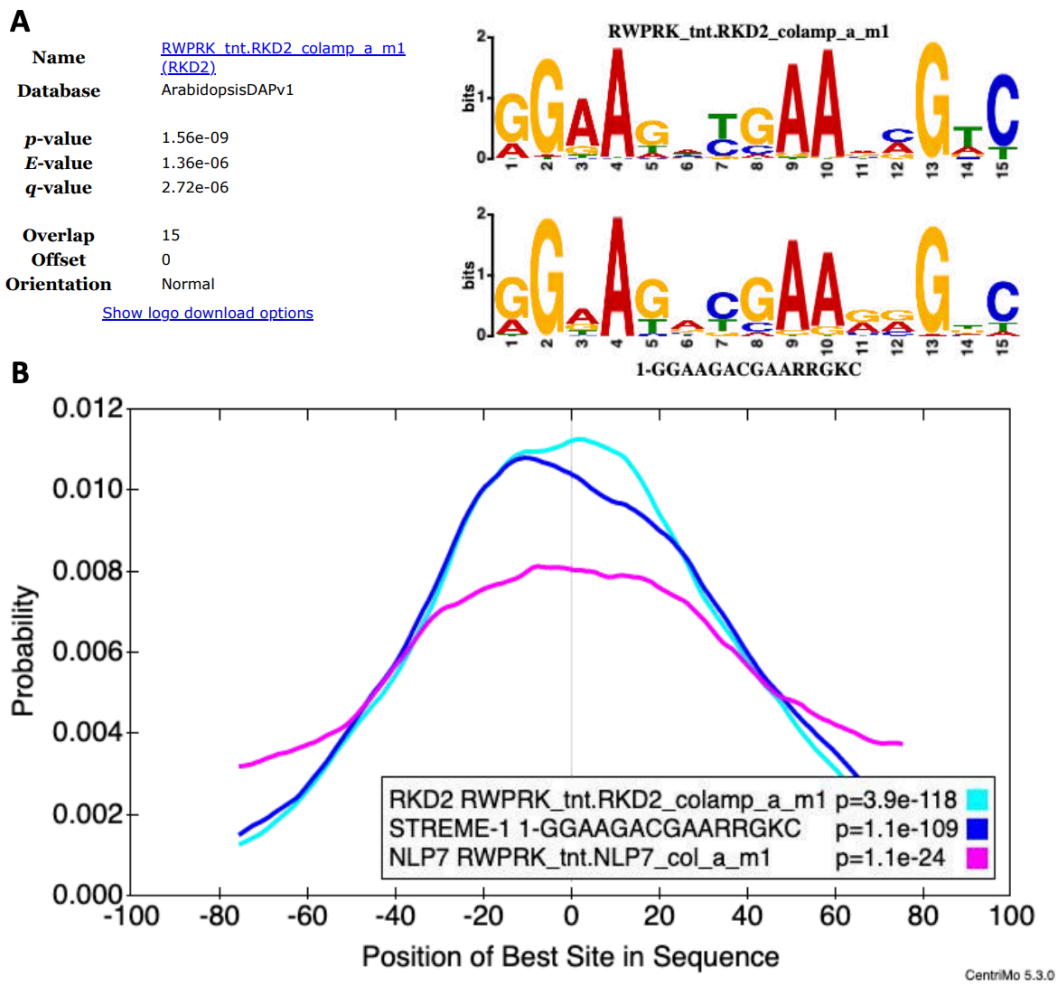


Figure 4.22. Motif similarity and central enrichment of TaRKD1-7D binding motif. (A) TomTom motif similarity analysis of TaRKD1-7D binding motif is very significantly like the *Arabidopsis thaliana* RKD2 binding motif. Calculations are given on the right while motif probability values are given by left with 2-bit denoting 100% confidence and 0-bits denoting 0% confidence of a given nucleotide at this position. **(B)** Central enrichment of a binding motif in the centre of the inputted sequence calculated by Centrimo. Different coloured line plots correspond to instances of motifs denoted in the legend on the bottom right of the plot. Beside each colour is the statistical significance of the central enrichment of this motif.

suggesting TaRKD1-7D can act as both an activator and a repressor of gene expression through direct binding (Figure 4.25). To identify transcriptionally regulated genes with TaRKD1-7D binding site in the proximal promoter region we overlapped DAPseq genes with the *indTaRKD1-7D* #2.28 immature embryo shared DEGs giving a list of 477 and narrowed this down to 78 genes by overlapping them with ChIPseq genes

(Supplementary Figure 4.11, Figure 4.22). Both upregulated and downregulated *indTaRKD1-7D* #2.28 immature embryo shared DEGs with a ChIPseq and DAPseq peak were visualised to confirm concentrated coverage in the promoter region. Two upregulated trehalose phosphate synthase genes with a DAPseq TaRKD1-7D binding motif were selected for visualisation due to their implicated roles in meristem determination and two downregulated genes with a DAPseq TaRKD1-7D binding motif involved in regulated cell death and DNA repair as these were seen to be consistent with increased tiller development and plant growth overall (Figure 4.26, Figure 4.27).

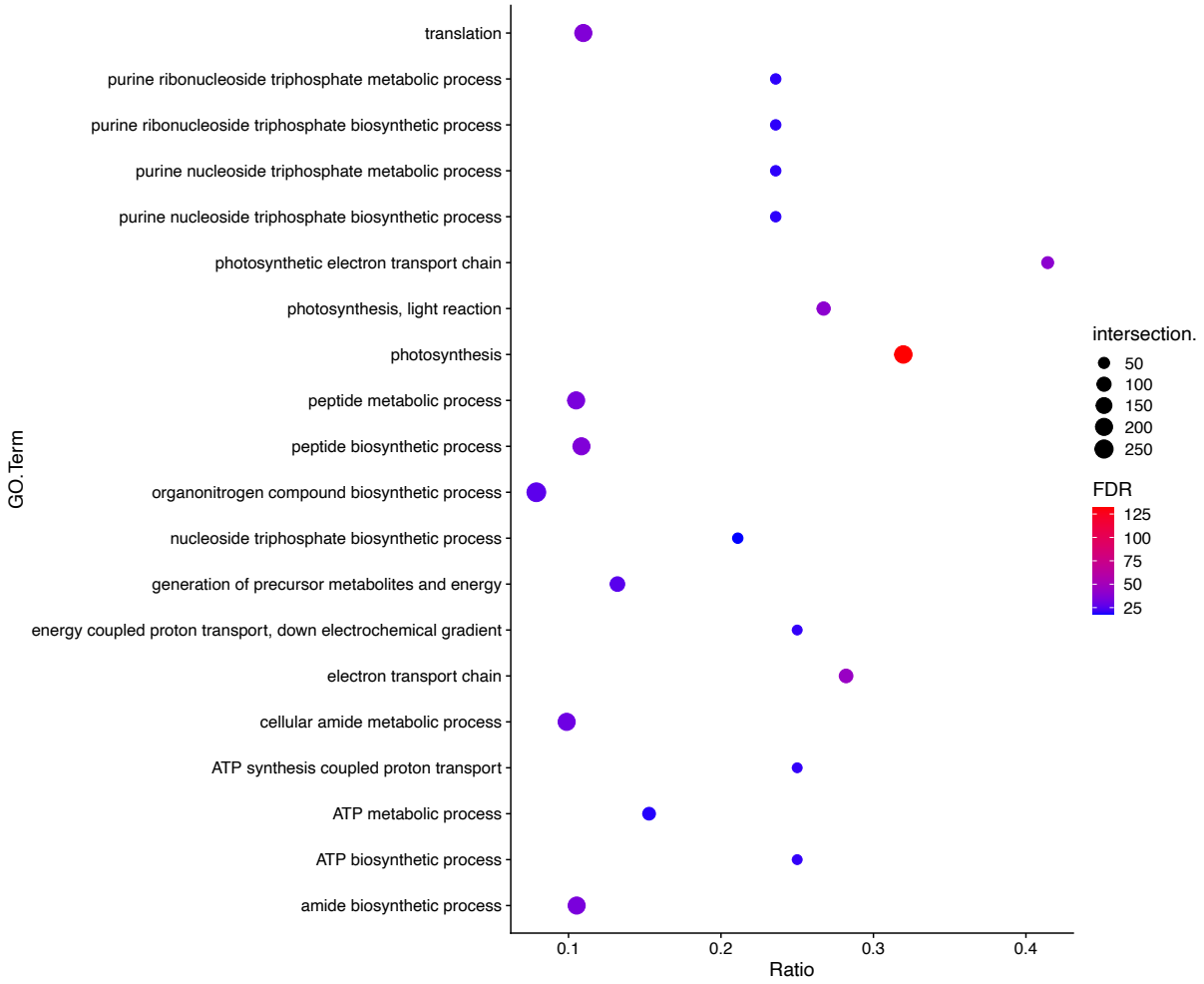


Figure 4.23. GO term analysis of genes with a MACS2 ChIPseq peak called within 10kb of the TSS shows enrichment of photosynthetic and translation terms. The top 20 most significant gene ontology (GO) terms of genes with a MACS2 ChIPseq peak called within 10kb of the TSS are plotted here. GO terms are labelled on the y axis and the ratio of genes for a GO term in the genes with a MACS2 ChIPseq peak called within 10kb of the TSS compared to the total number of genes corresponding to that GO term in the genome is plotted on the x axis. Point size is determined by the number of genes from that GO term in the genes with a MACS2 ChIPseq peak called within 10kb of the TSS as denoted on the legend on the right. False discovery rate (FDR) is computed by taking the $-\log_{10}$ of the p value determined from an enrichment test with a Benjimini-Hochberg correction and is coloured as denoted on the legend on the right.

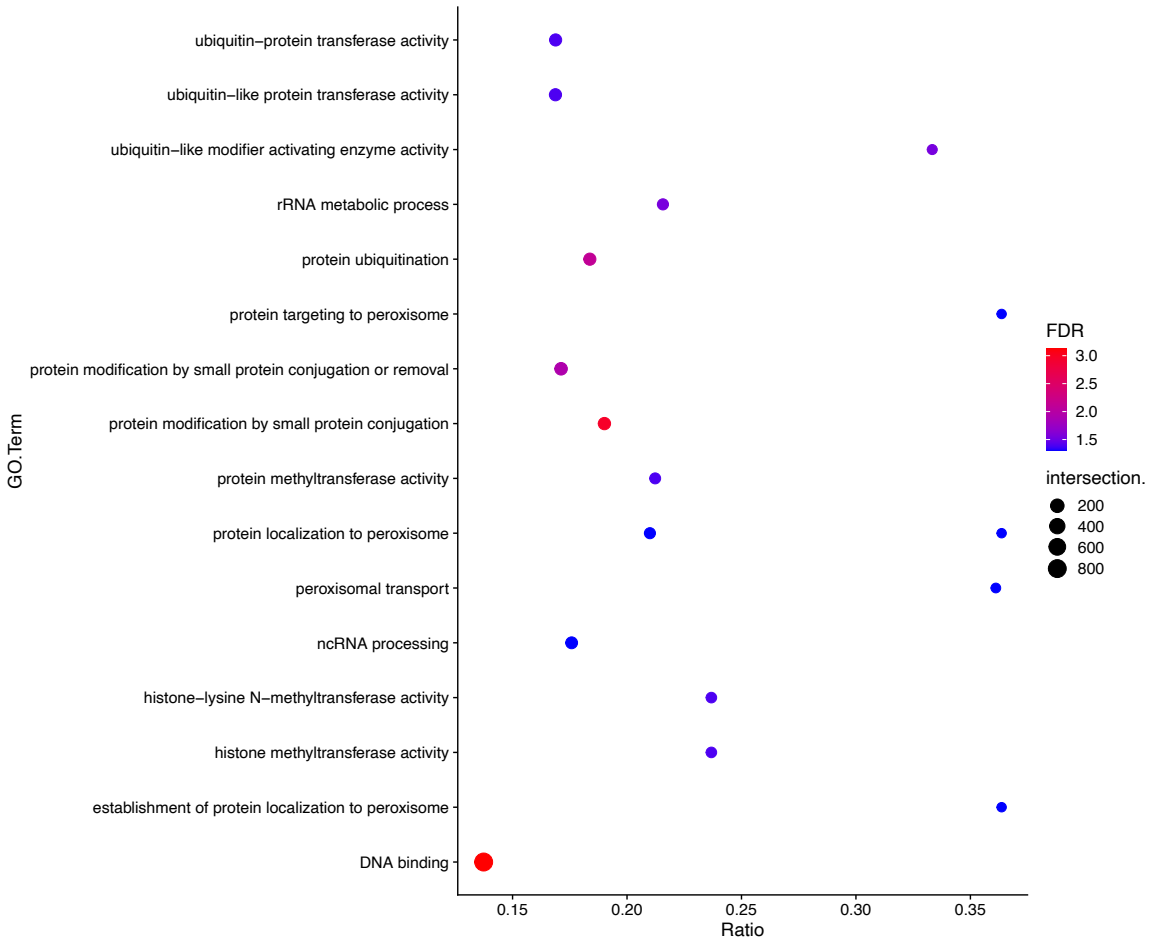


Figure 4.24. GO term analysis of genes with a MACS2 DAPseq peak called within 10kb of the TSS shows enrichment of histone methylation and ubiquitination terms. The top 20 most significant gene ontology (GO) terms of genes with a MACS2 DAPseq peak called within 10kb of the TSS are plotted here. GO terms are labelled on the y axis and the ratio of genes for a GO term in the genes with a MACS2 DAPseq peak called within 10kb of the TSS compared to the total number of genes corresponding to that GO term in the genome is plotted on the x axis. Point size is determined by the number of genes from that GO term in the genes with a MACS2 DAPseq peak called within 10kb of the TSS as denoted on the legend on the right. False discovery rate (FDR) is computed by taking the $-\log_{10}$ of the p value determined from an enrichment test with a Benjimini-Hochberg correction and is coloured as denoted on the legend on the right.

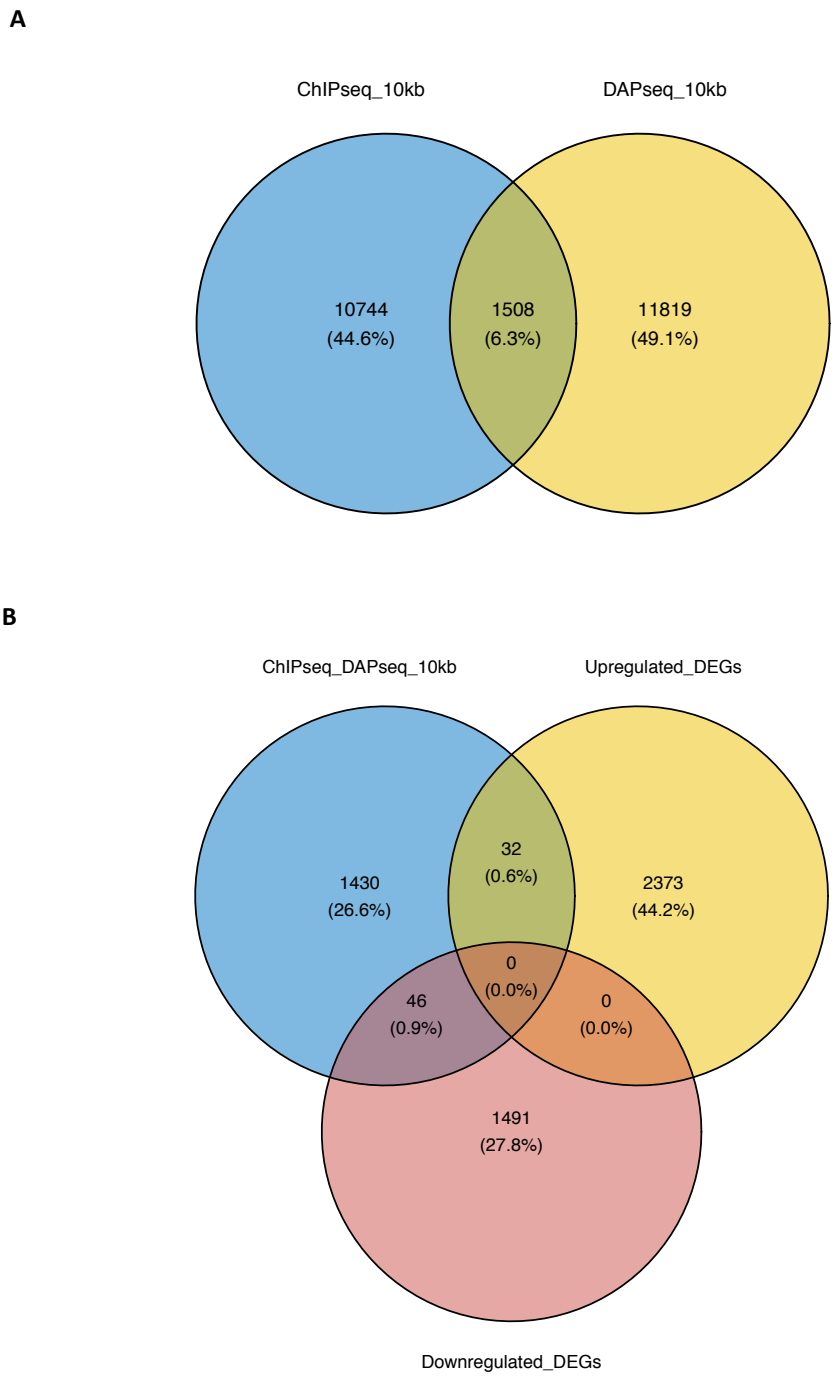


Figure 4.25. Venn diagrams of *indTaRKD1-7D* ChIPseq, DAPseq and immature embryo RNAseq shared DEGs identify high confidence targets of TaRKD1-7D. (A) Venn diagram overlap of genes with a MACS2 ChIPseq or DAPseq peak called within 10kb of the TSS. Percentages denote the proportion of the total genes represented. **(B)** Venn diagram overlap of genes with a MACS2 ChIPseq and DAPseq peak called within 10kb of the TSS and upregulated or downregulated immature embryo *indTaRKD1-7D #2.28* shared DEGs.

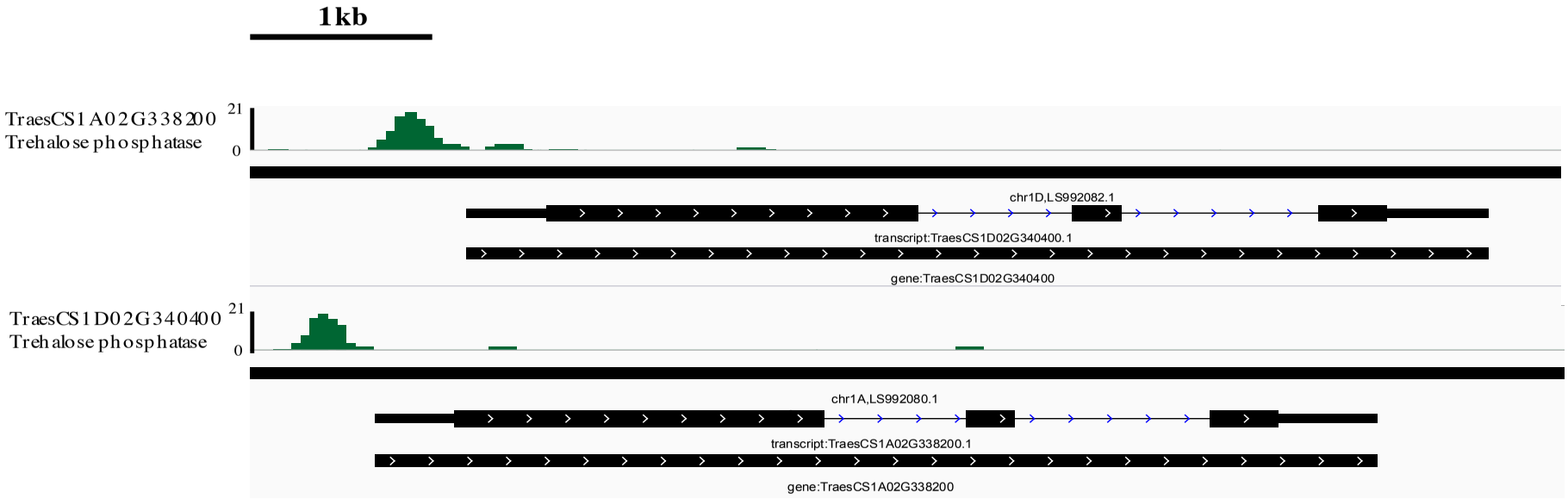


Figure 4.26. indTaRKD1-7D DAPseq peak within 10kb of two upregulated *indTaRKD1-7D #2.28* immature embryo shared DEGs relevant to meristem determinacy. IGV snapshots of a DAPseq bigwig file over an annotation of two upregulated transcription factors. The scale bar on the left denotes read coverage and the bars below represent exons in the gene structure.

Figure 4.27. *indTaRKD1-7D* DAPseq peak within 10kb of two downregulated *indTaRKD1-7D #2.28* immature embryo shared DEG TFs relevant to cell death and DNA repair. IGV snapshots of a DAPseq bigwig file over an annotation of two upregulated transcription factors. The scale bar on the left denotes read coverage and the bars below represent exons in the gene structure.

4.3. Discussion

4.3.1. Characterising transcriptional cascades using wheat bioinformatics resources

The development of the International Wheat Genome Sequencing Consortium (IWGSC) annotated reference for wheat has allowed computational researchers to identify and characterise genes implicated in wheat development to increase yield and accelerate genetic gain in this agronomically important species (Jia et al., 2018; IWGSC, 2018). While many researchers have characterised the transcriptional landscape and networks involved in processes such as meiosis and senescence, none have integrated it with ChIPseq and DAPseq to fully elucidate binding motifs and targets implicated in transcription factor pathways (Ramírez-González et al., 2018; Borrill et al., 2019). This marks a significant shift in wheat bioinformatics, using sequencing experiments that are commonplace in model organisms such arabidopsis, maize and rice. The landscape of wheat research has quickly changed throughout this research project with other groups we collaborate with using state-of-the-art techniques for assessing chromatin topology such as ATACseq, Hi-C, and Hi-ChIP in wheat to show how local chromatin structure is arranged and dynamically changes through transcription (Concia et al., 2020). This shows how many researchers are not only investigating wheat for its agronomic importance but as a genetically complex organism capable of adapting to multiple environments through the variation present from the hybridisation of three separate grass species to understand more about plant genetics and development.

4.3.2. TaRKD1-7D binds histone and histone methylation genes, upregulating genes involved in chromatin accessibility, cell cycle and meristematic determinacy.

Temporal dynamics in meristem determinacy have been shown to be regulated by histone methylation marks in *Arabidopsis thaliana*. The active histone mark H3K4me3 and the repressive histone mark H3K27me3 correlate strongly with transcriptional changes in the shoot apical meristem implicated with flowering time (You et al., 2017). Somatic embryogenesis, a process that dedifferentiates somatic tissue into rapidly dividing callus, has been found to be regulated by the histone methylation mark H3K27me3 through deposition around somatic embryogenesis regulators like *LEAFY COTYLEDON1* (*LEC1*) and *BABY BOOM1* (*BBM1*) (Nic-Can et al., 2013).

Here, we've found that TaRKD1-7D binds to histone and histone methylation genes in our TaRKD1-7D DAPseq experiment and were upregulated in our immature embryo RNAseq data (Figure 4.6, Figure 4.24). Further to this, histones show increased localisation in RNA in situ experiments in axillary meristems relative to wild type and act as regulatory hubs in *our indTaRKD1-7D #2.28* immature embryo network (Figure 4.11, Figure 4.26). We also confirmed that gene ontology (GO) terms for the DNA packaging complex were tiller-related through re-analysis of publicly available data from high tiller number NILs (Figure 4.9). Our hypothesis suggests that *TaRKD1-7D* upregulation of chromatin accessibility factors in pre-meristematic immature tissue is causative of increased axillary meristem formation and higher tiller number in our *indTaRKD1-7D* transgenic lines (Figure 3.7). As highlighted in Chapter 3, we speculate that TaRKD1-7D induced higher tiller number acts independently of

many endogenous inhibitory signals related to nutrient availability and soil volume through increasing meristem formation by upregulating genes involved in pluripotency and epigenetic regulation. Others have shown that increased expression of epigenetic factors such as *TROGEN GROWTH RESPONSE GENE 5* (*NGR5*) through recruiting *POLYCOMB REPRESSIVE COMPLEX 2* (*PRC2*) which increases H3K27me3 deposition on shoot inhibitory genes resulting in higher tiller number and increased nitrogen response in rice suggesting this isn't unique to ectopic expression of *TaRKD1-7D* (Wu et al., 2020).

Our speculation that ectopic *TaRKD1-7D* can also cause increased plant height and spikelet number per spike further suggested from some findings involving meristematic determinacy factors in this chapter (Figure 3.8, Figure 3.11). This contrasts with semi-dwarf ideotypes typically associated with higher tiller number in the literature such as the *Rht1* mutant alleles (Allan et al., 1959). The BBR/BPC family, a transcription factor family implicated in stem cell niche maintenance, is differentially upregulated in *indTaRKD1-7D* immature embryo and leaf tissue as well as tiller nodes from high tiller number lines (Figure 4.10) (Theune et al., 2019). More interestingly, we see a large subset of genes involved in cell cycle and meristematic determinacy that are upregulated in *indTaRKD1-7D* immature embryo tissue that have a DAPseq peak within 10kb of their promoter region (Supplementary Table 4.1). Further to this, two trehalose-6-phosphate synthase genes, which have been shown to be involved in meristematic determinacy in maize, also have binding peaks in our *indTaRKD1-7D* hypocotyl tissue ChIPseq experiment (Claeys et al., 2019). This suggests that TaRKD1-7D can bind these loci in shoot apical meristem tissue and potentially inflorescence and floral meristem tissue. However, it should be noted that

these genes are not upregulated in our *indTaRKD1-7D* leaf tissue RNAseq experiment suggesting chromatin surrounding these loci is not accessible in vegetative tissue. Our increased height phenotype could be a result of the delayed flowering time we see in *indTaRKD1-7D* lines as this would allow the plant more time for internodal elongation and to accumulate more total plant biomass. One way to differentiate whether increased height was due to SAM indeterminacy or delayed flowering time would have been to measure whether there was an increase in the number of nodes in the stem or an increase in internodal length, but unfortunately these were not measured at the time of the trial. However, it would be interesting in future experiments to confirm this and see if decreased meristematic determinacy could be regulated by ectopic *TaRKD1-7D* expression to manipulate plant height as other have tried by modulating Gibberellin metabolism genes (Bhattacharya et al., 2010).

4.3.3. *indTaRKD1-7D* leaf tissue shows upregulation of photosynthesis genes but not histone methylation genes

The model of the epigenetic landscape from Conrad Waddington describes how cell fate decisions are analogous to a ball going down a set of contours that it is unable to go back up, an early effort to explain how a zygote can give rise to all the cells and tissues in the body (Baedke et al., 2013). Dedifferentiation has shown that this model isn't completely correct with somatic reprogramming through the discovery of the Yamanaka factors in mammalians systems and somatic embryogenesis in plant systems (Yamanaka, 2008). However, Waddington's theory helps us understand how much epigenetic reprograming must take place to dedifferentiate a leaf cell as opposed to immature embryos, the further a ball rolls down a contour the greater effort that must be exerted to push it back up.

Despite higher expression of *indTaRKD1-7D* in leaf tissue than immature embryos, we see no upregulation of binding targets of TaRKD1-7D suggesting that the chromatin accessibility is blocked at histone methylation genes in leaf tissue. This prevents the transcriptional cascade necessary for *TaRKD1-7D* to induce dedifferentiation and the transition to a more rapidly differentiating tissue, which may explain why no significant increase in flag leaf area was observed in the tiller traits experiment (Supplementary figure 3.11). *AtRKD4* induced somatic embryogenesis has been shown to induce somatic embryos in the zone of division and elongation in root tissue but not the zone maturation in agreement with this (Waki et al., 2011). Upregulation of photosynthetic genes and terms in leaf tissue is likely due to the transcriptional cascade of TaRKD1-7D but is more consistent with higher energy demand due to increased vegetative growth of tillers in the plant. While photosynthetic terms are also upregulated in the NIL RNAseq, this is most likely due to vegetative tissue in tiller node samples as without laser microdissection, this tissue is very difficult to get in high quantities without contamination from surrounding tissues.

We see in Chapter 5 below that the *TaRKD1-7D* ortholog *OsRKD3* is incapable of inducing transcriptional cascades to upregulate chromatin accessibility genes in leaf tissue. This highlights that ectopic expression of zygotic RKDs can only manipulate plant architecture when expressed in appropriate tissue contexts as some tissues become terminally differentiated and where *TaRKD1-7D* targets are in heterochromatic states. The potential of zygotic RKD, or more generally epigenetic master regulators, to manipulate plant architecture by expressing them in specific tissue contexts could apply to a range of plant species in multiple breeding programs as a novel approach for ideotype breeding.

Chapter 5

5.1. Introduction

5.1.1. Rice as a model organism of the grasses

Rice was the first model plant whose whole genome was sequenced amongst the cereal crops and has had a tremendous amount of bioinformatics resources expression browsers, gene coexpression networks and genome variation resources developed for researchers (IRGS Project, 2005; Garg and Jaiswal, 2016). Its relatively small and less complex diploid genome allowed the development of advanced functional genomic resources much earlier compared to larger complex grass genomes like barley and wheat (Li et al., 2018). Additionally, somatic embryogenesis and plant transformation in rice is highly efficiency making it highly amenable to genetic analysis (Heyser et al., 1983; Hiei et al., 1994). Mutational analysis of rice genes has allowed orthologous genes implicated in fundamental biological processes and yield-determining traits to be identified in other agronomically important species making it a useful model for comparative genomics of BEP clade cereal crops (Shimamoto et al., 2002). However, the Oryzoideae (It's ahartoideae) subfamily diverged from the Triticeae around 50 million years ago making rice and wheat one of the most evolutionarily distant species pair amongst cereal crops in the BEP clade (Chalupska et al., 2008). The Oryzoideae is closest to the Bambusoideae, which contains bamboo, while the Triticeae contains many other agronomically important cereal crops like barley and rye (Figure 1.4). Identifying evolutionarily conserved genes across rice and wheat that increase yield in both species can allow the characterisation of breeding candidates that can be adaptive across multiple species in the BEP clade.

5.1.2. Wild diversity in rice and domestication

Rice is a major staple crop and is grown by millions of small-holder farmers throughout Asia but is also known for its within species and within genus genetic diversity. Phylogenetic evidence has suggested that the two main subspecies of cultivated rice *japonica* and *indica* has arisen from a single domestication event from wild rice (Cheng et al., 2003; Londo et al., 2006). Advanced functional and genomic resources have been built for both these subspecies, with the *japonica* cultivar Nipponbare being chosen for the International Rice Genome Sequencing Project (IRGSP) and consequently all major bioinformatics databases such as the Rice Annotation Project Database (RAP-DB) and the Michigan State University (MSU) rice genome annotation project. While many cultivars of *Oryza sativa* share selection in well-characterised domestication genes like *An-1* (awn development number, grain size and grain number), *bh4* (grain colouration and seed shattering) and *OsC1* (anthocyanin accumulation) they can also be distinguished on this basis (Saitoh et al., 2004; Zhu et al., 2011; Luo et al., 2013). Pigmented Indonesian rice cultivars have recently been characterised with molecular markers and shown that many red and black rice varieties can be either *sativa* and *japonica* (Kurniasih et al., 2019). *Cempo Ireng* is a *japonica* cultivar of black rice that is characterised by low tiller number, long panicles and high production of grain anthocyanin. High demand of black rice over white rice (anthocyanin has been shown to have health benefits it's as a potent anti-oxidant) has resulted in many Indonesian research groups focusing on improving yields in *Cempo Ireng* through breeding programs using conventional mutagenesis methods and introgression of transgenes (Purwestri et al., 2015; Masruroh et al., 2016; Apridamayanti et al., 2017).

Chapter aims and hypothesis

The overall aim of this chapter is to investigate whether ectopic *RKD* expression can induce higher tiller number and seed number per panicle in rice, followed by characterisation of the molecular pathways downstream of zygotic *RKDs*. This aim will be addressed through the phenotypic characterisation of ectopic *RKD* expression in two different subspecies of rice with two different zygotic *RKDs*. Additionally, we will analyse whole genome RNAseq experiments from lines showing ectopic *RKD* expression and tissues relevant to tillering from publicly available datasets. The hypotheses being tested in this chapter are as follows:

- Can plant architecture be manipulated through ectopic expression of zygotic RKDs in rice?
- What molecular mechanisms are involved in ectopic expression of zygotic RKDs in rice?

We aim to investigate whether our novel strategy for inducing higher tiller number in wheat also works in rice. Wheat and rice are one of most evolutionarily divergent pairs of grass species in the BEP clade. If we can induce higher tiller number in both these species, its likely RKD-induced tiller number can be used as a breeding strategy for crop improvement in all BEP clade cereal crops.

5.2. Results

5.2.1. Ectopic expression of *indAtRKD4* in *Oryza sativa ssp. japonica* 'Nipponbare' increases vegetative tiller number

To investigate whether ectopic *RKD* expression could be used as a breeding strategy in other cereal crops, we tested whether ectopic expression of the *A. thaliana* embryonic *RKD* gene *AtRKD4* or its rice ortholog *OsRKD3,* which phylogenetically groups with *TaRKD1* homeologues, could increase tiller number in the *O. sativa ssp. japonica* cultivar 'Nipponbare'. Two-component dexamethasone inducible constructs were transformed into the *Oryza sativa ssp. japonica* cultivar 'Nipponbare' for the inducible expression *AtRKD4* and *OsRKD3* (Supplemental Table 5.1). Transgenic lines were screened for high copy T-DNA insertion number using DNA profiling and 46 *indAtRKD4* and 46 *indOsRKD3* T1 lines were transferred and propagated at the phytobiology facility (PBF) at the University of Warwick. T2 lines that showed ectopic expression of *indAtRKD4* and *indOsRKD3* in vegetative tissue were selected and propagated to T3. Homozygous *indAtRKD4* and *indOsRKD3* T3 lines were selected and grown in a randomised trial for 70 days (Figure 5.1). Plants from *indOsRKD3* T3 lines showed poor growth and not enough samples were present for robust statistical analysis. However, increased vegetative tiller number was observed in *indAtRKD4* T3 lines despite poor growth (Figure 5.2). No significant difference in height was observed in *indAtRKD4* as was seen in *indTaRKD1-7D* (Figure 5.3). A summary of these results can be found in Supplementary Table 5.2. As facilities in PBF were not suited to rice growth, our Indonesian collaborators at the Research Institute for Agriculture Technology, Indonesia transformed and phenotyped *indOsRKD3* lines to investigate whether ectopic *OsRKD3* expression induced higher tiller number in rice.

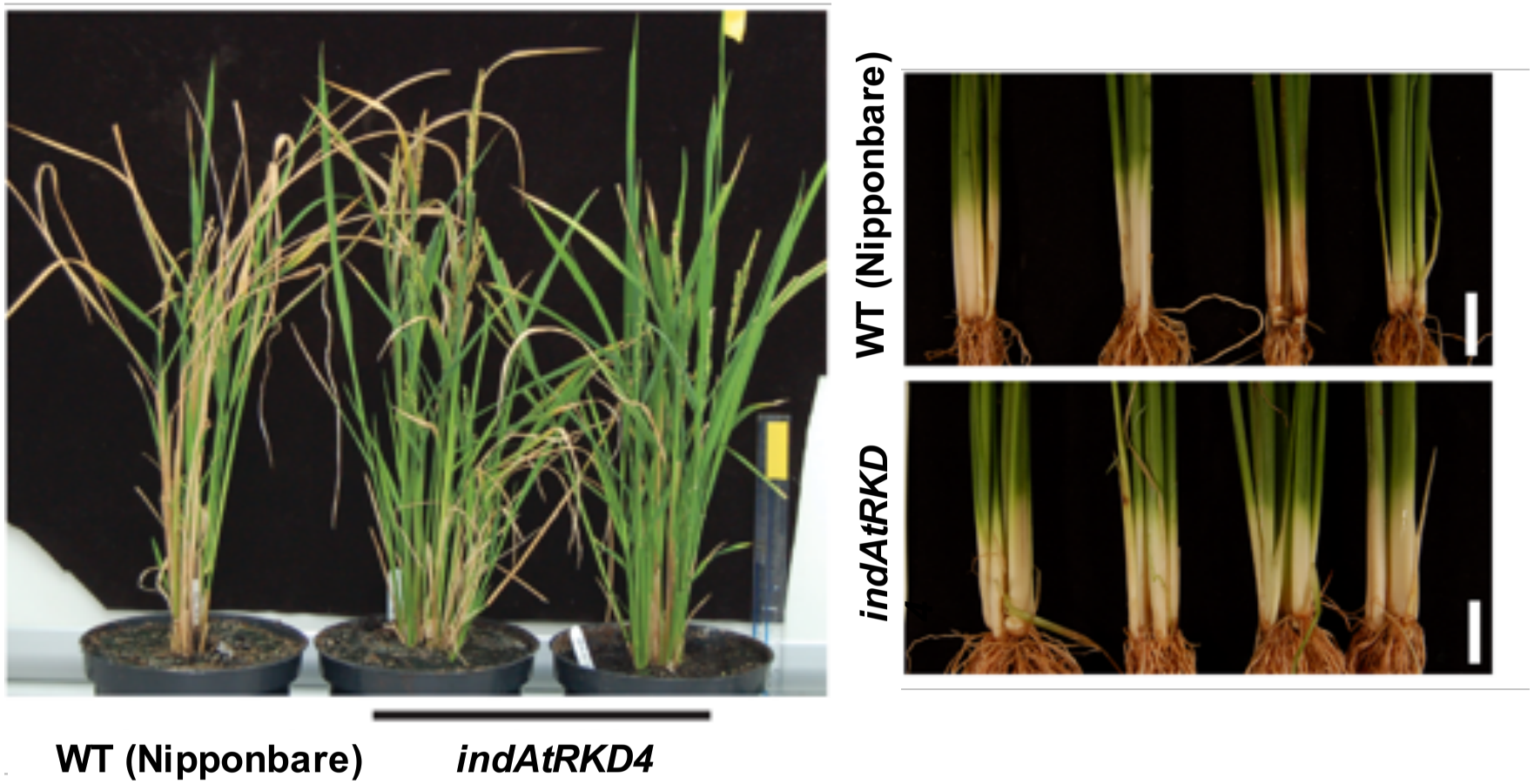


Figure 5.1. *indAtRKD4* Nipponbare lines show increased vegetative tiller number and delayed senescence. Phenotypic images of indOsRKD3. **(A)** *indAtRKD4* lines show delayed senescence and higher vegetative tiller number at 70 days after germination. Yellow scale bar is 5cm. **(B)** *indAtRKD4* lines show higher tiller initiation at 30 days and a higher number of adventitious roots. White scale bar is 2cm.

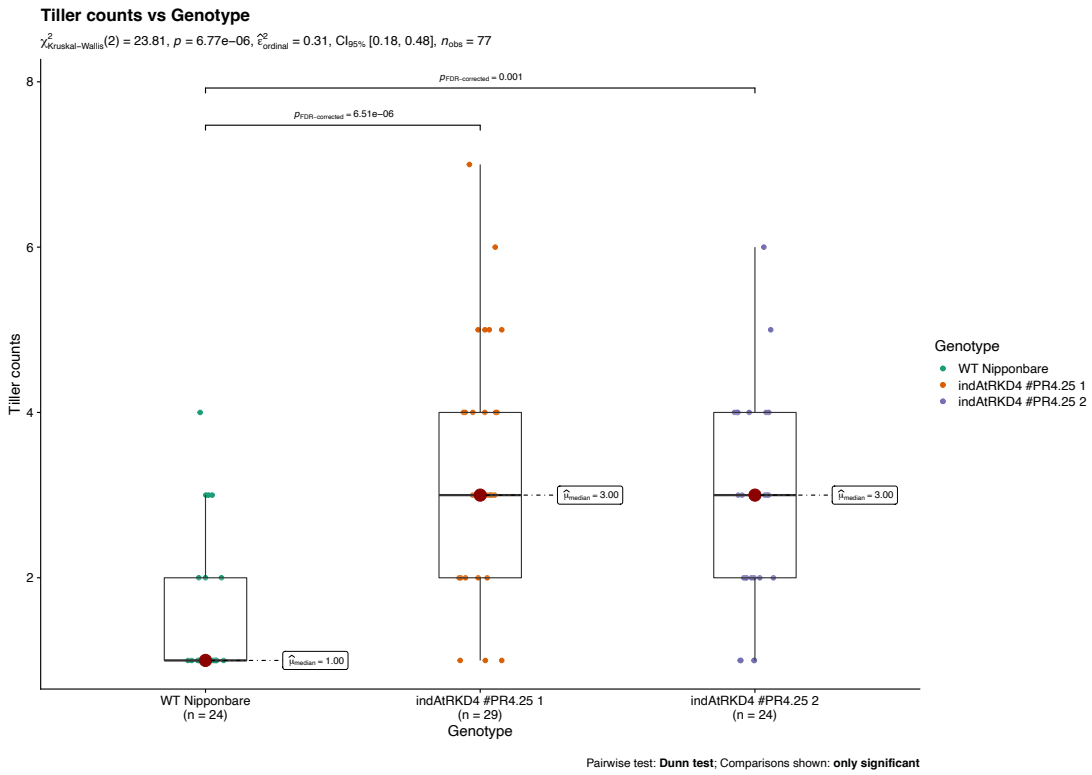


Figure 5.2. Vegetative tiller number in two *indAtRKD4* T3 Nipponbare lines. Vegetative tiller number from plants measured from counting the number of tillers at 70 days after germination. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

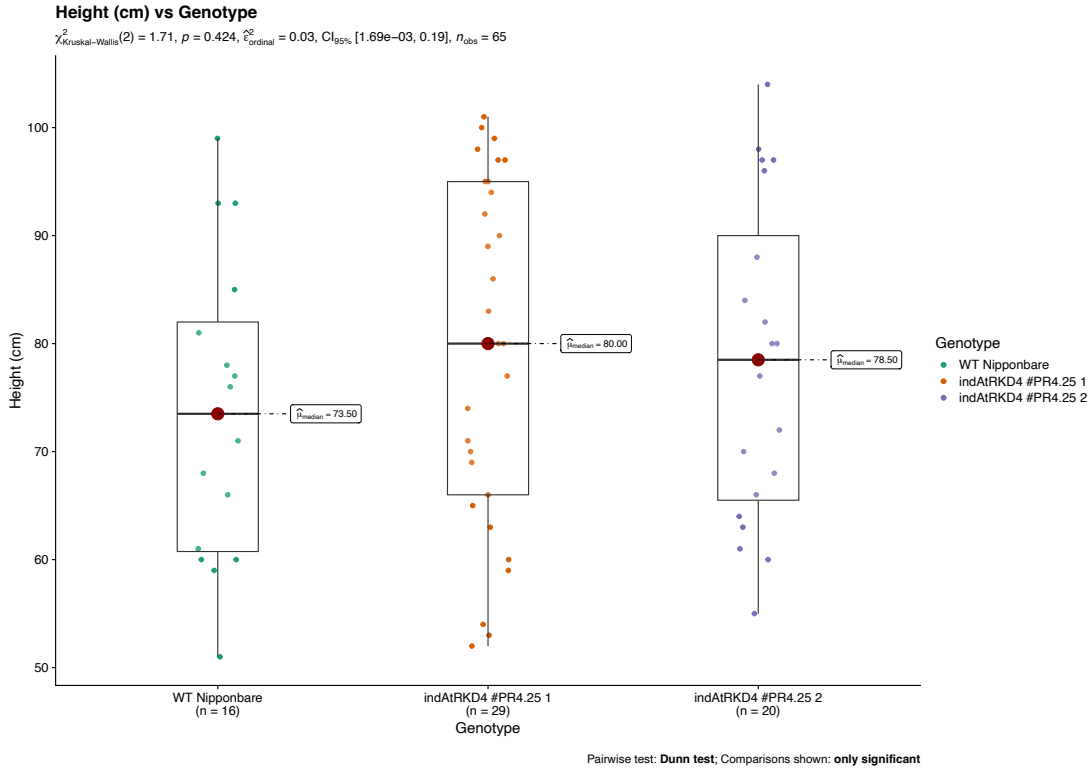


Figure 5.3. Plant height in *indAtRKD4* T3 Nipponbare lines. Height measured from the base of the plant to the tallest panicle at senescence. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Additional statistics can be found on the top left of the plot.

5.2.2 Ectopic expression of *OsRKD3* in *Oryza sativa Cempo Ireng* increases tiller number and seed weight in two *indOsRKD3* T3 lines

Two homozygous *indOsRKD3* T3 lines were selected and grown in randomised design for 120 days at the Research Institute for Agriculture Technology, Yogyakarta, Indonesia. Increased reproductive tiller number was observed in both i*ndOsRKD3* T3 lines and an increase in plant height was also observed in contrast to most characterised high tiller number phenotypes in the literature (supplementary figure 5.2). Increased seed weight per plant was observed in both *indOsRKD3* T3 lines suggesting that ectopic expression of *OsRKD3* might work as a breeding strategy for increasing yield in rice, however, it should be noted that a decrease in seed weight per 1000 was also observed (Figure 5.6, Supplementary Figure 5.3). Increased panicle branch number and seed number per panicle were also observed in both *indOsRKD3* T3 lines (Figure 5.7, Figure 5.8, Supplementary Figure 5.4). A summary of these results can be found in Table 5.1. In summation we found comparable phenotypes in between *indOsRKD3* and *indTaRKD1-7D* as could be expected between two evolutionarily distant cereal crops (Supplementary Table 3.6, Supplementary Table 3.9, Table 5.1). Interesting aspects of this unconventional phenotype such as increased tiller number and height were shared across both crops along with similar increases in inflorescence size and seed yield. However, it should be noted that seed weight per 1000 showed an inverse relationship across both species suggesting these phenotypic traits need to be confirmed with a field trial. All of this confirmed that zygotic RKDs could induce increased tiller formation in variety of grass species. Additionally, a drastic increase in the number of internodal tiller branches were observed in both *indOsRKD3* T3 lines (Figure 5.9). Adventitious roots were seen at the base of each internodal tiller branch suggesting a homeotic transformation from an inflorescence meristem into a tiller had

taken place (Figure 5.10). This suggested that the inflorescence meristem was being reprogrammed through ectopic *OsRKD3* expression leading to a developmentally earlier meristematic region, the axillary meristem and suggesting a rice-specific phenotype that wasn't observed in wheat. To investigate the molecular pathways involved in *OsRKD3* induced higher tiller number and increase homeotic transformation of tillers we obtained and analysed i*ndOsRKD3* leaf tissue RNAseq data.

A

B

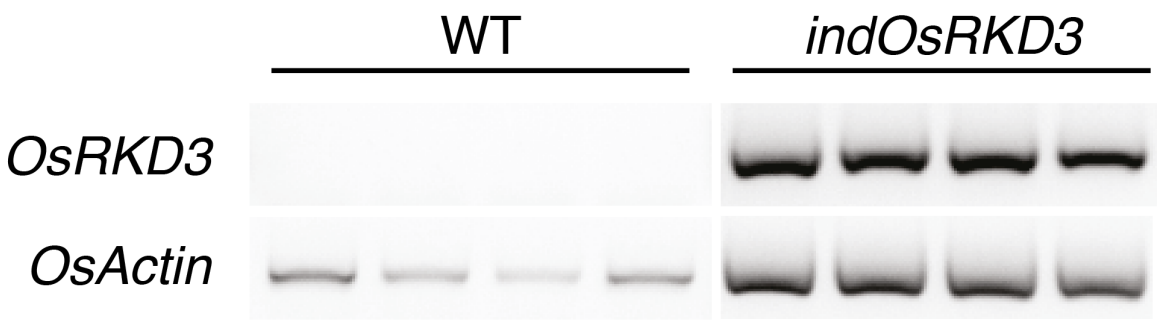


Figure 5.4. Increased reproductive tiller number in two *indOsRKD3 Cempo Ireng* T2 lines. Identification of high tiller number *Cempo Ireng* lines with ectopic expression *indOsRKD3* expression. **A.** Ectopic expression of *indOsRKD3* in *Cempo Ireng* wild type and *indOsRKD3* leaf tissue with *OsActin*, a housekeeping gene, as a control for RNA extraction and RT-PCR and WT *Cempo Ireng* as a control for ectopic expression. **B.** Representative samples from genotypes labelled below each plant had photographs taken on a black background at early generative stage. Scale bar in the bottom right of the plot is for 30cm. *OsRKD3 T1* and *OsRKD3 T2* refer to two T2 *indOsRKD3 Cempo Ireng* lines from independent transgenic events.

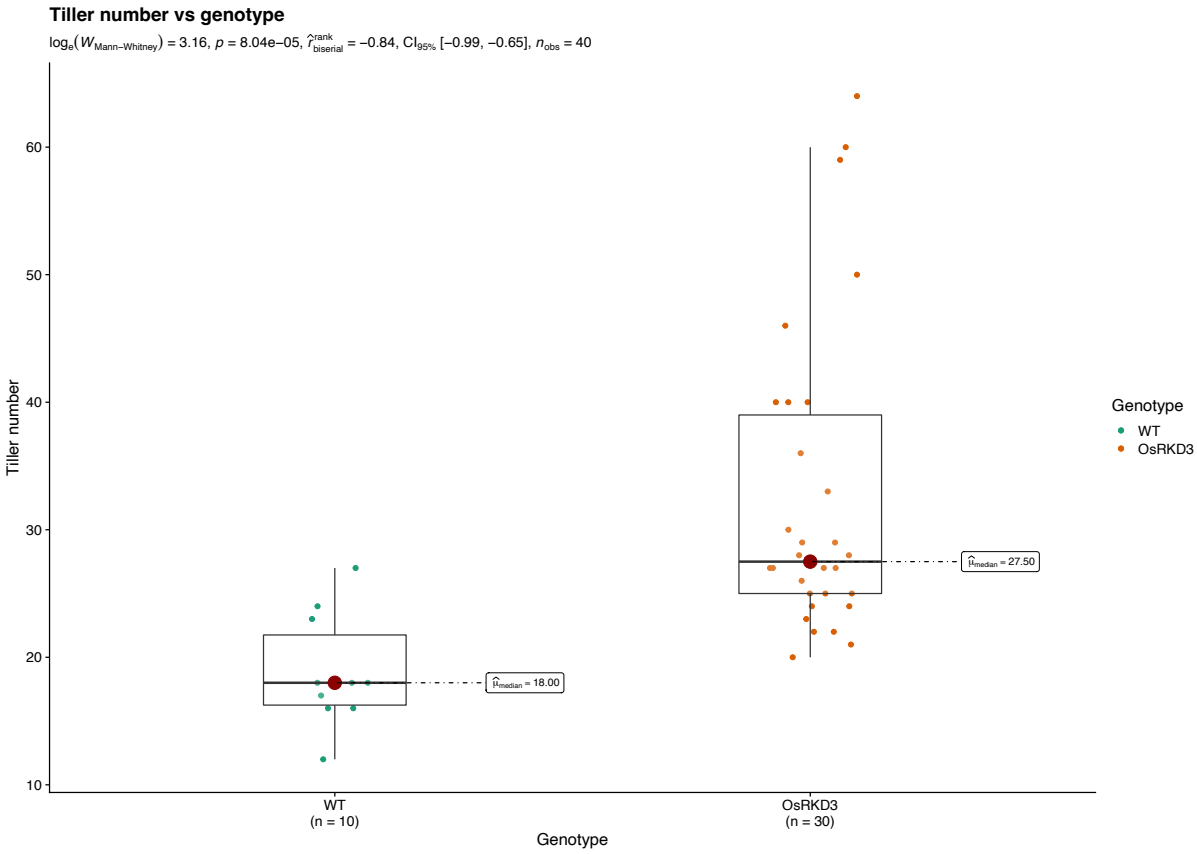


Figure 5.5. Reproductive tiller number in an *indOsRKD3 Cempo Ireng* T2 line. Reproductive tiller number from plants measured from counting the number of fully developed panicles at senescence. T3 lines were merged for more statistically robust tests. Sample number of each transgenic line is described in the (n=x) brackets below the label. The the mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. iSignificant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.

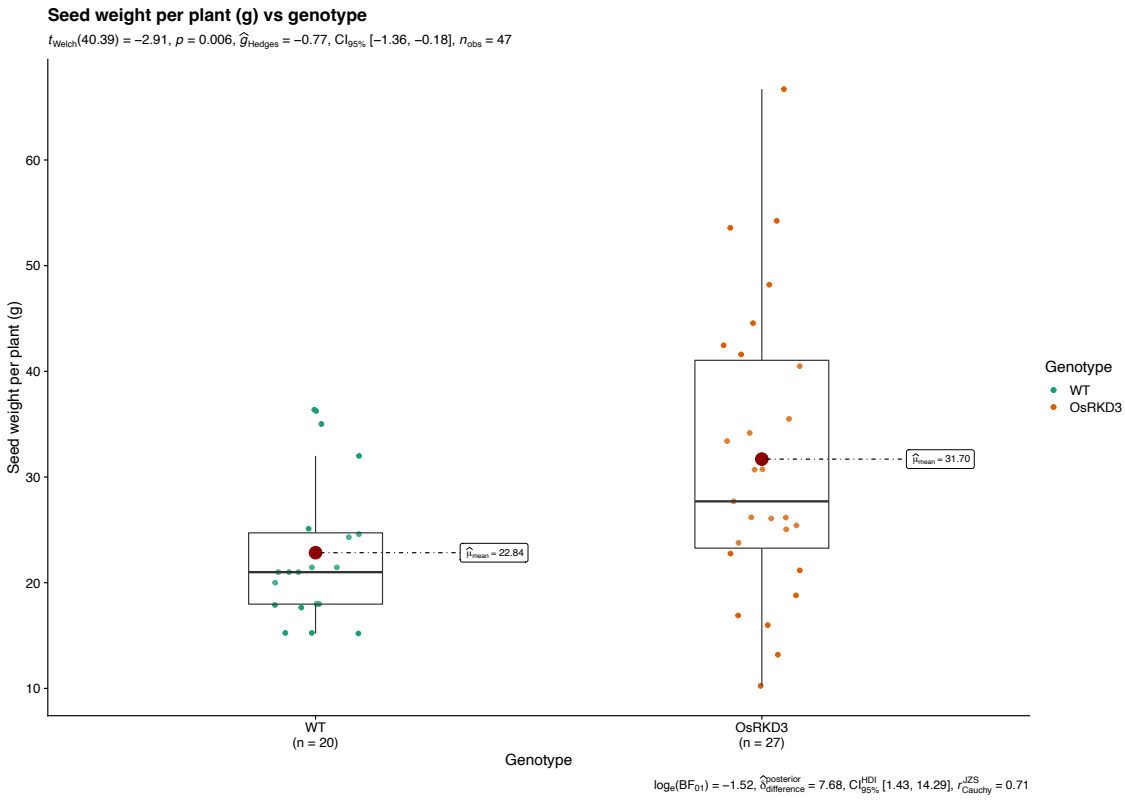


Figure 5.6. Seed weight per plant in an *indOsRKD3 Cempo Ireng* T2 line. The seed weight of the plant's seeds was measured after threshing fully dried panicle. T3 lines were merged for more statistically robust tests. The label of each transgenic line is described in the (n=x) brackets below the label. The the mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.

Figure 5.7. Panicle size, panicle branch number and seed number per panicle in two T2 *indOsRKD3 Cempo Ireng* lines. Representative panicles from each genotype labelled on the left were photographed on a white background. Black scale bar is set at 10cm and applies to all images.

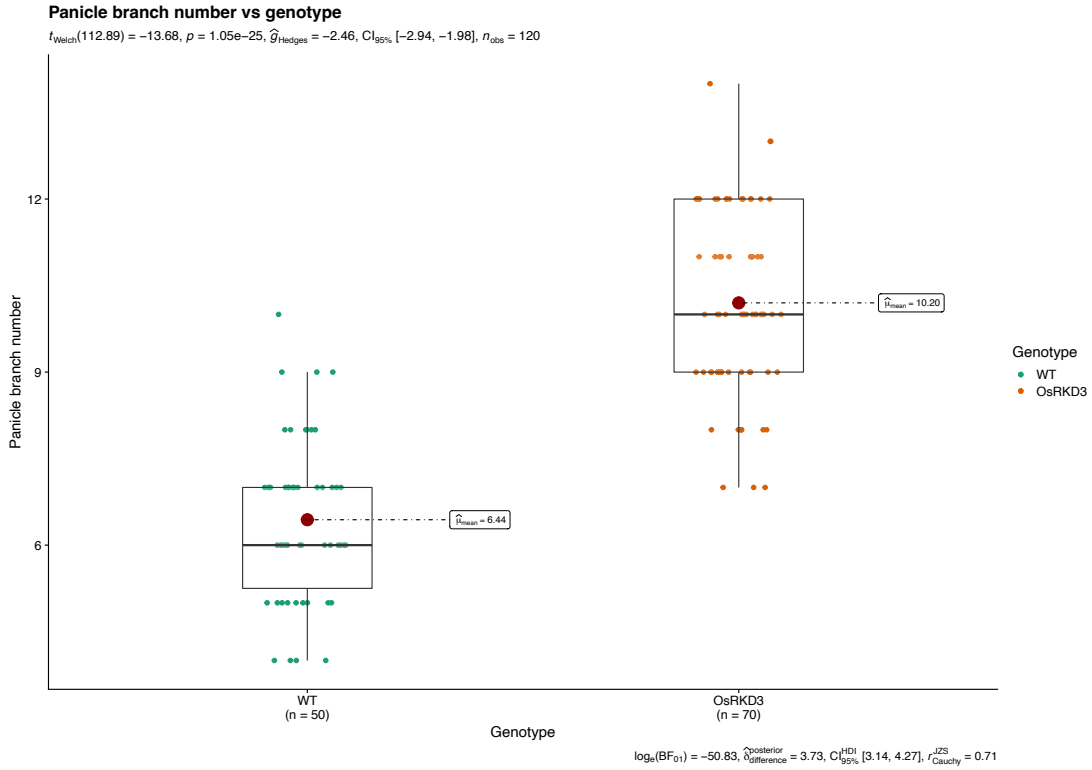


Figure 5.8. Panicle branch number in an *indOsRKD3 Cempo Ireng* T2 line. The average panicle branch number of each line was measured at senescence. T3 lines were merged for more statistically robust tests. The label of each transgenic line is described in the (n=x) brackets below the label. The the mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.

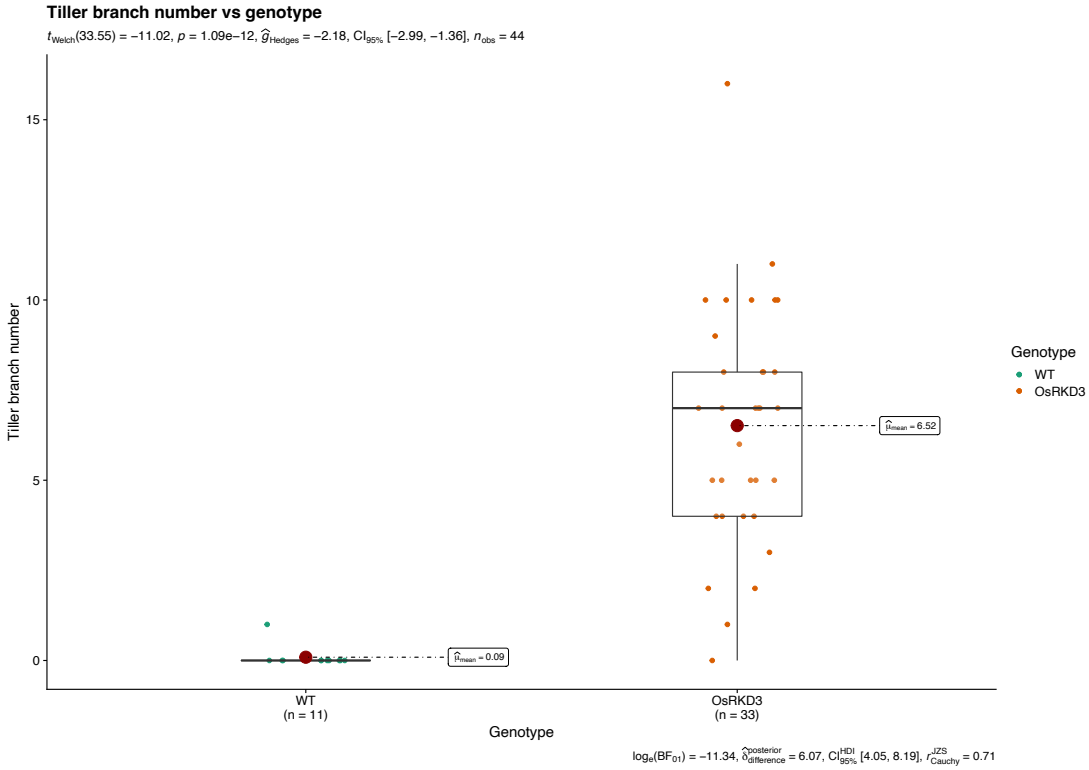


Figure 5.9. Increased tiller branch number in an *indOsRKD3 Cempo Ireng* T2 line. Tiller branch number, ~~a~~ homeotic transformation of an inflorescence meristem into a tiller, was counted in each plant. T3 lines were merged for more statistically robust tests. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.

Figure 5.10. Increased inflorescence branches in two *indOsRKD3 Cempo Ireng* T2 lines. Inflorescence branches are homeotic transformations of inflorescence meristems to form tillers. Representative samples from genotypes labelled above had photographs taken at 50 days after germination with a black background. White scale bars are 5cm. *OsRKD3 T1* and *OsRKD3 T2* refer to two T2 *indOsRKD3 Cempo Ireng* lines from independent transgenic events.

Table 5.1. Statistics for phenotypic traits in *indOsRKD3 Cempo Ireng* trial.

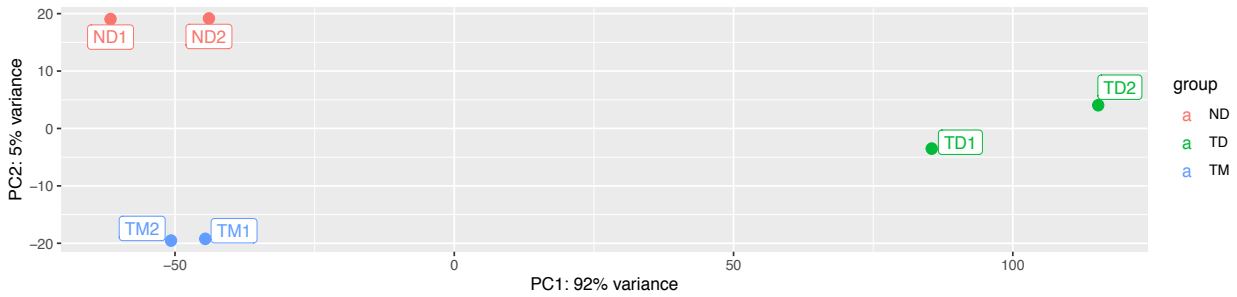
Traits	*indOsRKD3* FC	*indOsRKD3* p.val
Height	1.246987952	5.88E-05
Reproductive tiller number	1.527777778	8.04E-05
Days to maturity	1.981818182	7.93E-07
Tiller branch number	70>	1.72E-06
Panicle branch number	1.583850932	1.05E-25
Seed number per panicle	1.704042412	6.15E-18
Seed weight per plant	1.319047619	1.10E-02
Seed weight per 1000	0.887301587	2.91E-04

5.2.3. *indOsRKD3* leaf tissue shows upregulation of translational and metabolic procesess.

To investigate the transcriptional networks induced by ectopic expression of *indOsRKD3*, a RNAseq analysis was carried out in *indOsRKD3 Cempo Ireng* leaf tissue. Whole genome RNAseq data from *indOsRKD3* leaf tissue were obtained using four sample conditions: transgenic DEX (TD), transgenic mock I, and non-transgenic DEX (ND). *indOsRKD3* expression was verified in all transgenic samples, significantly different between TD and TM conditions, and replicate concordance was confirmed using principal component analysis (PCA) (Supplementary Figure 5.5, Figure 5.10). Pairwise comparisons for differentially genes (DEGs) were made between *indOsRKD3* and wild type samples to obtain a high confidence list of highly expressed *indOsRKD3* DEGs by taking DEGs specific to the TD vs ND comparison and removing DEGs shared in the TM vs ND list (TD vs ND only DEGs) (Figure 5.11).

TD vs ND DEGs showed an equal amount of up and downregulation (51% up) and TM vs ND DEGs showed greater downregulation (41% up) (Supplemental Figure 5.7). TD samples clustered together and showed a distinct expression profile to ND (Figure 5.12). Gene ontology (GO) term analysis was carried out on upregulated shared DEGs to test for functional enrichment of biological processes, cellular components, and molecular functions, showing significant enrichment of translational and metabolic terms (Figure 5.13). Like GO term analysis of *indTaRKD1-7D #1.4* leaf tissue it was difficult to deduce whether GO term enrichment was the result upregulating translational genes or consistent with energy demands for increased vegetative tissue growth.

A



B

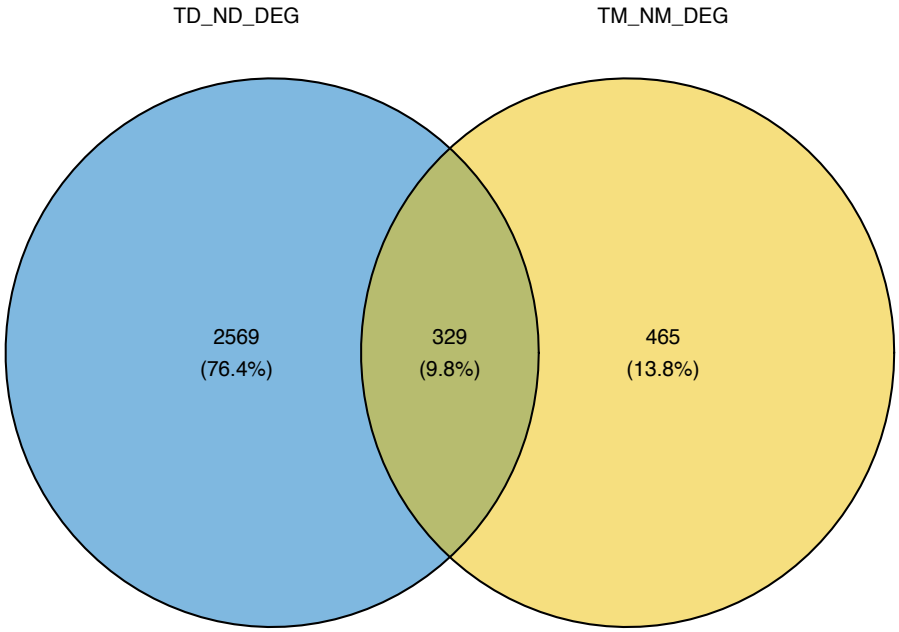


Figure 5.11. PCA of samples and Venn diagram of DEGs from *indOsRKD3 Cempo Ireng* leaf RNAseq experiment. TD = transgenic DEX, ND = non-transgenic DEX, TM = transgenic mock **(A)** Principal component analysis (PCA) of *indOsRKD3 Cempo Ireng* leaf samples coloured according to genotype and treatment condition as described in the legend on the right. Percentage of variance accounted for in each principal component is denoted beside the label. **(B)** Venn diagram of differentially expressed genes (DEGs) shared between the pairwise differential comparisons TD vs ND and TM vs ND. Percentages highlight proportion of total genes between both pairwise differential comparisons.

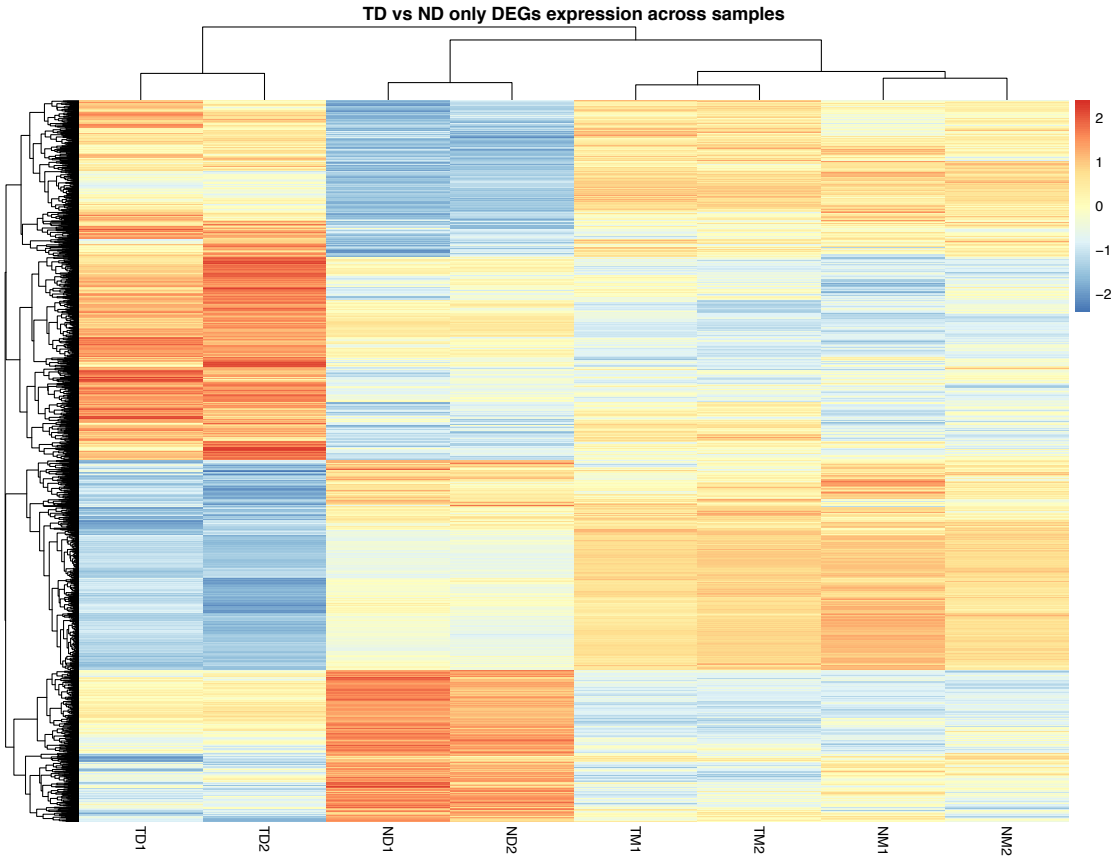


Figure 5.12. Heatmap of TD vs ND only DEGs from *indOsRKD3 Cempo Ireng* leaf RNAseq experiment. TD = transgenic DEX, ND = non-transgenic DEX, TM = transgenic mock. Scale displayed is row scaled logarithmically transformed DESeq2 normalised count values from htseq-count. Hierarchical clustering was performed on *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs and is displayed as a dendrogram on the side of the heatmap. Hierarchical clustering was also performed on samples and is displayed as a dendrogram on the top of the heatmap.

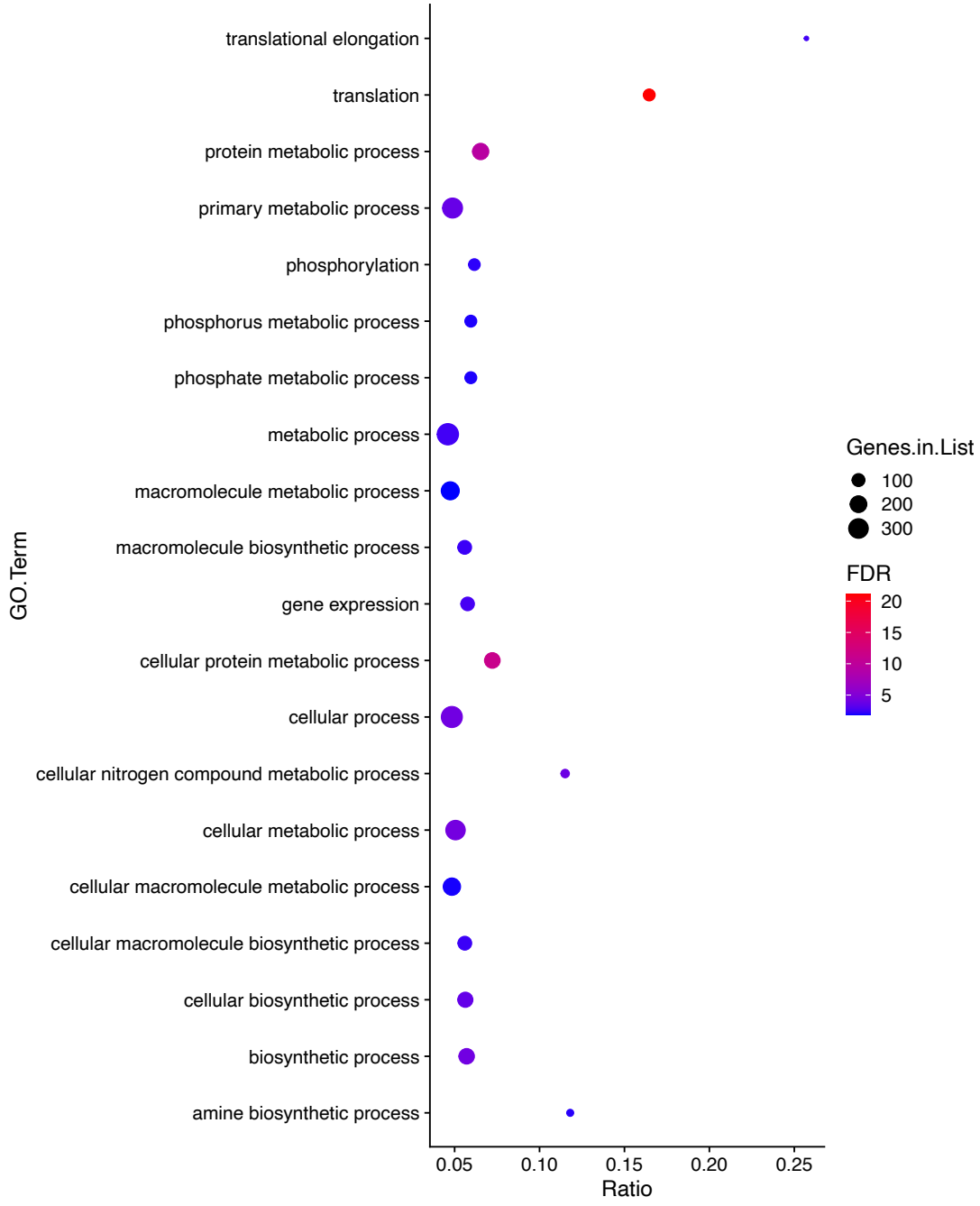


Figure 5.13. GO term analysis of upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs. The top 20 most significant gene ontology (GO) terms from the upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs are plotted here. GO terms are labelled on the y axis and the ratio of genes for a GO term in the upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs compared to the total number of genes corresponding to that GO term in the genome is plotted on the x axis. Point size is determined by the number of genes from that GO term in the upregulated NIL tiller node shared DEGs list as denoted on the legend on the right. False discovery rate (FDR) is computed by taking the -$\log_{10}$ of the p value determined from an enrichment test with a Benjimini-Hochberg correction and is coloured as denoted on the legend on the right.

5.2.4. Upregulated TD vs ND only DEGs with an AtRKD2 motif show enrichment for functional terms involved in transcriptional and metabolic processes

Transcriptomics analysis in leaf tissue was not sufficient to determine whether *indOsRKD3* upregulated translational GO terms directly, to solve for this the AtRKD2 motif, an evolutionary conserved and verified zygotic RKD motif, was used to predict TD vs ND only DEGs with an OsRKD3 binding motif in their 3kb promoter region. TD vs ND only DEGs were filtered for significant predicted AtRKD2 binding motifs, the only zygotic RKD in the DAPseq database, within their 3kb promoter region. It was confirmed there was a higher density of the AtRKD2 binding motifs upstream of the TSS than downstream (Figure 5.15). GO term analysis was carried out on TD vs ND only leaf DEGs with a predicted AtRKD2 binding motif in their 3kb upstream promoter region to test for functional enrichment of biological processes, cellular components, and molecular functions, showing significant enrichment of transcriptional and metabolic process terms, but not translational terms (Figure 5.16). Following this we decided to intersect our TD vs ND only DEGs with publicly available RNAseq data from developmental timecourses in meristematic and reproductive tissues to confirm genes that may be relevant to tillering.

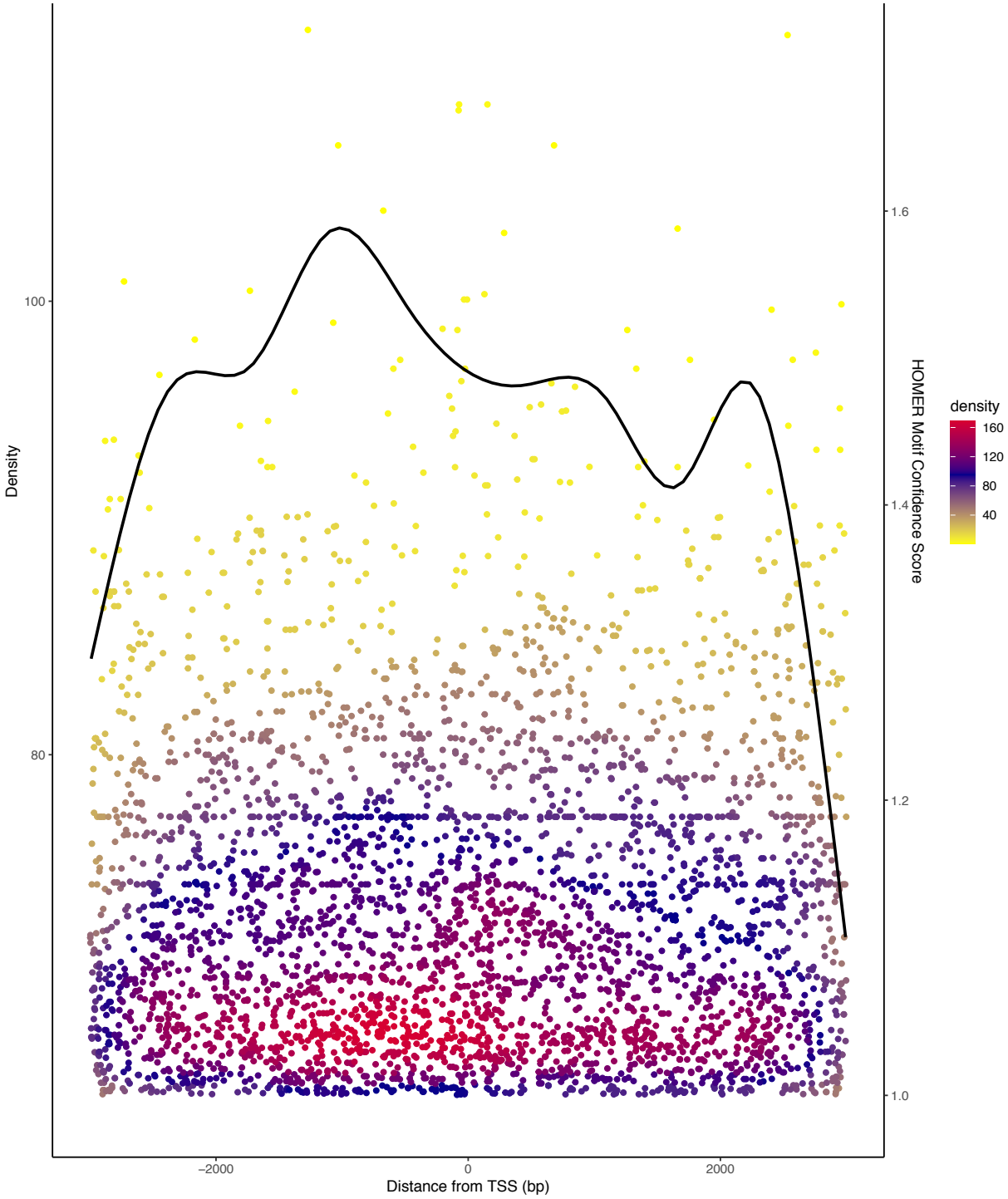


Figure 5.14. Increased incidence of predicted AtRKD2 binding motif in 3kb upstream promoter regions of upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs. Density values for predicted AtRKD2 binding motifs are given by the smooth line plot and the y-axis on the left. Points on the graph are also plotted along the y-axis on the right for a normalised HOMER motif score. Points are coloured according to density values denoted on the legend to the right.

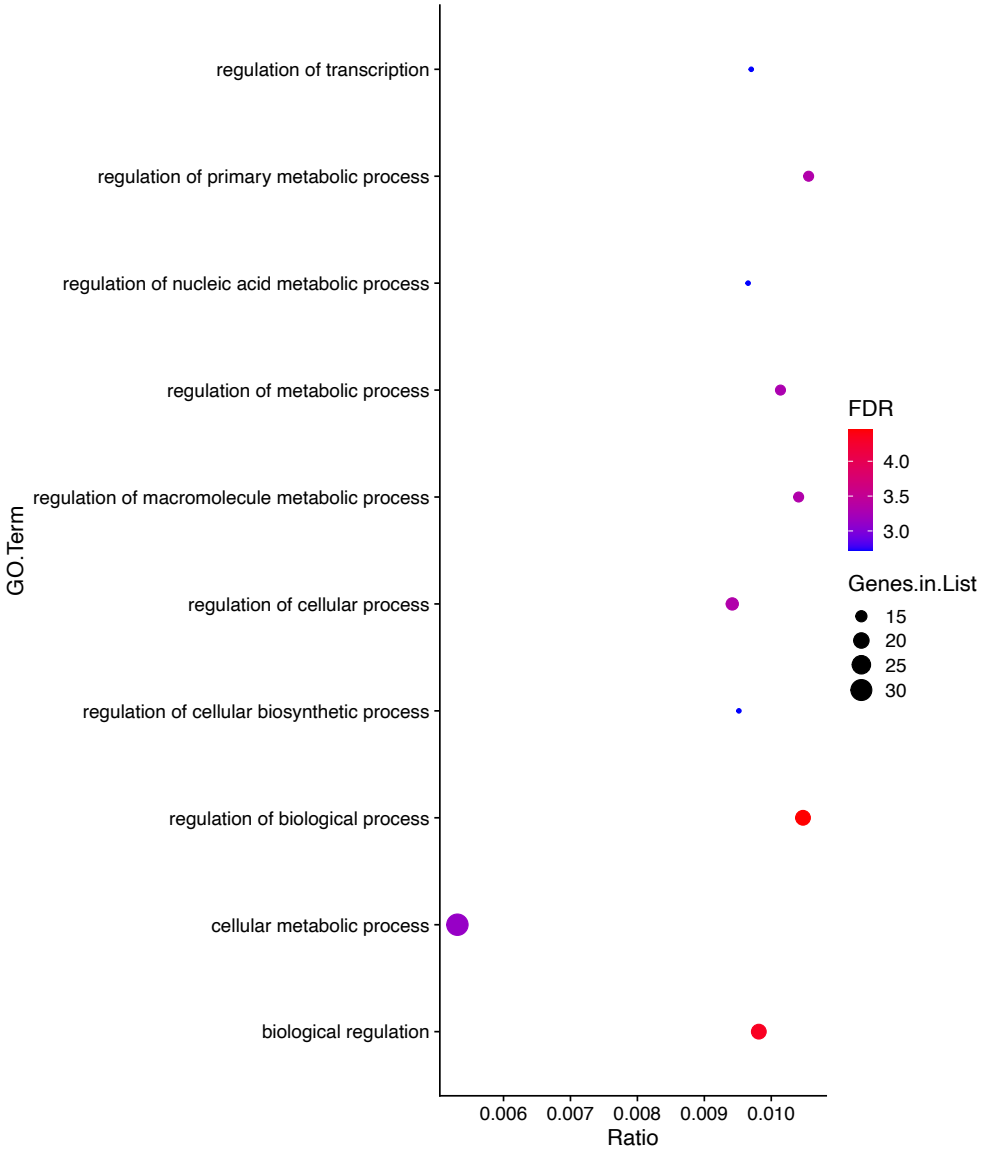


Figure 5.15. GO term analysis of upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs with a predicted AtRKD2 binding motif in their 3kb upstream promoter region shows enrichment of terms involved in transcription and metabolic process. The top 10 most significant gene ontology (GO) terms from the upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs with a predicted AtRKD2 binding motif in their 3kb upstream promoter region are plotted here. GO terms are labelled on the y axis and the ratio of genes for a GO term in the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs with a predicted AtRKD2 binding motif in their 3kb upstream promoter region compared to the total number of genes corresponding to that GO term in the genome is plotted on the x axis. Point size is determined by the number of genes from that GO term in the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs with a predicted AtRKD2 binding motif in their 3kb upstream promoter region list as denoted on the legend on the right. False discovery rate (FDR) is computed by taking the $-\log_{10}$ of the p value determined from an enrichment test with a Benjimini-Hochberg correction and is coloured as denoted on the legend on the right.

5.2.5. Upregulated TD vs ND only DEGs expressed in meristematic and reproductive tissue are involved in DNA binding and kinase signalling

To identify TD vs ND only DEGs that showed tissue specific expression in meristematic and reproductive tissues transcript per million values (tpm) were obtained from the MSU database for a variety of vegetative and reproductive tissues from *O. sativa ssp. japonica* 'Nipponbare'. TPM values for TD vs ND only DEGs were visualised with TD samples, vegetative tissues, and reproductive tissues (Supplemental figure 5.8). Hierarchical clustering split the TD vs ND only DEGs tpm values into 14 distinct clusters with different expression profiles (Figure 5.16). Cluster 12 showed an expression profile relevant to tillering with high expression of DEGs in TD, shoot, anther, and post-emergence inflorescence samples (Figure 5.17). GO term analysis was carried out on cluster 12 DEGs to test for functional enrichment of biological processes, cellular components, and molecular functions, showing some significant enrichment of DNA binding and kinase signalling terms, but showing very significant enrichment of lipid transport that has been strongly implicated to play a major role in wheat and rice inflorescence development, consistent with their high expression seen in anther and post-emergence inflorescence tissue (Figure 5.17, Figure 5.18) (Zhang et al., 2010; Ma et al., 2018). These terms were not relevant to other molecular pathways we had elucidated in *indTaRKD1-7D* transcriptomics analyses, so we decided to investigate the transcriptional cascade of *indOsRKD3* using a coexpression network as this analysis might identify coregulated genes more relevant to tillering.

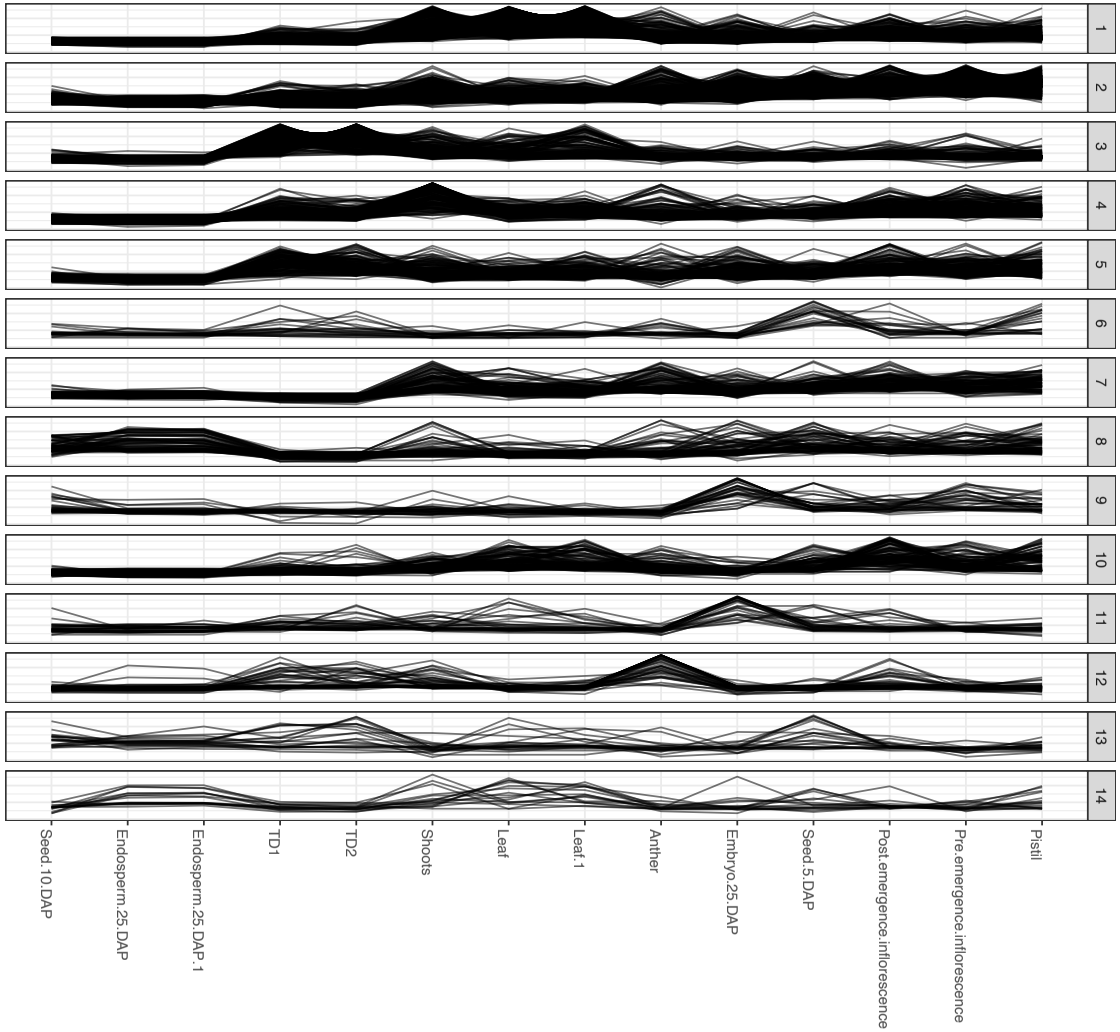


Figure 5.16. Clustered expression plot shows tissue specific expression amongst *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs. TD = transgenic DEX. Heatmap in supplementary figure 5.5 was split into clusters from a dendrogram value of 6.7. Lines represent normalised expression values at samples and labels on the right are cluster numbers. Cluster 12 shows high TD, shoot, anther, and post-emergence inflorescence expression.

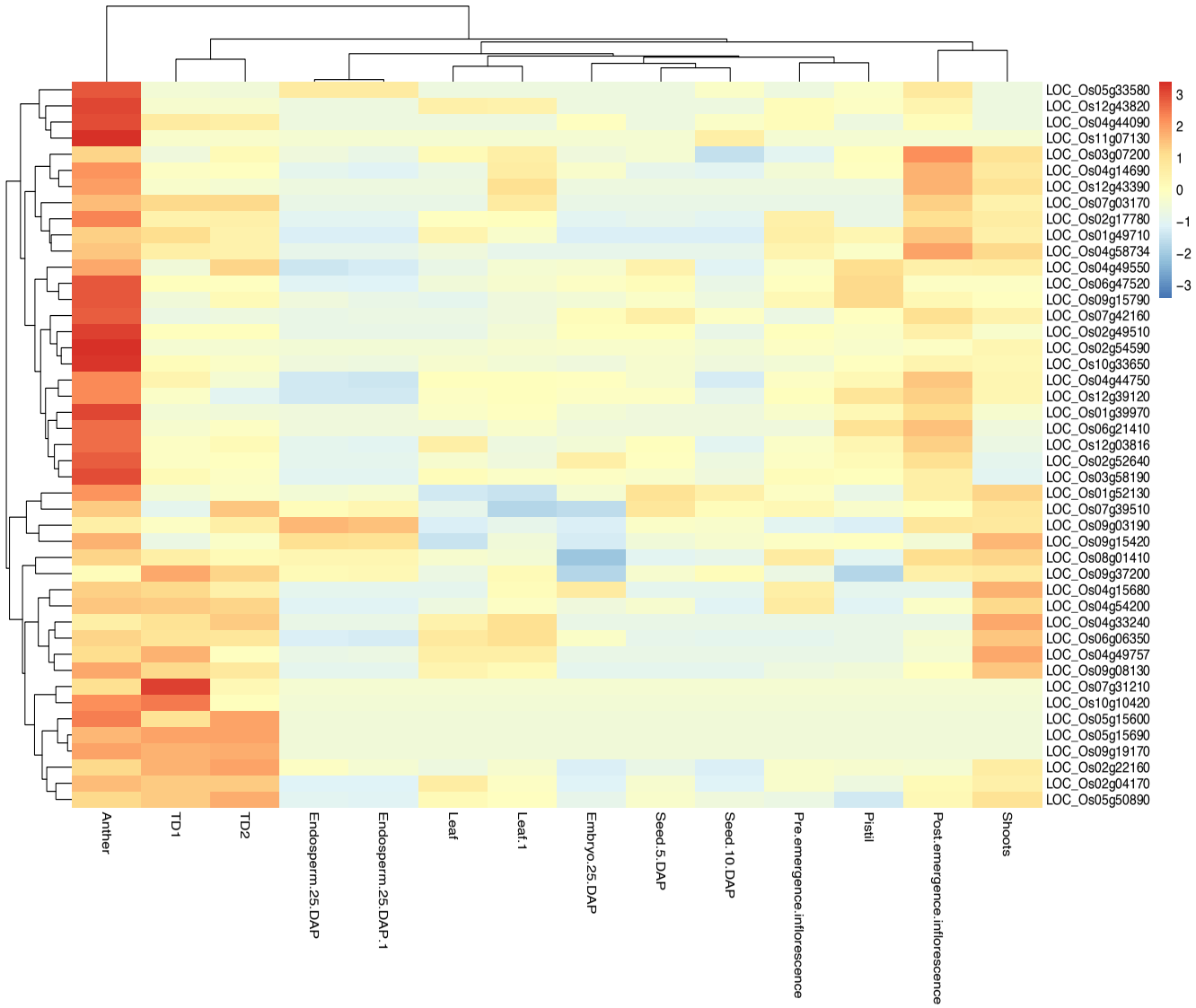


Figure 5.17. Heatmap of Cluster 12 shows list of genes that show high expression in tissues like those involved in tiller initiation. Heatmap aims to illustrate the upregulation of one distinct cluster of genes across inflorescence tissues to highlight that a subset of the OsRKD3 transcriptional cascade is implicated in meristem development. TD = transgenic DEX. Scale displayed is row scaled logarithmically transformed DESeq2 normalised count values from htseq-count. Hierarchical clustering was performed on *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs labelled on the right and is displayed as a dendrogram on the side of the heatmap. Hierarchical clustering was also performed on samples and is displayed as a dendrogram on the top of the heatmap.

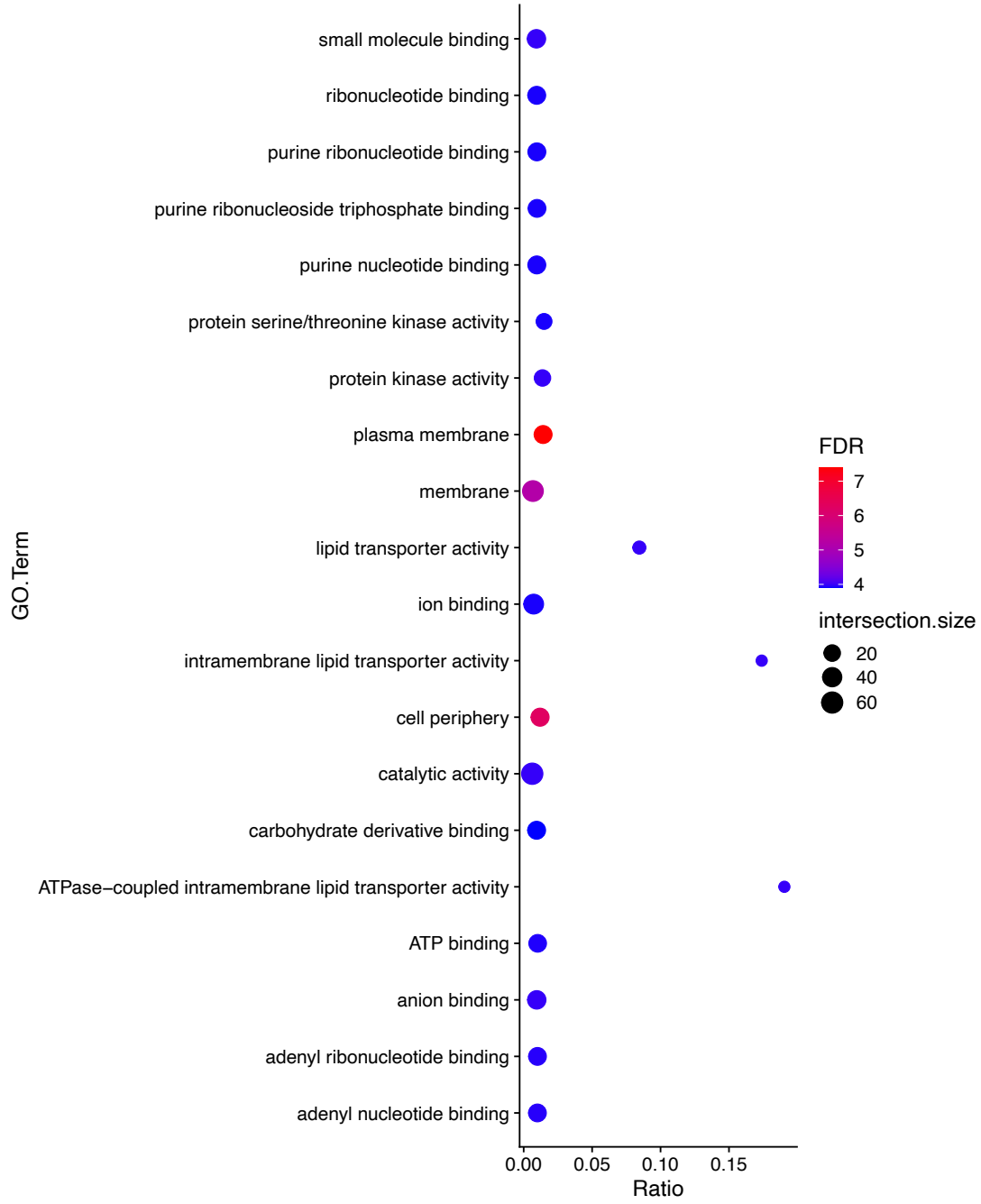


Figure 5.18. GO term analysis of *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs Cluster 12 shows enrichment of terms involved in DNA binding and signalling. The top 20 most significant gene ontology (GO) terms from the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs Cluster 12 are plotted here. GO terms are labelled on the y axis and the ratio of genes for a GO term in the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs Cluster 12 compared to the total number of genes corresponding to that GO term in the genome is plotted on the x axis. Point size is determined by the number of genes from that GO term in the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs Cluster 12 list as denoted on the legend on the right. False discovery rate (FDR) is computed by taking the -$\log_{10}$ of the p value determined from an enrichment test with a Benjimini-Hochberg correction and is coloured as denoted on the legend on the right.

5.2.6. An upregulated *indOsRKD3* TD vs ND only DEGs coexpression network shows functional enrichment for kinase signalling, stress response and cell cycle

To identify coexpressed TD vs ND only DEGs that were involved in tillering, a publicly available global coexpression network was filtered for high confidence coexpression correlations with at least one upregulated *indOsRKD3* TD vs ND only DEGs and separated into 6 coexpression clusters. GO term analysis of coexpression clusters showed significant enrichment of functional terms implicated in protein phosphorylation, translation, cell cycle and stress related processes (Figure 5.17). This network is available at nDEX for visualisation under 'indOsRKD3 coexpression network GO term'. Redundant GO terms in two clusters were removed and plotted in semantic space showing terms for transcription, signalling and microtubule-based processes signalling in the red 'protein phosphorylation' cluster and terms for cytoplasmic translation and cellular component organisation in the lilac 'translation' cluster. (Figure 5.18, Figure 5.19). The coexpression network was annotated by log2 fold change from the *indOsRKD3* TD vs ND pairwise comparison and showed that most genes were upregulated (Figure 5.20). This network is available at nDEX for visualisation under 'indOsRKD3 coexpression network logfold change'. Network analysis showed nodes with the highest betweenness centrality, a measure of how much information passes through a node, were kinase signalling and stress related genes (Figure 5.21). While GO terms enriched in clusters like cell cycle and fertilisation were relevant to tillering, no high confidence and consistent transcriptional trends could be matched between *indTaRKD1-7D* and *indOsRKD3* RNAseq experiments.

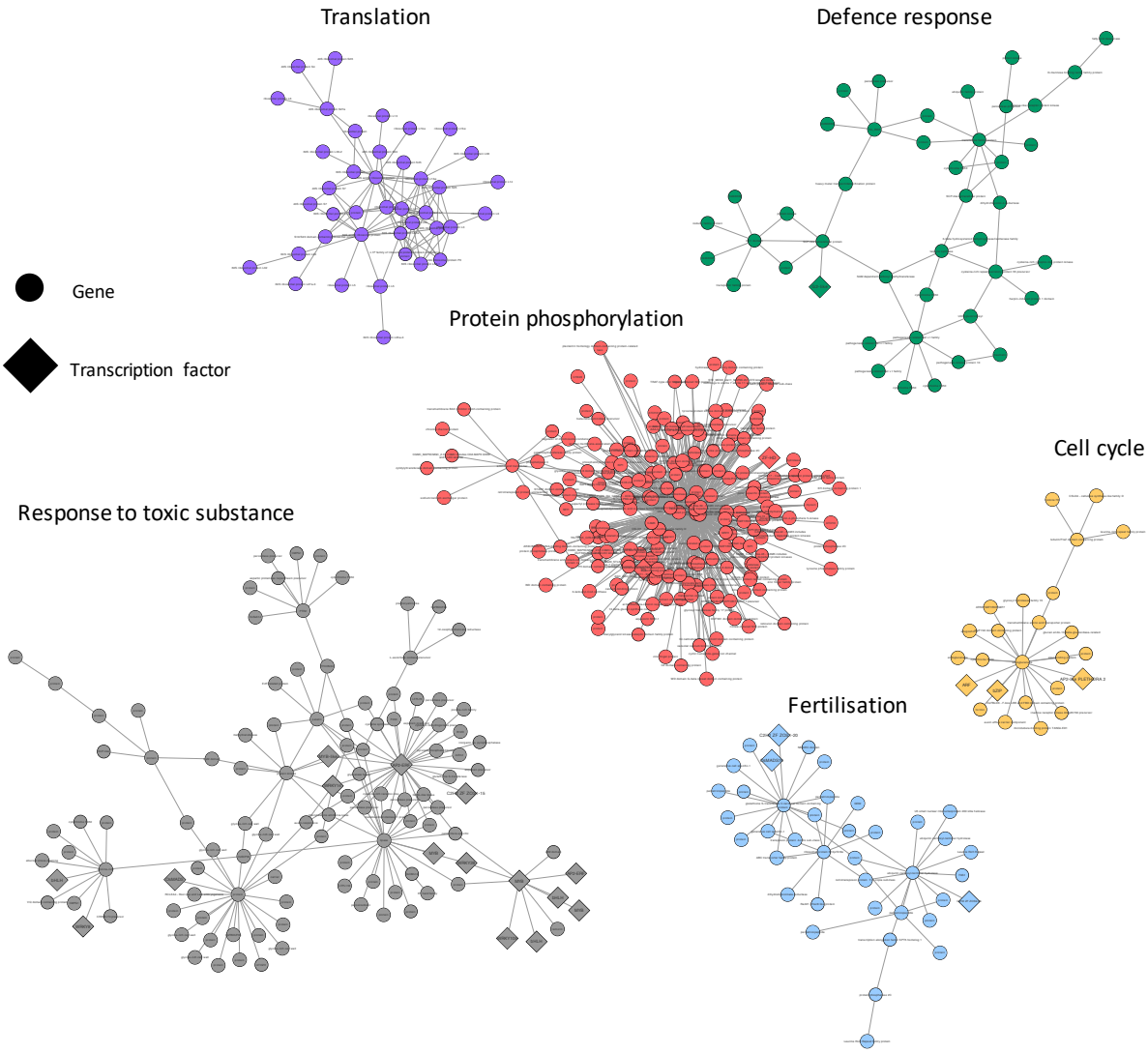


Figure 5.19. Upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs coexpression network clustered and functionally annotated. Nodes (genes) and edges (coexpression relationships) are spatially arranged with a prefuse force-directed layout in cytoscape and clustered according to how they separated in semantic space. Larger diamond nodes represent transcription factors. Each colour represents a cluster with a significant GO term that is labelled above the cluster.

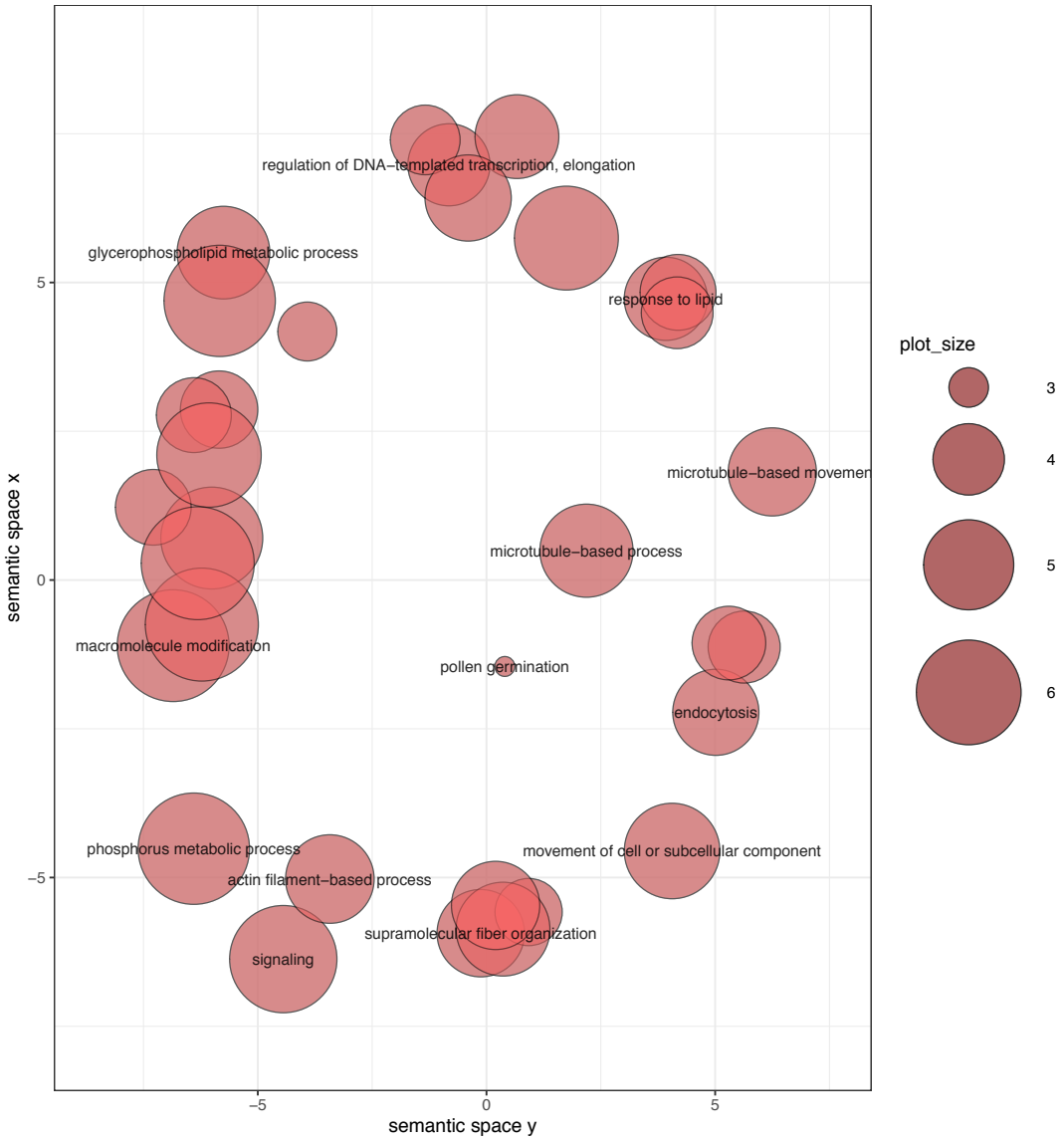


Figure 5.20. Revigo analysis of GO terms from protein phosphorylation cluster. GO terms coloured as found in the red 'protein phosphorylation cluster of figure 5.18 plotted in semantic space based on GO term similarity with the term size represented on the right of the plot.

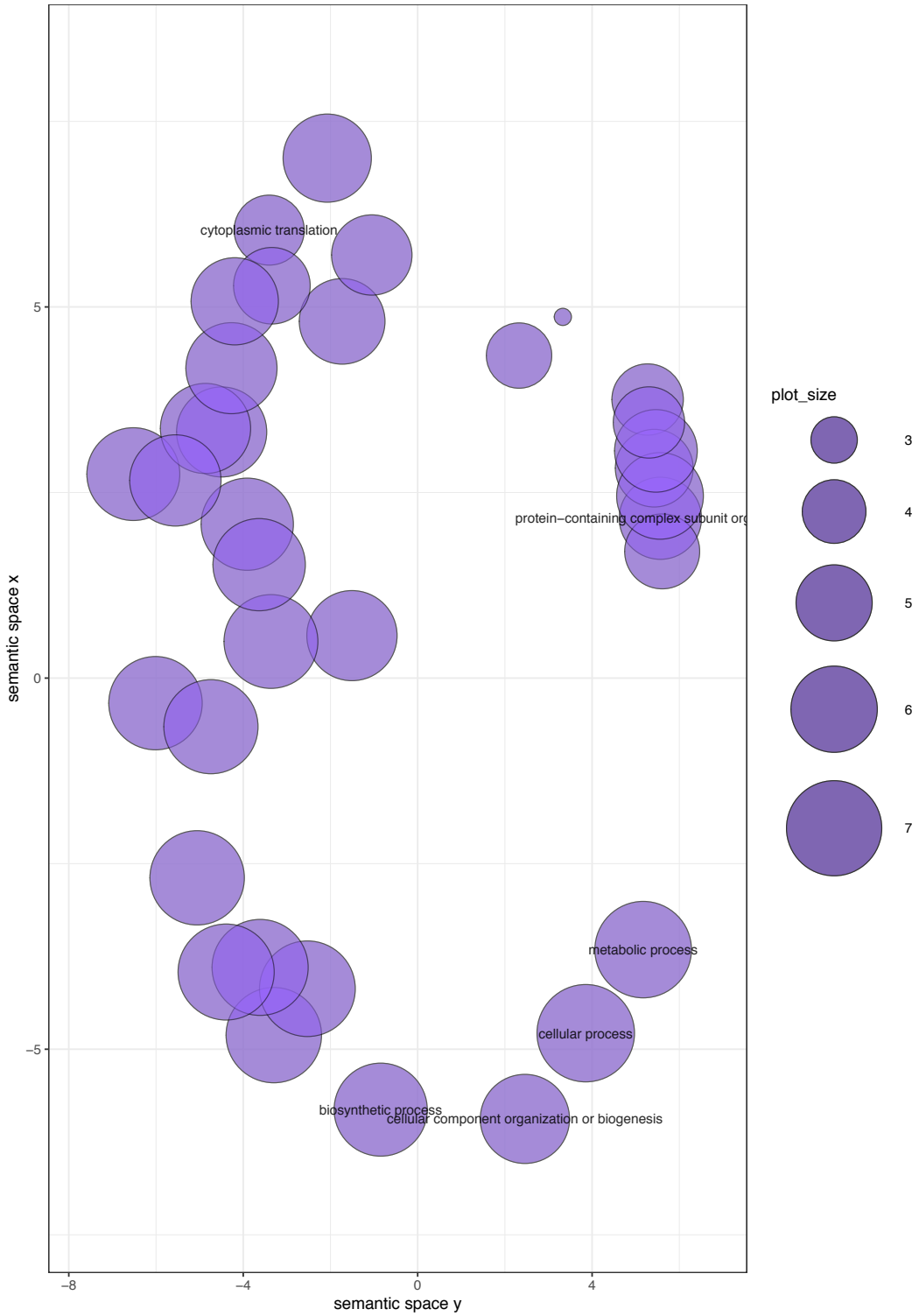


Figure 5.21. Revigo analysis of GO terms from translation cluster. GO terms coloured as found in the lilac 'translation' cluster of figure 5.18 plotted in semantic space based on GO term similarity with the term size represented on the right of the plot.

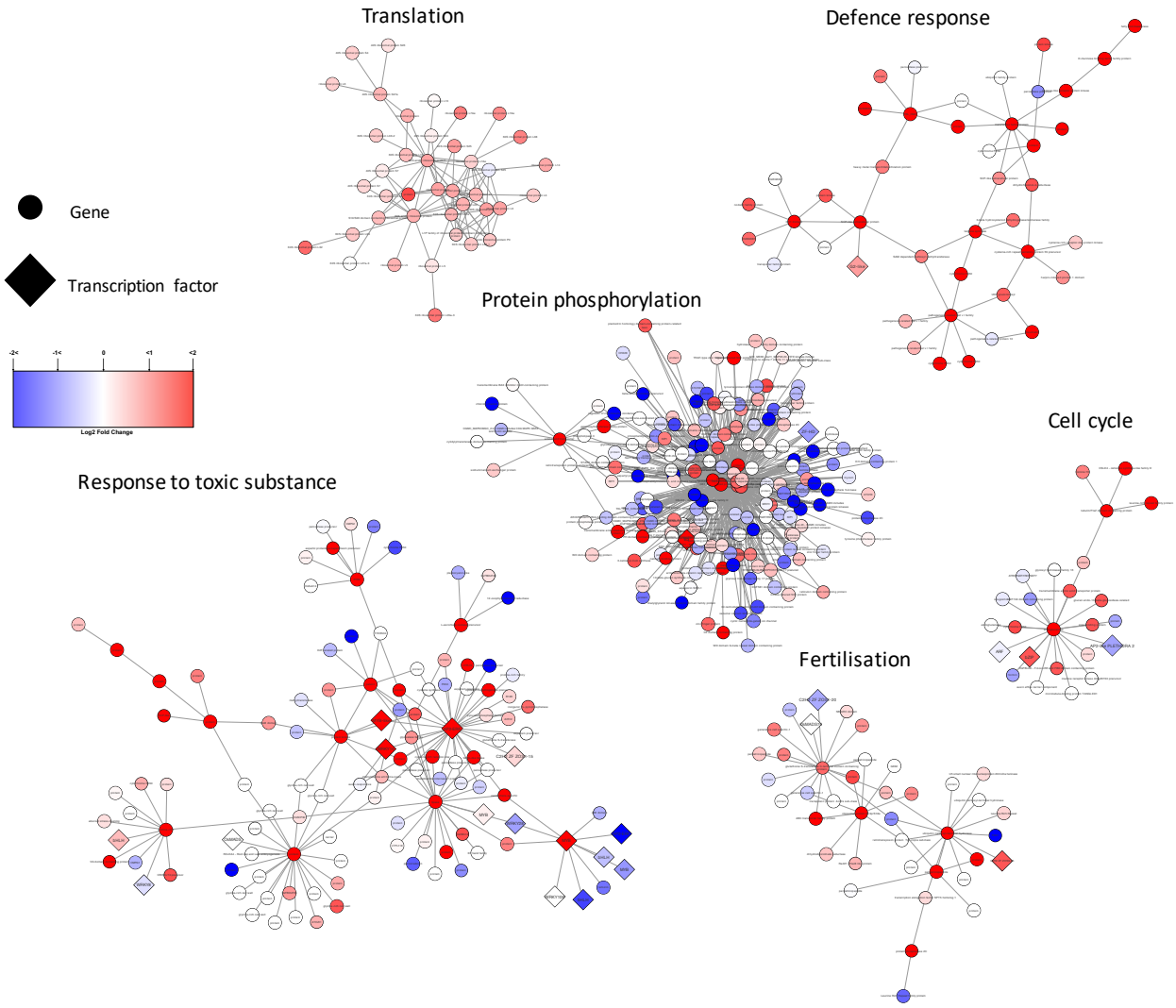


Figure 5.22. Upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs coexpression network annotated by differential expression. Nodes (genes) and edges (coexpression relationships) are spatially arranged with a prefuse force-directed layout in cytoscape and clustered according to spatial separation. Larger diamond nodes represent transcription factors. Nodes are coloured by log2 fold change in expression values from the *IndOsRKD3 Cempo Ireng* TD vs ND only comparison as denoted in the top left.

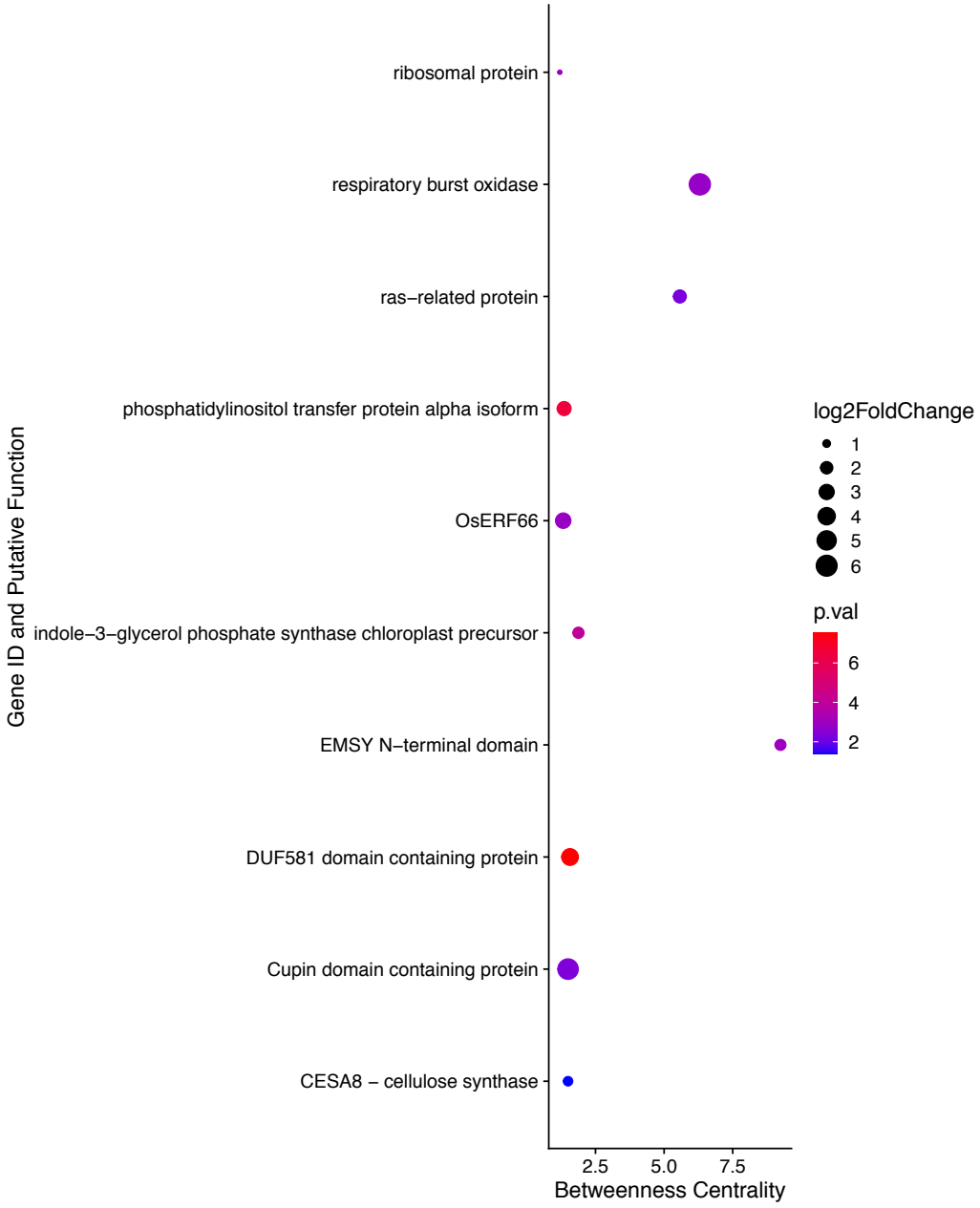


Figure 5.23. Enrichment of AP2 and ras-related genes in nodes with highest betweenness centrality in the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs coexpression network. Top 10 nodes with the highest betweenness centrality of the upregulated *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs coexpression network are plotted with a log10 transformation. Betweenness centrality is a measure of the number of coexpression correlation values that map in and out of a node. Point size is determined by the differential log2 fold change of that gene in the *indOsRKD3 Cempo Ireng* TD vs ND only pairwise comparison as denoted on the legend on the right. The adjusted p value is computed by taking the -log10 of the differential expression p value from the *indOsRKD3 Cempo Ireng* TD vs ND only pairwise comparison for that gene and is coloured as denoted on the legend on the right.

5.2.7. *OsRKD3* induces higher tiller formation in *Cempo Ireng* in an expression level dependent manner

To finalise our molecular analysis of *indOsRKD3* lines, acid fuschin, a histological staining method for assessing cell density and meristem formation was used to assess axillary meristem formation in rice seedling in the presence and absence of the *indOsRKD3* chemical inducer DEX. *indOsRKD3* and wild type plants were germinated on media in the presence or absence of DEX and were stained with acid fuschin to identify axillary meristem formation 15 and 21 days after sowing. Tiller formation was increased when comparing *indOsRKD3* to the wild type, with *indOsRKD3* showing higher tiller formation in the presence of DEX (Figure 5.22). This confirmed that *OsRKD3* expression increased tiller formation in *Cempo Ireng* in an expression level dependent manner.

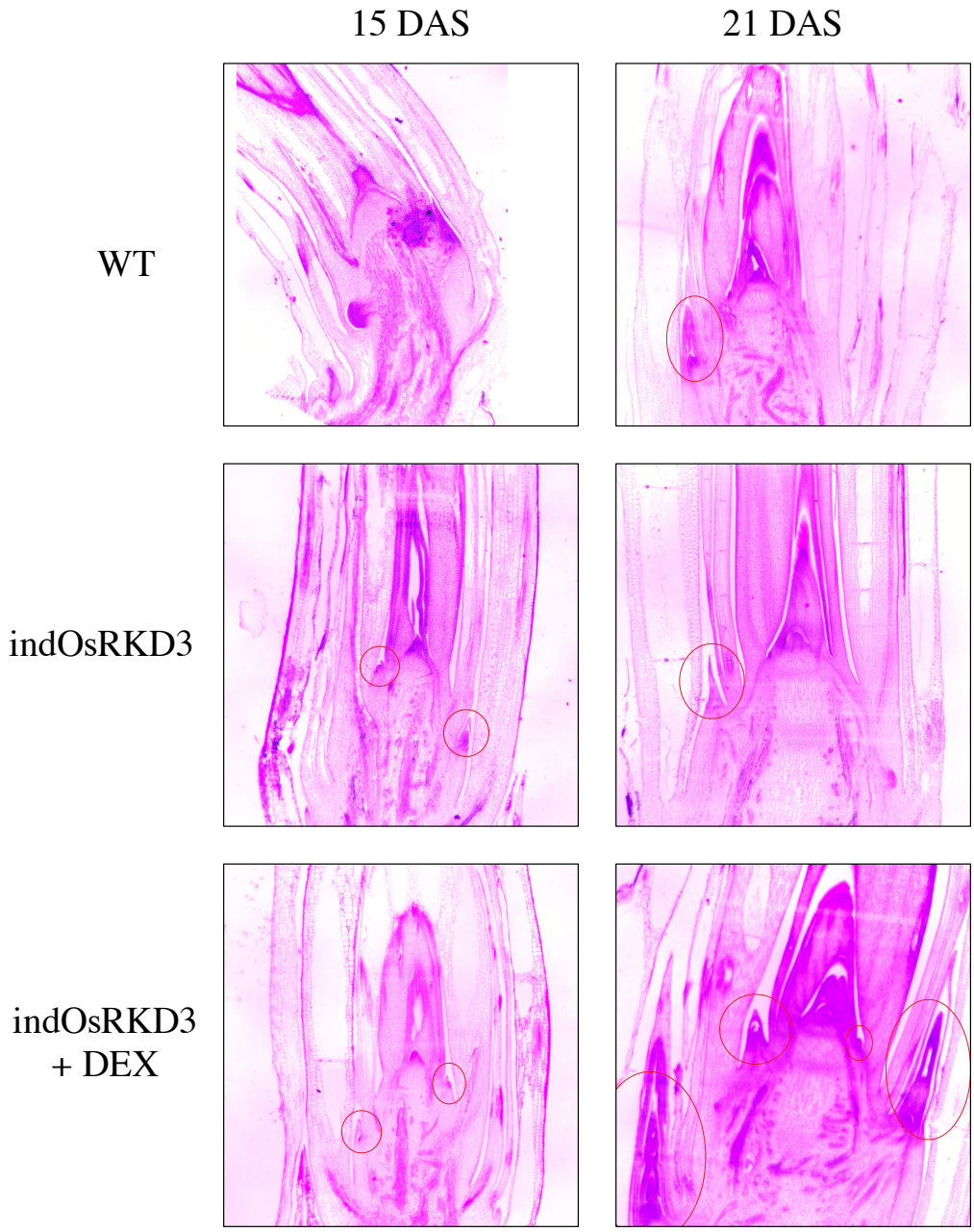


Figure 5.24. Increased tiller initiation in *indOsRKD3 Cempo Ireng* T2 line in a dose dependent manner. Tiller buds are circled in red in all images and days after sowing (DAS) is given by the labels at the top of the image. DEX is the chemical inducer of *indOsRKD3 Cempo Ireng* and results in higher *OsRKD3* expression, consistent with an increase in tiller bud initiation in *indOsRKD3* + DEX compared to *indOsRKD3*.

5.3. Discussion

5.3.1. Ectopic expression of zygotic RKDs induces higher tiller number in several BEP clade grass species

Here, we've shown that ectopic expression of the zygotic transcription factors, *AtRKD4* and *OsRKD3*, induce higher tiller number in 'Nipponbare', a cultivar of *O. sativa ssp. japonica* and *O. sativa* '*Cempo Ireng'*, a black Indonesian rice whose subspecies hasn't been determined, respectively. Ectopic expression of *OsRKD3* in *Cempo Ireng* led to an increase in plant height and seed weight per plant as in *indTaRKD1-7D* 'Fielder' lines. Inflorescence architecture was also drastically affected in *indOsRKD3 Cempo ireng* with an increase in the panicle branches and seeds per panicle. Two major differences between *indOsRKD3 Cempo Ireng* and *indTaRKD1-7D* 'Fielder' phenotypes was the seed weight per 1000 and an internodal tiller branching phenotype. Measurements seems contradictory in seed weight experiments with *indTaRKD1-7D* 'Fielder' showing an increase and *indOsRKD3* 'Cempo Ireng' showing a decrease in seed weight per 1000. This could be due to multiple factors such as incomplete grain filling due to a high number of unproductive tillers or species-specific differences in vegetative and inflorescence architecture. Field trials with large replication will be needed to investigate this further to clarify how ectopic expression of zygotic RKDs effects grain filling and seed weight per 1000. The internodal tiller branching phenotype has been previously described in the literature through homeotic mutations of AP1/FUL-Like MADS box and SEPALLATA MADS box genes (Kobayashi et al., 2012). However, in these studies floral meristems (FM) are not formed. Interestingly, the adventitious roots that can be seen at the nodal junctions suggests that this is a full transformation of an inflorescence meristem (IM) to an axillary meristem (AXM), possibly meaning that *OsRKD3* is reprograms IM to its

earlier AXM state. This also suggests that ectopic *OsRKD3* expression can influence IM and potentially FM meristems in a similar manner to how it increases AXM formation and tiller bud outgrowth.

5.3.2. Transcriptomics analysis of *indOsRKD3 Cempo Ireng* leaf tissues shows upregulation of terms relevant to meristem and inflorescence development

Dexamethasone (DEX) conditions in *indOsRKD3* samples significantly increased ectopic *OsRKD3* expression by a factor of 10-fold, much higher compared to Estradiol conditions in *indTaRKD1-7D* leaf and immature embryo tissue. This suggests that DEX appears to be a more potent chemical inducer of gene expression which may be because Estradiol has poor intercellular kinetics, although most of the literature shows it is appropriate for plant systems. Tissue specific differences between rice and wheat also could play a role, particularly the waxy cuticle of wheat leaf tissue, but this would not explain the lack of induction in immature embryos. As highlighted in the discussion of Chapter 4, leaf tissue doesn't appear to have an appropriate chromatin accessibility context for zygotic RKD-induced dedifferentiation. However, through hierarchical clustering analysis using our samples and publicly available time course data we overcame this obstacle to identify a subset of genes that were upregulated in anther and inflorescence tissues (Figure 5.16, Figure 5.17). This cluster showed very significant enrichment of lipid transport GO terms that have been shown to play a major role in inflorescence development in rice and wheat (Zhang et al., 2010; Ma et al., 2018). This suggests a new class of genes implicated in the spike and panicle phenotypes observed in *indTaRKD1-7D* and *indOsRKD3* respectively. This was likely not detected in wheat due to the poor GO term annotation in the wheat genome compared to rice but could also represent a rice specific transcriptional pathway that

isn't present here. Additionally, through filtering for a high confidence list of *indOsRKD3* upregulated leaf DEGs for predicted *AtRKD2* motifs we constructed a coexpression network that nicely surmised the balance of signal to noise in this transcriptomics analysis. Here, we saw upregulation of cell cycle and translation coexpression clusters in the *indOsRKD3* TD vs ND only DEGs coexpression network that suggests high ectopic expression of *OsRKD3* can induce some transcriptional changes in rice leaf tissue. However, we also saw that the biggest cluster in the network is response to toxic substance, which suggests that this level of filtering data does not increase the signal to noise ratio in the dataset in most cases. The consistency in phenotypic results between *indTaRKD1-7D* and *indOsRKD3* haven't been replicated in transcriptomics analysis here, but we see some indication that there are signatures of transcriptomics processes relevant to zygotic RKD-induced meristem formation. To conclude, while some RKD relevant processes can be identified in leaf tissue, it can be clearly stated that leaf tissue does not contain much of the chromatin accessibility necessary for RKD induced meristem formation.

Chapter 6

6.1 General discussion

"We may observe that the plant is capable of taking this sort of backward step, reversing the order of growth. This makes us all the more aware of nature's regular course; we will familiarize ourselves with the laws of metamorphosis by which nature produces one part through another, creating a great variety of forms through the modification of a single organ" – The Metamorphosis of Plants by Johann Wolfgang Von Goethe – 1790.

Plant development requires the continuous initiation of new organs through pluripotent stem cell niches called meristems. Meristem formation and maintenance determines plant architecture through integration of environmental signals over their lifecycle, which can last from weeks to centuries. The domestication of major crops through selection of plant architecture traits amenable to agriculture frequently selected quantitative trait loci (QTL) that controlled meristem determinacy such as *Teosinte Branched 1* (*TB1*) for tiller bud outgrowth in maize and CLAVATA 3 (*CLV3*) whose misexpression increases fruit sizes in tomato through enlarged floral meristems (FM) (Doebley et al., 1995; Xu et al., 2015). Natural diversity amongst wild relatives is an invaluable resource for variants that modulates plant architecture that can be used for crop improvement, safeguarded by organisations like ICARDA which have maintained over 80,000 wheat accessions despite disruption with the Syrian civil war (Westengen et al., 2020). However, the pressure of climate change on agriculture requires the acceleration of genetic gain in cereal crops through the identification and generation of novel mutations not necessarily present in nature through novel

technologies like genome editing. We've shown here that ectopic expression of zygotic RKD can increase tiller number and has the potential to improve yield in a range of cereal crops like wheat and rice. While we've extensively characterised the molecular pathways in *TaRKD1* induced tiller formation, we didn't manage to show this for ectopic expression of *OsRKD3*. The following is a model for *TaRKD1* induced higher tiller number and spikelet per spike number (summarised in Figure 6.1):

1. TaRKD1 binds promoter regions and upregulates genes involved in chromatin accessibility and cell cycle in pre-meristem tissue.
2. Reprogramming of pre-meristem to axillary meristem (AXM) tissue increases the number of outgrowing tiller buds in the plant.
3. AXMs are maintained through *TaRKD1* promoting meristematic indeterminacy through upregulation of cell cycle genes, maintaining a dividing pluripotent AXM. This allows AXMs to develop into mature tillers.
4. We also speculate that ectopic *TaRKD1* expression promotes meristematic indeterminacy in the shoot apical meristem, increasing plant height, and in the inflorescence meristem, increasing spikelet number per spike via cell cycle genes and meristematic determinacy factors such as trehalose-6-phosphate synthases.

To expand on our *TaRKD1* induced tiller number model with our phenotypic data, we've shown that *TaRKD1* induced tiller formation is not inhibited by endogenous signals that temper tiller outgrowth with nutrient availability to the same degree as wild type plants. Interestingly, we also see that at lower nitrogen availability there is a greater increase in tiller number relative to the wild type. With fertiliser production

being recognised as one of the greatest contributors to greenhouse gas emissions, there has been a demand for making wheat cultivars less reliant on high nutrient inputs (Hawkesford 2014; Chai et al., 2019). Many research groups have tackled the goal of sustainable agriculture by increasing tiller number as it appears that ideotype breeding can maximise nutrient use, with some using RWP-RK transcription factors as well (Li et al., 2018; Wu et al., 2020; Yu et al., 2021). The high evolutionary conservation of zygotic RWP-RK transcription factors across the plant kingdom means this strategy may be used to improve nutrient use in non-grass breeding programs through increasing lateral branch formation (Koi et al., 2016). However, it is possible that increased nitrogen response in rice and wheat is due to adventitious roots formed with each additional tiller, which would not translate to increased nutrient use in non-grass breeding programs. To confirm this, it would be useful in future experiments to investigate the root architecture of *indTaRKD1-7D* to confirm an increase in adventitious root formation through new phenotypic imaging techniques that have become standard in the field (Atkinson et al., 2019).

This finding opens the possibility of a novel breeding strategy though adjusting the level of ectopic *TaRKD1* expression by genome editing to maximise nitrogen response while minimising the number of unproductive tillers. Further to this, ectopic expression of epigenomic regulators has been shown to improve yield-determining traits like *NITROGEN GROWTH RESPONSE GENE 5* (*NGR5*) through recruiting *POLYCOMB REPRESSIVE COMPLEX 2* (*PRC2*) to repress expression of shoot branching inhibitory genes, increasing nitrogen use efficiency (NUE) (Wu et al., 2020). This suggests that increasing tiller number by circumventing inhibitory endogenous signal can also lead to increased nitrogen response. This is historically

evident by high tiller number ideotypes bred through the *Rht-1* mutant alleles increasing NUE during the second green revolution (Conway, 2019).

We also observed the *indTaRKD1-7D* lines induced higher tiller number in very restricted soil volumes suggesting the *TaRKD1-7D* transcriptional cascade can override many of the negative feedback loops tempering tiller number. Environmental insensitivity has been exploited by pathogens in cereal crops, with the fungus *Gibberella fuikuroi* secreting GA inducing spindled growth and to disable JA-media□exaploidy□roph resistance (Phiney and West, 1959; Navarro et al., 2008). This highlights the need for plants to actively modulate their shoot architecture through sensing the available soil volume and nutrients. The SL pathway is fundamental to sensing soil volume and has been suggested to be analogous to plant-plant communication or plant seeding density in field conditions (Wheeldon et al., 2021). Epigenetic factors have also been previously shown to regulate plant density and growth through the release of histone deacetylases inhibitors to suppress competitor growth (Venturelli et al., 2015). Classical tillering mutants like the *Rht1* mutant alleles have been shown to be subject to soil volume constraints with no increased tiller number being observed at lower plant spacings (Jobson et al., 2019). Ectopic expression of *TaRKD1-7D* has a unique ability to induce higher tiller number regardless of soil volume may also apply to lower plant spacing in future field trials.

One of the main hypotheses of the model suggests that *indTaRKD1-7D* decreases meristematic determinacy in both SAM and IM tissue to effect height and SNS respectively. We were unable to confirm whether increased plant height was caused by meristematic indeterminacy or increased stem elongation from delayed flowering

time as we hadn't recorded the number or length of the internodes during the trial. Other researchers continuing this study will investigate this in due course. However, for IM indeterminacy, many examples of axillary meristem factors effecting tiller and spike formation exist in the literature with *MOC* loss of function resulting in no AXM formation for tillering or spikes in rice and the *TaMOC1-7A HapH* overexpression haplotype leading to increased SNS in wheat (Li et al., 2003; Zhang et al., 2015). These previous findings give context to our hypothesis that ectopic *TaRKD1-7D* expression increases AXM formation and increases SNS. More complex relationships exist with ectopic expression of morphogenic regulators depending on their tissue context such as *ZmTB1* overexpression promoting lateral bud dormancy, but *TaTB1* overexpression leading to supernumerary spikelet formation in wheat, which is also like the homeotic transformations we've seen in internodal tiller branches in our *indOsRKD3* lines (Lewis et al., 2008; Dixon et al., 2018).

While we've phenotypically characterised spike phenotypes in *indTaRKD1-7D* extensively with automated image analysis, SEM imaging and CT scans we haven't conclusively proven that *indTaRKD1-7D* expression causes decreased inflorescence meristem determinacy. Future experiments could try to establish that increased SNS isn't a consequence of delayed flowering time as these two processes are intricately linked as discussed in Chapter 1. In contrast to *indTaRKD1-7D* induced higher tiller formation, no difference in SNS increase was observed across nitrogen conditions suggesting that this developmental process isn't affected as much in *indTaRKD1-7D* as tiller formation.

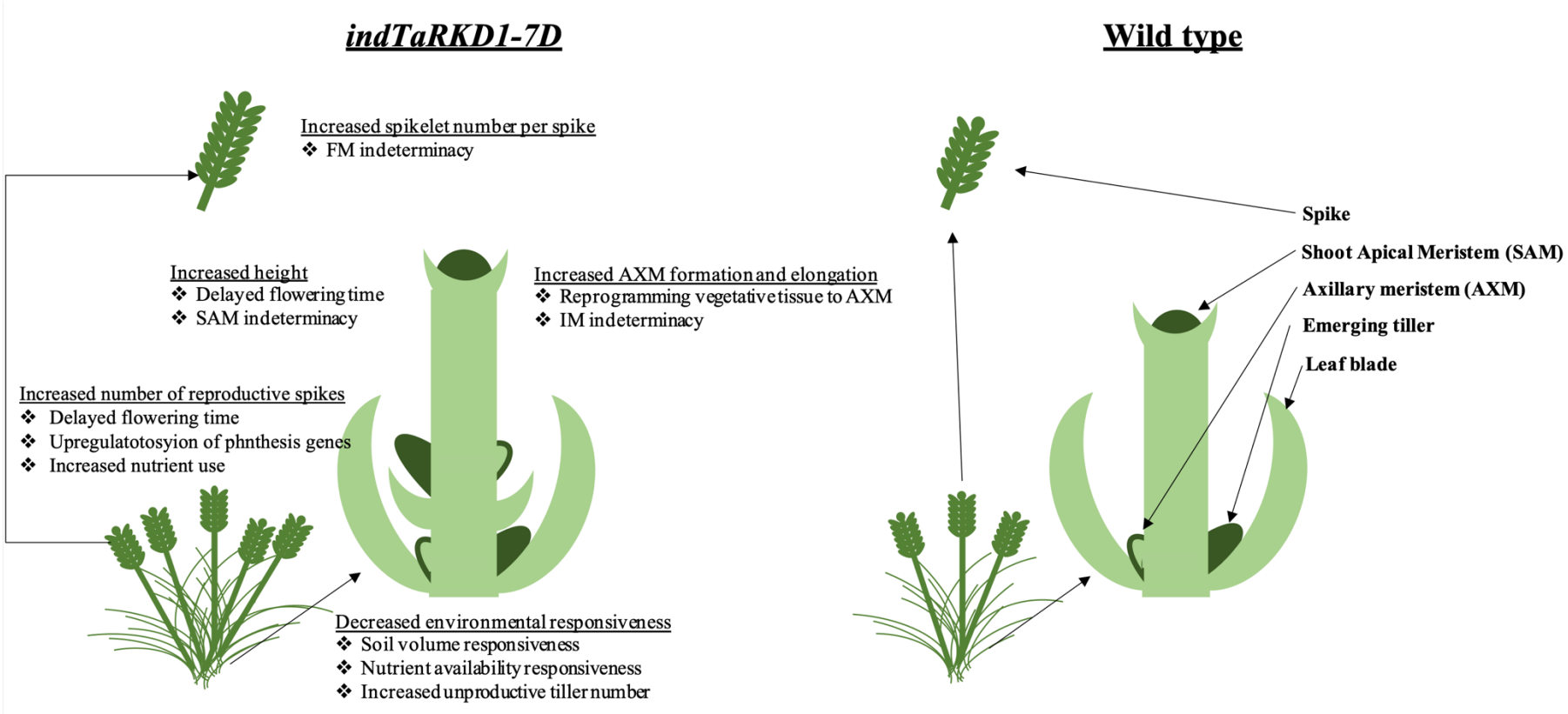


Figure 6.1. Model for TaRKD1 induced higher tiller and spikelet per spike number.

Through extensive bioinformatics analysis integrating DAPseq, ChIPseq and RNAseq we suggest that these processes occur through the binding and upregulation of genes related to cell cycle and meristematic determinacy. Tillering has long been known to be regulated by cell cycle factors in wheat. The *tiller inhibition* (*tin*) gene was mapped to the short arm of chromosome 1A after showing reduced tillering, thicker stems, and large spike and has been functionally characterised through map-based cloning of a cellulose synthase-like (Csl) protein at this locus [ref[. Early cessation of internode elongation of lateral culms in this genotype has been associated with hollow basal internodes due to reduced Suc content in dormant *tin* buds (Hyles et al., 2017).

Markers of DNA replication like histone H4 and cyclin-dependent kinase B were shown to be downregulated in dormant tiller buds in the tin mutant suggesting cell cycle regulators are key to tiller outgrowth (Kebrom et al., 2012). Moreover, this suggests that many endogenous signals can inhibit cell cycle genes necessary for AXM outgrowth and tiller formation. External conditions can also limit these processes with nitrogen deficiency inhibiting cell division-determined elongation, but not initiation, of rice tiller buds (Luo et al., 2017). This shows the potential of how ectopic expression of *TaRKD1-7D* can increase AXM initiation but is still subject to endogenous signal for tiller outgrowth. It should be noted however, that high ectopic expression does result in a significant increase in unproductive tillers and so is at least partially independent of this endogenously tempered outgrowth.

Meristem development is tightly controlled by multiple pathways that tweak cell proliferation and the timing of termination of this pluripotent stem cell niche. In maize, loss of function mutations in *RAMAOSA3* (*RA3*), which encodes a trehalose 6-

phosphate phosphatase, results in decreased meristematic determinacy leading to increased tassel branching and ectopic branching in ears (Satoh-Nagasawa et al., 2006). Additionally, trehalose 6-phosphate has been shown to trigger AXM bud outgrowth in garden pea (*Pisum sativum L.*) (Fichtner et al., 2017). However, more recently no correlation has been found between trehalose 6-phosphate phosphatases (TPPs) enzymatic activity and meristematic determinacy but do show nuclear localisation suggesting a moonlighting function for TPPs. In conclusion, trehalose 6-phosphate and genes that regulate its abundance have been shown to be involved in meristematic development, but their exact mechanism is more complicated than initially thought. In our wheat bioinformatics analysis, we see upregulation of trehalose-6-phosphate synthase (TPPS) genes with ChIPseq and DAPseq peaks in their promoter regions suggesting they play a role in *TaRKD1* induced axillary meristem formation. This is only a small subset of the TPPSs (~25) within the wheat genome suggesting that others do not have a TaRKD1-7D binding motif or can be expressed in the tissues we've analysed.

Our bioinformatics analysis showed upregulation of many cell cycle genes with a DAPseq peak within 10kb of their TSS and as nodes with the highest in and outdegree in our co-expression networks suggesting this is one of the main pathways in *indTaRKD1-7D* induced higher tiller number. Additionally, we found significant enrichment of terms relating to histone methylation and histone H4 which have both been shown to be play key roles in cell division and dedifferentiation of cells into more pluripotent tissues (Fobert et al., 1994; Nic-can et al., 2013). However, much of the bioinformatics analysis has relied heavily on GO term analysis which has its limitations (Khatri and Drăghici, 2005). While model organisms like arabidopsis and

rice have GO terms and verified biological functions associated with most genes, plants with large complex genomes such as wheat don't have the same level of annotation. This greatly narrows the scope for discovery in wheat, with some specific GO terms like 'somatic embryogenesis' having 3 genes associated with it. Additionally, in many cases throughout this thesis we see non-specific parent terms over-represented that are difficult to interpret such as 'catalytic activity'. While other less annotation heavy strategies exist such as searching for overrepresentation of protein domains using Pfam (although it should be noted GO terms are also based on this approach) the issue of annotation and characterisation of genes in wheat compared will remain until better systematic annotations for wheat are developed (Mistry et al., 2021).

Beyond the use of GO terms to identify biological processes and genes implicated in the *TaRKD1-7D* transcriptional cascade we employed a few strategies to narrow down lists of candidates such as overlapping different NGS techniques to get around ~150 genes with a DAPseq binding site and RNAseq differential expression. Many of these genes were also involved in epigenetic and cell cycle related processes which suggests that these terms are better annotated than others. Overlapping of orthologs between our *indTaRKD1-7D* and *indOsRKD3* lines was most likely unsuccessful due to the poor quality of the rice RNAseq dataset and the evolutionary divergence of these species within the BEP clade (~55 MYA).

Zygotic RKDs has been mainly characterised in the literature as regulators of zygotic elongation and as inducers of somatic embryogenesis when ectopically expressed. Somatic embryogenesis was revolutionary for academic and industry researchers as it

allowed mutational analysis and higher yielding varieties to be developed to gain a greater understanding of plant development and agriculture. A wide plethora of somatic embryogenesis factors (SEFs) like *BABY BOOM* (*BBM*), *WUSHCEL* (*WUS*) and *RWP-RK DOMAIN-CONTAINING 4* (*RKD4*) were discovered for multiple plant species to increase transformation efficiencies but has ended with the open-source GRF4-GIF1 construct being shown to have high transformation efficiencies across multiple agronomically important plant species (Debernardi et al., 2020). Mutational analysis of SEFs has been shown to effect plant architecture with the *WUS* ortholog *TILLERS ABSENT 1* (*TAB1*) being implicated in AXM formation (Tanaka et al., 2015). Other researchers have shown the potential of SEFs to modulate inflorescence architecture in agronomically important crops by tuning the expression of one of those classically studied meristem formation and maintenance pathways, the WUSCHEL-CLAVATA feedback loop, using genome editing (Rodríguez-Leal et al., 2017; Liu et al., 2021). Ectopic SEF expression has been shown to be present in nature and maybe perhaps an untapped resource in breeding programs, particularly considering the number of SEF ectopic expression lines exist in many research institutes (Shimada et al., 2018). This will hopefully pave the way for other researchers to investigate whether these lines or others can be used to modulate plant architecture to understand more about plant development and improve yield by manipulating agronomically important traits.

Breeding for high yielding cereal varieties will require the introgression of available natural variation in wild landraces to ensure major staple crops can adapt to climate change and improve food production.□exaploidyloid wheat, the D genome derived from *A. tauschii* has been highlighted as a diversity bottleneck from bread wheat

domestication, with many wild accessions bearing natural variation unrepresented in modern elite cultivars, showing that even in a crop domesticated ~12,000 years ago the full potential of natural diversity remains untested (Wang et al., 2014). Recently the use of k-mer based GWAS has been used to capture the structural variation available in wild populations of *A. tauschii* to improve plant breeding (Gaurav et al., 2021). However, natural variation for ectopic SEF expression may be restricted as we've seen that *indTaRKD1-7D* plants are less receptive to environmental factors that may restrict tiller vigour like nutrient availability and soil volume and may not be as adaptive in wild populations as they are in agricultural varieties grown in nutrient rich conditions.

Following this, manipulating meristem formation through ectopic expression of SEFs may require the generation of synthetic variation through genome editing as opposed to transgene integration to increase cereal crop yields in the current agricultural policy environment. Allelic diversity in the promoter regions of morphogenic regulators generated by CRISPR/Cas has been shown to be capable of adjusting expression levels of the *CLAVATA3/Embryo Surrounding Region-Related* (*CLE*) genes which determine maize inflorescence architecture. Promoter bashing using CRISPR/Cas9 and multiple guide RNAs can decrease expression through deletion of regulatory regions or can increase expression as evidenced through the $CLE7^{CR\text{-}pro3}$ promoter inversion event generated (Liu et al., 2021). The development of new genome editing technologies such as prime editing has opened the possibility of high efficiency insertion events into promoter regions, perhaps allowing the introduction of motifs that induce ectopic expression of SEFs in a tissue specific manner (Anzalone et al., 2019).

Here, we've shown that ectopic expression of *TaRKD1* induces higher tiller number, increased SNS, and seed yield in rice and wheat. This is potentially a novel strategy for improving yield across a range of cultivars and cereal crop species. A model has been developed using extensive molecular data for how *TaRKD1* ectopic expression can be used to manipulate axillary meristem formation and potentially apical and inflorescence meristem determinacy. We're excited about future experiments that may be conducted based on these results to understand more about plant architecture and meristem development to improve crop yield to tackle the challenges of food security in the 21st century.

Appendices

Supplemental Information – Chapter 2

Table 2.1. Oligonucleotide sequences

Name	sequence
synTaRKD1-7D_attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGATGCAACAACAATAC
synTaRKD1-7D_without_stop_attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTGAAATCCAAAGAATCCCATTTG
synTaRKD1-7D_with_stop_attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTGCtAAAATCCAAAGAATCCCATTTG
pERV1-XVEU-FOR	CGGAATTCCCAGAGATGTTAGCTG
pERV1-T3A-REV	CGGTCATTAGAGGCCACGATTTGAC
synTaRKD1-7D_FP_1	ATGTCTCCCTATGGCAGCAG
synTaRKD1-7D_RP_4	ACATCTGTTCCTCAGGGCGTATC
TaGAPDH_FWD	TTAGACTTGCGAAGCCAGCA
TaGAPDH_REV	AAATGCCCTTGAGGTTTCCC
synTaRKD1-7D_RTF2	GTCAGAGGGCTCATGTCTCC
synTaRKD1-7D_RTR2	ACAGAGCGTCGTCAAGTCCT
synTaRKD1-7D_RTF3	GGCAGGCTTGTTTCAAAGAG
synTaRKD1-7D_RTR3	CATTTGTGGGAGAGGCATTT
TaRKD1-7D/AetRKD1-7D_RTF2	ACCGGAAGATGAAGAGCCTC
TaRKD1-7D/AetRKD1-7D_RTR2	CCGCTCTCCTCCTCTTGTAG
TaRKD1-7D/AetRKD1-7D_RTF4	AAGCCTCGATCACATAGGGT
AetRKD1-7D_RTR4	AGCAATGGTCGAGCAAGTTG
Hygromycin-Fwd	GAGCCTGACCTATTGCATCTCC
Hygromycin-Rev	GGCCTCCAGAAGAAGATGTTGG
BAR PB FWD	caggaaccgcaggagtggac
BAR PB REV	ggtgggtgtagagcgtggag

Synthetic DNAs

>synOsRKD3

ggacaagtttgtacaaaaaagcaggctcgagATGGAAATGCACGAGTGCTGCTACTACGGGGGTT
CCATCGGAGGCGACTGGCTTAATCCGCTCGCGGCAATACCGCCGCCGTGTTCAT
CATCCAGCTCATCGTGGTCTAGCCAGCTGCTGCTGCTCTCGGACCATGACGATGT
GTTGCTCCACAGCGCCGGCGATCACGGAGGGGCGGTGGCGGGAATTGGCGGTG
CCTGTATGACTGCTGATCTGGTGGTGAGGGACGAGGAGATGGAGATGGCCGCC
GGCTACCTCCCTGTCGCTGCGTCCGCAGCCGCGGCGGCGGATGTCGATCACTAC
ATGTACCAGCAGTTCCAGCTTGAGCCCGATCAGTTCGTGTCAACCTTGCCGGCT
GTTGCCGTTGCCGTGGCCGCGACAGCAGGGGGAGGATCACACGACGACGAACT
CCTCAGGATGCCATTCACCGACATCGATTTGGATGCATTTGCCGATGCCAGAGA
TGTGGTCGTCGGAGTTGGGGAGCCGAAGCCCTCTCCCCAGCATACCCTTGACGC
TGCAATCGCGCTTCCTGCAGTAGGCGGCGGCGGCGCTCATCATTTCGGAACGCA
AGACGATGACGTGAAGTTCGACGTAACCAAACAGCGGAACGATGCTGCGCTCG
CTGGCGATGATAGTCTGAGTATGGTGATCGTGGAGTCGTACGAGATGGGAATGA
GAAGACATGCGGCTGAGCAAGAGCAGGAACAGAAGCCAAAGATCATTACCTCA
GCAGCCACTACCCTCACGCCGCTCCCGCTGCCGCCTCCTCCTCCACCACCACCTA
GAGTGACGCGGTCAAGGAGAGATGGTAGCTCAGCAGCAACCGCCGGGGGAAAA
ACGCGCCTTGACCACATAGGATTCGAGGATCTTAGAAGATATTTTTACATGCCC

ATCACAAAGGCAGCGCGCGAGATGAATGTGGGGCTTACAGTCTTGAAGAAGAG
GTGCAGAGAACTGGGCGTTGCGCGGTGGCCACACCGCAAAATGAAGAGCTTGA
AAAGCTTGATCCTTAATGAAATGGGGAGCAAGGGTATGTCTGCTGCCGCGATGA
GGCGCGAGCTGGAAGCCCTTGAAAACTGCTGCGCACTGATGGAGAGGAACCCC
GCGGTGGAACTTACAGAACGGACTAAGAAGTTGCGCCAGGCCTGCTTCAAGGA
AAATTATAAGCGCAGGCGCGCGGCTGCTGTCGACGTGCTGGATCTCGACCACTG
CTTTTCTTTCGCGGCTGGTCACTGCCACAGACACCACCACCAACAGCTTGCGCTG
CCCCCACCACCTGCGGCGGCTGCTGATCACCGTCGTCGCGATTTTTTCGGCTACg
actagtacccagctttcttgtacaaagtggtgg

>synTaRKD1-7D

ggacaagtttgtacaaaaaagcaggctATGGAGATGCAACAACAATACTTCGGGGGGGACGGC
GATGCGGACTGGTTCCATCAACTCGCATTGCTTCCCCCACTTCCAATCTCATCGT
CTCTCCCCCCACTCCCGATGTCAGAGGGCTCATGTCTCCCTATGGCAGCAGCAG
CTGCAGCTGCACTCCCCCTTGGCGATTGCTCGAGCGCCCTCATGATACGCCCTGA
GGAACAGATGTCTTGCCTTCCAATGAACCCCTCTCCAGCGGTCGTCGACGATGT
CTACTCTTCCTACGCACCGAACAATGTCGACGTGTTGCCGCCATTCCCGGCAGG
ACTTGACGACGCTCTGTTGATGGAGTCTTTTTCTGACATCGACCTCGAGGAGTTT
GCTGACGCATTTGGCCACAAGATCAAGACAGAACCCCTCGACGATGCCATGGTC
CCCGCGGACCACGACTTCGCGGCTCAAGCCCAACAGGCCTGCCCTGTGGTCATC
ATGAATCAGCAACAACTCAACGCACCCAGAGACGTGCGCCTGCTCATTGACCCG
GATGATGATGACAGCACCGTGGTGGCCGGGGGCTATGAAGCTGCAGCGGTGGG
GTGCGCCGAGCAGAAACAGGTCAGGCCAGCACCACGTAGGGTGAGAAAGAGCT
CAGGCGGCGCAAGACCAGCCGCGGGAGGAAAGTCCCTCGATCACATCGGATTC
GAGGAACTCAGGACCTATTTCTATATGCCAATCACCAAGGCAGCGAGGGAAAT
GAACGTGGGGCTGACAGTCCTGAAGAAGAGATGCCGGGAACTGGGGGTGGCGC
GCTGGCCACACAGAAAGATGAAGTCTCTGAGAAGCCTGATCCTCAACATTCAGG
AGATGGGGAAGGGCGCAACATCTCCCGCAGCCGTGCAGGGGGAACTTGAAGCG
CTTGAGAGGTATTGCGCCATTATGGAGGAGAACCCGGCTATAGAGCTCACCGAG
CAAACGAAGAAGCTCAGGCAGGCTTGTTTCAAAGAGAATTATAAGCGGCGTAG
AGCCGCCGCTTCTGTTAATCTTCTCGATCACTGCTATAACGATCTGGCATCTCAT
GAGCAGCAAATGCCTCTCCCACAAATGGGATTCTTTGGATTTTaGacccagctttcttgtac
aaagtggtcc

2.7.2. Pre-processing statistics

All re-analysed dataset statistics can be found–- https://github.com/meehanca.

Below is a list of all the analysed dataset generated for this project.

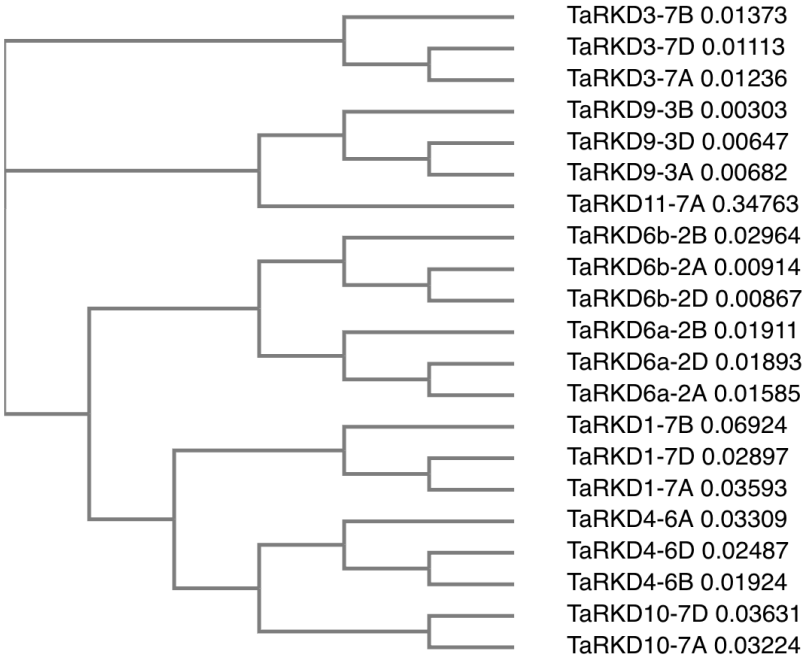
Table 2.2. Pre-processing statistics for next generation sequencing analysis

Library name	Species	Sample name	Condition	Raw reads	Trimmed Reads	Percentage trimmed	No of targets	No pseudoaligned	No uniquely pseudoaligned	Percentage pseudoaligned	Percentage uniquely pseudoaligned
YL100	Triticum aestivum	IE_TE1	Immature embryo indTaRKD1-7D + Estradiol	29618456	29098807	98.2	133747	21094024	14937518	72.5	51.3
YL101	Triticum aestivum	IE_TE2	Immature embryo indTaRKD1-7D + Estradiol	21117450	20892422	98.9	133747	15355144	10747040	73.5	51.4
YL103	Triticum aestivum	IE_TE3	Immature embryo indTaRKD1-7D + Estradiol	19721766	19512334	98.9	133747	14130936	9847983	72.4	50.5
YL104	Triticum aestivum	IE_TM1	Immature embryo indTaRKD1-7D + Mock	24166254	23926036	99	133747	17299381	11978589	72.3	50.1
YL105	Triticum aestivum	IE_TM2	Immature embryo indTaRKD1-7D + Mock	20240726	20039046	99	133747	14655877	10205710	73.1	50.9
YL107	Triticum aestivum	IE_TM3	Immature embryo indTaRKD1-7D + Mock	27225448	26950183	99	133747	19621777	13468018	72.8	50
YL117	Triticum aestivum	IE_WE1	Immature embryo Wild Type + Estradiol	22596492	22368139	99	133747	16189942	11208977	72.4	50.1
YL118	Triticum aestivum	IE_WE2	Immature embryo Wild Type + Estradiol	21791622	21562836	99	133747	15955804	11008093	74	51.1
YL120	Triticum aestivum	IE_WE3	Immature embryo Wild Type + Estradiol	21983694	21751789	98.9	133747	15736861	10757263	72.3	49.5
YC01	Triticum aestivum	TE1	Leaf indTaRKD1-7D + Estradiol	54483209	53997544	99.1	133747	42556019	28544977	78.8	52.9
YC02	Triticum aestivum	TE2	Leaf indTaRKD1-7D + Estradiol	54483209	51627003	94.8	133747	40047186	27240270	77.6	52.8
YC04	Triticum aestivum	TE3	Leaf indTaRKD1-7D + Estradiol	41481395	41167928	99.2	133747	31523688	20596588	76.6	50
YC05	Triticum aestivum	TM1	Leaf indTaRKD1-7D + Mock	47400481	46997017	99.1	133747	36679073	24622751	78	52.4
YC06	Triticum aestivum	TM2	Leaf indTaRKD1-7D + Mock	41041160	40601872	98.9	133747	30843716	20886387	76	51.4
YC08	Triticum aestivum	TM3	Leaf indTaRKD1-7D + Mock	48957321	48424443	98.9	133747	38632662	25419362	79.8	52.5
YC10	Triticum aestivum	WE1	Wild Type + Estradiol	47274400	46873411	99.2	133747	34721348	23544124	74.1	50.2
YC11	Triticum aestivum	WE2	Wild Type + Estradiol	45875311	45485856	99.2	133747	33260676	22476280	73.1	49.4
YC12	Triticum aestivum	WE3	Wild Type + Estradiol	43010373	42726335	99.3	133747	28243028	19331683	66.1	45.2

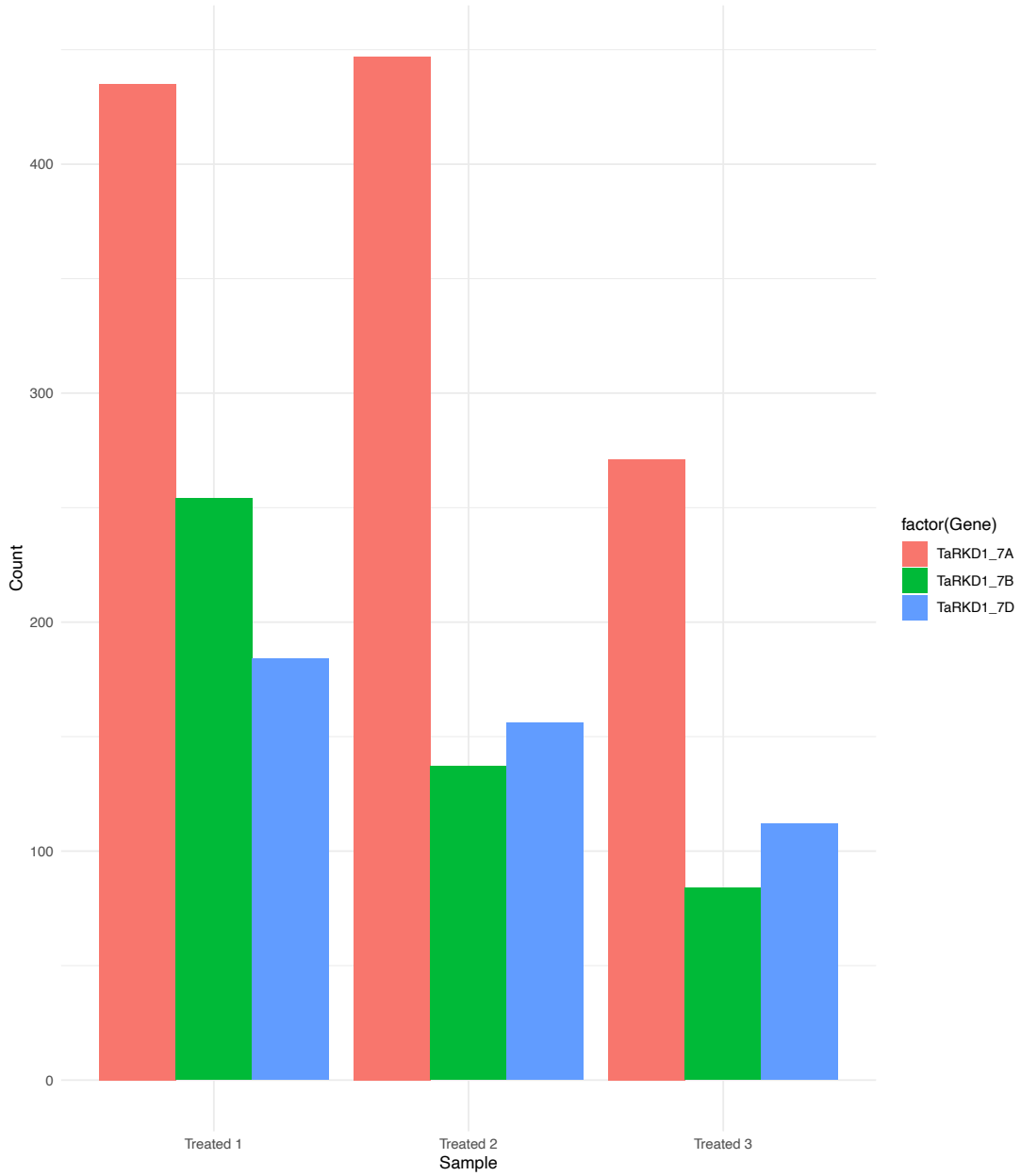
Library name	Species	Sample name	Condition	Raw reads	Trimmed reads	Trimming percentage	Reads aligned	Alignment percentage	Deduplicated aligned reads	Percentage of duplicates
TaRKD2_DAPseq	Triticum aestivum	DAPseq_TaRKD2	DAPseq of TaRKD1-7D bound to Fielder T. aestivum gDNA	29146311	26899130	92.3	22491462	83.6	17819948	20.8
HALO_DAPseq	Triticum aestivum	DAPseq_HALO	DAPseq of HALO bound to Fielder T. aestivum gDNA	34139092	31761475	93	21684756	68.3	18148926	16.3
Wh_Cut_RKD2_RKD2_1_S8	Triticum aestivum	ChIPseq_RKD1-7D_1	ChIPseq of TaRKD1-7D bound to hypocotyl T. aestivum chromatin	197775679	181219122	91.6	156794826	86.5	156794826	0
Wh_Cut_RKD2_Wt_S7	Triticum aestivum	ChIPseq_RKD1-7D_2	ChIPseq of TaRKD1-7D bound to hypocotyl T. aestivum chromatin	149546596	137109062	91.7	118753399	86.6	118753399	0

Library name	Species	Sample name	Condition	Raw reads	Trimmed reads	HISAT2 Aligned reads	HISAT2 Alignment percentage	HTseq counted reads	Htseq counted percentage
BR1	Oryza sativa (Cempo Ireng)	TM1	Leaf tissue of indOsRKD3 + Mock	17250149	17112831	15829369	92.5	11647430	73.6
BR2	Oryza sativa (Cempo Ireng)	TM2	Leaf tissue of indOsRKD3 + Mock	33116484	32866652	30927520	94.1	23967704	77.5
BR3	Oryza sativa (Cempo Ireng)	NM1	Leaf tissue of Wild Type + Mock	62605122	62108441	56767115	91.4	21271487	37.5
BR4	Oryza sativa (Cempo Ireng)	NM2	Leaf tissue of Wild Type + Mock	57820908	57357613	50876203	88.7	20115537	39.5
BR5	Oryza sativa (Cempo Ireng)	TD1	Leaf tissue of indOsRKD3 + DEX	28477195	28265171	25806101	91.3	15038636	58.3
BR6	Oryza sativa (Cempo Ireng)	TD2	Leaf tissue of indOsRKD3 + DEX	25278394	25092891	18066882	72	16057097	88.9
BR9	Oryza sativa (Cempo Ireng)	ND1	Leaf tissue of Wild Type + DEX	30808634	30567792	27266470	89.2	16992134	62.3
BR10	Oryza sativa (Cempo Ireng)	ND2	Leaf tissue of Wild Type + DEX	28993756	28710582	25983077	90.5	15603086	60.1

Supplemental Information – Chapter 3



Supplemental figure 3.1. Confirmation of *TaRKD1* triad verified through protein sequence phylogeny. Gene tree of alignment of amino acid sequences from the *TaRKD* genes using Clustal Omega. All sequences were taken from loci in Supplemental table 3.1. Values on the right indicate evolutionary distance to nearest branch node.



Supplementary figure 3.2. TaRKD1 homeologue triad shows high expression in dividing microspore tissue. DESeq2 normalised counts of the *TaRKD1* family genes *TaRKD1-7A* (*TraesCS7A02G135900*), *TaRKD1-7B* and *TaRKD1-7D* (Supplementary table 3.1.). Colour fill denotes homeologue as described in the legend on the right. Each group of three barplots represents a different sample from the RNAseq experiment.

Supplementary Table 3.1. List of TaRKD genes in IWGSC RefSeq v1.1. Assembly

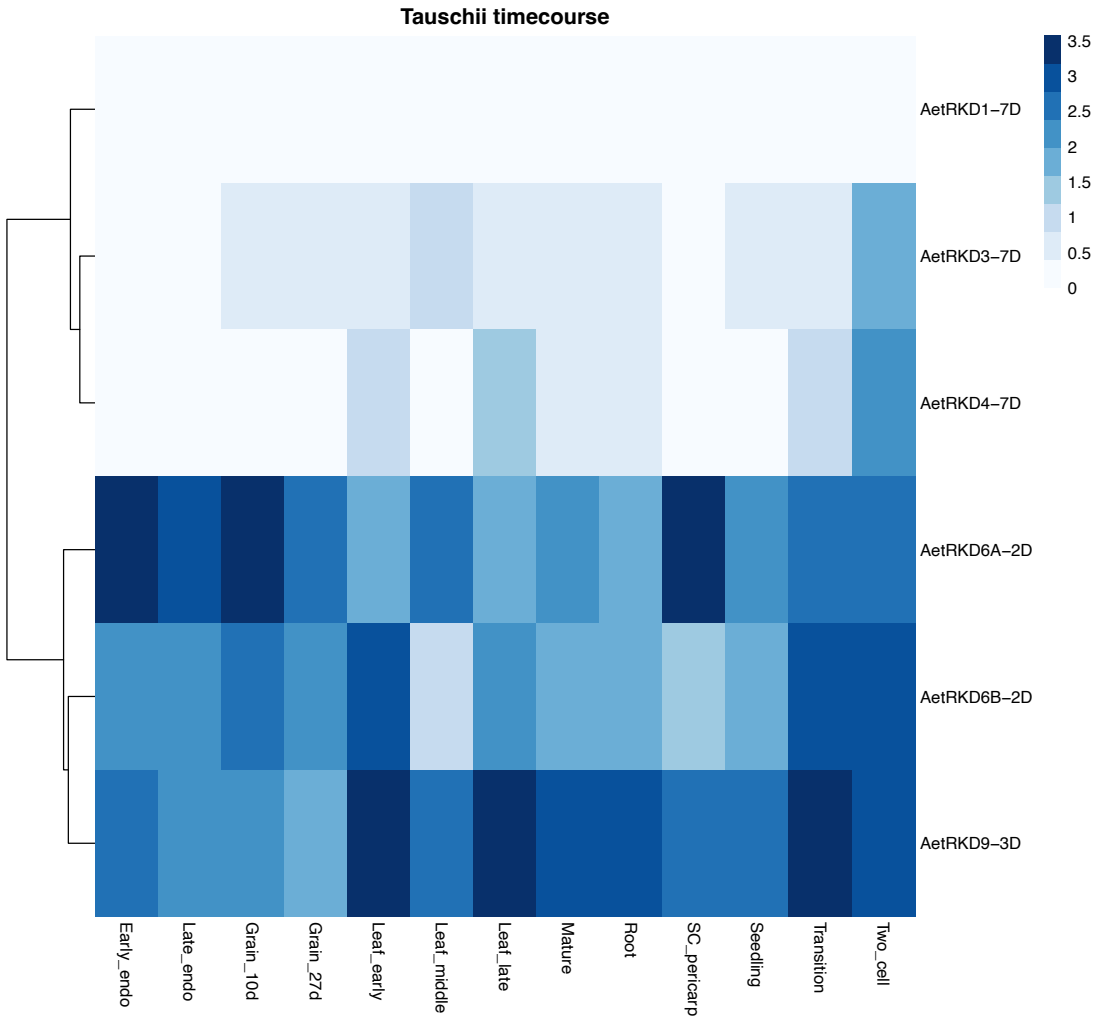
RefSeq v1.1 gene locus ID	RKD clade	Genomic co-ordinates
TraesCS7A02G135900	*TaRKD1-7A*	88824822-88827991
NA	*TaRKD1-7B*	35709553-35711044
NA	*TaRKD1-7D*	87251488-87252946
TraesCS7A02G188500	*TaRKD3-7A*	144,609,188-144,611,233
TraesCS7B02G093300.3	*TaRKD3-7B*	106945174-106949546
TraesCS7D02G189600	*TaRKD3-7D*	142,617,139-142,619,184
TraesCS6A02G307500	*TaRKD4-6A*	541595645-541596814
TraesCS6B02G336100	*TaRKD4-6B*	591264113-591265265
TraesCS6D02G286800	*TaRKD4-6D*	395390576-395391730
TraesCS2A02G404500	*TaRKD6a-2A*	659903623-659906083
TraesCS2B02G422400	*TaRKD6a-2B*	607548491-607552287
TraesCS2D02G401400.2	*TaRKD6a-2D*	516356306-516358927
TraesCS2A02G404400	*TaRKD6b-2A*	659890344-659895450
TraesCS2B02G422300	*TaRKD6b-2B*	607535903-607541394
TraesCS2D02G401300	*TaRKD6b-2D*	516326169-516331225
TraesCS3A02G167700	*TaRKD9-3A*	174283498-174288783
TraesCS3B02G198700	*TaRKD9-3B*	226085095-226090862
TraesCS3D02G173800	*TaRKD9-3D*	155555673-155560729
TraesCS7A02G188400	*TaRKD10-7A*	144606378-144607441
TraesCS7D02G189500	*TaRKD10-7D*	142612636-142613723
TraesCS7A02G497400	*TaRKD11-7A*	688153529-688155092

Supplementary table 3.2. List of AetRKD genes in INSDC Aet v4.0. Assembly

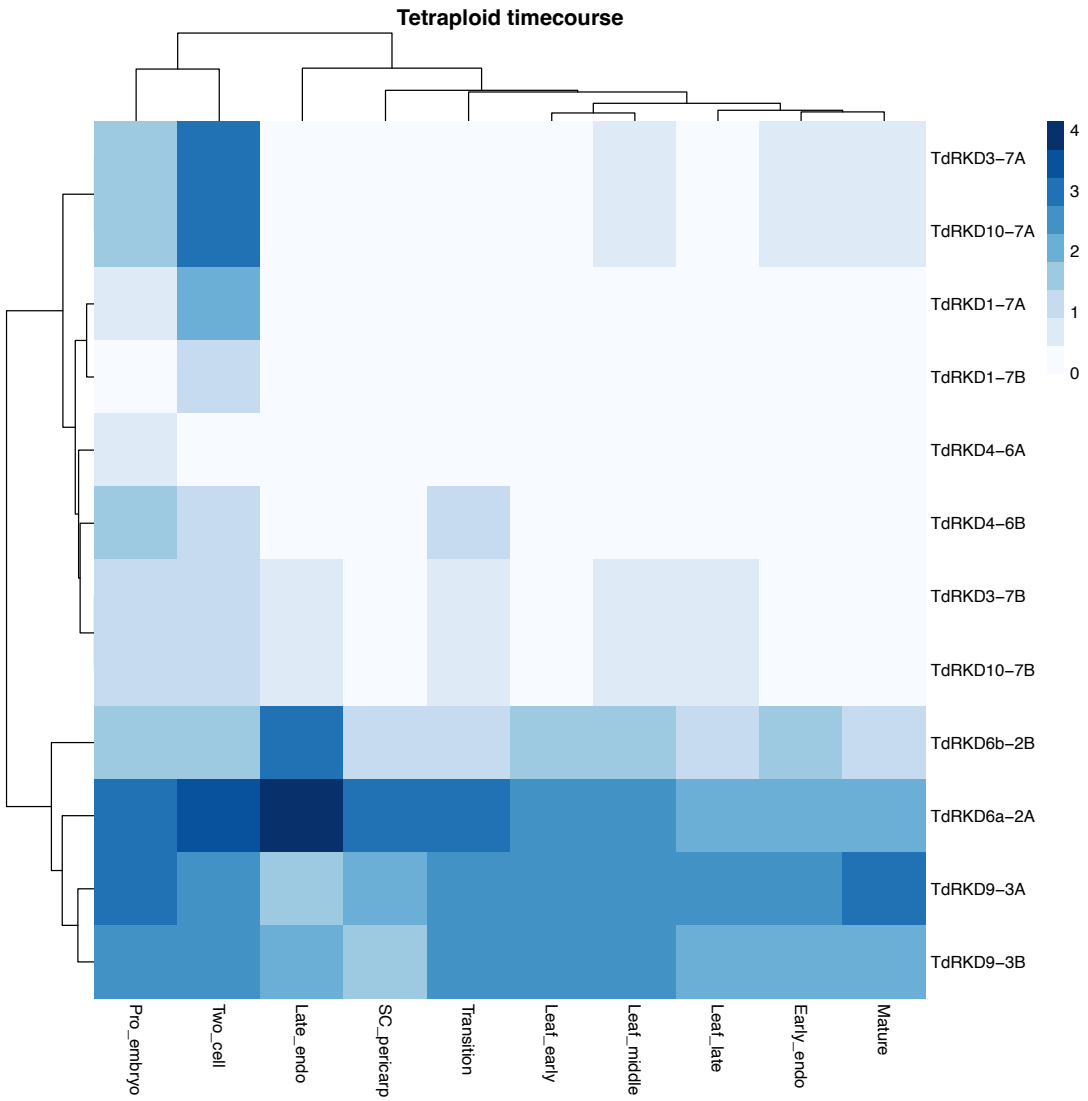
RefSeq v1.1 gene locus ID	RKD family and chromosome	Genomic co-ordinates
NA	*AetRKD1-7D*	88093618-88095096
AET7Gv20481900	*AetRKD3-7D*	144212500-144221715
AET6Gv20761000	*AetRKD4-6D*	395390576-395391730
AET2Gv20898300	*AtRKD6a-2D*	516356035-516359296
AET2Gv20898100	*AtRKD6b-2D*	516326100-516331430
AET3Gv20387300	*AtRKD9-3D*	158447705-158453266
AET7Gv20481900	*AetRKD10-7D* (*AetRKD3-7D)*	144212500-144221715

Supplementary table 3.3. List of TdRKD genes in INSDC WEWSeq v1.0. Assembly

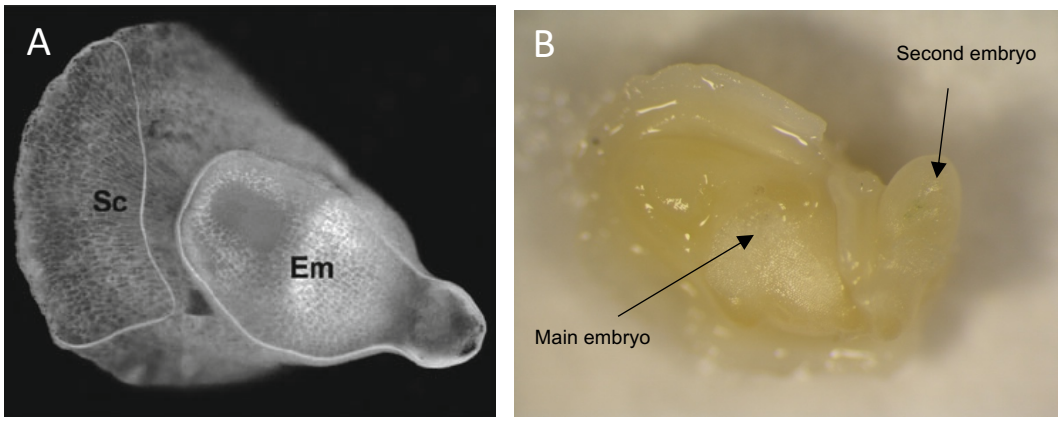
RefSeq v1.1 gene locus ID	RKD clade	Genomic co-ordinates
NA	*TdRKD1-7A*	86268325-86269661
NA	*TdRKD1-7B*	39204367-39204657
TRIDC7AG023400	*TdRKD3-7A*	142205949-142211000
TRIDC7BG013860	*TdRKD3-7B*	111017208-111021569
NA	*TdRKD4-6A*	541589639-541591143
NA	*TdRKD4-6B*	575554679-575555831
TRIDC2AG058460	*TdRKD6a-2A*	653,821,196-653,825,771
TRIDC2BG061420	*TdRKD6a-2B*	605,083,244-605,086,077
TRIDC2BG061400	*TdRKD6b-2B*	605,069,667-605,074,982
TRIDC3AG022940	*TdRKD9-3A*	659903623-659906083
TRIDC3BG027100	*TdRKD9-3B*	213,627,747-213,638,915
TRIDC7AG023400	*TdRKD10-7A*	142,205,949-142,211,000
TRIDC7BG013860	*TdRKD10-7B*	111,017,208-111,021,569



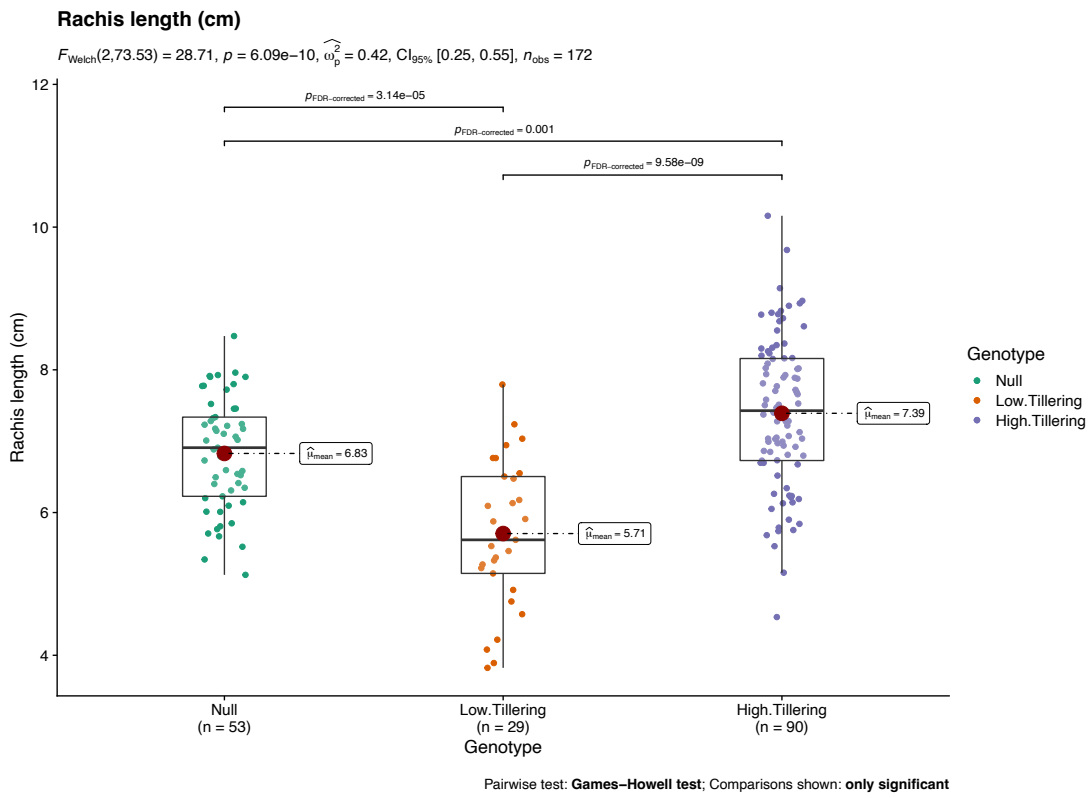
Supplementary figure 3.3. Heatmap of *AetRKD* expression from publicly available transcriptome data across tissues. Heatmap of re-analysed publicly available developmental timecourse transcriptome samples shows two distinct expression profiles of *AetRKD* genes from Supplementary table 3.2: a global vegetative expression profile (bottom) and a restricted embryonic expression profile (top). Scale displayed is logarithmically transformed transcript per million values from Kallisto quant. Hierarchical clustering was performed on *TaRKD* genes and is displayed as a dendrogram on the side of the heatmap.



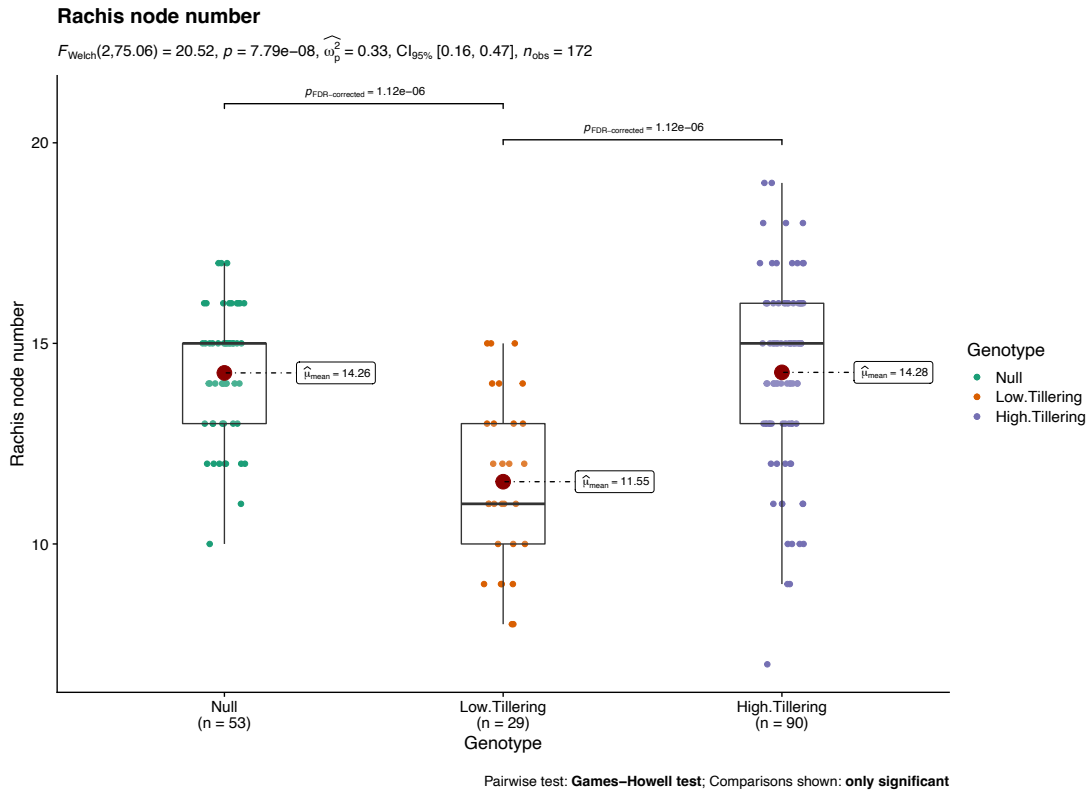
Supplementary figure 3.4. Heatmap of *TdRKD* expression from publicly available transcriptome data across tissues. Heatmap of re-analysed publicly available developmental timecourse transcriptome samples shows two distinct expression profiles of *TdRKD* genes from Supplementary table 3.3: a global vegetative expression profile (bottom) and a restricted embryonic expression profile (top). Scale displayed is logarithmically transformed transcript per million values from Kallisto quant. Hierarchical clustering was performed on *TaRKD* genes and is displayed as a dendrogram on the side of the heatmap.



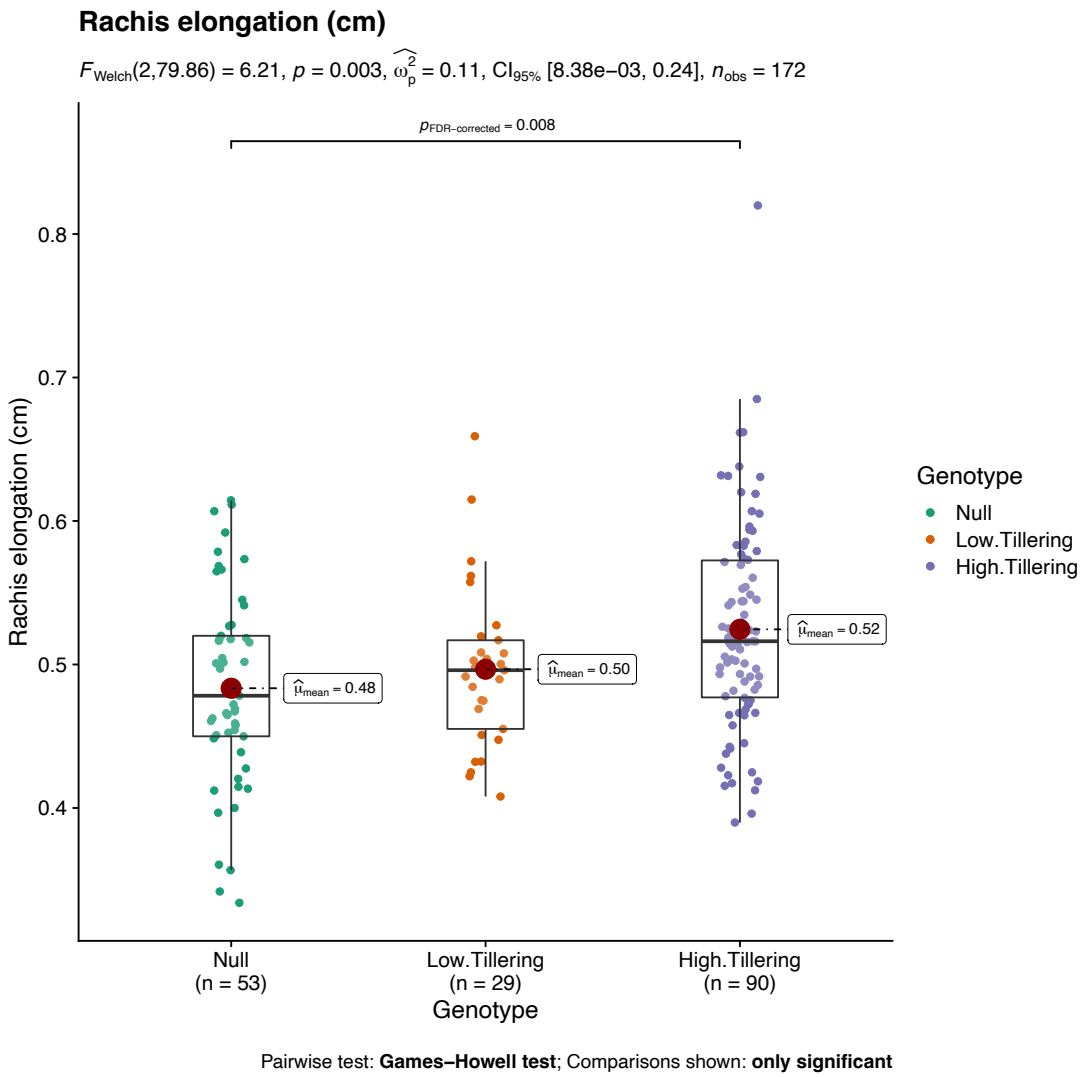
Supplementary figure 3.5. Polyembryonic structures in *indTaRKD1-7D* #2.28. Immature embryos were phenotyped under a light microspore for aberrant embryo development in *indTaRKD1-7D* #2.28, a line which exhibited high immature embryo expression of *TaRKD1-7D* in figure 3.4. **(A)** Diagram detailing where the embryo and scutellum are in an immature wheat embryo excised from the caryopsis 12-18 days after anthesis (Gils et al., 2012). (**B)** Polyembryonic wheat formed from 2% (4/200) of phenotyped embryos in *indTaRKD1-7D* #2.28.



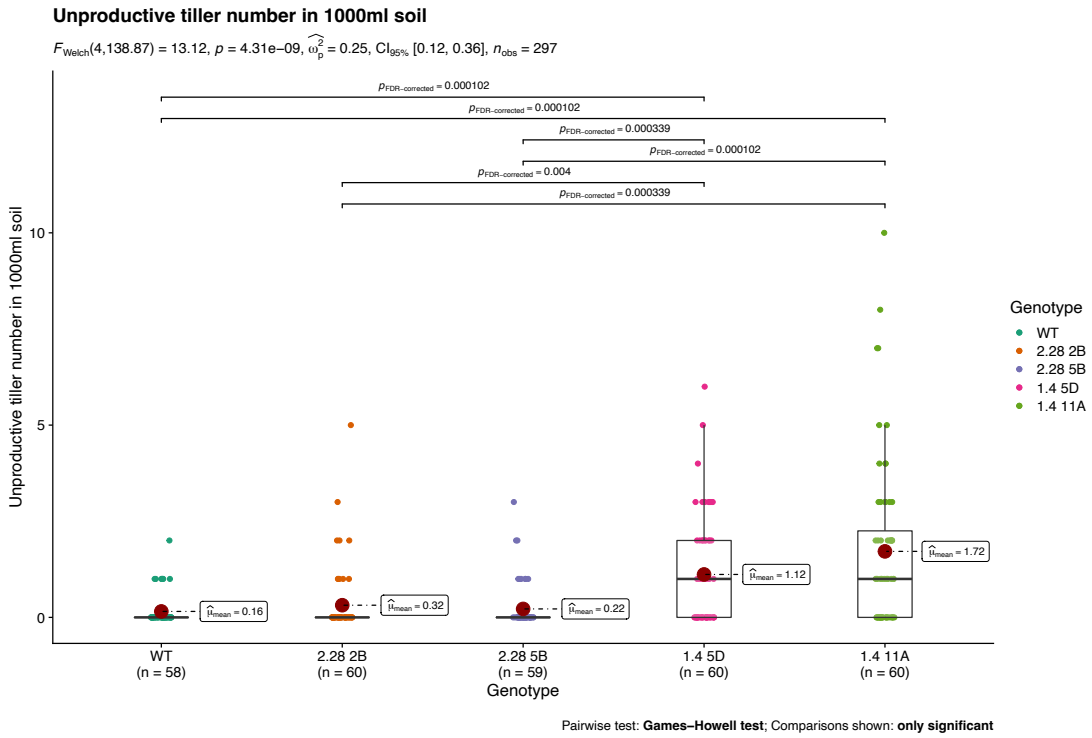
Supplementary Figure 3.6. Increased rachis length in high tiller number *indTaRKD1-7D* T3 lines. Fully dried spikes from *indTaRKD1-7D* T3 lines that showed high and low tiller number had their grain removed and the rachis was scanned, and length was measured in imageJ. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



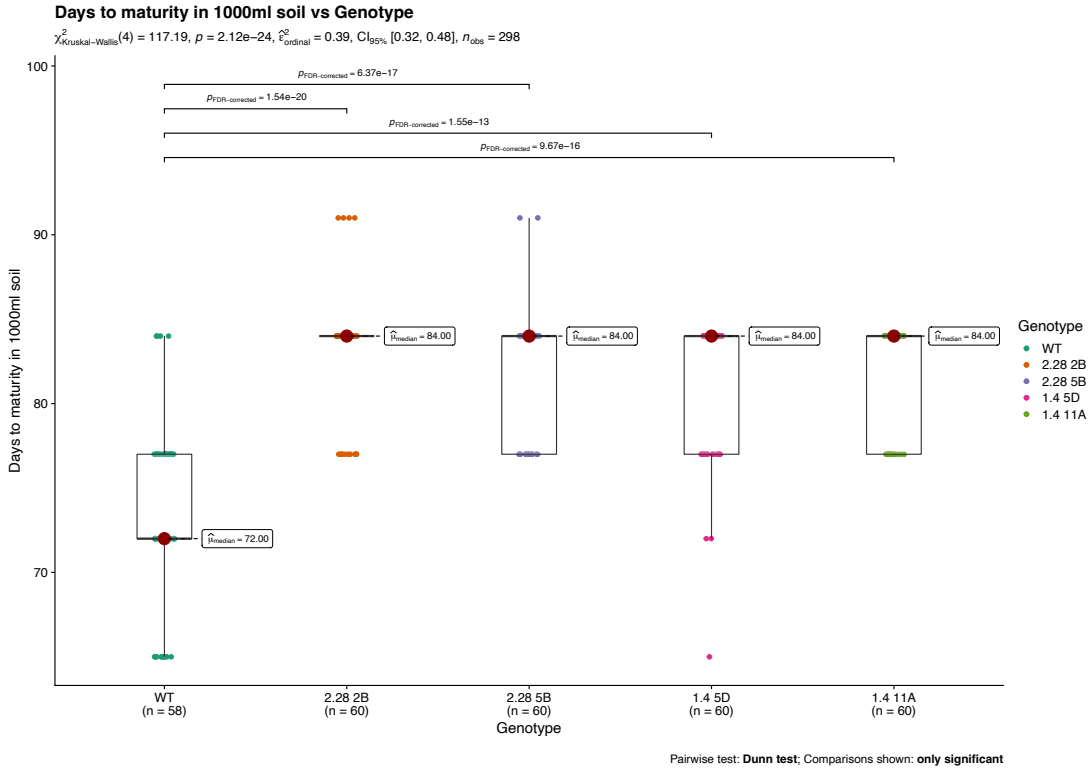
Supplementary Figure 3.7. Increased rachis node number in high tiller number *indTaRKD1-7D* T3 lines. Fully dried spikes from *indTaRKD1-7D* T3 lines that showed high and low tiller number had their grain removed and the rachis was scanned, and node number was measured in imageJ. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



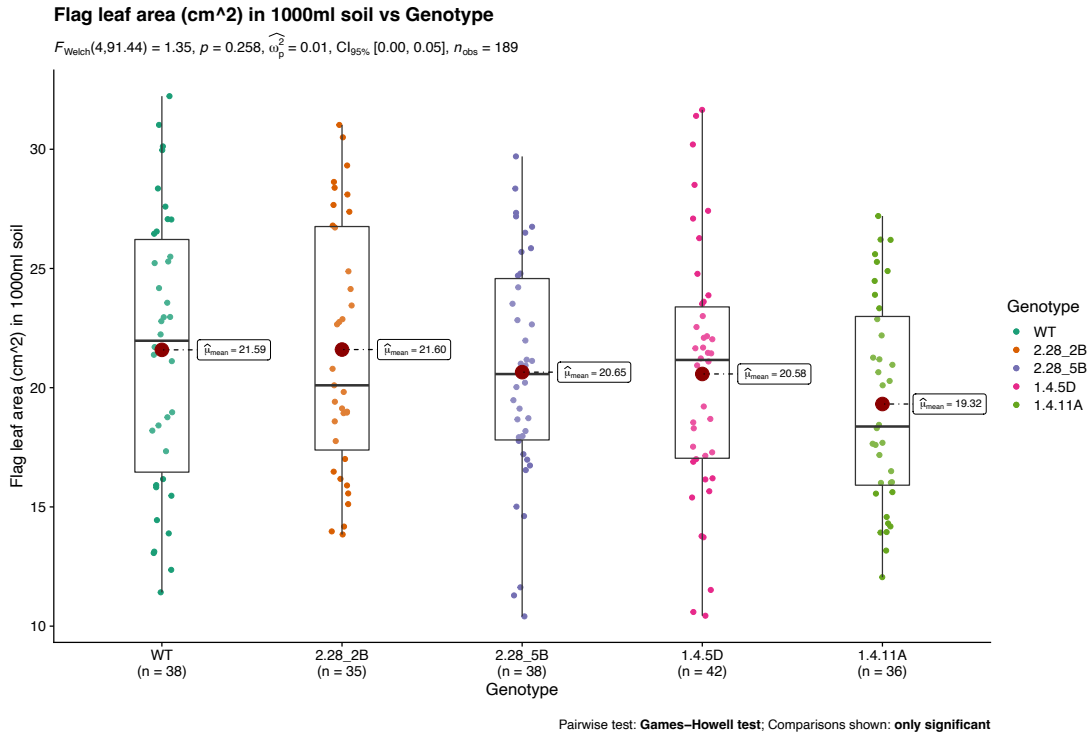
Supplementary Figure 3.8. Increased rachis elongation in *indTaRKD1-7D* T3 lines. Fully dried spikes from *indTaRKD1-7D* T3 lines that showed high and low tiller number had their grain removed and the rachis was scanned, elongation was measured by dividing the rachis length by the rachis node number (Supplementary figures 3.14 and 3.15). Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



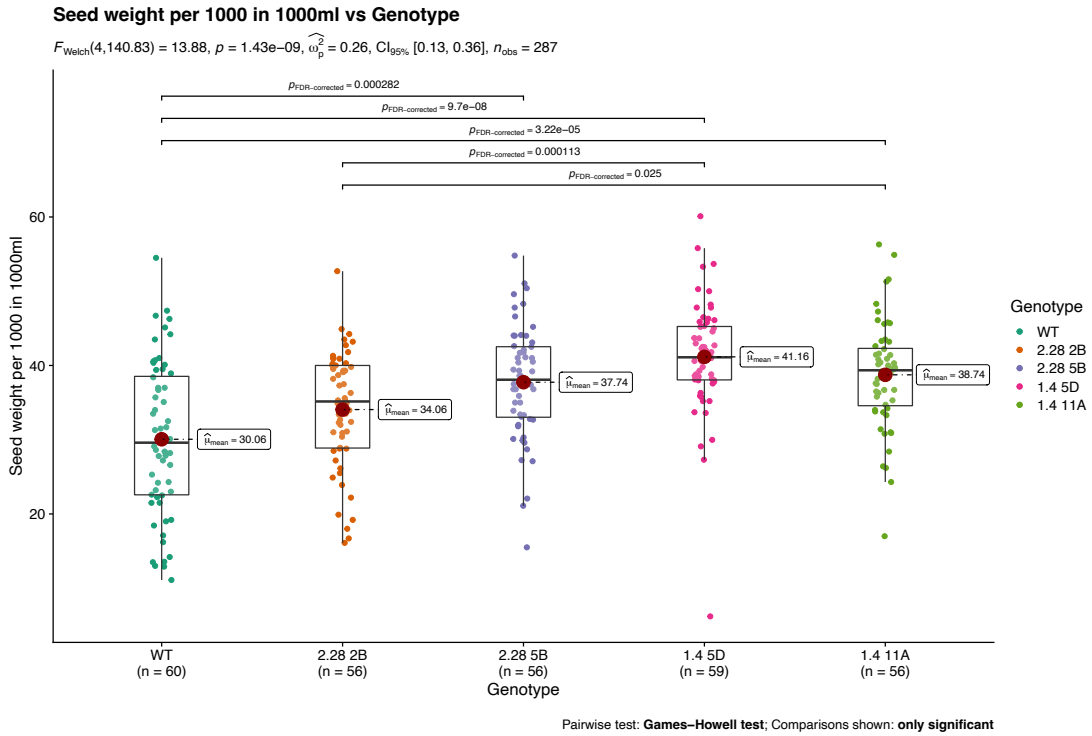
Supplementary figure 3.9. Increased unproductive tiller number of two *indTaRKD1-7D* T3 lines. Unproductive tiller number from plants measured from counting the number of vegetative and reproductive tillers and subtracting the number of fully developed reproductive tillers at senescence. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



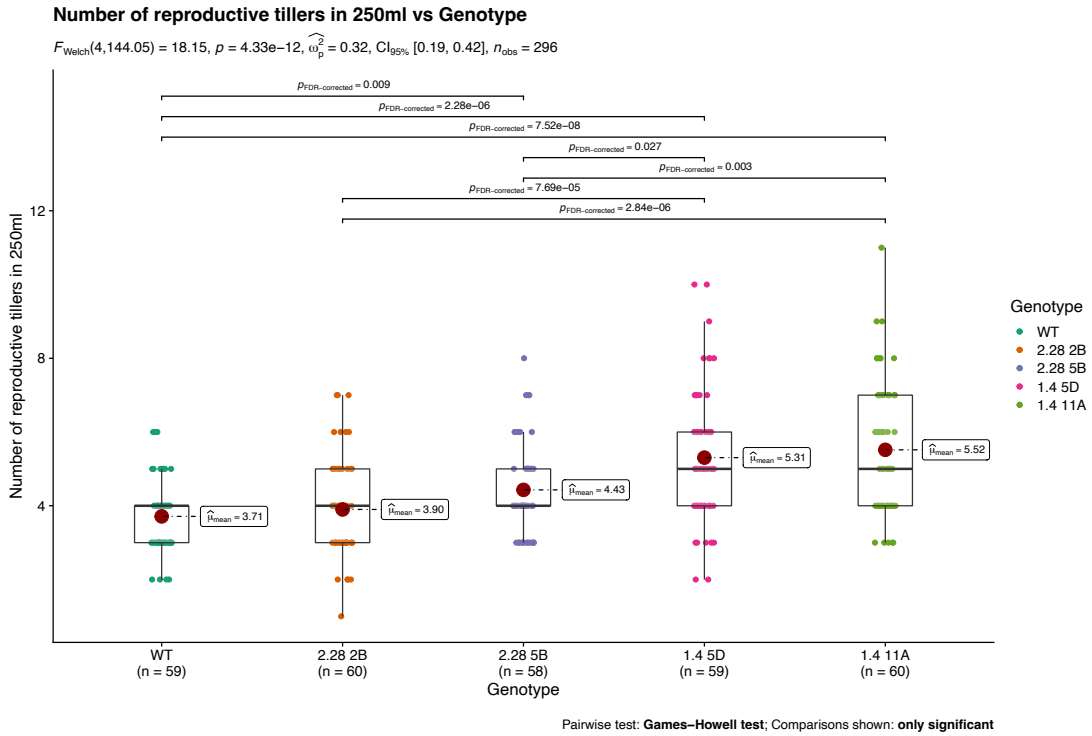
Supplementary figure 3.10. Delayed flowering time of two *indTaRKD1-7D* T3 lines. Days to maturity was measured when half of the spikes of the plant had senesced. The label of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



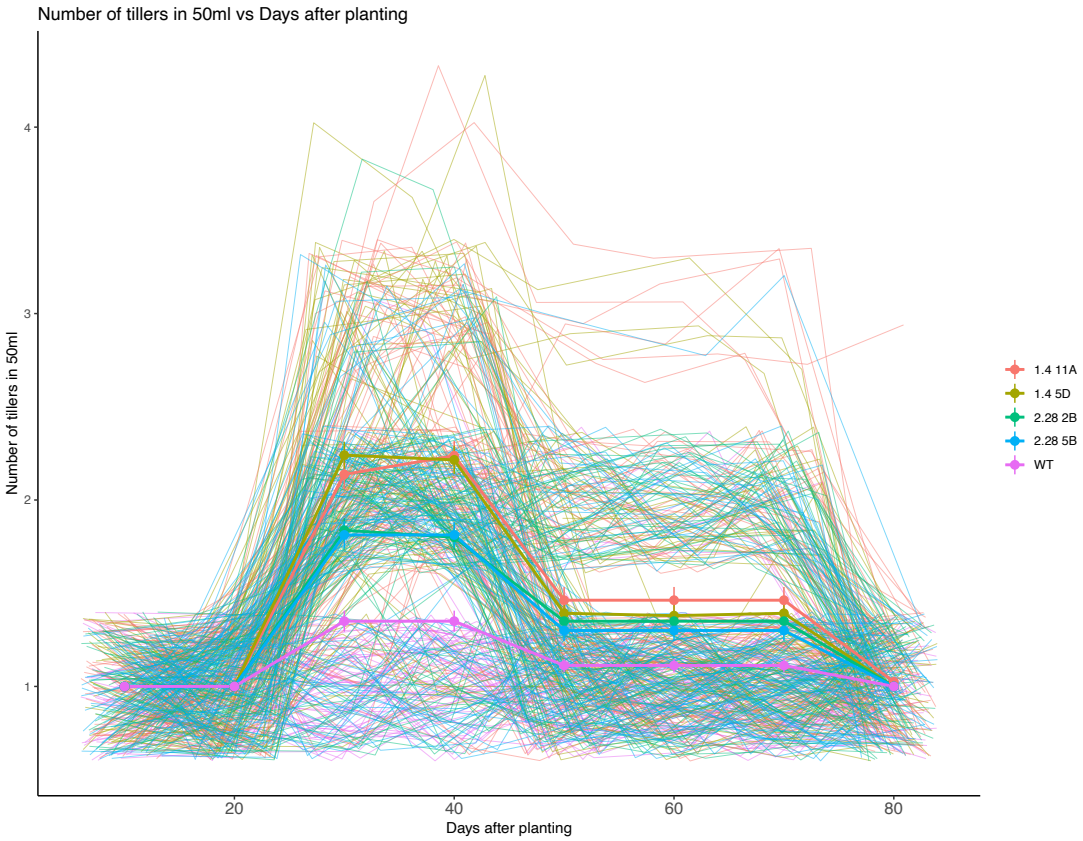
Supplementary figure 3.11. No difference in flag leaf area of two *indTaRKD1-7D* T3 lines. Flag leaves were taken once plants had begun senescence and were chosen at random. Flag leaves were scanned and measured in ImageJ. The label of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Additional statistics can be found on the top left of the plot.



Supplementary figure 3.12. Increased seed weight per 1000 of two *indTaRKD1-7D* T3 lines. The seed weight of 1000 seeds were measured after threshing fully dried spikes, where 1000 seed was not available this value was extrapolated. The label of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



Supplementary Figure 3.13. Increased reproductive tiller number of two *indTaRKD1-7D* T3 lines in lower soil volume. Reproductive tiller number from plants measured from counting the number of fully developed spikes at senescence. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



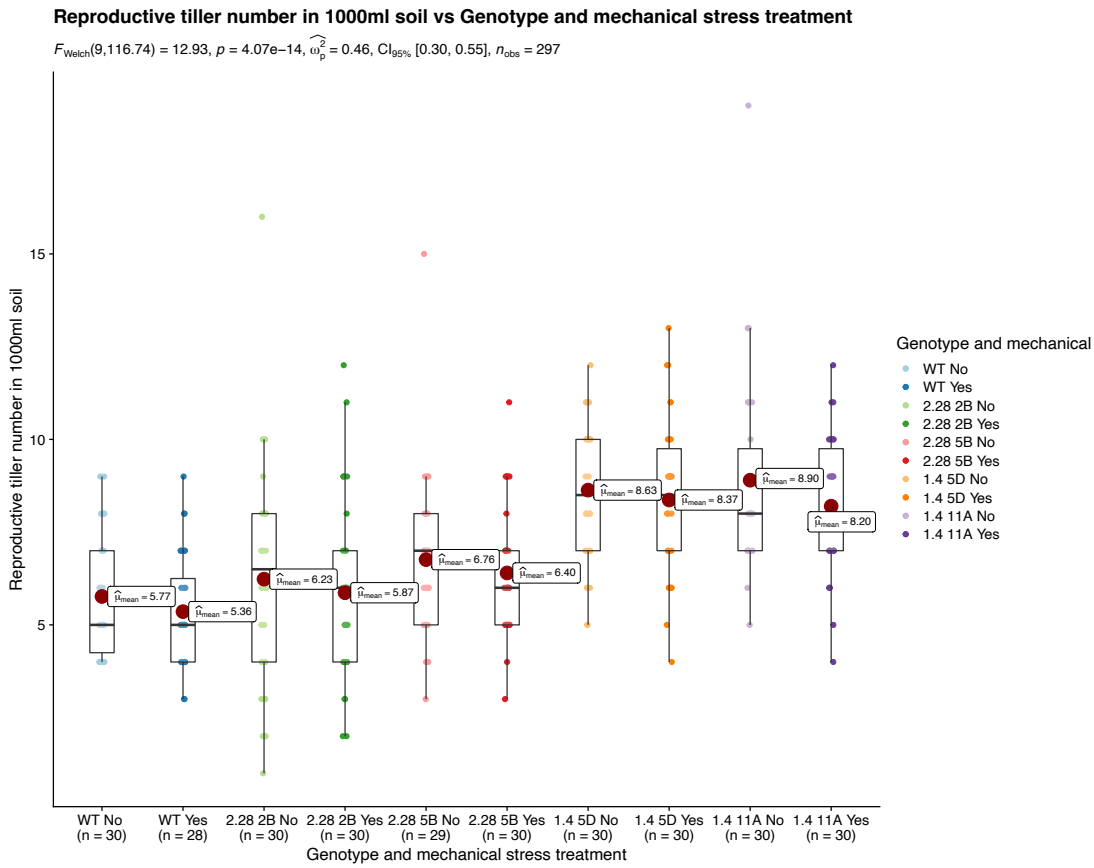
Supplementary figure 3.14. Increased tiller number independent of soil volume of two *indTaRKD1-7D* T3 lines. Vegetative tiller number was counted once every 10 days of plants grown in a restrictive 50ml soil volume condition. Line plots of all samples are plotted and coloured according to their genotype as indicated on the legend on the right. A mean line of regression is plotted in bold and coloured according to genotype.

Supplementary Table 3.4. Fold changes in phenotypic traits with mechanical stress treatment in tiller traits tirals compared to WT.

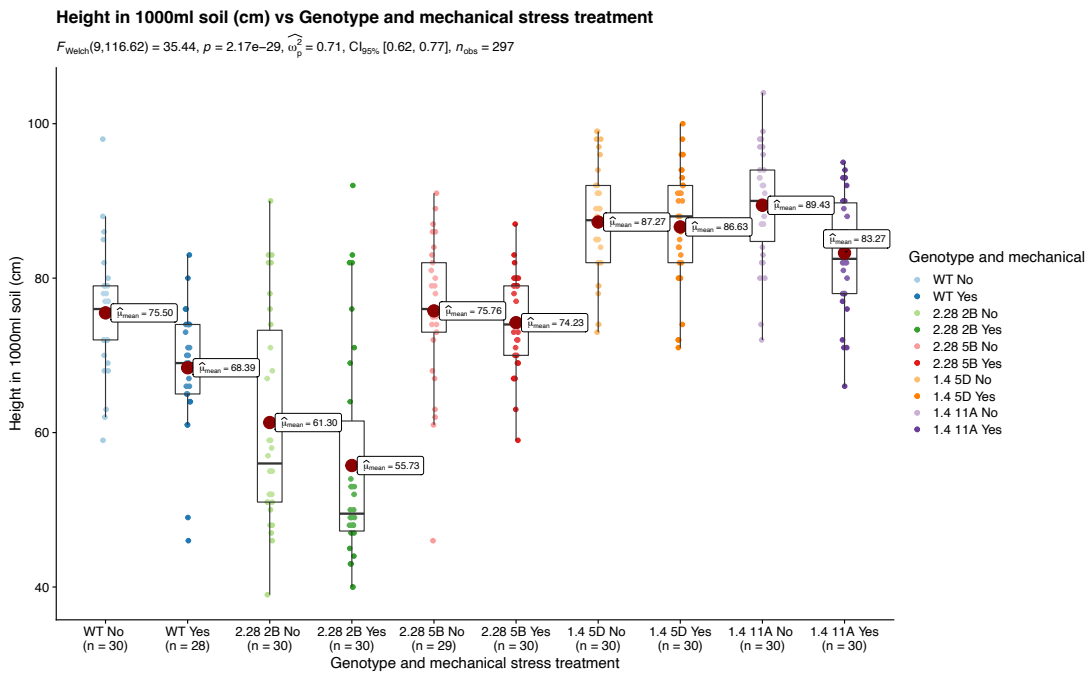
Traits	1.4 5D	1.4 11A	2.28 2B	2.28 5B	WT
Reproductive tiller number	0.92134831	0.96987254	0.94674556	0.94221509	0.92894281
Height	0.99266644	0.93111931	0.9091354	0.97980465	0.90450331

Supplementary Table 3.5. P values in phenotypic traits with mechanical stress treatment in tiller traits tirals compared to WT.

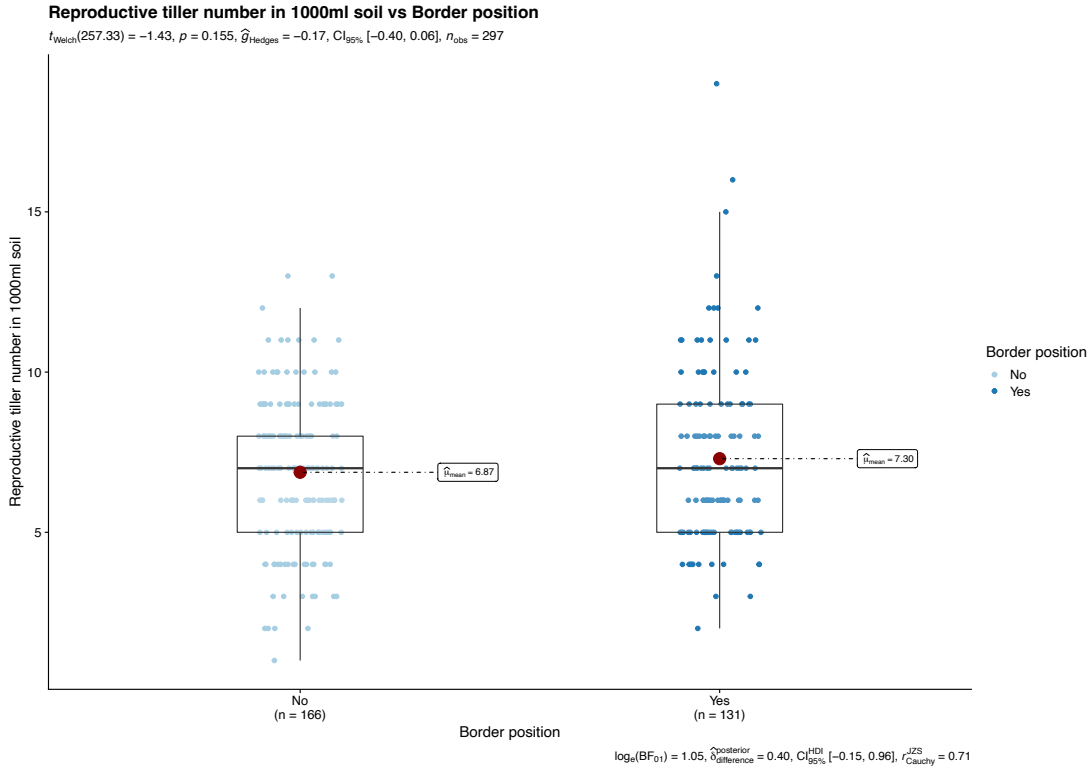
Traits	1.4 5D	1.4 11A	2.28 2B	2.28 5B	WT
Reproductive tiller number	n.s.	n.s.	n.s.	n.s.	n.s.
Height	n.s.	n.s.	n.s.	n.s.	n.s.



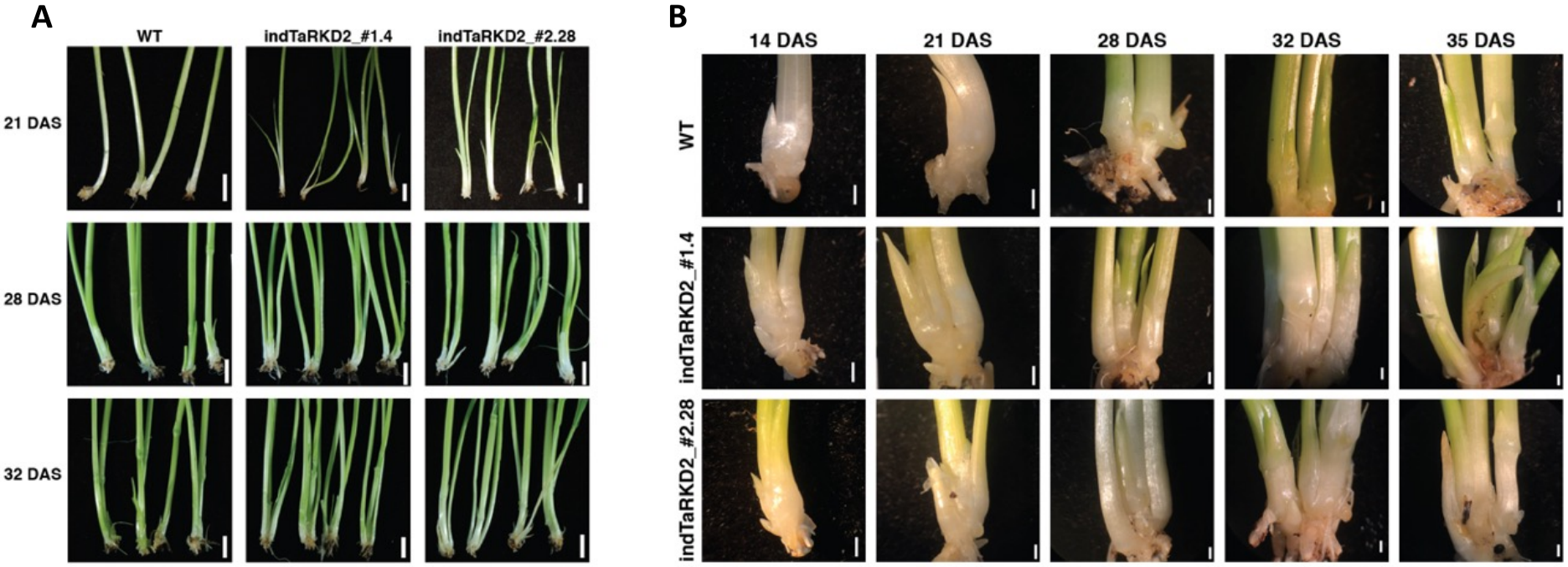
Supplementary Figure 3.15. No change in reproductive tiller number of two *indTaRKD1-7D* T3 lines or wild type with mechanical stress treatment. Reproductive tiller number from plants measured from counting the number of fully developed spikes at senescence. Mechanical stress treatment is indicated by 'Yes' or 'No' on label below boxplots. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Additional statistics can be found on the top left of the plot.



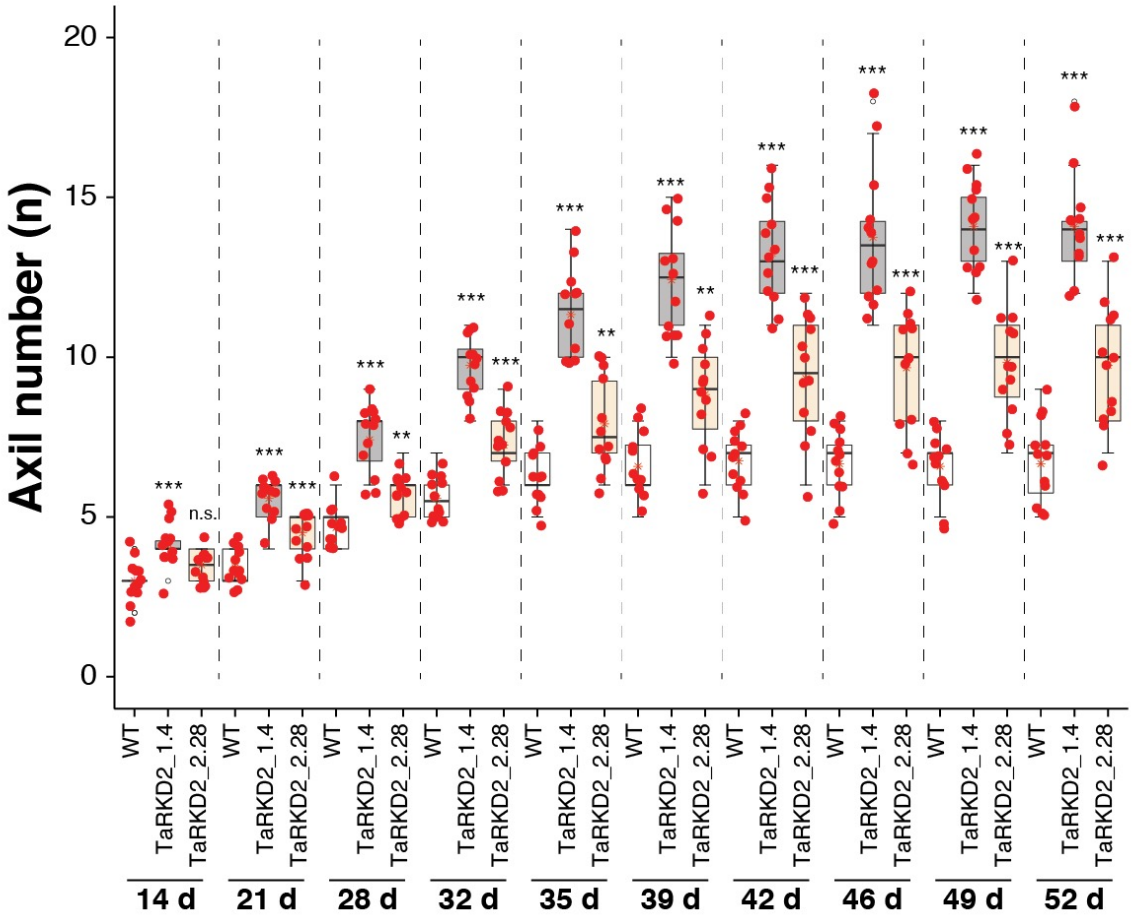
Supplementary Figure 3.16. No change in height of two *indTaRKD1-7D* T3 lines or wild type with mechanical stress treatment. Height measured from the base of the plant to the tallest spike at senescence. Sample number of each transgenic line is described in the (n=x) brackets below the label. Mechanical stress treatment is indicated by 'Yes' or 'No' on label below boxplots. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Additional statistics can be found on the top left of the plot.



Supplementary figure 3.17. Increased reproductive tiller number of two *indTaRKD1-7D* T3 lines at border positions of the trial. Reproductive tiller number from plants measured from counting the number of fully developed spikes at senescence. Border position is given by the 'Yes' or 'No' label below the boxplots. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.



Supplementary Figure 3.18. Higher axillary meristem formation in two *indTaRKD1-7D* T3 lines timecourse images. All instances labelled 'TaRKD2' are *indTaRKD1-7D*. **(A)** Images of full plants taken at timepoints in the independent confirmation of increased axillary meristem formation trial. **(B)** Images of plants with outer epidermal layer stripped away to highlight axillary meristem formation taken at timepoints in the independent confirmation of increased axillary meristem formation trial.



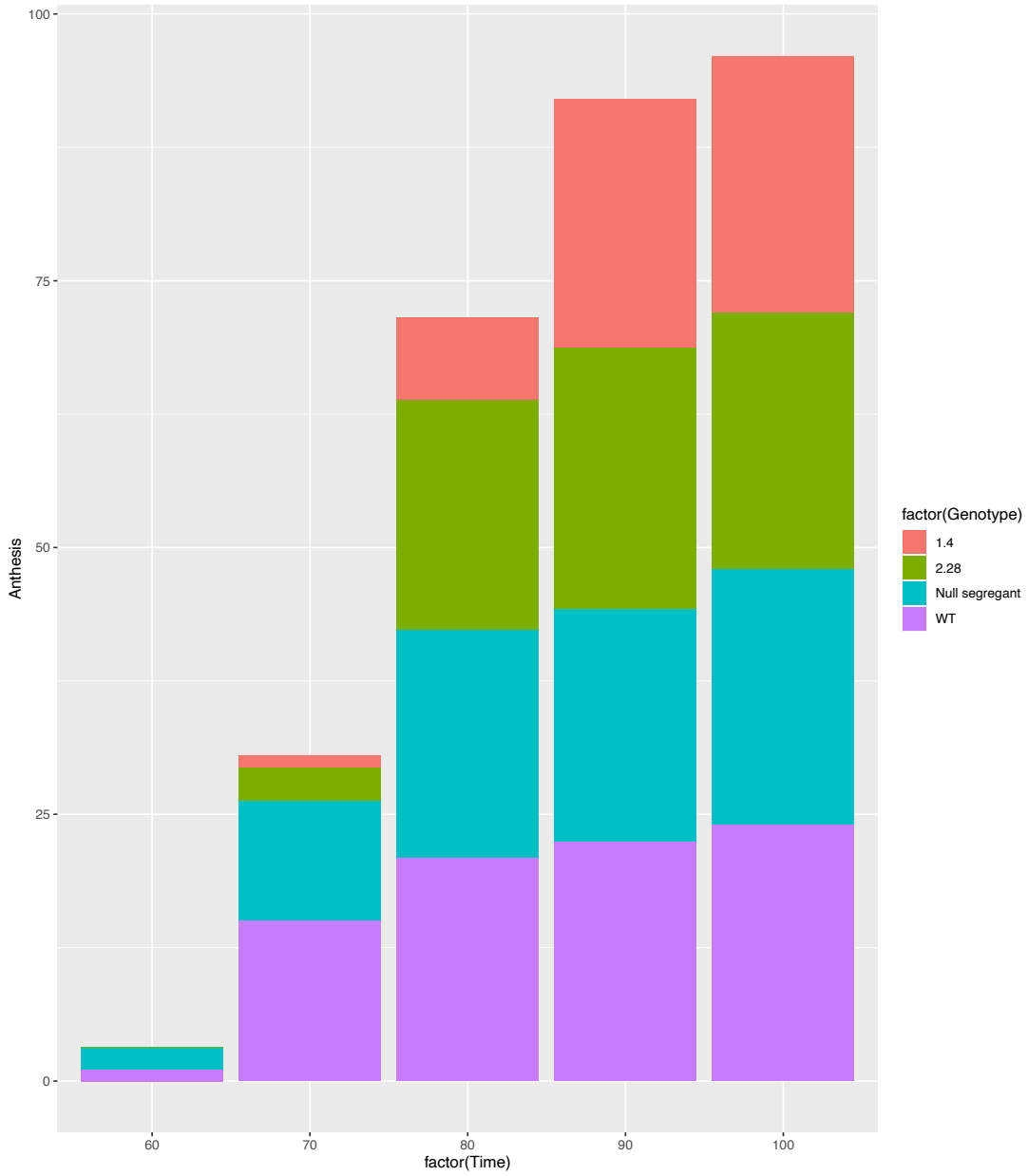
Supplementary Figure 3.19. Higher axillary meristem formation in two *indTaRKD1-7D* T3 lines timecourse boxplots. Trial was performed independently by Dr. Yang-Seok Lee to confirm tiller formation phenotype. All instances labelled 'TaRKD2' are *indTaRKD1-7D*. Asterix above boxplots indicate statistically significant differences between that genotype at that timepoint when compared to wild type. Dashed lines denote different timepoints.

Supplementary Table 3.6. Fold changes in phenotypic traits in tiller traits trial compared to WT.

Traits	Soil volume	1.4 5D FC	1.4 11A FC	2.28 2B FC	2.28 5B FC
Reproductive tiller number	1000ml	1.526032316	1.535008977	1.086175943	1.181328546
Height (cm)	1000ml	1.206465936	1.198140697	0.811988345	1.040377411
Days to maturity	1000ml	1.111532728	1.103270457	1.128782738	1.117790745
Flag leaf area (cm^2)	1000ml	0.894858731	0.953219083	1.000463177	0.956461325
Unproductive tiller number	1000ml	7	10.75	2	1.375
Spike length (cm)	1000ml	1.185393258	1.213483146	1.146067416	1.22752809
Seed number per plant	1000ml	1.883789474	2.00477193	1.232070175	1.543859649
Seed weight per 1000 (g)	1000ml	1.369261477	1.288755822	1.133067199	1.255489022
Seed weight per plant (g)	1000ml	2.512471655	2.471655329	1.358276644	1.845804989
Reproductive tiller number	250ml	1.48787062	1.431266846	1.051212938	1.194070081
Height (cm)	250ml	1.128081458	1.117765273	0.812299035	0.956457663
Days to maturity	250ml	1.107719204	1.110812265	1.125336202	1.095615923

Supplementary Table 3.7. P values in phenotypic traits in tiller traits trial compared to WT.

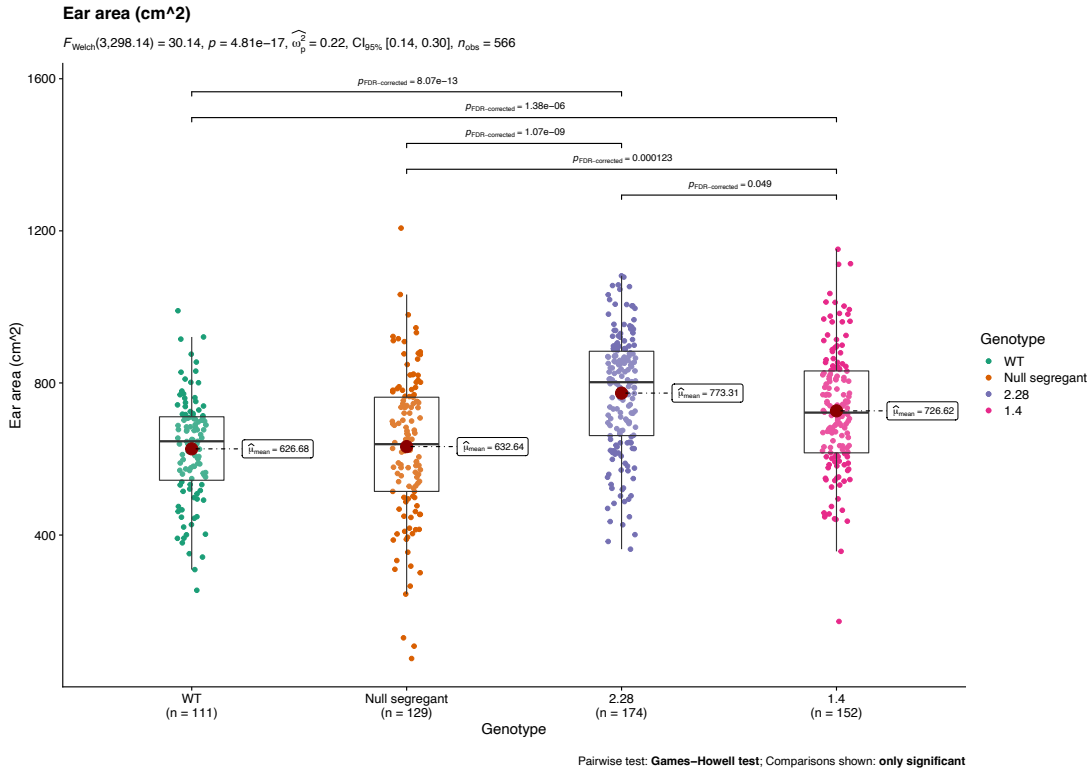
Traits	Soil volume	1.4 5D p.val	1.4 11A p.val	2.28 2B p.val	2.28 5B p.val
Reproductive tiller number	1000ml	3.10E-12	3.93E-11	n.s.	0.037
Height (cm)	1000ml	3.40E-13	3.40E-13	n.s.	n.s.
Days to maturity	1000ml	1.18E-09	1.35E-12	7.22E-10	8.85E-13
Flag leaf area (cm^2)	1000ml	n.s.	n.s.	n.s.	n.s.
Unproductive tiller number	1000ml	0.000102	0.000102	n.s.	n.s.
Spike length (cm)	1000ml	1.18E-09	1.00E-15	1.00E-15	1.00E-15
Seed number per plant	1000ml	3.24E-13	1.49E-09	n.s.	1.49E-07
Seed weight per 1000 (g)	1000ml	9.70E-08	3.22E-05	n.s.	0.000282
Seed weight per plant (g)	1000ml	8.94E-14	0.00E+00	n.s.	3.45E-10
Reproductive tiller number	250ml	7.52E-08	2.28E-06	n.s.	0.009
Height (cm)	250ml	4.38E-07	6.41E-08	1.54E-07	n.s.
Days to maturity	250ml	1.65E-13	1.65E-13	1.65E-13	5.39E-13



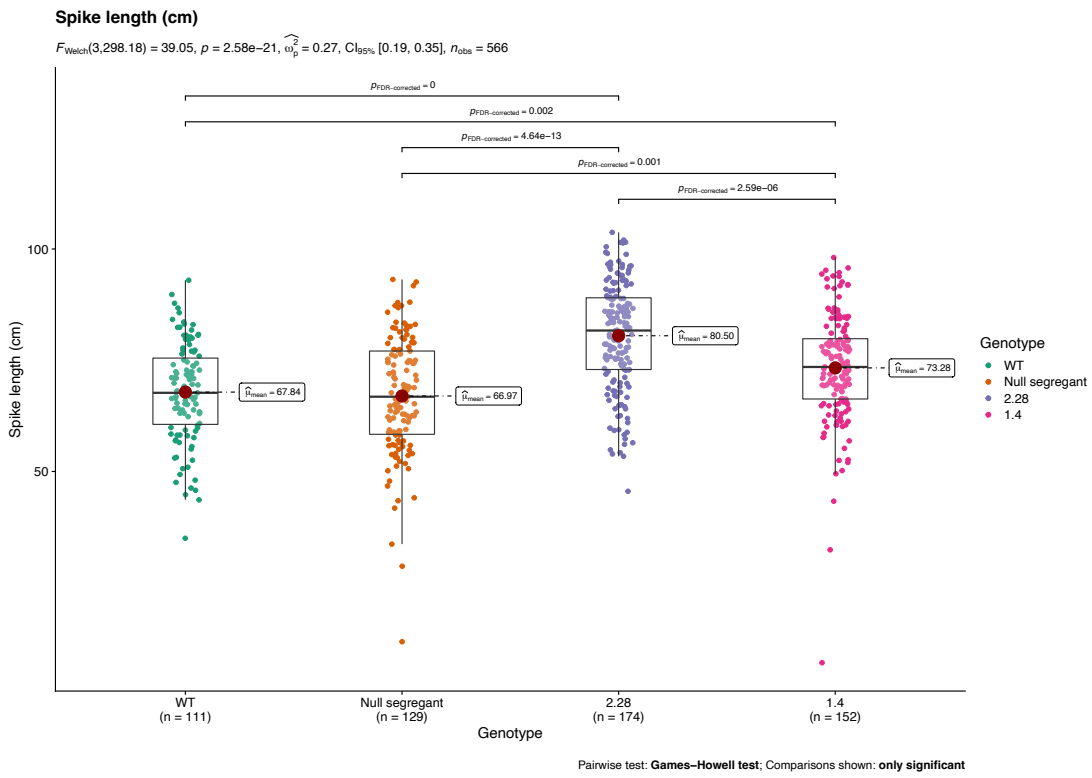
Supplementary Figure 3.20. Delayed anthesis time of two *indTaRKD1-7D* T3 lines. Ratio of plant spikes that had undergone anthesis was measured every 10 days and plotted. Colour denotes genotype as described in the figure on the right. Statistically significant difference as measured by a non-parametric tukey test between genotypes at time points is shown in Supplementary table 3.4.

Supplementary Table 3.8. Statistical significance test of anthesis time between genotypes from spike traits trial.

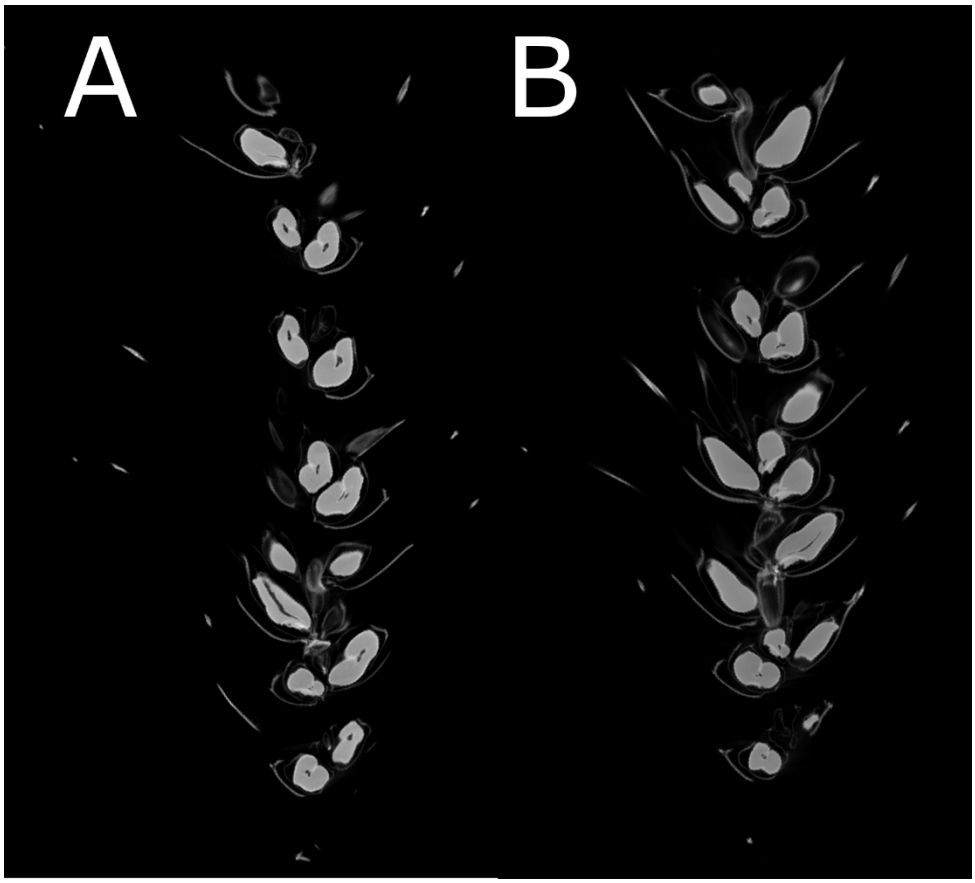
Samples	Days after sowing	Statistical difference
1.4	60	Significant
1.4	70	Significant
1.4	80	Significant
1.4	90	Non signficant
1.4	100	Non signficant
2.28	60	Significant
2.28	70	Significant
2.28	80	Significant
2.28	90	Significant
2.28	100	Non signficant
Null segregant	60	Significant
Null segregant	70	Significant
Null segregant	80	Non signficant
Null segregant	90	Non signficant
Null segregant	100	Non signficant



Supplementary figure 3.21. Increased spike area (cm^2) in two *indTaRKD1-7D* T3 lines. Ear area of fully dried spikes calculated by WERecognizer. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



Supplementary Figure 3.22. Increased spike length of two *indTaRKD1-7D* T3 lines. Spike length of fully dried spikes was calculated by WERecognizer. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



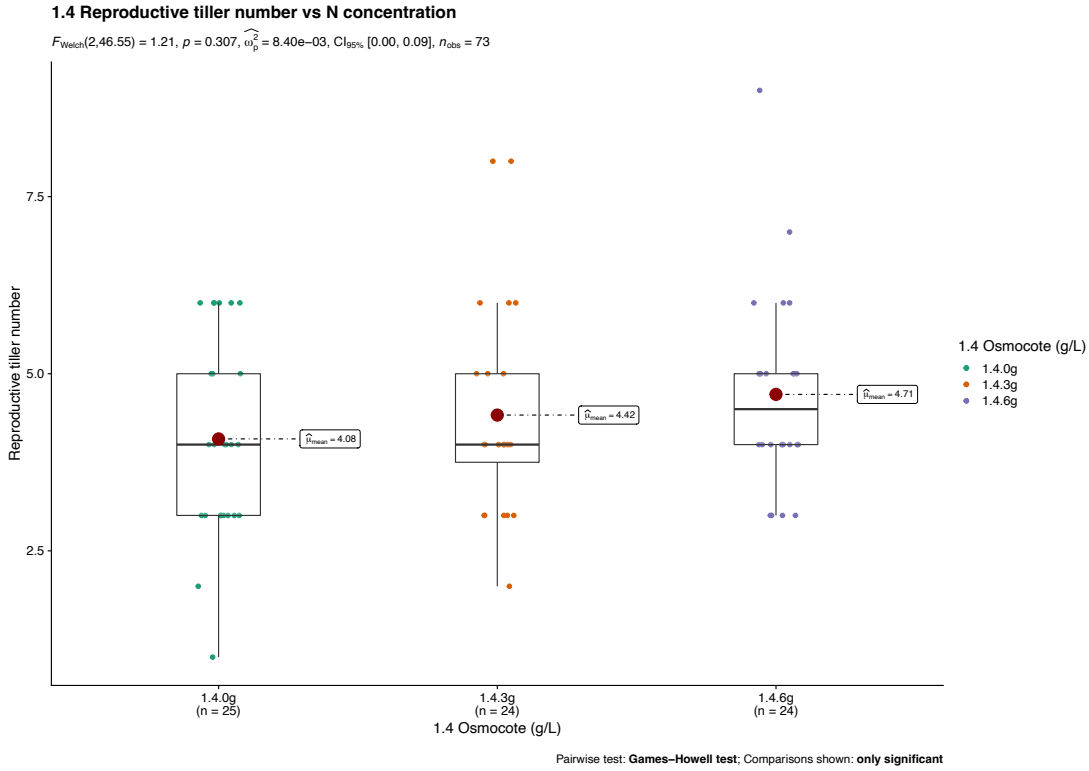
Supplemental figure 3.23. CT scan averaging to reduce background noise. Averaged images of computerized tomography (CT) scans of wheat spikes segmented for grain volumetric analysis. **(A)** Segmented CT scan image of Wild type wheat spike. **(B)** Average of 4 segmented CT scan images of Wild type wheat spike to reduce background noise.

Supplementary 3.9. Fold changes in phenotypic spike traits trial compared to WT.

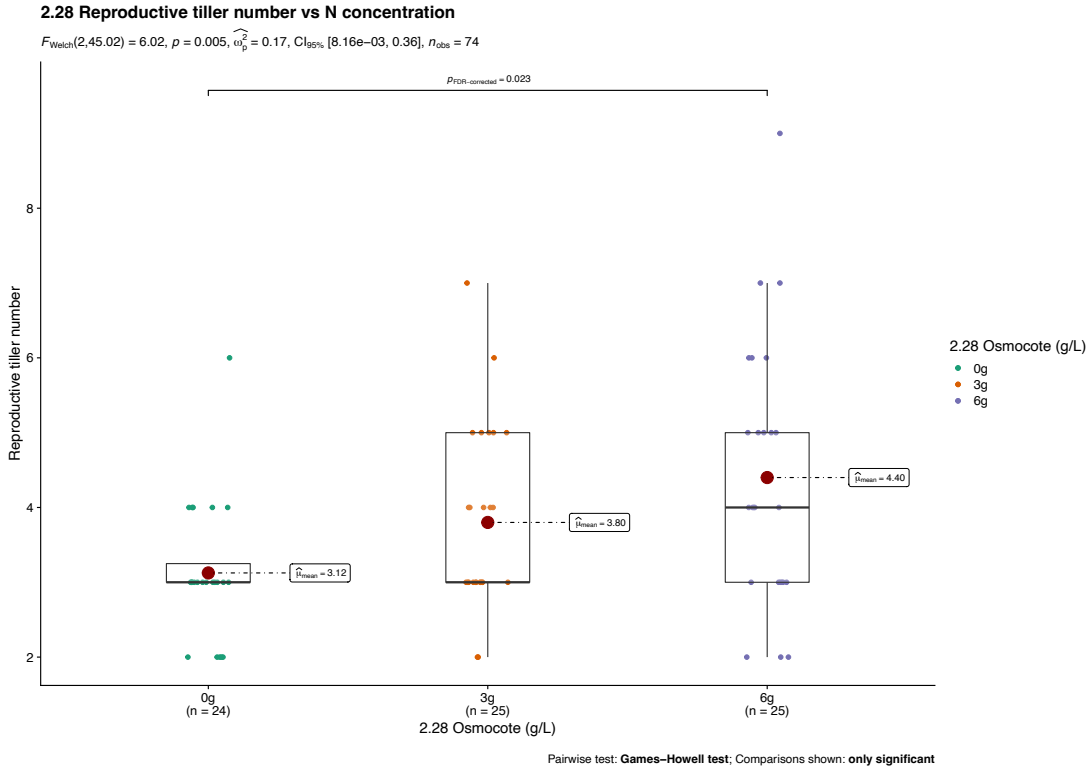
Traits	1.4 FC	2.28 FC	Null segregant FC
Spikelet number per spike	1.15359741	1.22635408	0.997574778
Rachis elongation	0.97735849	0.99622642	1.003773585
Spike length	1.08018868	1.18661557	0.987175708
Ear area (cm^2)	1.15947533	1.23397906	1.009510436
Awns area (cm^2)	1.11450094	1.8173258	1.013559322
Roundness	0.9375	0.875	1
Rugosity	0.98675497	0.97350993	0.960264901
Soiidity	1.01315789	1	1.013157895
Circularity index	1	0.93333333	1.066666667
Perimeter	1.06372421	1.14498999	0.967833203

Supplementary 3.10. P values in phenotypic spike traits trial compared to WT.

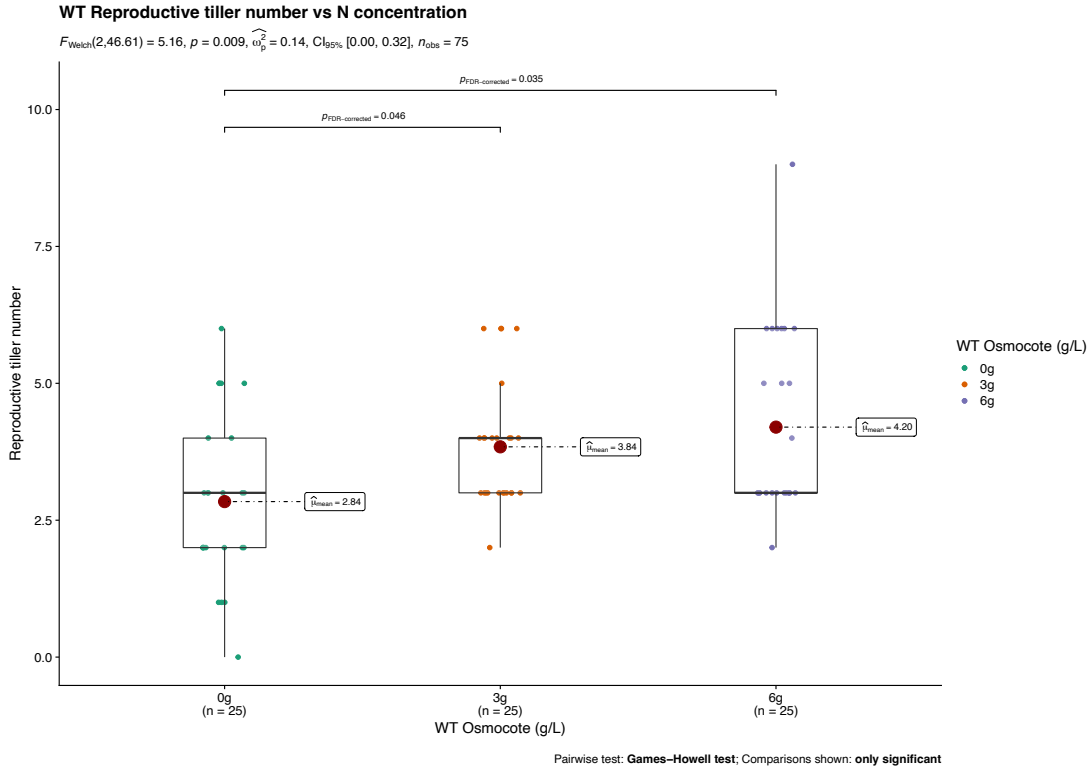
Traits	1.4 p.val	2.28 p.val	Null segregant p.val
Spikelet number per spike	0.003	2.26E-07	n.s.
Rachis elongation	n.s.	n.s.	n.s.
Spike length	0.002	1.00E-15	n.s.
Ear area (cm^2)	1.36E-06	8.07E-13	n.s.
Awns area (cm^2)	n.s.	0.001	n.s.
Roundness	0.004	2.31E-12	n.s.
Rugosity	n.s.	n.s.	n.s.
Soiidity	n.s.	n.s.	n.s.
Circularity index	n.s.	n.s.	n.s.
Perimeter	0.01	7.51E-10	n.s.



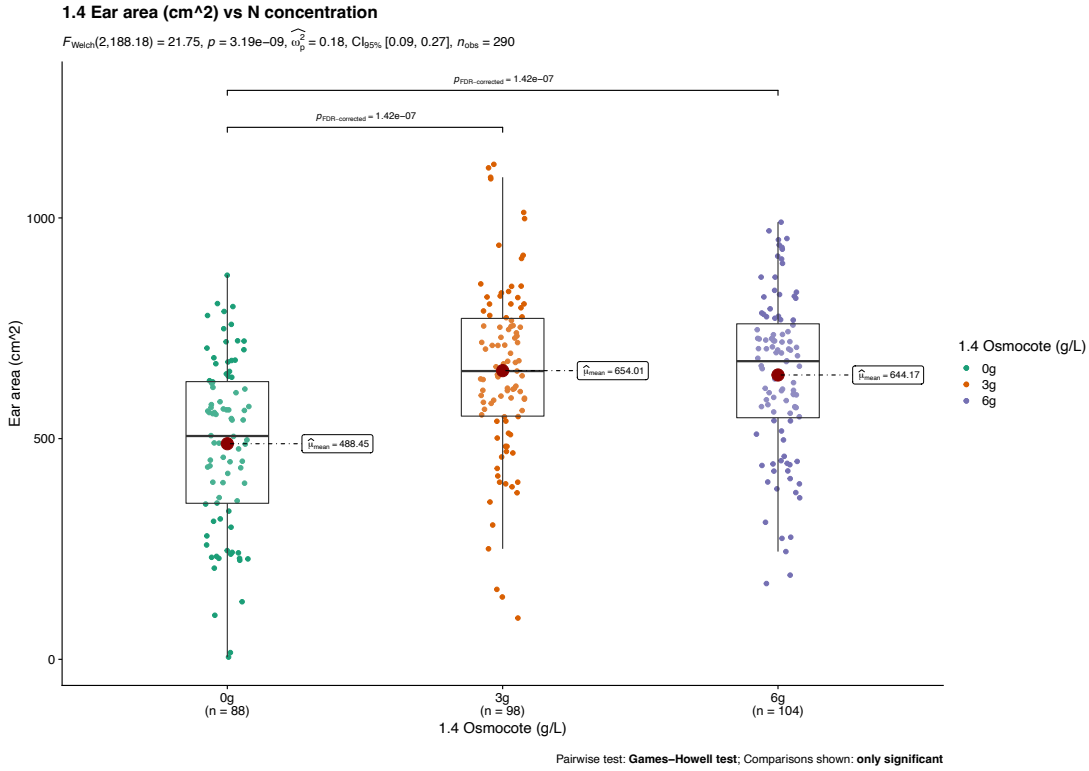
Supplementary Figure 3.24. Nitrogen response trial shows increase in reproductive tiller number of *indTaRKD1-7D #1.4* independent of nitrogen availability. Reproductive tiller number from plants was measured from counting the number of fully developed spikes at senescence. 0g, 3g and 6g represent the amount of Osmocote added to M2 soil. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Additional statistics can be found on the top left of the plot.



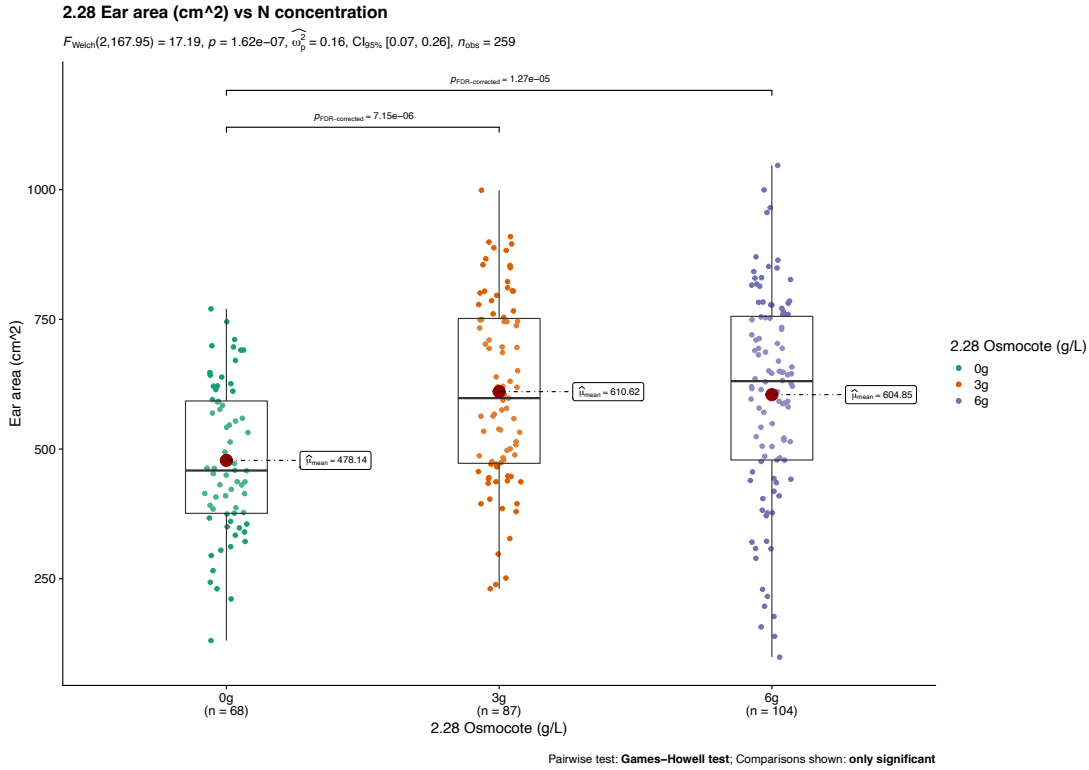
Supplementary Figure 3.25. Nitrogen response trial shows increase in reproductive tiller number of *indTaRKD1-7D #2.28* partially dependent on nitrogen availability. Reproductive tiller number from plants was measured from counting the number of fully developed spikes at senescence. 0g, 3g and 6g represent the amount of Osmocote added to M2 soil. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



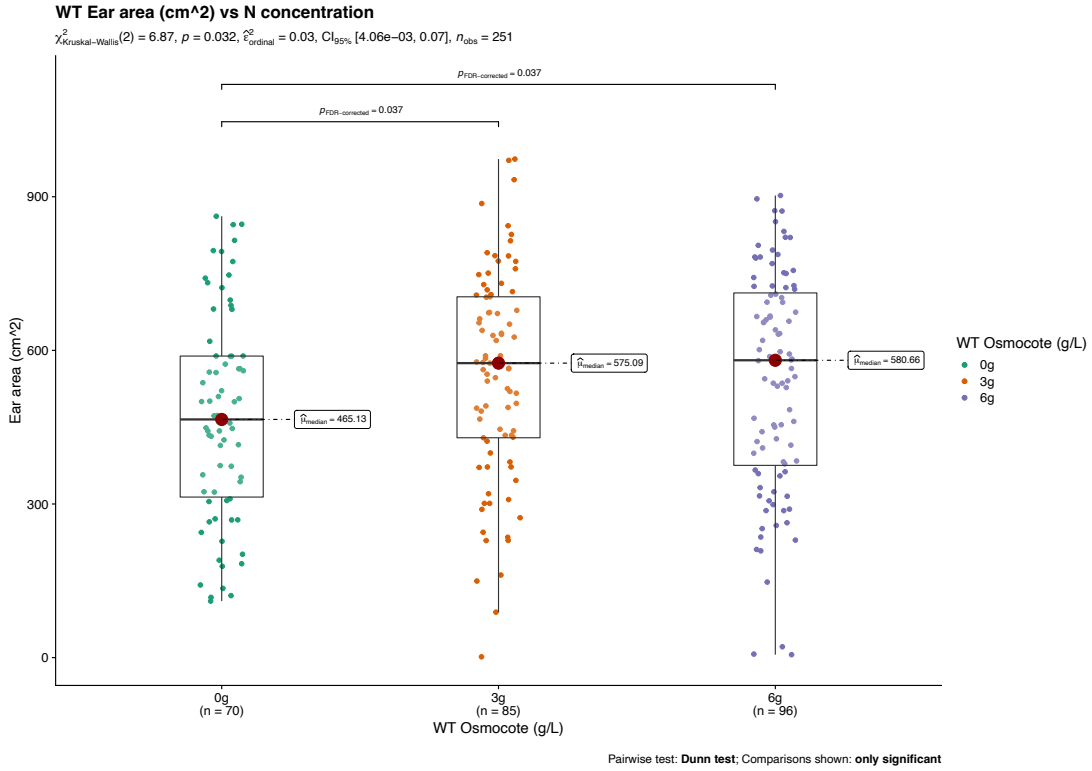
Supplementary Figure 3.26. Nitrogen response trial shows increase in reproductive tiller number of Wild Type dependent on nitrogen availability. Reproductive tiller number from plants was measured from counting the number of fully developed spikes at senescence. 0g, 3g and 6g represent the amount of Osmocote added to M2 soil. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



Supplementary Figure 3.27. Nitrogen response trial shows no increase in nitrogen use efficiency for ear area in *indTaRKD1-7D #1.4*. Ear area of fully dried spikes calculated by WERecognizer. 0g, 3g and 6g represent the amount of Osmocote added to M2 soil. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



Supplementary Figure 3.28. Nitrogen response trial shows no increase in nitrogen use efficiency for ear area in *indTaRKD1-7D #2.28*. Ear area of fully dried spikes calculated by WERecognizer. 0g, 3g and 6g represent the amount of Osmocote added to M2 soil. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



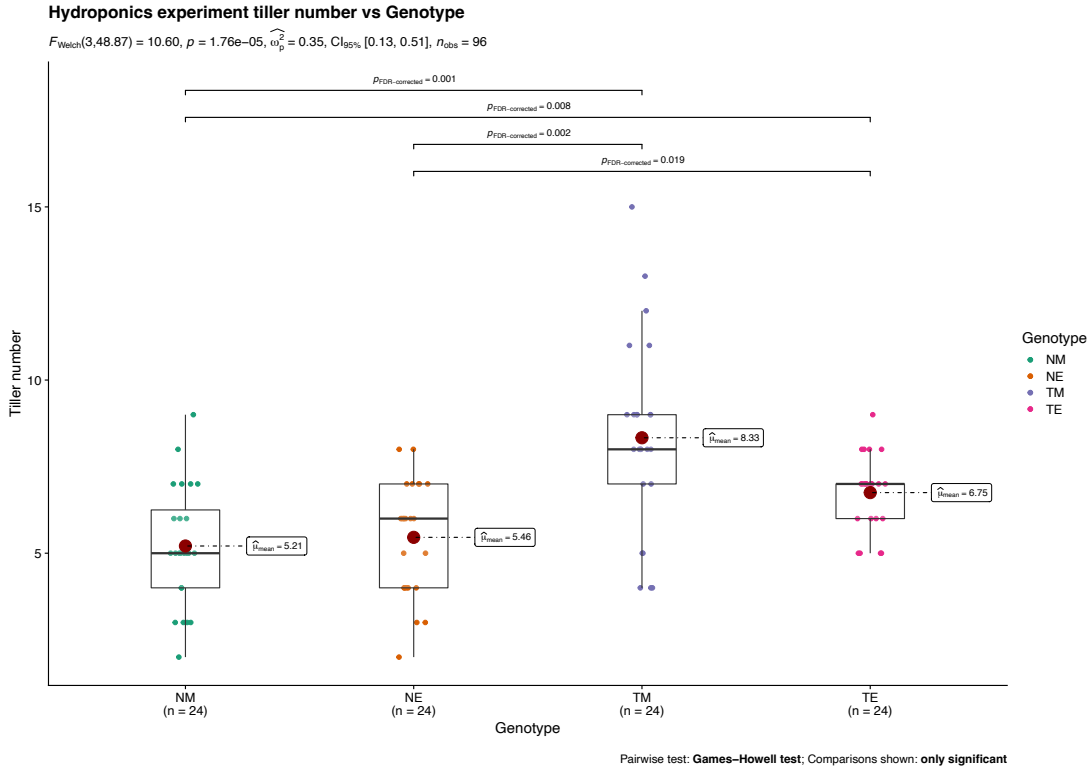
Supplementary Figure 3.29. Nitrogen response trial shows no increase in nitrogen use efficiency for ear area in Wild type. Ear area of fully dried spikes calculated by WERecognizer. 0g, 3g and 6g represent the amount of Osmocote added to M2 soil. The the mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

Supplementary Table 3.11. Fold changes in phenotypic traits in nitrogen response trial compared to WT.

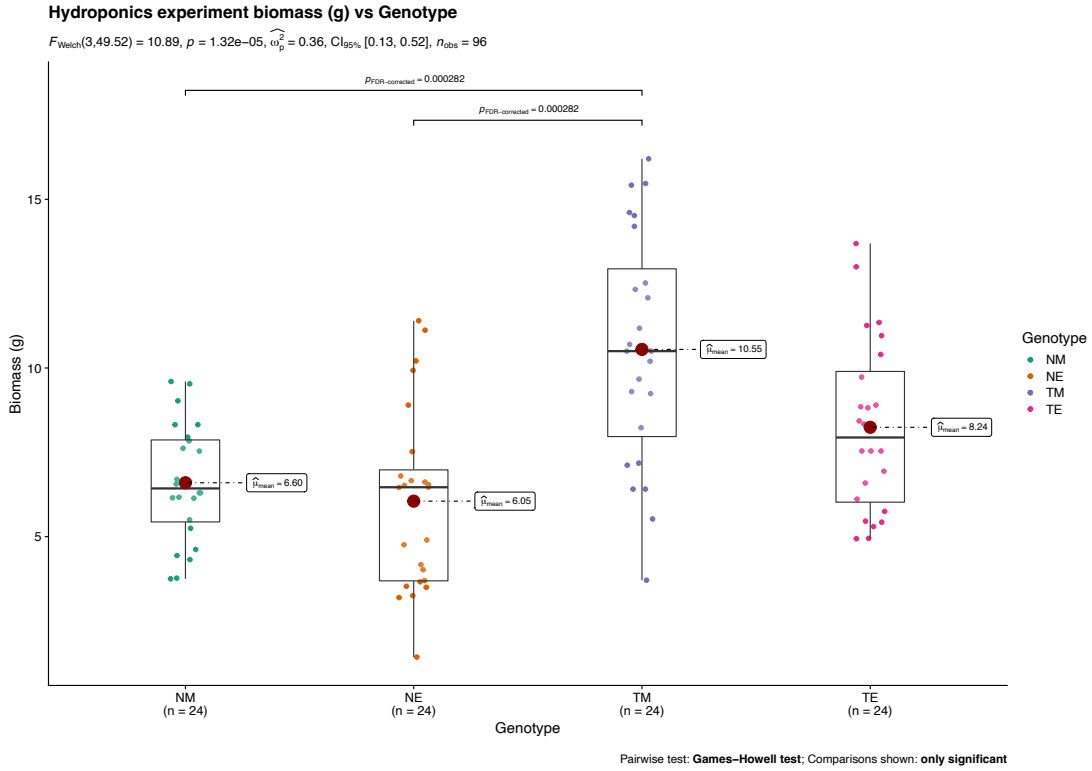
Traits	1.4 3g FC	1.4 6g FC	2.28 3g FC	2.28 6g FC	WT 3g FC	WT 6g FC
Reproductive tiller number	1.08333333	1.15441176	1.21794872	1.41025641	1.35211268	1.47887324
Ear area (cm^2)	1.33894974	1.31880438	1.27749195	1.26500607	1.23640703	1.24838217

Supplementary Table 3.12. P values in phenotypic traits in nitrogen response trial compared to WT.

Traits	1.4 3g p.val	1.4 6g p.val	2.28 3g p.val	2.28 6g p.val	WT 3g p.val	WT 6g p.val
Reproductive tiller number	n.s.	n.s.	n.s.	0.023	0.046	0.035
Ear area (cm^2)	1.42E-07	1.42E-07	7.15E-06	1.27E-05	3.70E-02	3.70E-02



Supplementary Figure 3.30. No increase in vegetative tiller number in inducer treated *indTaRKD1-7D #1.4* in hydroponics trial. Vegetative tiller number was measured 50 days after sowing. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.



Supplementary Figure 3.31. No increase in total plant biomass in Estradiol treated *indTaRKD1-7D #1.4* in hydroponics trial. TE = transgenic estradiol, TM = transgenic mock, NE = non-transgenic estradiol, NM = non-transgenic mock. Total plant biomass was measured 50 days after sowing after cleaning any vermiculite or perlite from the roots. Sample number of each transgenic line is described in the (n=x) brackets below the label. The the mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the horizontal brackets above the samples and values corresponding to these significant differences are indicated on top of the bracket. Additional statistics can be found on the top left of the plot.

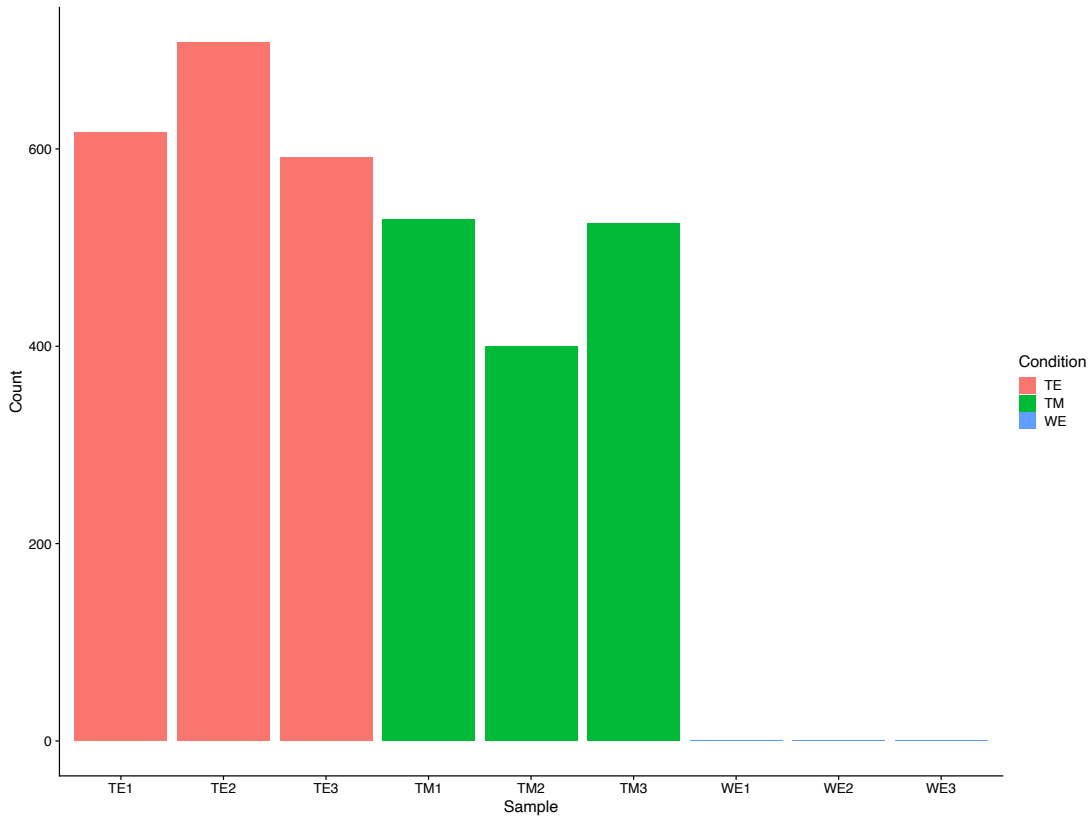
Supplementary Table 3.13. Fold changes in phenotypic traits in hydroponics traits tirals compared to mock solution.

Traits	TE vs TM (indTaRKD1-7D #1.4)	NE vs NM (WT)
Vegetative tiller number	0.81032413	1.047984645
Biomass (g)	0.781042654	0.889705882

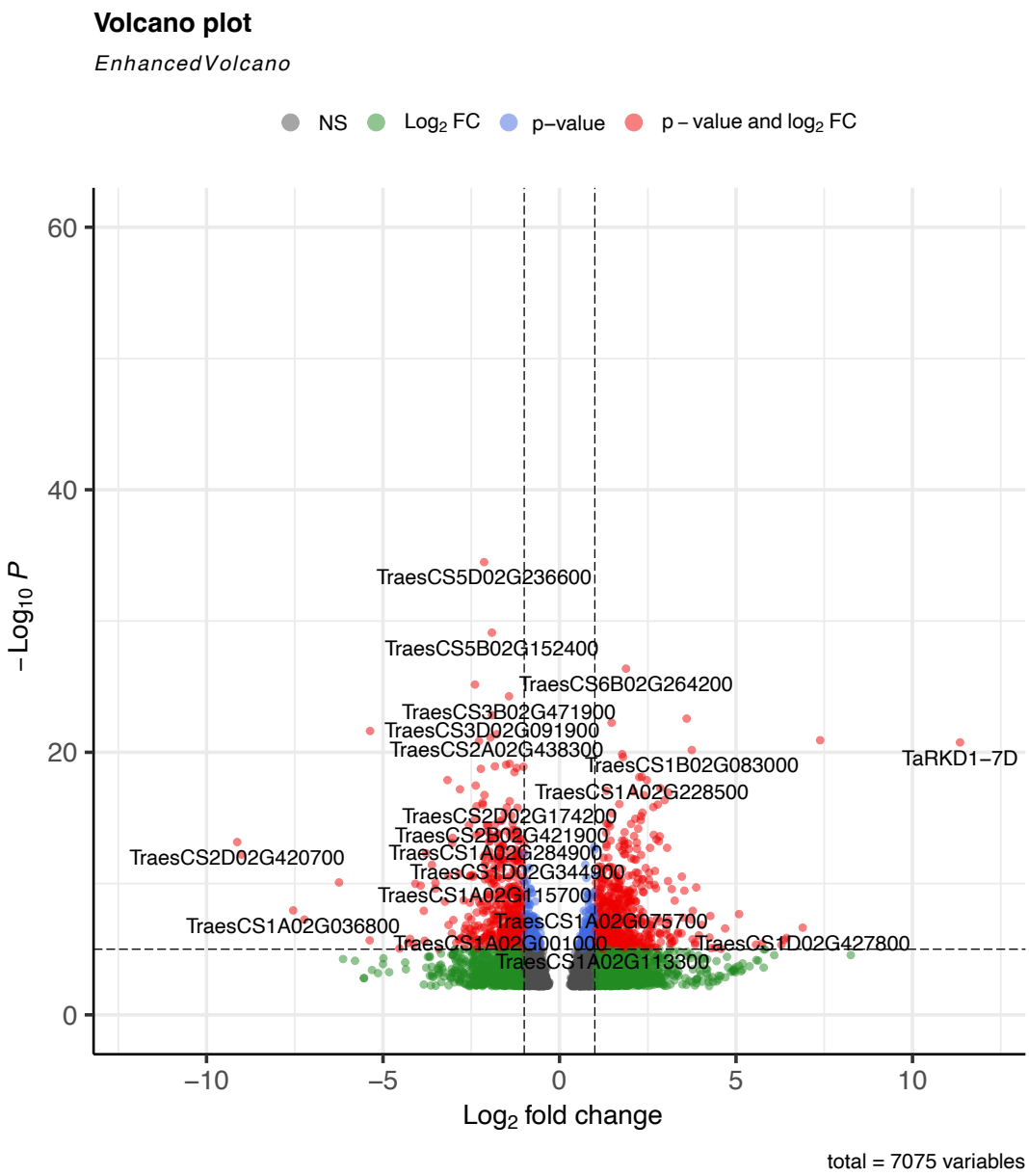
Supplementary Table 3.14. P values in phenotypic traits in hydroponics traits tirals compared to mock solution.

Traits	TE vs TM (indTaRKD1-7D #1.4)	NE vs NM (WT)
Vegetative tiller number	n.s.	n.s.
Biomass (g)	n.s.	n.s.

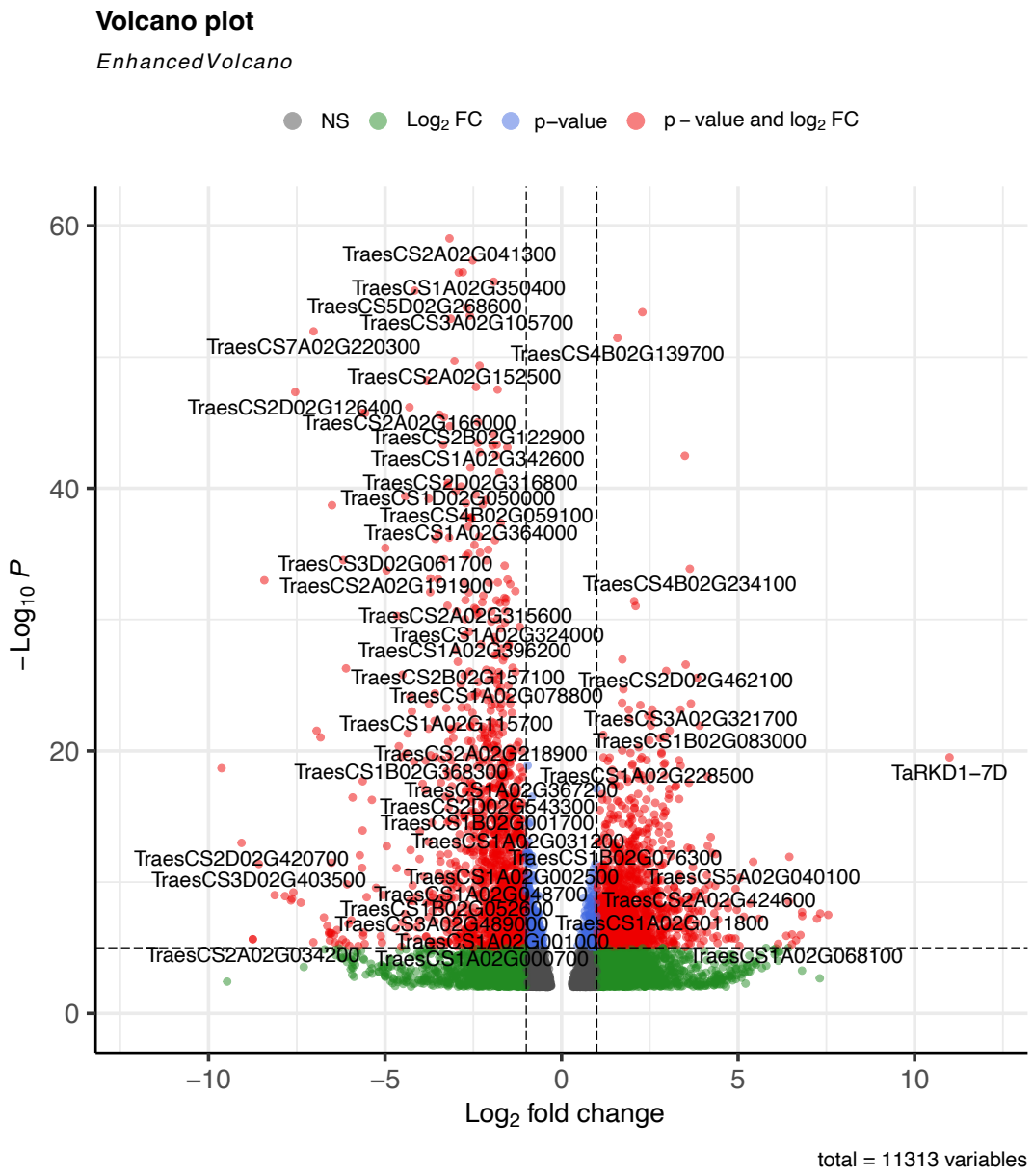
Supplementary Information – Chapter 4



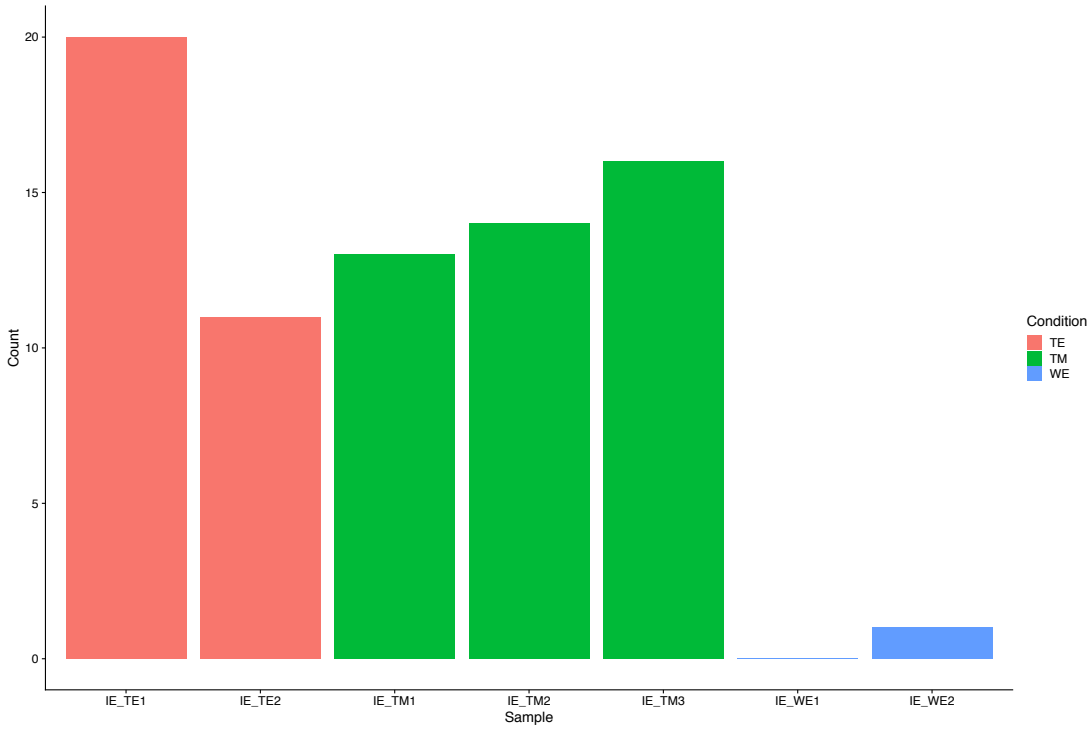
Supplementary figure 4.1. TaRKD1-7D expression in the *indTaRKD1-7D #1.4* leaf RNAseq experiment shows high expression amongst transgenic samples. TE = Transgenic □stradiol leaf, TM = Transgenic mock leaf, and WE = non-transgenic □stradiol leaf. Bar colour denotes sample condition as described in the legend on the right.



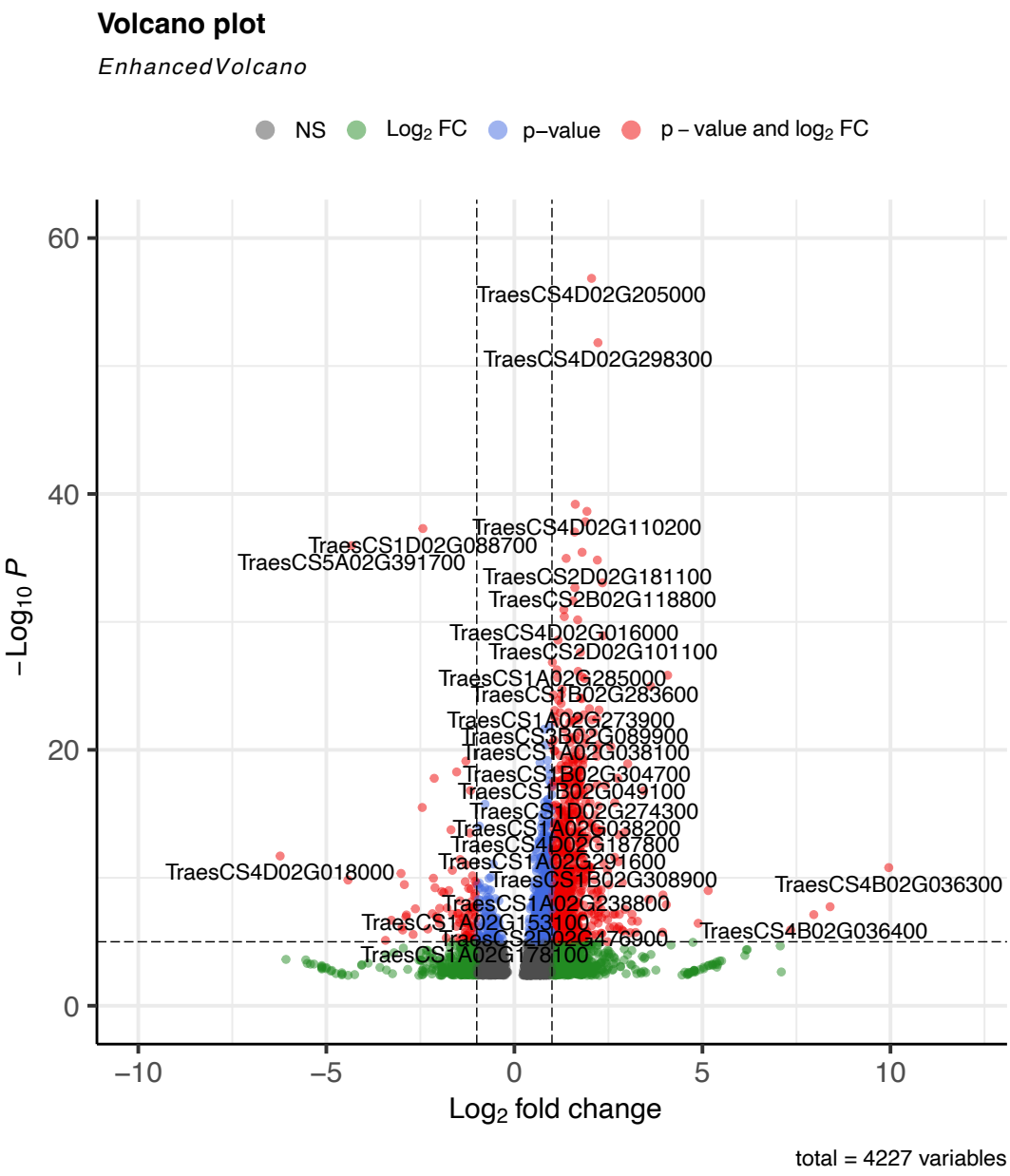
Supplemental figure 4.2. Volcano plot of the *indTaRKD1-7D #1.4* leaf TE vs NE DEGs shows greater upregulation. TE = Transgenic □stradiol leaf and NE = non-transgenic □stradiol leaf. 56% DEGs upregulated and 44% DEGs downregulated. Genes included in the TE vs WE pairwise comparison are plotted across their $\log_2$ fold change and -$\log_{10}$ p values. DEGs are labelled depending on spatial distribution. Red points represent genes that show sufficient $\log_2$ fold change p value to be annotated. Blue points represent genes that show sufficient p value but not $\log_2$ fold change to be annotated. Green points represent genes that show sufficient $\log_2$ fold change but not p value to be annotated. Grey points represent genes that show neither sufficient $\log_2$ fold change or p value to be annotated.



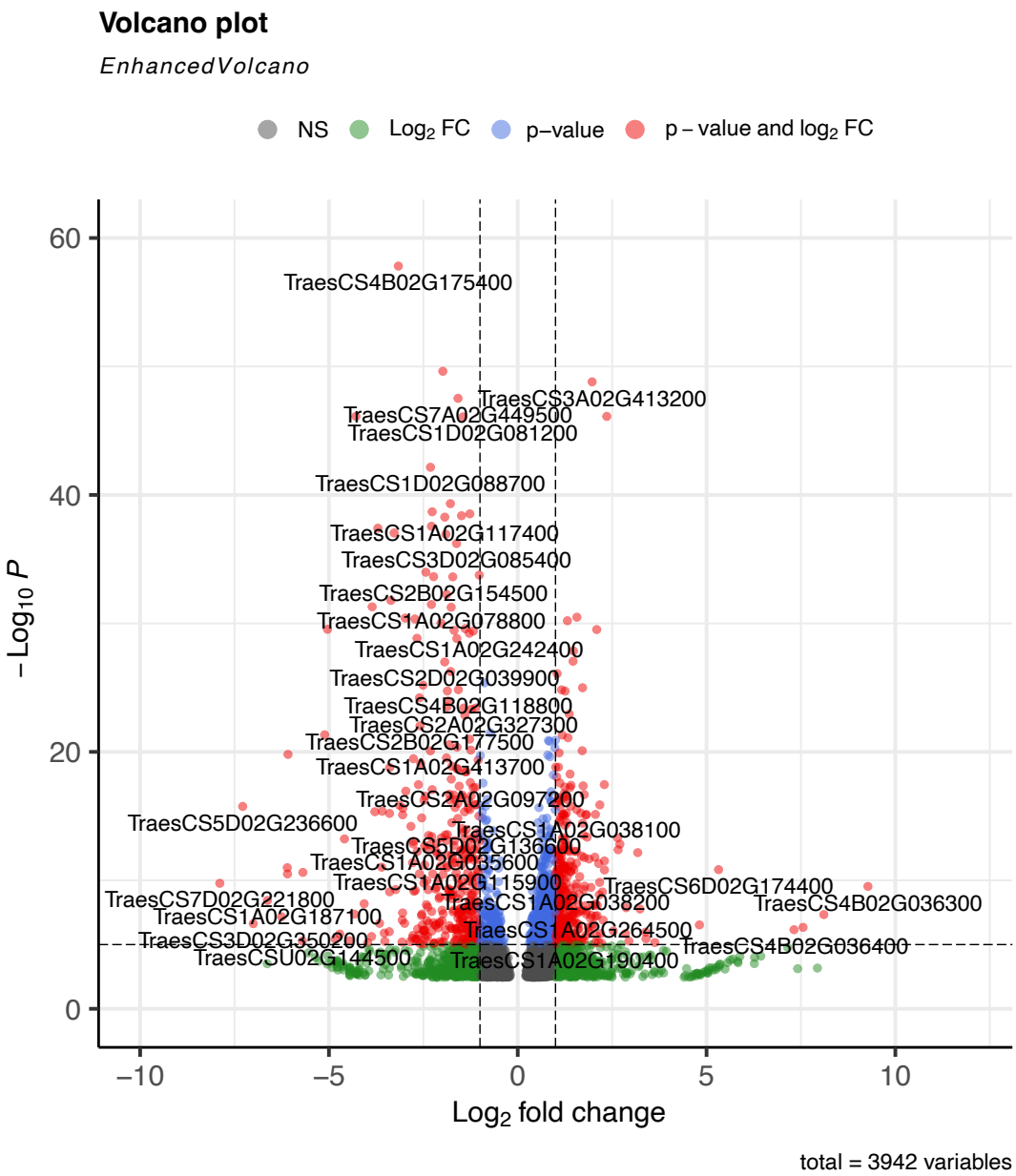
Supplemental figure 4.3. Volcano plot of the *indTaRKD1-7D #1.4* leaf TM vs NE DEGs and shows greater downregulation. TM = Transgenic mock leaf and NE = non-transgenic □stradiol leaf. 53% DEGs upregulated, 47% DEGs downregulated. Genes included in the TM vs WE pairwise comparison are plotted across their $\log_2$ fold change and $-\log_{10}$ p values. DEGs are labelled depending on spatial distribution. Red points represent genes that show sufficient $\log_2$ fold change p value to be annotated. Blue points represent genes that show sufficient p value but not $\log_2$ fold change to be annotated. Green points represent genes that show sufficient $\log_2$ fold change but not p value to be annotated. Grey points represent genes that show neither sufficient $\log_2$ fold change or p value to be annotated.



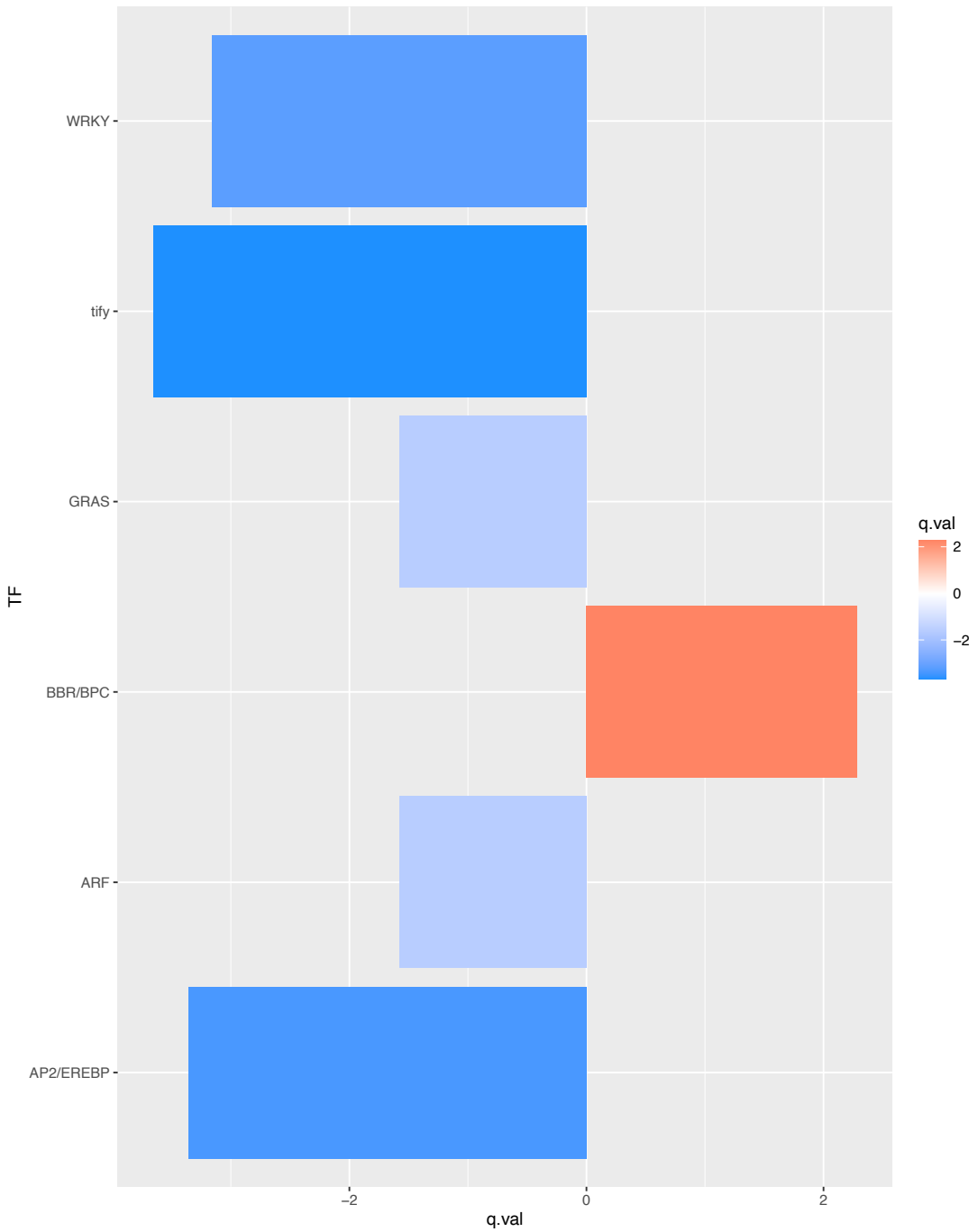
Supplementary figure 4.4. TaRKD1-7D expression in the *indTaRKD1-7D #2.28* immature embryo RNAseq experiment shows high expression amongst transgenic samples. IE_TE = Immature embryo transgenic □stradiol, IE_TM = Immature embryo transgenic mock, and IE_WE = Immature embryo wild type □stradiol. Bar colour denotes sample condition as described in the legend on the right.



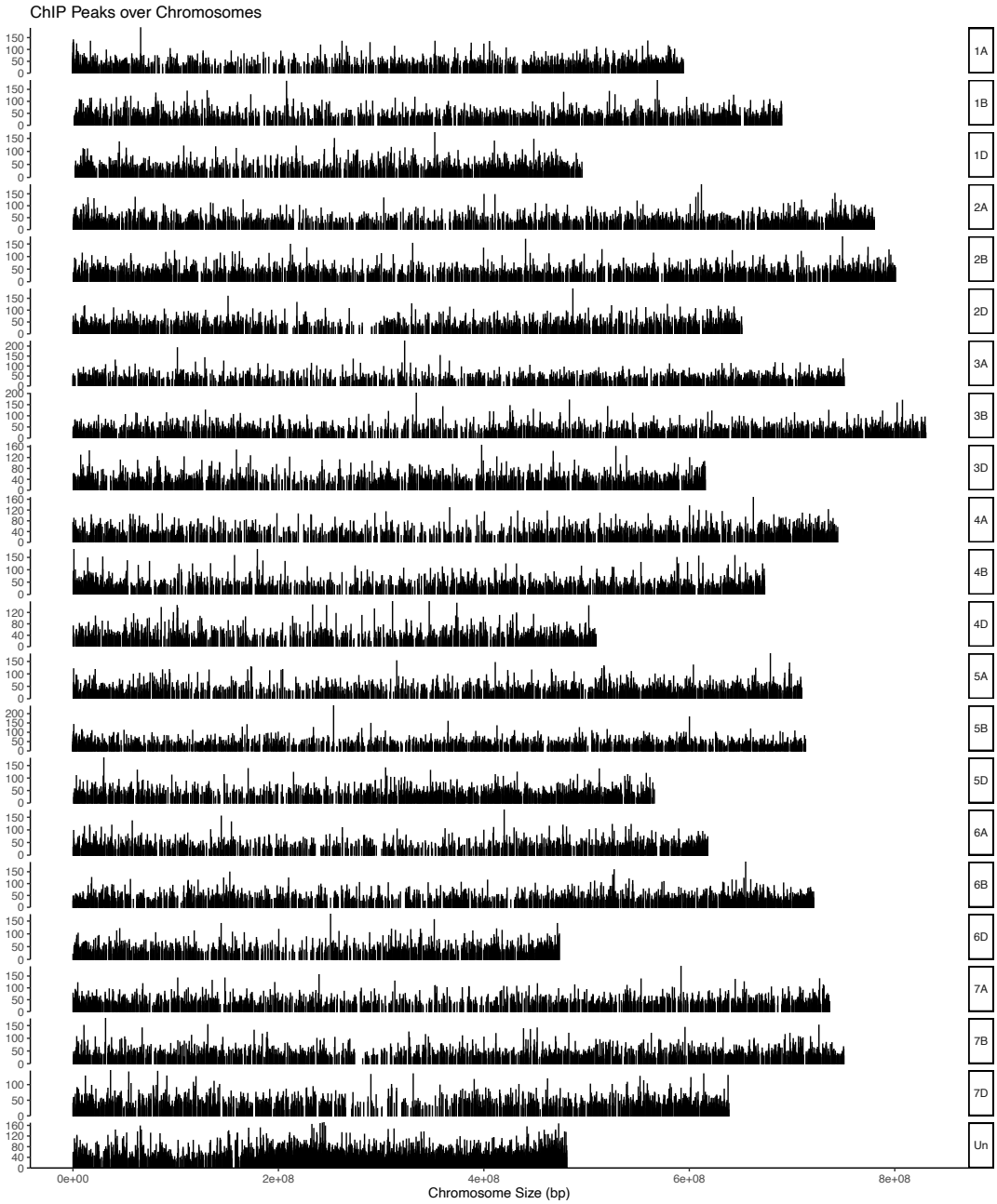
Supplemental figure 4.5. Volcano plot of the *indTaRKD1-7D #2.28* immature embryo IE_TE vs IE_WE DEGs shows greater upregulation. IE_TE = Immature embryo transgenic □stradiol and IE_WE = Immature embryo wild type □stradiol. 75% DEGs upregulated, 25% DEGs downregulated. Genes included in the IE_TE vs IE_WE pairwise comparison are plotted across their log2 fold change and -log10 p values. DEGs are labelled depending on spatial distribution. Red points represent genes that show sufficient log2 fold change p value to be annotated. Blue points represent genes that show sufficient p value but not log2 fold change to be annotated. Green points represent genes that show sufficient log2 fold change but not p value to be annotated. Grey points represent genes that show neither sufficient log2 fold change or p value to be annotated.



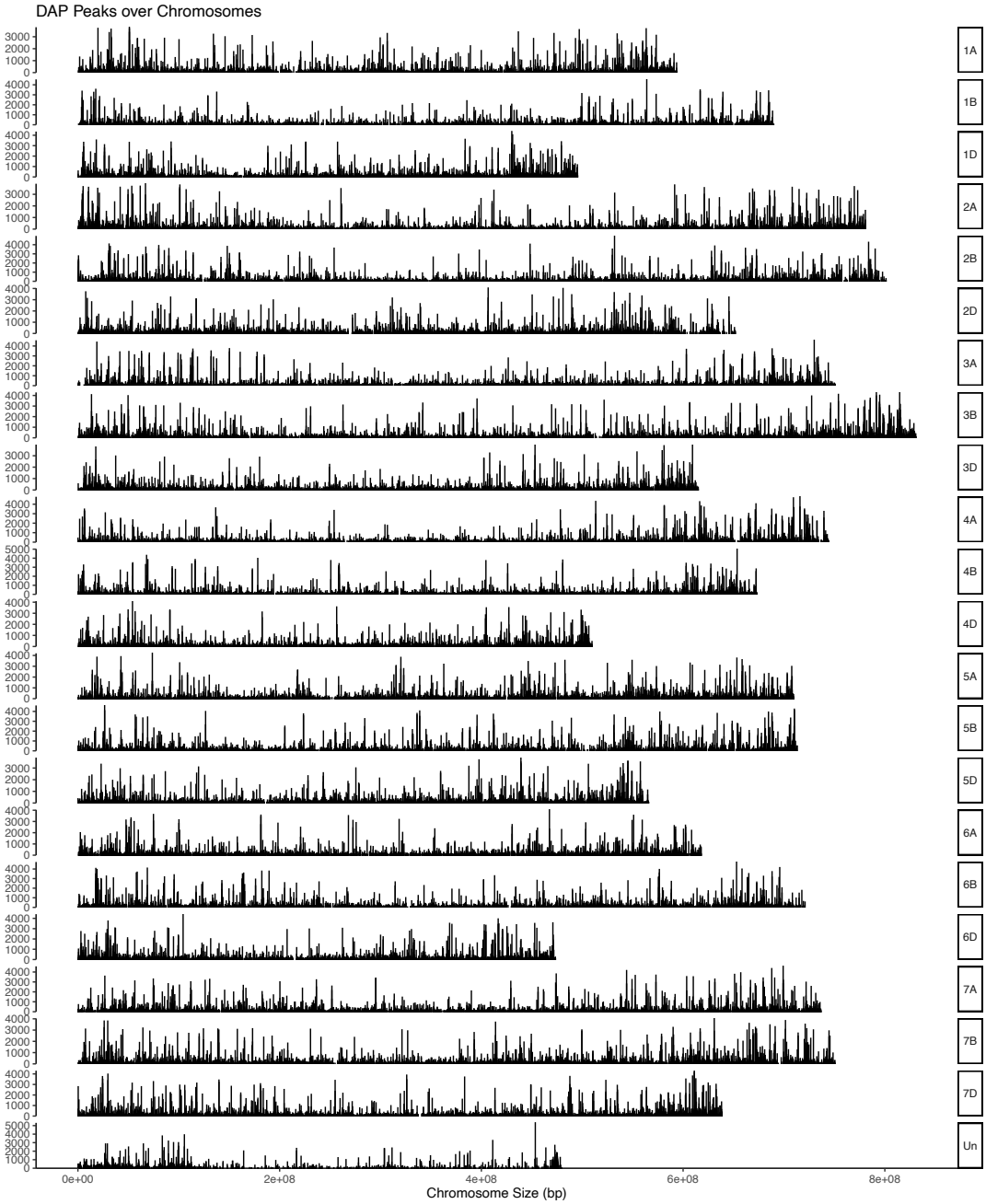
Supplemental figure 4.6. Volcano plot of the *indTaRKD1-7D #2.28* immature embryo IE_TM vs IE_WE DEGs shows greater downregulation. IE_TM = Immature embryo transgenic mock and IE_WE = Immature embryo wild type □stradiol. 61% DEGs upregulated and 39% DEGs downregulated. Genes included in the IE_TM vs IE_WE pairwise comparison are plotted across their $\log_2$ fold change and -$\log_{10}$ p values. DEGs are labelled depending on spatial distribution. Red points represent genes that show sufficient $\log_2$ fold change p value to be annotated. Blue points represent genes that show sufficient p value but not $\log_2$ fold change to be annotated. Green points represent genes that show sufficient $\log_2$ fold change but not p value to be annotated. Grey points represent genes that show neither sufficient $\log_2$ fold change or p value to be annotated.



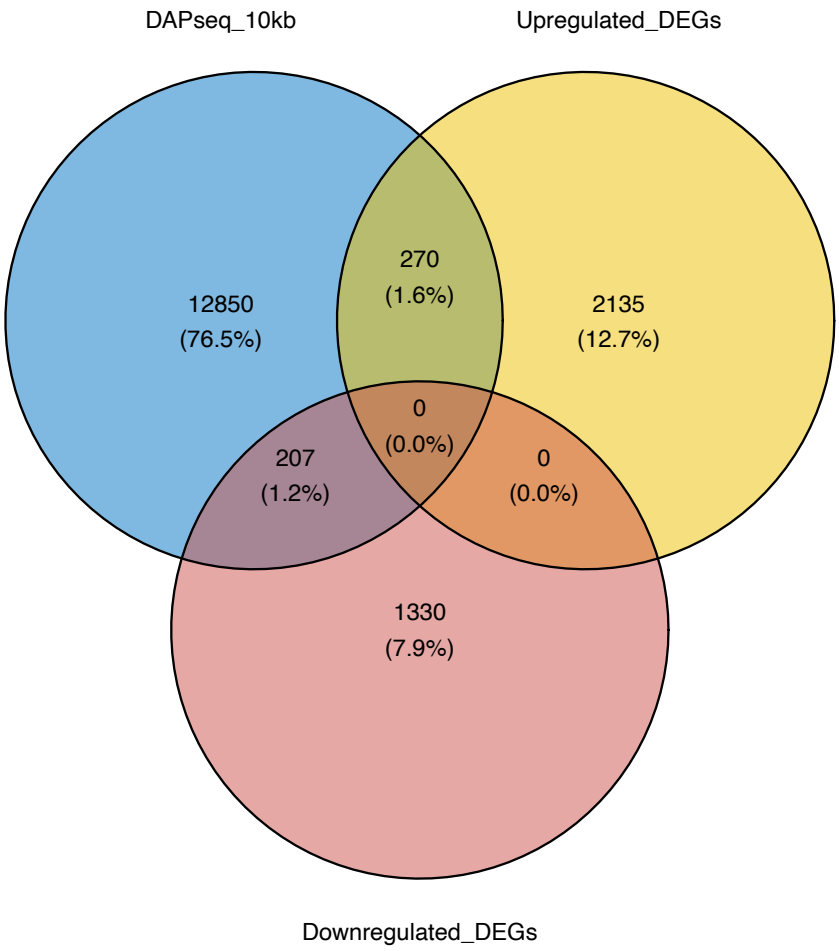
Supplemental figure 4.7. Significant up and downregulation of TF families in *indTaRKD1-7D #2.28* immature embryos. Barplots show significantly upregulated (red) or downregulated (blue) transcription factor (TF) families in *indTaRKD1-7D #2.28* immature embryos by taking the -$\log_{10}$ of the q value (adjusted p value) for differential regulation of a TF. Upregulation values are positive and downregulation values are negative.



Supplemental figure 4.8. MACS2 ChIPseq peaks are distributed evenly across all 21 wheat chromosomes but with a high amount called in the unassembled region.
The number of MACS2 peaks in a sliding window are shown by lines where the left y-axis shows the number of peaks called at each locus. Chromosome name is denoted on the right y-axis.



Supplemental figure 4.9. MACS2 DAPseq peaks are distributed evenly across all 21 wheat chromosomes. The number of MACS2 peaks in a sliding window are shown by lines where the left y-axis shows the number of peaks called at each locus. Chromosome name is denoted on the right y-axis.



Supplemental figure 4.10. Venn diagrams of *indTaRKD1-7D* DAPseq and immature embryo RNAseq shared DEGs identifies expression validated targets of TaRKD1-7D. Venn diagram overlap of genes with a MACS2 DAPseq peak called within 10kb of the TSS and upregulated or downregulated immature embryo *indTaRKD1-7D #2.28* shared DEGs.

Supplemental table 4.1. *indTaRKD1-7D #2.28* immature embryo shared DEGs with a DAPseq peak 10kb within TSS relevant to cell cycle and meristems.

Gene ID	Putative function	Biological process
TraesCS3A02G188400	Arognaute protein	RISC complex
TraesCSU02G081900	Arognaute protein	RISC complex
TraesCS3B02G181500	AUX/IAA response gene	Auxin signalling
TraesCS2B02G168900	B3 transcription factor	Panicle number
TraesCS3D02G080400	Cell division cycle associated 7	Cell division cyle
TraesCS3D02G326600	Cyclin	Cell division cyle
TraesCS4D02G199700	Elongator Acetyltransferase Complex	Transcription
TraesCS2B02G232500	F box domain	Ubiquitination for degradation
TraesCS4D02G097000	F box domain	Ubiquitination for degradation
TraesCS6D02G018600	F box domain	Ubiquitination for degradation
TraesCSU02G027800	F box domain	Ubiquitination for degradation
TraesCSU02G081900	F box domain	Ubiquitination for degradation
TraesCS2A02G076800	F box domain	Ubiquitination for degradation
TraesCS3D02G220100	GRAS transcription factor	Embryogenesis/grain size
TraesCS4D02G181700	Helicase	DNA replication
TraesCS4D02G041400	Helicase	DNA replication
TraesCS5A02G352100	High mobility box domain	Transcription and DNA replication
TraesCS7A02G329300	High mobility box domain	Transcription and DNA replication
TraesCS7D02G326200	High mobility box domain	Transcription and DNA replication
TraesCS5B02G448400	Histone H1	DNA packaging
TraesCS5A02G443800	Histone H1	DNA packaging
TraesCS4D02G205000	Histone H2A	DNA packaging
TraesCS6A02G062600	Histone H2A	DNA packaging
TraesCS6B02G049000	Histone H2A	DNA packaging
TraesCS6B02G049400	Histone H2A	DNA packaging
TraesCS6D02G107000	Histone H2A	DNA packaging
TraesCS7A02G438000	Histone H2A	DNA packaging
TraesCS7A02G501100	Histone H2A	DNA packaging
TraesCS7D02G427900	Histone H2A	DNA packaging
TraesCS1B02G386200	Histone H2B	DNA packaging
TraesCS3A02G357400	Histone H2B	DNA packaging
TraesCS3D02G351300	Histone H2B	DNA packaging
TraesCS3D02G351400	Histone H2B	DNA packaging
TraesCS6B02G084900	Histone H2B	DNA packaging
TraesCS6D02G040200	Histone H2B	DNA packaging
TraesCS7B02G394100	Histone H2B	DNA packaging
TraesCS7D02G428300	Histone H2B	DNA packaging
TraesCSU02G077400	Histone H2B	DNA packaging
TraesCS6A02G063400	Histone H3	DNA packaging
TraesCS7B02G016900	Histone H3	DNA packaging
TraesCS7D02G097600	Histone H3	DNA packaging
TraesCS4A02G067800	Histone H4	DNA packaging
TraesCS5B02G356200	Histone H4	DNA packaging
TraesCS5B02G348900	Histone H4	DNA packaging
TraesCS6A02G055400	Histone H4	DNA packaging
TraesCS6A02G062800	Histone H4	DNA packaging
TraesCS6B02G074000	Histone H4	DNA packaging
TraesCS6B02G078900	Histone H4	DNA packaging
TraesCS6D02G062100	Histone H4	DNA packaging
TraesCSU02G142200	Histone H4	DNA packaging
TraesCS3A02G434900	MADS box transcription factor	Floral organ identity
TraesCS1D02G375700	Methyltransferase	DNA/Histone methylation
TraesCS7D02G239400	Methyltransferase	DNA/Histone methylation
TraesCS2A02G157700	Origin of replication complex 6	Transcription initiation
TraesCS5D02G130400	RNA polymerase A	Transcription
TraesCS4D02G355000	SKP1-like	Ubiquitination for degradation
TraesCS4D02G157300	TCP chaperonin	Plant architecture
TraesCS1A02G338200	Trehalose phosphotase	Meristematic determinancy
TraesCS1D02G340400	Trehalose phosphotase	Meristematic determinancy
TraesCS4D02G189800	Wound induced 1 like protein	Cell division cyle

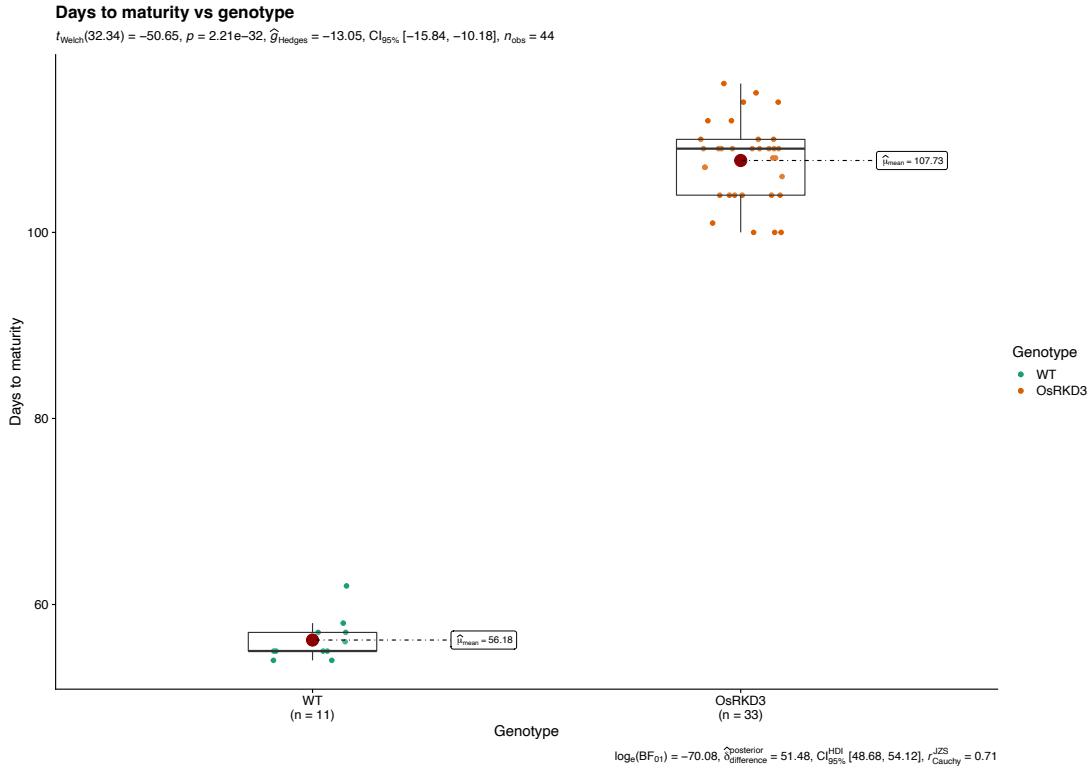
Supplementary Information – Chapter 5

Supplementary Table 5.1. Transgenic rice lines provided by NIAB Cambridge.

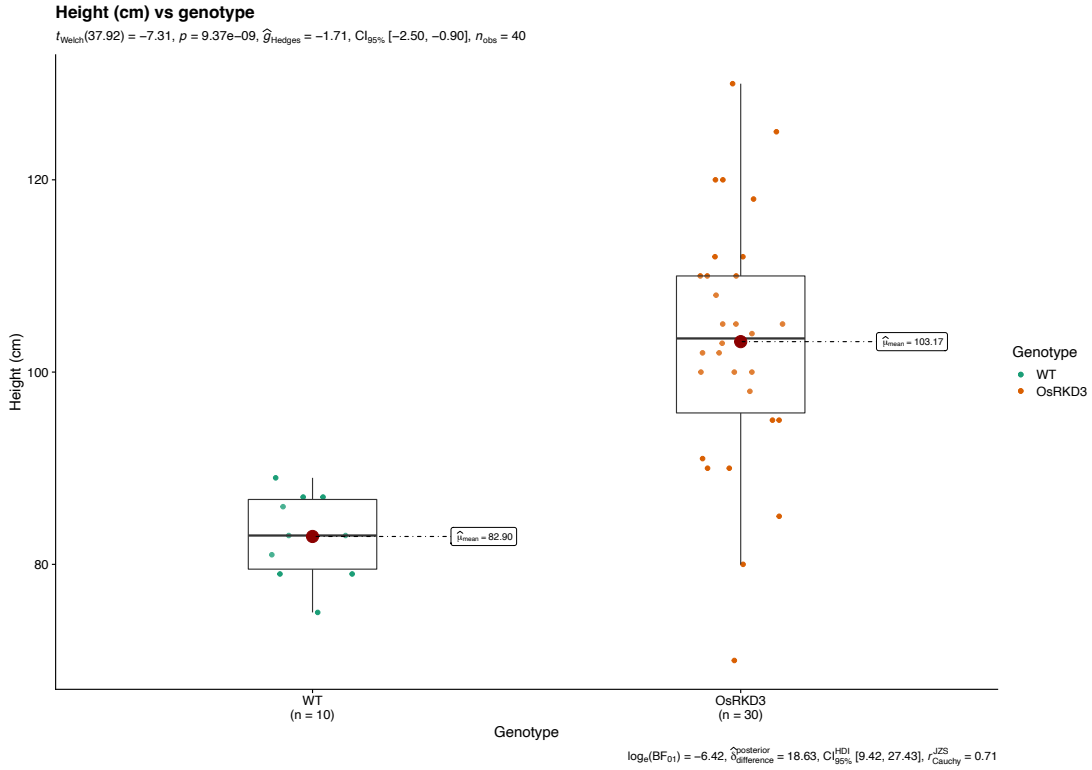
pERV1-hyg-AtRKD4 (Nipponbare)			pERV1-hyg-OsRKD3 (Nipponbare)		
Plant number	Hyg PCR (+/-)	GOI PCR (+/-)	Plant number	Hyg PCR (+/-)	GOI PCR (+/-)
PR2A.1	+	+	PR1.1	+	+
PR2A.2	+	(+)	PR1.2	+	+
PR2A.3	+	+	PR1.3	+	+
PR2A.4	+	+	PR1.4	+	+
PR2A.5	+	+	PR1.5	+	+
PR2A.7	+	+	PR1.6	+	+
PR2A.8	+	+	PR1.8	+	+
PR2A.9	+	+	PR1.13	+	+
PR2A.10	+	+	PR1.14	+	+
PR2A.11	+	+	PR1.15	+	+
PR2A.12	+	+	PR1.16	+	+
PR2A.13	+	+	PR1.17	+	+
PR2A.14	+	+	PR1.18	+	+
PR2A.15	+	+	PR1.19	+	+
PR2A.16	+	+	PR1.20	+	+
PR2A.19	+	+	PR1.22	+	+
PR2A.20	+	+	PR1.24	+	+
PR2A.21	+	+	PR1.25	+	+
PR2A.22	+	+	PR1.26	+	+
PR2A.23	+	+	PR1.27	+	+
PR2A.con 1	-	-	PR1.29	+	+
PR2A.con 2	-	-	PR1.30	+	+
PR2B.1	+	+	PR1.31	+	+
PR2B.2	+	+	PR1.33	+	+
PR2B.4	+	+	PR1.34	+	+
PR2B.5	+	+	PR1.38	+	+
PR2B.6	+	+	PR1.39	+	+
PR2B.8	+	+	PR1.40	+	+
PR2B.9	+	+	PR1.con1	-	-
PR2B.11	+	+	PR1.con2	-	-
PR2B.12	+	+	PR1.con3	-	-
PR2B.13	+	+	PR1.con4	-	-
PR2B.14	+	+	PR1.con5	-	-
PR2B.15	+	+			
PR2B.16	+	+			
PR2B.17	+	+			
PR2B.18	+	+			
PR2B.19	+	+			
PR2B.20	+	+			
PR2B.21	+	+			
PR2B.22	+	+			
PR2B.23	+	+			
PR2B.con 1	-	-			
PR2B.con 2	-	-			
PR2B.con 3	-	-			
PR2B.con 4	-	-			

Supplementary Table 5.2. Statistics for phenotypic traits in *indAtRKD4* Nipponbare trial.

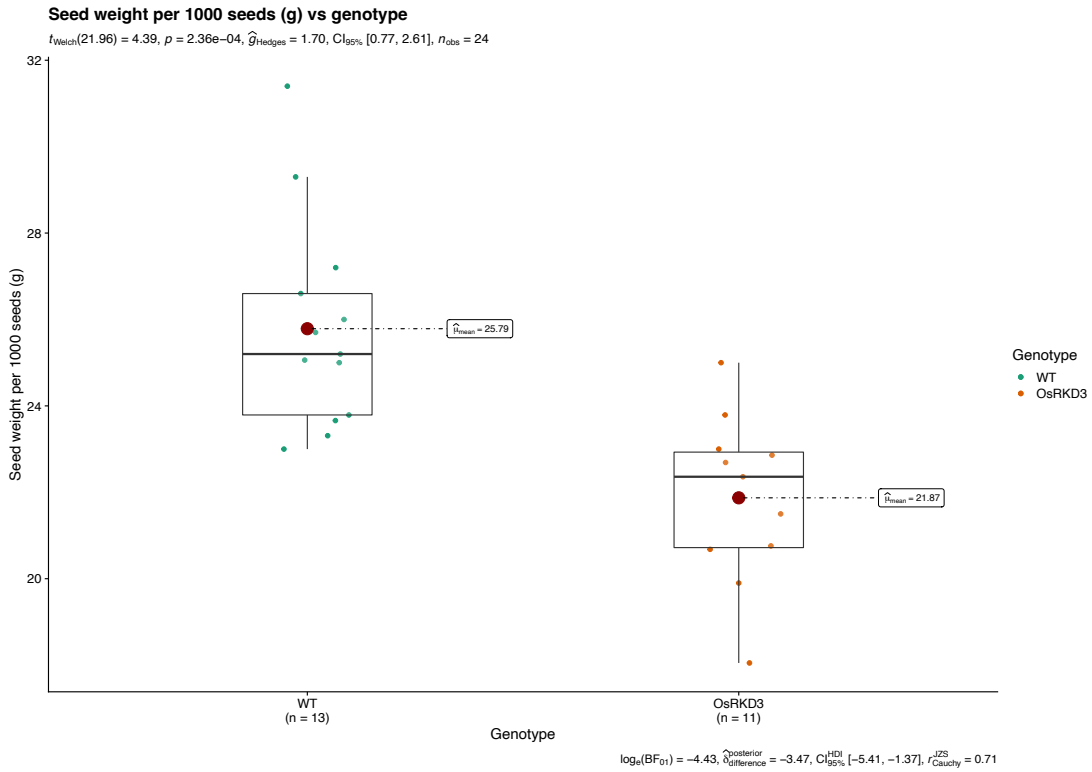
Traits	indAtRKD #PR4.25 1 FC	indAtRKD #PR4.25 1 p.val	indAtRKD #PR4.25 2 FC	indAtRKD #PR4.25 2 p.val
Vegetative tiller number	3	n.s.	3	n.s.
Height (cm)	1.088435374	n.s.	1.068027211	n.s.



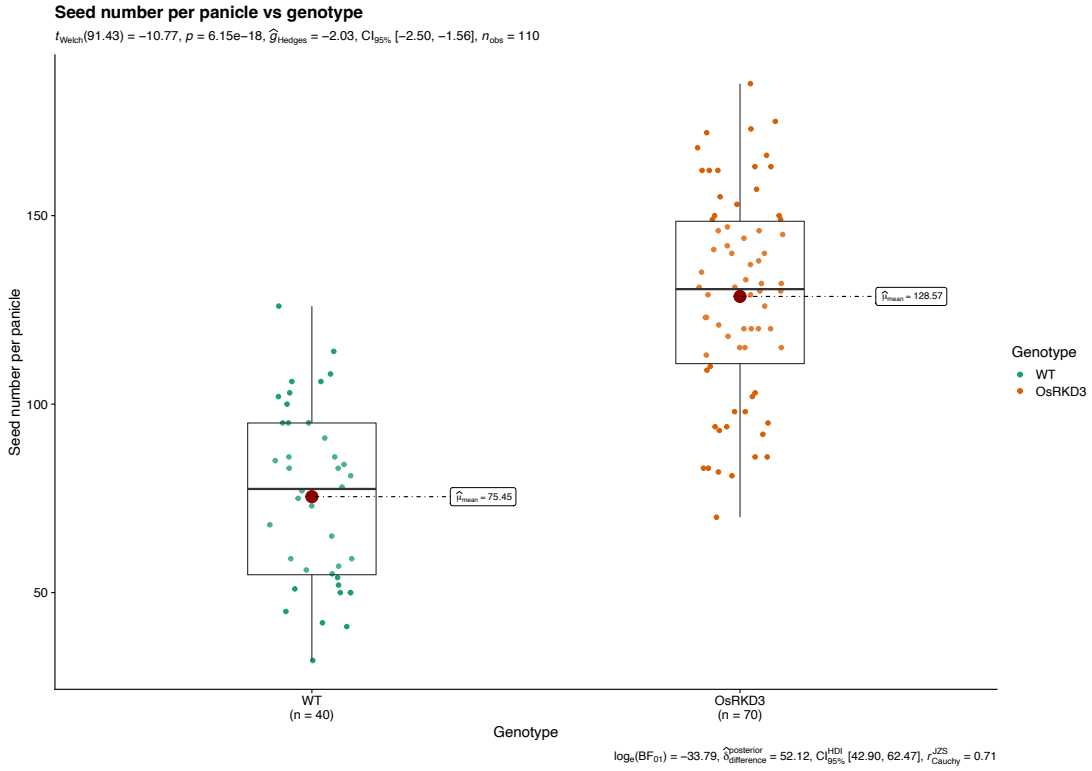
Supplementary Figure 5.1. Increased days to maturity in an *indOsRKD3 Cempo Ireng* T2 line. Days to senescence was measured from germination date to when half of the panicles had senesced. T3 lines were merged for more statistically robust tests. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.



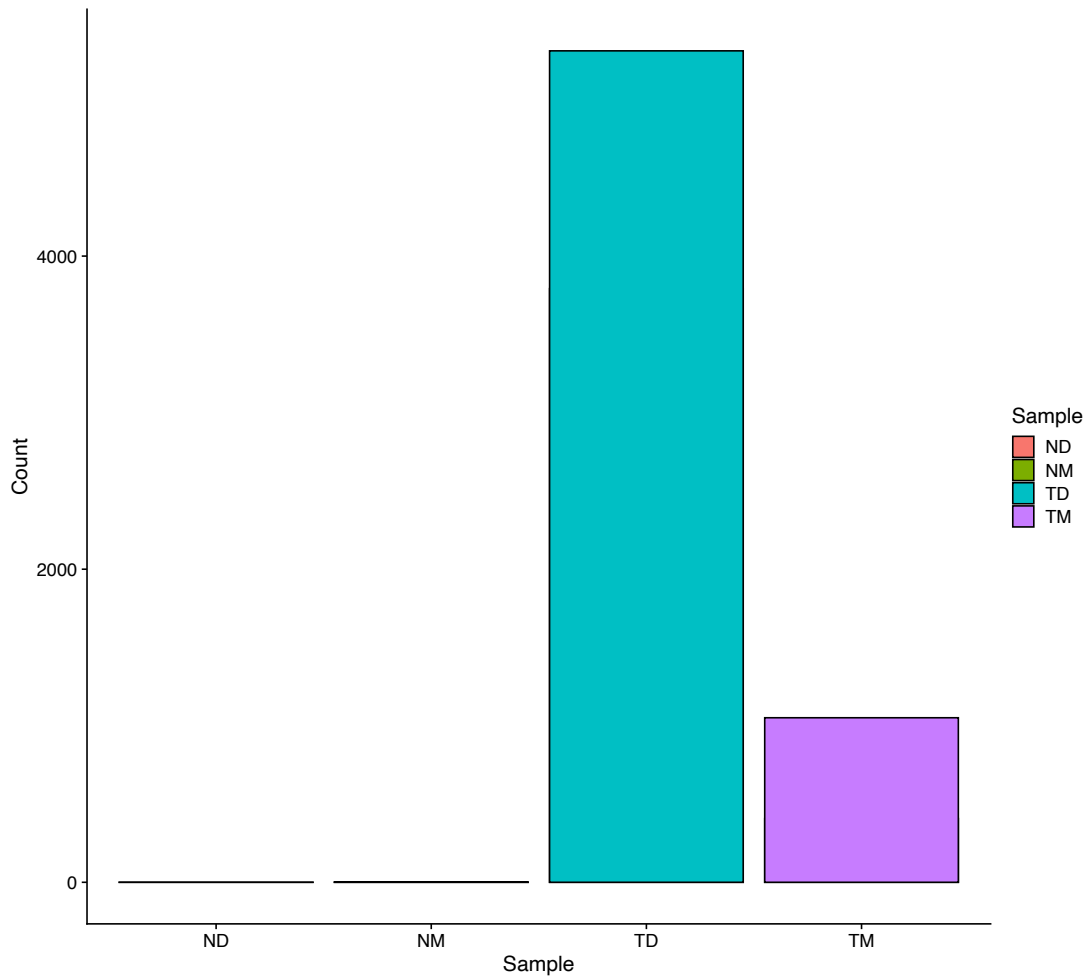
Supplementary Figure 5.2. Increased height in an *indOsRKD3 Cempo Ireng* T2 line. Height measured from the base of the plant to the tallest spike at senescence. T3 lines were merged for more statistically robust tests. Sample number of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.



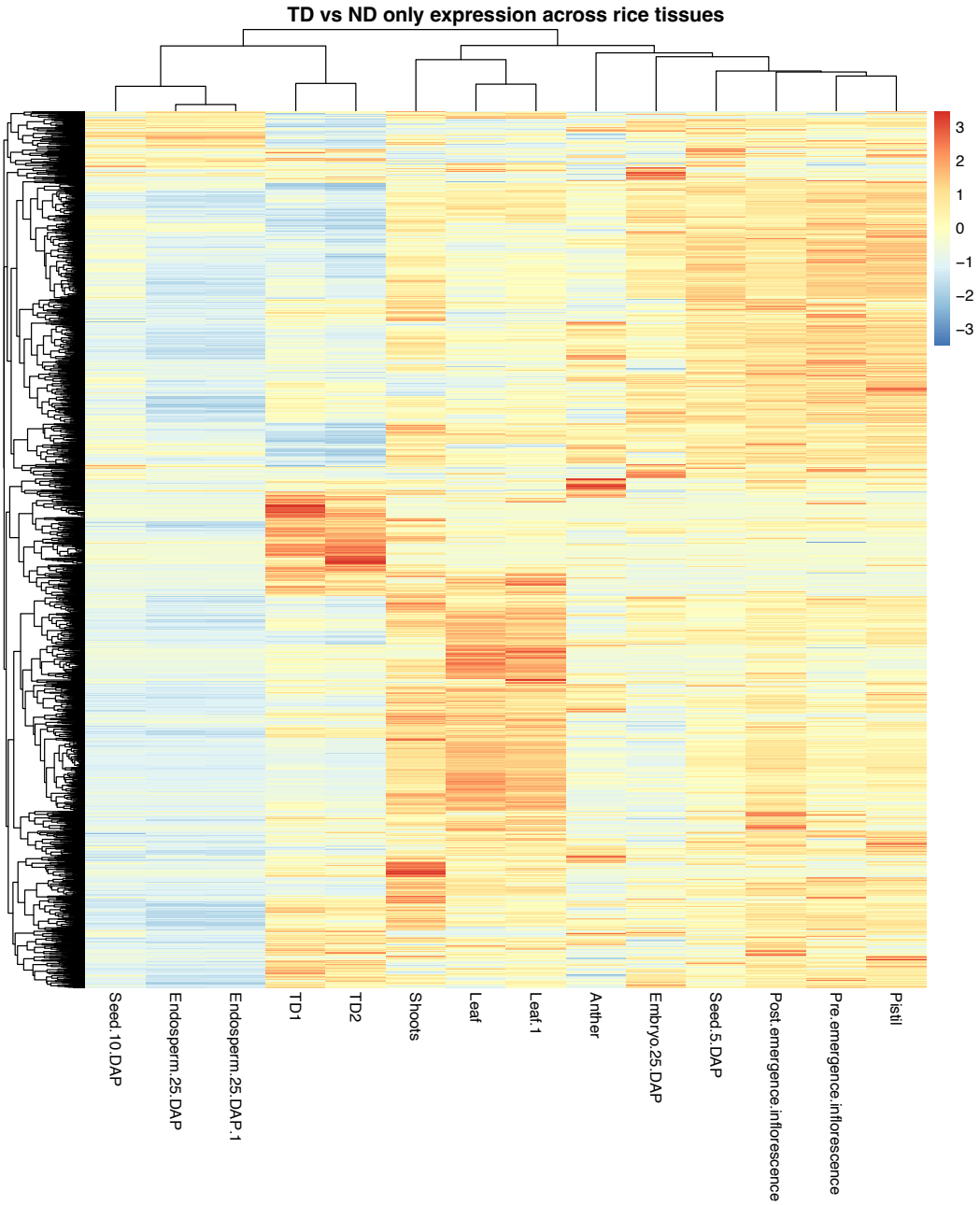
Supplementary figure 5.3. Decreased seed weight per 1000 seeds in an *indOsRKD3 Cempo Ireng* T2 line. The seed weight of the 1000 seeds was measured after threshing fully dried panicles. T3 lines were merged for more statistically robust tests. Where 1000 seeds were not available, this value was extrapolated. The label of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.



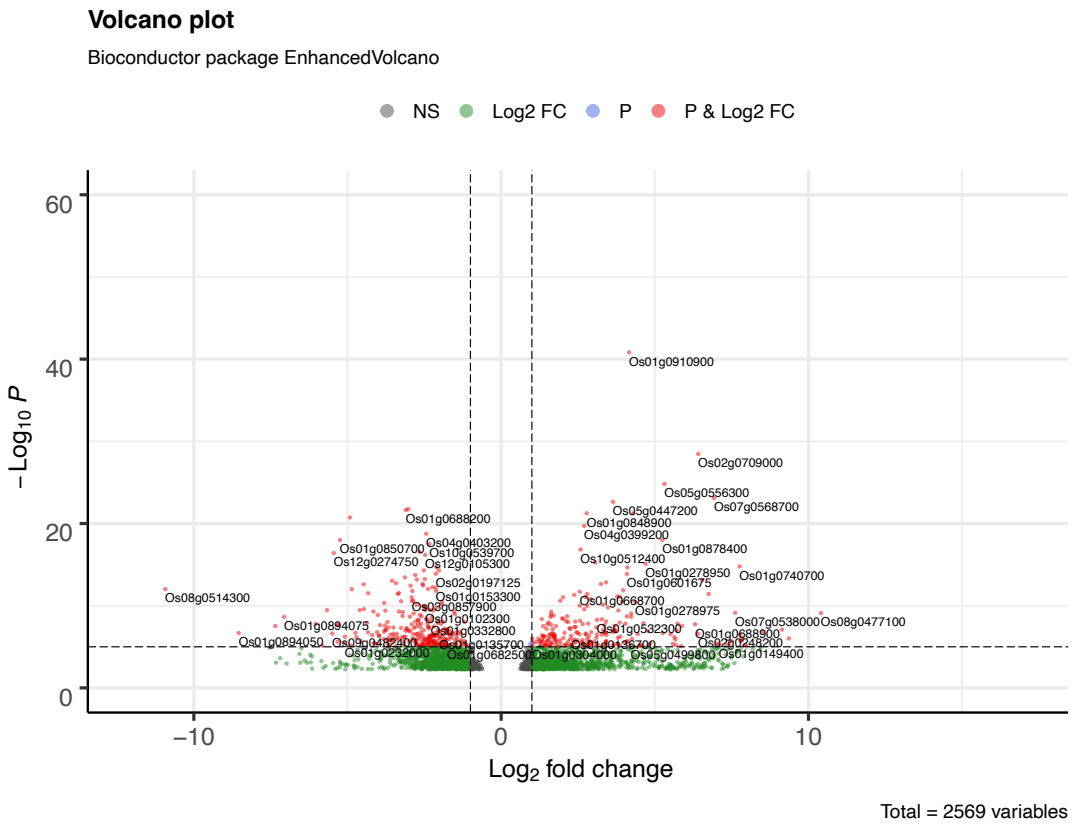
Supplementary figure 5.4. Increased seed number per panicle in an *indOsRKD3 Cempo Ireng* T2 line. The seed number per panicle was measured after threshing fully dried panicles separately. T3 lines were merged for more statistically robust tests. The label of each transgenic line is described in the (n=x) brackets below the label. The mean value is given by μ in the centre of each boxplot and sample points are highlighted in their condition's respective colour. Details of tests used for statistical analysis is on the bottom right of the plot. Significant difference between conditions is indicated by the p value in the top left of the plot. Additional statistics can be found on the top left of the plot.



Supplementary figure 5.5. *indOsRKD3* expression in the *indOsRKD3* leaf RNAseq experiment shows high expression amongst transgenic samples. TD = transgenic DEX, TM = transgenic mock, ND = non-transgenic DEX and NM = non-transgenic mock. Bar colour denotes sample condition as described in the legend on the right.



Supplementary figure 5.6. Heatmap of TD vs ND only DEGs from *indOsRKD3 Cempo Ireng* leaf RNAseq experiment across tissues. TD = Transgenic DEX leaf and ND = non-transgenic DEX leaf. Tissue tpm values were taken from the MSU database. Scale displayed is row scaled logarithmically transformed tpm values. Hierarchical clustering was performed on *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs and is displayed as a dendrogram on the side of the heatmap. Hierarchical clustering was also performed on samples and is displayed as a dendrogram on the top of the heatmap.



Supplemental figure 5.7. Volcano plot of the *IndOsRKD3 Cempo Ireng* TD vs ND only leaf DEGs equal shows up and downregulation. 51% upregulated DEGs, 49% downregulated DEGs. Genes included in the *IndOsRKD3 Cempo Ireng* TD vs ND only pairwise comparison are plotted across their $\log_2$ fold change and -$\log_{10}$ p values. DEGs are labelled depending on spatial distribution. Red points represent genes that show sufficient $\log_2$ fold change p value to be annotated. Blue points represent genes that show sufficient p value but not $\log_2$ fold change to be annotated. Green points represent genes that show sufficient $\log_2$ fold change but not p value to be annotated. Grey points represent genes that show neither sufficient $\log_2$ fold change or p value to be annotated.

References

Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C. and Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome biology*, *12*(2), pp.1-14.

Allan, R.E., Vogel, O.A. and Craddock Jr, J.C., 1959. Comparative response to gibberellic acid of dwarf, semidwarf, and standard short and tall winter wheat varieties 1. Agronomy Journal, 51(12), pp.737-740.

Amin, M.R., Hamid, A., Choudhury, R.U., Raquibullah, S.M. and Asaduzzaman, M., 2006. Nitrogen fertilizer effect on tillering, dry matter production and yield of traditional varieties of rice. Intl. J. Suatain. Crop Prod, 1(1), pp.17-20.

Andaya, V. C., and T. H. Tai. "Fine mapping of the qCTS12 locus, a major QTL for seedling cold tolerance in rice." Theoretical and Applied Genetics 113, no. 3 (2006): 467-475.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. and Liu, D.R., 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature, 576(7785), pp.149-157.

Apridamayanti, P., Pratiwi, R., Purwestri, Y.A., Tunjung, W.A.S. and Rumiyati, R., 2017. Anthocyanin, nutrient contents, and antioxidant activity of black rice bran of Oryza sativa L.'Cempo Ireng'from Sleman, Yogyakarta, Indonesia. Indonesian Journal of Biotechnology, 22(1), pp.49-54.

Arite, T., Umehara, M., Ishikawa, S., Hanada, A., Maekawa, M., Yamaguchi, S. and Kyozuka, J., 2009. d14, a strigolactone-insensitive mutant of rice, shows an accelerated outgrowth of tillers. Plant and Cell Physiology, 50(8), pp.1416-1424.

Atkinson, J.A., Pound, M.P., Bennett, M.J. and Wells, D.M., 2019. Uncovering the hidden half of plants using new advances in root phenotyping. Current opinion in biotechnology, 55, pp.1-8.

Babb, S. and Muehlbauer, G., 2003. Genetic and morphological characterization of the barley uniculm2 (cul2) mutant. Theoretical and Applied Genetics, 106(5), pp.846-857.

Baccarella, A., Williams, C.R., Parrish, J.Z. and Kim, C.C., 2018. Empirical assessment of the impact of sample number and read depth on RNA-Seq analysis workflow performance. BMC bioinformatics, 19(1), pp.1-12.

Baedke, J., 2013. The epigenetic landscape in the course of time: Conrad Hal Waddington's methodological impact on the life sciences. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *44*(4), pp.756-773.

Baroux, C., Raissig, M.T. and Grossniklaus, U., 2011. Epigenetic regulation and reprogramming during gamete formation in plants. Current opinion in genetics & development, 21(2), pp.124-133.

Bartlett, A., O'Malley, R.C., Huang, S.S.C., Galli, M., Nery, J.R., Gallavotti, A. and Ecker, J.R., 2017. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature protocols*, *12*(8), p.1659.

Becker-Andre, M. and Hahlbrock, K., 1989. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). Nucleic acids research, 17(22), pp.9437-9446.

Benke, K. and Tomkins, B., 2017. Future food-production systems: vertical farming and controlled-environment agriculture. Sustainability: Science, Practice and Policy, 13(1), pp.13-26.

Bhattacharya, A., Kourmpetli, S. and Davey, M.R., 2010. Practical applications of manipulating plant architecture by regulating gibberellin metabolism. Journal of plant growth regulation, 29(2), pp.249-256.

Blighe K, Rana S, Lewis M (2020). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.8.0,

Boden, S.A., Cavanagh, C., Cullis, B.R., Ramm, K., Greenwood, J., Finnegan, E.J., Trevaskis, B. and Swain, S.M., 2015. Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat. Nature Plants, 1(2), pp.1-6.

Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), pp.2114-2120.

Booker, J., Auldridge, M., Wills, S., McCarty, D., Klee, H. and Leyser, O., 2004. MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule. Current biology, 14(14), pp.1232-1238.

Borrill, P., Harrington, S.A., Simmonds, J. and Uauy, C., 2019. Identification of transcription factors regulating senescence in wheat through gene regulatory network modelling. Plant Physiology, 180(3), pp.1740-1755.

Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M., 1989. Genes directing flower development in Arabidopsis. The Plant Cell, 1(1), pp.37-52.

Brand, L., Hörler, M., Nüesch, E., Vassalli, S., Barrell, P., Yang, W., ... & Curtis, M. D. (2006). A versatile and reliable two-component system for tissue-specific gene induction in Arabidopsis. Plant physiology, 141(4), 1194-1204.

Brassac, J., Muqaddasi, Q.H., Plieske, J., Ganal, M.W. and Röder, M.S., 2021. Linkage mapping identifies a non-synonymous mutation in FLOWERING LOCUS T (FT-B1) increasing spikelet number per spike. Scientific reports, 11(1), pp.1-12.

Bray, N., Pimentel, H., Melsted, P. and Pachter, L., 2015. Near-optimal RNA-Seq quantification. arXiv preprint arXiv:1505.02710.

Broughton, W.J., 1970. Methods in legume-rhizobium technology: plant nutrient solutions. Handbook for rhizobia, pp.245-249.

Chai, R., Ye, X., Ma, C., Wang, Q., Tu, R., Zhang, L. and Gao, H., 2019. Greenhouse gas emissions from synthetic nitrogen manufacture and fertilization for main upland crops in China. Carbon balance and management, 14(1), pp.1-10.

Chalupska, D., Lee, H.Y., Faris, J.D., Evrard, A., Chalhoub, B., Haselkorn, R. and Gornicki, P., 2008. Acc homoeoloci and the evolution of wheat genomes. Proceedings of the National Academy of Sciences, 105(28), pp.9691-9696.

Chardin, C., Girin, T., Roudier, F., Meyer, C. and Krapp, A., 2014. The plant RWP-RK transcription factors: key regulators of nitrogen responses and of gametophyte development. Journal of experimental botany, 65(19), pp.5577-5587.

Chen, L., Zhao, Y., Xu, S., Zhang, Z., Xu, Y., Zhang, J. and Chong, K., 2018. Os MADS 57 together with Os TB 1 coordinates transcription of its target Os WRKY 94 and D14 to switch its organogenesis to defense for cold adaptation in rice. New Phytologist, 218(1), pp.219-231.

Chen, S., Zhou, Y., Chen, Y. and Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), pp.i884-i890.

Cheng, C., Motohashi, R., Tsuchimoto, S., Fukuta, Y., Ohtsubo, H. and Ohtsubo, E., 2003. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Molecular Biology and Evolution, 20(1), pp.67-75.

Chung, H.S. and Howe, G.A., 2009. A critical role for the TIFY motif in repression of jasmonate signaling by a stabilized splice variant of the JASMONATE ZIM-domain protein JAZ10 in Arabidopsis. *The Plant Cell*, *21*(1), pp.131-145.

Civán, Peter, Hayley Craig, Cymon J. Cox, and Terence A. Brown. "Multiple domestications of Asian rice." Nature plants 2, no. 4 (2016): 1-1.

Claeys, H., Vi, S.L., Xu, X., Satoh-Nagasawa, N., Eveland, A.L., Goldshmidt, A., Feil, R., Beggs, G.A., Sakai, H., Brennan, R.G. and Lunn, J.E., 2019. Control of meristem determinacy by trehalose 6-phosphate phosphatases is uncoupled from enzymatic activity. Nature plants, 5(4), pp.352-357.

Clark, L.G., Zhang, W. and Wendel, J.F., 1995. A phylogeny of the grass family (Poaceae) based on ndhF sequence data. Systematic Botany, pp.436-460.

Clayton, W.D. and Renvoize, S.A., 1986. Genera graminum. Grasses of the world. Genera graminum. Grasses of the World., 13.

Colgan, D.F. and Manley, J.L., 1997. Mechanism and regulation of mRNA polyadenylation. Genes & development, 11(21), pp.2755-2766.

Concia, L., Veluchamy, A., Ramirez-Prado, J.S., Martin-Ramirez, A., Huang, Y., Perez, M., Domenichini, S., Granados, N.Y.R., Kim, S., Blein, T. and Duncan, S., 2020. Wheat chromatin architecture is organized in genome territories and transcription factories. Genome biology, 21(1), pp.1-20.

Conway, G., 2019. The doubly green revolution: food for all in the twenty-first century. Cornell University Press.

Dahlgren, R.M.T., Clifford, H.T. and Yeo, F.T., 1985. The Families of the Monocotyledons Springer Verlag. Berlin, Heidelberg, New York, Tokyo.

de Matos Simoes, R., Emmert-Streib, F., de Matos Simoes, M.R. and Emmert-Streib, B.S.N.I., 2012. Package 'bc3net'

de Mendiburu, F. and de Mendiburu, M.F., 2019. Package 'agricolae'. R Package, Version, pp.1-2.

Debernardi, J.M., Lin, H., Chuck, G., Faris, J.D. and Dubcovsky, J., 2017. microRNA172 plays a crucial role in wheat spike morphogenesis and grain threshability. Development, 144(11), pp.1966-1975.

Debernardi, J.M., Tricoli, D.M., Ercoli, M.F., Hayta, S., Ronald, P., Palatnik, J.F. and Dubcovsky, J., 2020. A GRF–GIF chimeric protein improves the regeneration efficiency of transgenic plants. Nature Biotechnology, 38(11), pp.1274-1279.

Dixon, L.E., Greenwood, J.R., Bencivenga, S., Zhang, P., Cockram, J., Mellers, G., Ramm, K., Cavanagh, C., Swain, S.M. and Boden, S.A., 2018. TEOSINTE BRANCHED1 regulates inflorescence architecture and development in bread wheat (Triticum aestivum). The Plant Cell, 30(3), pp.563-581.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), pp.15-21.

Doebley, J., Stec, A. and Gustus, C., 1995. teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics, 141(1), pp.333-346.

Doebley, J., Stec, A. and Hubbard, L., 1997. The evolution of apical dominance in maize. Nature, 386(6624), pp.485-488.

Dong, Z., Xiao, Y., Govindarajulu, R., Feil, R., Siddoway, M.L., Nielsen, T., Lunn, J.E., Hawkins, J., Whipple, C. and Chuck, G., 2019. The regulatory landscape of a core maize domestication module controlling bud dormancy and growth repression. Nature communications, 10(1), pp.1-15.

Doyle, J.J. and Doyle, J.L., 1990. Isolation ofplant DNA from fresh tissue. Focus, 12(13), pp.39-40.

Dubcovsky, J. and Dvorak, J., 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. Science, 316(5833), pp.1862-1866.

Duggan, B.L., Richards, R.A., Van Herwaarden, A.F. and Fettell, N.A., 2005. Agronomic evaluation of a tiller inhibition gene (tin) in wheat. I. Effect on yield, yield components, and grain protein. Australian Journal of Agricultural Research, 56(2), pp.169-178.

Dyer, N.P., Shahrezaei, V. and Hebenstreit, D., 2019. LiBiNorm: an htseq-count analogue with improved normalisation of Smart-seq2 data and library preparation diagnostics. PeerJ, 7, p.e6222.

Edwards, E.J. and Smith, S.A., 2010. Phylogenetic analyses reveal the shady history of C4 grasses. Proceedings of the National Academy of Sciences, 107(6), pp.2532-2537.

Evans, L.T. and Fischer, R.A., 1999. Yield potential: its definition, measurement, and significance. Crop science, 39(6), pp.1544-1551.

FAO, FAOSTAT, 2018. In http://faostat.fao.org/site/291/default.aspx, Food and Agriculture Organization: 2018.

Feekes, Willem (1941). "De tarwe en haar milieu [Wheat and its environment]". Verslagen van de Technische Tarwe Commissie. (in Dutch and English). 17: 523–888.

Feldman, M. and Millet, E., 2001. The contribution of the discovery of wild emmer to an understanding of wheat evolution and domestication and to wheat improvement. Israel Journal of Plant Sciences, 49(sup1), pp.25-36.

Fennell, S., Bohorova, N., Van Ginkel, M., Crossa, J., & Hoisington, D. (1996). Plant regeneration from immature embryos of 48 elite CIMMYT bread wheats. Theoretical and Applied Genetics, 92(2), 163–169. https://doi.org/10.1007/BF00223371

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T. and Pasini, D., 2014. Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Molecular cell, 53(1), pp.49-62.

Flintham, J.E., Börner, A., Worland, A.J. and Gale, M.D., 1997. Optimizing wheat grain yield: effects of Rht (gibberellin-insensitive) dwarfing genes. The Journal of Agricultural Science, 128(1), pp.11-25.

Fobert, P.R., Coen, E.S., Murphy, G.J. and Doonan, J.H., 1994. Patterns of cell division revealed by transcriptional regulation of genes during the cell cycle in plants. The EMBO journal, 13(3), pp.616-624.

Garg, P. and Jaiswal, P., 2016. Databases and bioinformatics tools for rice research. Current Plant Biology, 7, pp.39-52.

Gaurav, K., Arora, S., Silva, P., Sánchez-Martín, J., Horsnell, R., Gao, L., Brar, G.S., Widrig, V., John Raupp, W., Singh, N. and Wu, S., 2021. Population genomic analysis of Aegilops tauschii identifies targets for bread wheat improvement. Nature biotechnology, pp.1-10.

Genaev, M.A., Komyshev, E.G., Smirnov, N.V., Kruchinina, Y.V., Goncharov, N.P. and Afonnikov, D.A., 2019. Morphometry of the wheat spike by analyzing 2D images. Agronomy, 9(7), p.390.

Glenn, P., Zhang, J., Brown-Guedira, G., DeWitt, N., Cook, J.P., Li, K., Akhunov, E. and Dubcovsky, J., 2021. Identification and characterization of a natural polymorphism in FT-A2 associated with increased number of grains per spike in wheat. Theoretical and Applied Genetics, pp.1-14.

Gooding, M.J., Addisu, M., Uppal, R.K., Snape, J.W. and Jones, H.E., 2012. Effect of wheat dwarfing genes on nitrogen-use efficiency. The Journal of Agricultural Science, 150(1), pp.3-22.

Greenwood, J.R., Finnegan, E.J., Watanabe, N., Trevaskis, B. and Swain, S.M., 2017. New alleles of the wheat domestication gene Q reveal multiple roles in growth and reproductive development. Development, 144(11), pp.1959-1965.

Guo, Y., Mahony, S. and Gifford, D.K., 2012. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. PLoS Comput Biol, 8(8), p.e1002638.

Guo, Z., Zhao, Y., Röder, M.S., Reif, J.C., Ganal, M.W., Chen, D. and Schnurbusch, T., 2018. Manipulation and prediction of spike morphology traits for the improvement of grain yield in wheat. Scientific reports, 8(1), pp.1-10.

Hamada, H., Linghu, Q., Nagira, Y., Miki, R., & Taoka, N. (2017). An in planta biolistic method for stable wheat transformation. Scientific Reports, 2–9. https://doi.org/10.1038/s41598-017-11936-0

Hawkesford, M.J., 2014. Reducing the reliance on nitrogen fertilizer for wheat production. Journal of cereal science, 59(3), pp.276-283.

Hedden, P., 2003. The genes of the Green Revolution. TRENDS in Genetics, 19(1), pp.5-9.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell, 38(4), pp.576-589.

Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B. and Salamini, F., 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. Science, 278(5341), pp.1312-1314.

Heyser, J.W., Dykes, T.A., DeMott, K.J. and Nabors, M.W., 1983. High frequency, long term regeneration of rice from callus culture. Plant Science Letters, 29(2-3), pp.175-182.

Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T., 1994. Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. The Plant Journal, 6(2), pp.271-282.

Hilbert, L., Neves, E.G., Pugliese, F., Whitney, B.S., Shock, M., Veasey, E., Zimpel, C.A. and Iriarte, J., 2017. Evidence for mid-Holocene rice domestication in the Americas. Nature ecology & evolution, 1(11), pp.1693-1698.

Hoogendoorn, J., Rickson, J.M. and Gale, M.D., 1990. Differences in leaf and stem anatomy related to plant height of tall and dwarf wheat (Triticum aestivum L.). Journal of Plant Physiology, 136(1), pp.72-77.

Horikoshi, M., Yamamoto, T., Ohkuma, Y., Weil, P.A. and Roeder, R.G., 1990. Analysis of structure-function relationships of yeast TATA box binding factor TFIID. Cell, 61(7), pp.1171-1178.

https://www.cerealsdb.uk.net/cerealgenomics/WheatBP/Documents/DOC_Evolution.php

https://www.worldbank.org/en/data/interactive/2020/11/11/covid-19-high-frequency-monitoring-dashboard

Hu, H., Scheben, A. and Edwards, D., 2018. Advances in integrating genomics and bioinformatics in the plant breeding pipeline. Agriculture, 8(6), p.75.

Huang, X., Kurata, N., Wang, Z.X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y. and Lu, Y., 2012. A map of rice genome variation reveals the origin of cultivated rice. Nature, 490(7421), pp.497-501.

Iida, H., 2014. Mugifumi, a beneficial farm work of adding mechanical stress by treading to wheat and barley seedlings. Frontiers in plant science, 5, p.453.

IWGSC, A.R., Eversole, K., Feuillet, C., Keller, B., Rogers, J. and Stein, N., 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science, 361(6403), pp.eaar7191-eaar7191.

Jain, N., Arora, P., Tomer, R., Mishra, S.V., Bhatia, A., Pathak, H., Chakraborty, D., Kumar, V., Dubey, D.S., Harit, R.C. and Singh, J.P., 2016. Greenhouse gases emission from soils under major crops in Northwest India. Science of the Total Environment, 542, pp.551-561.

Jantasuriyarat, C., Vales, M.I., Watson, C.J.W. and Riera-Lizarazu, O., 2004. Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (Triticum aestivum L.). Theoretical and Applied Genetics, 108(2), pp.261-273.

Jeong, S., Palmer, T.M. and Lukowitz, W., 2011. The RWP-RK factor GROUNDED promotes embryonic polarity by facilitating YODA MAP kinase signaling. Current Biology, 21(15), pp.1268-1276.

Jia, M., Guan, J., Zhai, Z., Geng, S., Zhang, X., Mao, L. and Li, A., 2018. Wheat functional genomics in the era of next generation sequencing: an update. The Crop Journal, 6(1), pp.7-14.

Jobson, E.M., Johnston, R.E., Oiestad, A.J., Martin, J.M. and Giroux, M.J., 2019. The impact of the wheat Rht-B1b semi-dwarfing allele on photosynthesis and seed development under field conditions. Frontiers in plant science, 10, p.51.

Kaneko-Suzuki, M., Kurihara-Ishikawa, R., Okushita-Terakawa, C., Kojima, C., Nagano-Fujiwara, M., Ohki, I., Tsuji, H., Shimamoto, K. and Taoka, K.I., 2018. TFL1-like proteins in rice antagonize rice FT-like protein in inflorescence development by competition for complex formation with 14-3-3 and FD. Plant and Cell Physiology, 59(3), pp.458-468.

Kebrom, T.H., Chandler, P.M., Swain, S.M., King, R.W., Richards, R.A. and Spielmeyer, W., 2012. Inhibition of tiller bud outgrowth in the tin mutant of wheat is associated with precocious internode development. Plant Physiology, 160(1), pp.308-318.

Kellogg, E.A. and Linder, H.P., 1995. Phylogeny of Poales.

Kellogg, E.A., 2000. The grasses: a case study in macroevolution. Annual review of ecology and systematics, 31(1), pp.217-238.

Khush, G., 2003. Productivity improvements in rice. Nutrition reviews, 61(suppl_6), pp.S114-S116.

Khush, G.S., 2001. Green revolution: the way forward. Nature reviews genetics, 2(10), pp.815-822.

Kim, D., Langmead, B. and Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. Nature methods, 12(4), pp.357-360.

Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology, 37(8), pp.907-915.

Kobayashi, K., Yasuno, N., Sato, Y., Yoda, M., Yamazaki, R., Kimizu, M., Yoshida, H., Nagamura, Y. and Kyozuka, J., 2012. Inflorescence meristem identity in rice is specified by overlapping functions of three AP1/FUL-like MADS box genes and PAP2, a SEPALLATA MADS box gene. *The Plant Cell*, *24*(5), pp.1848-1859.

Koi, S., Hisanaga, T., Sato, K., Shimamura, M., Yamato, K.T., Ishizaki, K., Kohchi, T. and Nakajima, K., 2016. An evolutionarily conserved plant RKD factor controls germ cell differentiation. Current Biology, 26(13), pp.1775-1781.

Kolde, R. and Kolde, M.R., 2015. Package 'pheatmap'. R package, 1(7), p.790.

Kolpakov, F.A., Ananko, E.A., Kolesov, G.B. and Kolchanov, N.A., 1998. GeneNet: a gene network database and its automated visualization. Bioinformatics (Oxford, England), 14(6), pp.529-537.

Kőszegi, D., 2008. RKD genes: a novel transcription factor family involved in the female gametophyte development of Arabidopsis and wheat.

Kőszegi, D., Johnston, A.J., Rutten, T., Czihal, A., Altschmied, L., Kumlehn, J., Wüst, S.E., Kirioukhova, O., Gheyselinck, J., Grossniklaus, U. and Bäumlein, H., 2011. Members of the

RKD transcription factor family induce an egg cell-like gene expression program. The Plant Journal, 67(2), pp.280-291.

Krizek, B.A. and Fletcher, J.C., 2005. Molecular mechanisms of flower development: an armchair guide. Nature Reviews Genetics, 6(9), pp.688-698.

Kucera, M., Isserlin, R., Arkhangorodsky, A. and Bader, G.D., 2016. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. F1000Research, 5.

Kumar, A., Batra, R., Gahlaut, V., Gautam, T., Kumar, S., Sharma, M., Tyagi, S., Singh, K.P., Balyan, H.S., Pandey, R. and Gupta, P.K., 2018. Genome-wide identification and characterization of gene family for RWP-RK transcription factors in wheat (Triticum aestivum L.). PloS one, 13(12), p.e0208409.

KURNIASIH, N.S., SUSANDARINI, R., SUSANTO, F.A., NURINGTYAS, T.R., JENKINS, G. and PURWESTRI, Y.A., 2019. Characterization of Indonesian pigmented rice (Oryza sativa) based on morphology and Single Nucleotide Polymorphisms. Biodiversitas Journal of Biological Diversity, 20(4), pp.1208-1214.

Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J.A., Brown-Guedira, G., DeWitt, N. and Haugrud, A.P., 2019. Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. Theoretical and Applied Genetics, 132(9), pp.2689-2705.

Langfelder, P. and Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics, 9(1), pp.1-13.

Langmead, B., 2010. Aligning short sequencing reads with Bowtie. Current protocols in bioinformatics, 32(1), pp.11-7.

Lawrence-Dill, C.J., Schnable, P.S. and Springer, N.M., 2019. Idea factory: the maize genomes to fields initiative. Crop Science, 59(4), pp.1406-1410.

Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D. and Lewis, B.A., 2005. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Molecular and cellular biology, 25(21), pp.9674-9686.

Lee, S., Jeon, J.S., Jung, K.H. and An, G., 1999. Binary vectors for efficient transformation of rice. Journal of Plant Biology, 42(4), pp.310-316.

Lewis, J.M., Mackintosh, C.A., Shin, S., Gilding, E., Kravchenko, S., Baldridge, G., Zeyen, R. and Muehlbauer, G.J., 2008. Overexpression of the maize Teosinte Branched1 gene in wheat suppresses tiller development. Plant cell reports, 27(7), pp.1217-1225.

Li, C. and Dubcovsky, J., 2008. Wheat FT protein regulates VRN1 transcription through interactions with FDL2. The Plant Journal, 55(4), pp.543-554.

Li, C., Lin, H., Chen, A., Lau, M., Jernstedt, J. and Dubcovsky, J., 2019. Wheat VRN1, FUL2 and FUL3 play critical and redundant roles in spikelet development and spike determinacy. Development, 146(14).

Li, S., Tian, Y., Wu, K., Ye, Y., Yu, J., Zhang, J., Liu, Q., Hu, M., Li, H., Tong, Y. and Harberd, N.P., 2018. Modulating plant growth–metabolism coordination for sustainable agriculture. Nature, 560(7720), pp.595-600.

Li, X., Qian, Q., Fu, Z., Wang, Y., Xiong, G., Zeng, D., Wang, X., Liu, X., Teng, S., Hiroshi, F. and Yuan, M., 2003. Control of tillering in rice. Nature, 422(6932), pp.618-621.

Li, Y., Xiao, J., Chen, L., Huang, X., Cheng, Z., Han, B., Zhang, Q. and Wu, C., 2018. Rice functional genomics research: past decade and future. Molecular plant, 11(3), pp.359-380.

Lin, H., Arrivault, S., Coe, R.A., Karki, S., Covshoff, S., Bagunu, E., Lunn, J.E., Stitt, M., Furbank, R.T., Hibberd, J.M. and Quick, W.P., 2020. A partial C4 photosynthetic biochemical pathway in rice. Frontiers in plant science, 11.

Lin, H., Wang, R., Qian, Q., Yan, M., Meng, X., Fu, Z., Yan, C., Jiang, B., Su, Z., Li, J. and Wang, Y., 2009. DWARF27, an iron-containing protein required for the biosynthesis of strigolactones, regulates rice tiller bud outgrowth. The Plant Cell, 21(5), pp.1512-1525.

Lin, J.S., Li, X., Luo, Z., Mysore, K.S., Wen, J. and Xie, F., 2018. NIN interacts with NLPs to mediate nitrate inhibition of nodulation in Medicago truncatula. Nature plants, 4(11), pp.942-952.

Liu, C., Xi, W., Shen, L., Tan, C. and Yu, H., 2009. Regulation of floral patterning by flowering time genes. Developmental cell, 16(5), pp.711-722.

Liu, H., Song, S. and Xing, Y., 2019. Beyond heading time: FT-like genes and spike development in cereals. Journal of experimental botany, 70(1), p.1.

Liu, L., Gallagher, J., Arevalo, E.D., Chen, R., Skopelitis, T., Wu, Q., Bartlett, M. and Jackson, D., 2021. Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. Nature Plants, 7(3), pp.287-294.

Liu, L., Gallagher, J., Arevalo, E.D., Chen, R., Skopelitis, T., Wu, Q., Bartlett, M. and Jackson, D., 2021. Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. Nature Plants, 7(3), pp.287-294.

Liu, Y., Gu, D., Ding, Y., Wang, Q., Li, G. and Wang, S., 2011. The relationship between nitrogen, auxin and cytokinin in the growth regulation of rice ('Oryza sativa'l.) Tiller buds. Australian journal of crop science, 5(8), pp.1019-1026.

Liu, Y., Zhang, L., Melzer, M., Wang, Z., Sun, Z., Schnurbusch, T. and Guo, Z., 2020. Ppd-1 Remodels Spike Architecture by Regulating Floral Development in wheat. bioRxiv.

Lo, S.F., Yang, S.Y., Chen, K.T., Hsing, Y.I., Zeevaart, J.A., Chen, L.J. and Yu, S.M., 2008. A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. The Plant Cell, 20(10), pp.2603-2618.

Londo, J.P., Chiang, Y.C., Hung, K.H., Chiang, T.Y. and Schaal, B.A., 2006. Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa. Proceedings of the National Academy of Sciences, 103(25), pp.9578-9583.

Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 15(12), p.550.

Luo, L., Pan, S., Liu, X., Wang, H. and Xu, G., 2017. Nitrogen deficiency inhibits cell division-determined elongation, but not initiation, of rice tiller buds. Israel Journal of Plant Sciences, 64(3-4), pp.32-40.

Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D. and Woolf, P.J., 2009. GAGE: generally applicable gene set enrichment for pathway analysis. BMC bioinformatics, 10(1), pp.1-17.

Ma, L., Ma, S.W., Deng, Q., Yuan, Y., Wei, Z., Jia, H. and Ma, Z., 2018. Identification of wheat inflorescence development-related genes using a comparative transcriptomics approach. *International journal of genomics*, *2018*.

Mäkelä, P. and Muurinen, S., 2012. Uniculm and conventional tillering barley accessions under northern growing conditions. The Journal of Agricultural Science, 150(3), pp.335-344.

Masruroh, F., Samanhudi, S. and Yunus, A., 2016. Improvement of rice (Oryza sativa L.) var. Ciherang and Cempo Ireng productivity using gamma irradiation. Journal of Agricultural Science and Technology B, 6, pp.289-294.

McLeay, R.C. and Bailey, T.L., 2010. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC bioinformatics, 11(1), pp.1-11.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L., Tosatto, S.C., Paladin, L., Raj, S., Richardson, L.J. and Finn, R.D., 2021. Pfam: The protein families database in 2021. Nucleic Acids Research, 49(D1), pp.D412-D419.

Miura, K., Ikeda, M., Matsubara, A., Song, X.J., Ito, M., Asano, K., Matsuoka, M., Kitano, H. and Ashikari, M., 2010. OsSPL14 promotes panicle branching and higher grain productivity in rice. Nature genetics, 42(6), pp.545-549.

Mizoi, J., Shinozaki, K. and Yamaguchi-Shinozaki, K., 2012. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1819*(2), pp.86-96.

Mizuta, Y., Harushima, Y. and Kurata, N., 2010. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. Proceedings of the National Academy of Sciences, 107(47), pp.20417-20422.

Mouradov, A., Cremer, F. and Coupland, G., 2002. Control of flowering time: interacting pathways as a basis for diversity. The Plant Cell, 14(suppl 1), pp.S111-S130.

Muqaddasi, Q.H., Brassac, J., Koppolu, R., Plieske, J., Ganal, M.W. and Röder, M.S., 2019. TaAPO-A1, an ortholog of rice ABERRANT PANICLE ORGANIZATION 1, is associated with total spikelet number per spike in elite European hexaploid winter wheat (Triticum aestivum L.) varieties. Scientific reports, 9(1), pp.1-12.

Murai, K., 2013. Homeotic genes and the ABCDE model for floral organ formation in wheat. Plants, 2(3), pp.379-395.

Nalam, V.J., Vales, M.I., Watson, C.J., Kianian, S.F. and Riera-Lizarazu, O., 2006. Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (Triticum turgidum L.). Theoretical and Applied Genetics, 112(2), pp.373-381.

Navarro, L., Bari, R., Achard, P., Lisón, P., Nemri, A., Harberd, N.P. and Jones, J.D., 2008. DELLAs control plant immune responses by modulating the balance of jasmonic acid and salicylic acid signaling. Current Biology, 18(9), pp.650-655.

Nesbitt, M., 1998. Where was einkorn wheat domesticated?. Trends in Plant Science (United Kingdom).

Nic-Can, G.I., López-Torres, A., Barredo-Pool, F., Wrobel, K., Loyola-Vargas, V.M., Rojas-Herrera, R. and De-la-Peña, C., 2013. New insights into somatic embryogenesis: LEAFY COTYLEDON1, BABY BOOM1 and WUSCHEL-RELATED HOMEOBOX4 are epigenetically regulated in Coffea canephora. PLoS One, 8(8), p.e72160.

Nic-Can, G.I., López-Torres, A., Barredo-Pool, F., Wrobel, K., Loyola-Vargas, V.M., Rojas-Herrera, R. and De-la-Peña, C., 2013. New insights into somatic embryogenesis: LEAFY COTYLEDON1, BABY BOOM1 and WUSCHEL-RELATED HOMEOBOX4 are epigenetically regulated in Coffea canephora. PLoS One, 8(8), p.e72160.Khatri, P. and Drăghici, S., 2005. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics, 21(18), pp.3587-3595.

Nilsson-Ehle H (1917) Untersuchungen uber speltoid mutationen beim weizen. Botan Notiser 305: 3020.

Nilsson-Ehle H (1920) Multiple allelomorphe und complex mutationen beim weizen. Hereditas 1: 277-311.

NSF (2011). Phenomics: genotype to phenotype. a report of the phenomics workshop sponsored by the usda and nsf 2011 (available at: https://www.nsf.gov/bio/pubs/reports/phenomics_workshop_report.pdf)

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R. L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., and Buell, C. R. 2007. The TIGR Rice Genome Annotation Resource: improvements and new features. NAR 35 Database Issue: D846-851.

Park, C.J., Canlas, P.E. and Ronald, P.C., 2012. Establishment of glucocorticoid-mediated transcriptional induction of the rice XA21 pattern recognition receptor. Journal of Plant Biology, 55(1), pp.43-49.Park, Y.B. and Cosgrove, D.J., 2012. A revised architecture of primary cell walls based on biomechanical changes induced by substrate-specific endoglucanases. Plant Physiology, 158(4), pp.1933-1943.

Patil, I., 2021. Visualizations with statistical details: The'ggstatsplot'approach. Journal of Open Source Software, 6(61).Patil I (2018). "ggstatsplot: 'ggplot2' Based Plots with Statistical Details." CRAN. doi: 10.5281/zenodo.2074621, https://CRAN.R-project.org/package=ggstatsplot.

Patro, R., Mount, S.M. and Kingsford, C., 2014. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nature biotechnology, 32(5), pp.462-464.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F. and Sudhakar, D., 1999. 'Green revolution'genes encode mutant gibberellin response modulators. Nature, 400(6741), pp.256-261.

Phukan, U.J., Jeena, G.S. and Shukla, R.K., 2016. WRKY transcription factors: molecular regulation and stress responses in plants. *Frontiers in plant science*, *7*, p.760.

Power, J.F. and Alessi, J., 1978. Tiller development and yield of standard and semidwarf spring wheat varieties as affected by nitrogen fertilizer. The Journal of Agricultural Science, 90(1), pp.97-108.

Project, I.R.G.S., 2005. The map-based sequence of the rice genome. Nature, 436(7052), pp.793-800.

Purwestri, Y.A., Sari, R.D.K., Anggraeni, L.N. and Sasongko, A.B., 2015. Agrobacterium tumefaciens mediated transformation of rolC:: Hd3a-GFP in black rice (Oryza sativa L. cv. Cempo Ireng) to promote early flowering. Procedia Chemistry, 14, pp.469-473.

Ramanathan, A., Robb, G.B. and Chan, S.H., 2016. mRNA capping: biological functions and applications. Nucleic acids research, 44(16), pp.7511-7526.

Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., Van Ex, F., Pasha, A. and Khedikar, Y., 2018. The transcriptional landscape of polyploid wheat. Science, 361(6403).

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. and Manke, T., 2014. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic acids research, 42(W1), pp.W187-W191.

Rasband, W.S., 1997. ImageJ.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. and Vilo, J., 2019. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic acids research, 47(W1), pp.W191-W198.

Raven, P.H., Evert, R.F. and Eichhorn, S.E., 2005. Biology of plants. Macmillan.
Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A., 2013. Yield trends are insufficient to double global crop production by 2050. PloS one, 8(6), p.e66428.

Rebetzke, G.J., Bonnett, D.G. and Reynolds, M.P., 2016. Awns reduce grain number to increase grain size and harvestable yield in irrigated and rainfed spring wheat. Journal of Experimental Botany, 67(9), pp.2573-2586.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H. and Vilo, J., 2016. g: Profiler—a web server for functional interpretation of gene lists (2016 update). Nucleic acids research, 44(W1), pp.W83-W89.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A. and Dynlacht, B.D., 2002. E2F integrates cell cycle progression with DNA repair, replication, and G2/M checkpoints. *Genes & development*, *16*(2), pp.245-256.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. and Volkert, T.L., 2000. Genome-wide location and function of DNA binding proteins. Science, 290(5500), pp.2306-2309.

Ribaut, J.M. and Ragot, M., 2019. Modernising breeding for orphan crops: tools, methodologies, and beyond. Planta, 250(3), pp.971-977.

Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E. and Lippman, Z.B., 2017. Engineering quantitative trait variation for crop improvement by genome editing. Cell, 171(2), pp.470-480.

Rogers, J. and Wall, R., 1980. A mechanism for RNA splicing. Proceedings of the National Academy of Sciences, 77(4), pp.1877-1879.

Sage R, Monson R (1999). "7". C4 Plant Biology. pp. 228–229.

Saitoh, K., Onishi, K., Mikami, I., Thidar, K. and Sano, Y., 2004. Allelic diversification at the C (OsC1) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. Genetics, 168(2), pp.997-1007.

Sánchez-Ken, J.G. and Clark, L.G., 2010. Phylogeny and a new tribal classification of the Panicoideae sl (Poaceae) based on plastid and nuclear sequence data and structural data. American Journal of Botany, 97(10), pp.1732-1748.

Sato, K., 2020. History and future perspectives of barley genomics. DNA Research, 27(4), p.dsaa023.

Satoh-Nagasawa, N., Nagasawa, N., Malcomber, S., Sakai, H. and Jackson, D., 2006. A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature, 441(7090), pp.227-230.

Schaefer, J., Opgen-Rhein, R., Strimmer, K. and Strimmer, M.K., 2015. Package 'GeneNet'.

Schauser, L., Roussis, A., Stiller, J. and Stougaard, J., 1999. A plant regulator controlling development of symbiotic root nodules. Nature, 402(6758), pp.191-195.

Schmitz, G. and Theres, K., 2005. Shoot and inflorescence branching. Current opinion in plant biology, 8(5), pp.506-511.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13(11), pp.2498-2504.

Shantz, H.L., 1954. The place of grasslands in the Earth's cover. Ecology, 35(2), pp.143-145.

Shaw, L.M., Lyu, B., Turner, R., Li, C., Chen, F., Han, X., Fu, D. and Dubcovsky, J., 2019. FLOWERING LOCUS T2 regulates spike development and fertility in temperate cereals. Journal of experimental botany, 70(1), pp.193-204.

Sherry, S., Xiao, C., Durbrow, K., Kimelman, M., Rodarmer, K., Shumway, M. and Yaschenko, E., 2012, January. Ncbi sra toolkit technology for next generation sequence data. In Plant and Animal Genome XX Conference (January 14-18, 2012). Plant and Animal Genome.

Shimada, T., Endo, T., Fujii, H., Nakano, M., Sugiyama, A., Daido, G., Ohta, S., Yoshioka, T. and Omura, M., 2018. MITE insertion-dependent expression of CitRKD1 with a RWP-RK domain regulates somatic embryogenesis in citrus nucellar tissues. BMC plant biology, 18(1), pp.1-19.

Shimada, T., Endo, T., Fujii, H., Nakano, M., Sugiyama, A., Daido, G., Ohta, S., Yoshioka, T. and Omura, M., 2018. MITE insertion-dependent expression of CitRKD1 with a RWP-RK domain regulates somatic embryogenesis in citrus nucellar tissues. BMC plant biology, 18(1), p.166.

Shimamoto, K. and Kyozuka, J., 2002. Rice as a model for comparative genomics of plants. Annual Review of Plant Biology, 53(1), pp.399-419.

Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y.S., Gill, B.S. and Faris, J.D., 2006. Molecular characterization of the major wheat domestication gene Q. Genetics, 172(1), pp.547-555.

Sreenivasulu, N. and Schnurbusch, T., 2012. A genetic playground for enhancing grain number in cereals. Trends in plant science, 17(2), pp.91-101.

Studer, A., Zhao, Q., Ross-Ibarra, J. and Doebley, J., 2011. Identification of a functional transposon insertion in the maize domestication gene tb1. Nature genetics, 43(11), pp.1160-1163.

Susanto, F.A., Wijayanti, P., Fauzia, A.N., Komalasari, R.D., Nuringtyas, T.R. and Purwestri, Y.A., 2020. Establishment of a plant tissue culture system and genetic transformation for

agronomic improvement of Indonesian black rice (Oryza sativa L.). Plant Cell, Tissue and Organ Culture (PCTOC), 141(3), pp.605-617.

Tanaka, W., Ohmori, Y., Ushijima, T., Matsusaka, H., Matsushita, T., Kumamaru, T., Kawano, S. and Hirano, H.Y., 2015. Axillary meristem formation in rice requires the WUSCHEL ortholog TILLERS ABSENT1. The Plant Cell, 27(4), pp.1173-1184.

Tavakol, E., Okagaki, R., Verderio, G., Shariati, V., Hussien, A., Bilgic, H., Scanlon, M.J., Todt, N.R., Close, T.J., Druka, A. and Waugh, R., 2015. The barley Uniculme4 gene encodes a BLADE-ON-PETIOLE-like protein that controls tillering and leaf patterning. Plant physiology, 168(1), pp.164-174.

Team, R.C., 2013. (2015). R: A language and environment for statistical computing.

Theune, M.L., Bloss, U., Brand, L.H., Ladwig, F. and Wanke, D., 2019. Phylogenetic analyses and GAGA-motif binding studies of BBR/BPC proteins lend to clues in GAGA-motif recognition and a regulatory role in Brassinosteroid signaling. Frontiers in plant science, 10, p.466.

Toki, S., Hara, N., Ono, K., Onodera, H., Tagiri, A., Oka, S. and Tanaka, H., 2006. Early infection of scutellum tissue with Agrobacterium allows high-speed transformation of rice. The Plant Journal, 47(6), pp.969-976.

United Nations Population Division, 2019. In https://population.un.org/wpp/, World Population Prospects 2019, Dept of Economic and Social Affairs, File: Total Fertility.

Vaucheret, H., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Mourrain, P., Palauqui, J.C. and Vernhettes, S., 1998. Transgene-induced gene silencing in plants. The Plant Journal, 16(6), pp.651-659.

Venturelli, S., Belz, R.G., Kämper, A., Berger, A., von Horn, K., Wegner, A., Böcker, A., Zabulon, G., Langenecker, T., Kohlbacher, O. and Barneche, F., 2015. Plants release precursors of histone deacetylase inhibitors to suppress growth of competitors. The Plant Cell, 27(11), pp.3175-3189.1

Von Goethe, J.W. and Miller, G.L., 2009. The metamorphosis of plants. Cambridge: MIT Press.

Waki, T., Hiki, T., Watanabe, R., Hashimoto, T. and Nakajima, K., 2011. The Arabidopsis RWP-RK protein RKD4 triggers gene expression and pattern formation in early embryogenesis. Current Biology, 21(15), pp.1277-1281.

Wang, M., Yu, Y., Haberer, G., Marri, P.R., Fan, C., Goicoechea, J.L., Zuccolo, A., Song, X., Kudrna, D., Ammiraju, J.S. and Cossu, R.M., 2014. The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nature genetics, 46(9), pp.982-988.

Wang, Z., Gerstein, M. and Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, *10*(1), pp.57-63.

Wang, Z., Liu, Y., Shi, H., Mo, H., Wu, F., Lin, Y., Gao, S., Wang, J., Wei, Y., Liu, C. and Zheng, Y., 2016. Identification and validation of novel low-tiller number QTL in common wheat. Theoretical and Applied Genetics, 129(3), pp.603-612.

Watson, L. and Dallwitz, M.J., 1992. onwards. The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology,

anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references.

West, C.A. and Phinney, B.O., 1959. Gibberellins from Flowering Plants. I. Isolation and Properties of a Gibberellin from Phaseolus vulgaris L. Journal of the American Chemical Society, 81(10), pp.2424-2427.

Westengen, O.T., Lusty, C., Yazbek, M., Amri, A. and Asdal, Å., 2020. Safeguarding a global seed heritage from Syria to Svalbard. Nature Plants, 6(11), pp.1311-1317.

Wheeldon, C.D., Walker, C.H., Hamon-Josse, M. and Bennett, T., 2021. Wheat plants sense substrate volume and root density to proactively modulate shoot growth. Plant, Cell & Environment, 44(4), pp.1202-1214.

Wheeldon, C.D., Walker, C.H., Hamon-Josse, M. and Bennett, T., 2021. Wheat plants sense substrate volume and root density to proactively modulate shoot growth. Plant, Cell & Environment, 44(4), pp.1202-1214.

Whipple, C.J., Kebrom, T.H., Weber, A.L., Yang, F., Hall, D., Meeley, R., Schmidt, R., Doebley, J., Brutnell, T.P. and Jackson, D.P., 2011. grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. Proceedings of the National Academy of Sciences, 108(33), pp.E506-E512.

Wickham, H., 2011. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics, 3(2), pp.180-185.

Wickland, D.P. and Hanzawa, Y., 2015. The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: functional evolution and molecular mechanisms. Molecular Plant, 8(7), pp.983-997.

Wilkinson, M.D. and Haughn, G.W., 1995. UNUSUAL FLORAL ORGANS controls meristem identity and organ primordia fate in Arabidopsis. The Plant Cell, 7(9), pp.1485-1499.

Wu, K., Wang, S., Song, W., Zhang, J., Wang, Y., Liu, Q., Yu, J., Ye, Y., Li, S., Chen, J. and Zhao, Y., 2020. Enhanced sustainable green revolution yield via nitrogen-responsive chromatin modulation in rice. Science, 367(6478).

Würschum, T., 2012. Mapping QTL for agronomic traits in breeding populations. Theoretical and Applied Genetics, 125(2), pp.201-210.

Xiang, D., Quilichini, T.D., Liu, Z., Gao, P., Pan, Y., Li, Q., Nilsen, K.T., Venglat, P., Esteban, E., Pasha, A. and Wang, Y., 2019. The transcriptional landscape of polyploid wheats and their diploid ancestors during embryogenesis and grain development. *The Plant Cell*, *31*(12), pp.2888-2911.

Xu, C., Liberatore, K.L., MacAlister, C.A., Huang, Z., Chu, Y.H., Jiang, K., Brooks, C., Ogawa-Ohnishi, M., Xiong, G., Pauly, M. and Van Eck, J., 2015. A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nature genetics, 47(7), pp.784-792.

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X. and Zhang, Q., 2008. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nature genetics, 40(6), pp.761-767.

Yamanaka, S., 2008. Induction of pluripotent stem cells from mouse fibroblasts by four transcription factors. Cell proliferation, 41, pp.51-56.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T. and Dubcovsky, J., 2003. Positional cloning of the wheat vernalization gene VRN1. Proceedings of the National Academy of Sciences, 100(10), pp.6263-6268.

Yu, H., Jiao, B., Lu, L., Wang, P., Chen, S., Liang, C. and Liu, W., 2018. NetMiner-an ensemble pipeline for building genome-wide and high-quality gene co-expression network using massive-scale RNA-seq samples. PloS one, 13(2), p.e0192613.

Yu, J., Xuan, W., Tian, Y., Fan, L., Sun, J., Tang, W., Chen, G., Wang, B., Liu, Y., Wu, W. and Liu, X., 2021. Enhanced OsNLP4-OsNiR cascade confers nitrogen use efficiency by promoting tiller number in rice. Plant biotechnology journal, 19(1), pp.167-176.

Yu, L., Rosegrant, M.W., Wood, S. and Sun, D., 2009. Impact of growing season temperature on wheat productivity in China. Agricultural and Forest Meteorology, 149(6-7), pp.1009-1014.

Zale, J.M., Borchardt-Wier, H., Kidwell, K.K. and Steber, C.M., 2004. Callus induction and plant regeneration from mature embryos of a diverse set of wheat genotypes. Plant Cell, Tissue and Organ Culture, 76(3), pp.277-281.

Zhang, B., Liu, X., Xu, W., Chang, J., Li, A., Mao, X., Zhang, X. and Jing, R., 2015. Novel function of a putative MOC1 ortholog associated with spikelet number per spike in common wheat. Scientific reports, 5(1), pp.1-13.

Zhang, B., Liu, X., Xu, W., Chang, J., Li, A., Mao, X., Zhang, X. and Jing, R., 2015. Novel function of a putative MOC1 ortholog associated with spikelet number per spike in common wheat. Scientific reports, 5(1), pp.1-13.

Zhang, D., Liang, W., Yin, C., Zong, J., Gu, F. and Zhang, D., 2010. OsC6, encoding a lipid transfer protein, is required for postmeiotic anther development in rice. Plant physiology, 154(1), pp.149-162.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. and Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome biology, 9(9), pp.1-9.

Zhao, S., Zhang, Y., Gamini, R., Zhang, B. and von Schack, D., 2018. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Scientific reports*, *8*(1), pp.1-12.

Zhu, B.F., Si, L., Wang, Z., Zhu, Y.Z.J., Shangguan, Y., Lu, D., Fan, D., Li, C., Lin, H., Qian, Q. and Sang, T., 2011. Genetic control of a transition from black to straw-white seed hull in rice domestication. Plant physiology, 155(3), pp.1301-1311.

Zhu, L.J., 2013. Integrative analysis of ChIP-chip and ChIP-seq dataset. In Tiling Arrays (pp. 105-124). Humana Press, Totowa, NJ.

Zuckerkandl, E. and Pauling, L., 1965. Evolutionary divergence and convergence in proteins. In Evolving genes and proteins (pp. 97-166). Academic Press.