The Role of Chromatin State Transitions in Modulating Early Phosphate-deficiency Response Genes in Plants

Maryam Foroozani
Louisiana State University

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Genetics Commons, Genomics Commons, and the Plant Biology Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/5044

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
THE ROLE OF CHROMATIN STATE TRANSITIONS IN MODULATING EARLY PHOSPHATE-DEFICIENCY RESPONSE GENES IN PLANTS

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

The Department of Biological Sciences

by

Maryam Foroozani
M.S., Shiraz University Iran, 2007
December 2019
To my family for their endless love, support, and encouragement
# Table of Contents

Abstract ................................................................................................................................................. iv

Chapter 1. Introduction .................................................................................................................................. 1
  Pi Deficiency in Plants .......................................................................................................................... 1
  H3K4me3, an Important Histone Modification Mark, Regulates Diverse Biological Processes .......................................................... 13
  Epigenetic Regulation of Pi Starvation Responses .............................................................................. 41

Chapter 2. Genome-wide Distribution of H3K4 tri-methylation, H2A.Z and RNAseq under Control and P deficiency Conditions ................................................................................................................. 44
  Introduction ........................................................................................................................................ 44
  Materials and Methods: ...................................................................................................................... 45
  Results ............................................................................................................................................... 51
  Discussion ....................................................................................................................................... 66

Chapter 3. Chromatin States Associated with Early Pi Deficiency Responses ........................................ 70
  Introduction ....................................................................................................................................... 70
  Materials and Methods ...................................................................................................................... 71
  Results ............................................................................................................................................... 74
  Discussion ....................................................................................................................................... 99

Chapter 4. The Role of SET DOMAIN GROUP2 in Pi Starvation Responses ....................................... 103
  Introduction ....................................................................................................................................... 103
  Materials and Methods ...................................................................................................................... 105
  Results ............................................................................................................................................... 106
  Conclusions ..................................................................................................................................... 112

References ........................................................................................................................................... 116

Vita ....................................................................................................................................................... 144
Abstract

Phosphorus (P) is an essential plant macronutrient vital to fundamental metabolic processes. Plant-available P is low in most soils, making it a frequent limiter of growth. Declining P reserves for fertilizer production exasperates this agricultural challenge. Plants modulate complex responses to fluctuating P levels via global transcriptional regulatory networks. Although chromatin structure plays a substantial role in controlling gene expression, the chromatin-level mechanisms involved in regulating P homeostasis have not been determined.

In this work, I used chromatin immunoprecipitation (ChIP) combined with next-generation DNA sequencing (ChIP-seq) to map the distribution of H3K4me3 in rice (*Oryza sativa* L.) during both control and P-deficient conditions, and also used RNA-seq analyses to investigate global gene expression under the same set of conditions. In addition, I integrated multiple aspects of chromatin structure to define distinct chromatin states across the rice genome and compared them with profiles of genes differentially expressed by P deficiency. Finally, I investigated the potential role of the Arabidopsis SDG2 histone methyltransferase in P homeostasis.

The results of this work reveal that chromatin structure plays a valuable role in modulating gene expression in response to P-deficiency. First, I showed that H3K4me3 is prominent at the 5’ end of rice protein-coding genes, that it co-localizes with the H2A.Z histone variant, and the abundance of both H3K4me3 and H2A.Z exhibit distinct correlations with gene expression. Second, I defined distinct chromatin states across the rice genome by integrating the distribution of H3K4me3, H2A.Z, and nucleosome positioning. In response to P deficiency, 40% of all protein-coding genes exhibit a
chromatin state transition, and several of these transitions are enriched in subsets of genes differentially expressed by P deficiency. The most prominent subset supports the presence of a coordinated signaling network that targets cell wall structure and is regulated in part via a loss of transcription start site-localized H3K4me3. Finally, I showed that there was significant overlap between genes differentially expressed by P deficiency and those differentially expressed in an Arabidopsis sdg2 mutant. The P-deficiency induced chromatin dynamics and correlated genes identified here will be used in future efforts to generate crop plants exhibiting enhanced P-use efficiency, which is critical for sustaining global agriculture.
Chapter 1. Introduction

Pi Deficiency in Plants

Phosphorus: a major nutrient for plant growth

Phosphorus (P) is an essential plant nutrient in light of its involvement as a component in nucleic acids, lipids, and energy transfer molecules (ATP and NADPH). In addition, it plays a key role in the maintenance of metabolism, enzymatic reactions, and signal transduction (Figure 1).

Figure 1. General scheme of Phosphorus roles.

Despite the high concentration of P in the soil, most of the plant-available form of P, inorganic Pi (Pi), is unavailable for plant uptake due to its slow diffusion in soil and high absorption capacity. Although the Pi concentration in soil has large fluctuations, plants maintain intracellular P levels by a series of complex responsive and adaptive mechanisms for recycling, enhancing acquisition, and remobilization of Pi. It is necessary to gain insights into Pi sensing and signaling mechanisms to develop a better
understanding of P homeostasis in plants (Secco et al., 2013b; Tiessen, 2008; L. Zhang et al., 2014).

**Pi sensing and signaling**

The integration of endogenous signals and environmental factors modulates plant growth. Local and systemic (i.e. long distance) Pi sensing and signaling pathways coordinate changes in internal and external Pi status with the appropriate response to maintain optimal intracellular Pi concentrations. Figure 2 illustrates root and shoot specific responses to Pi starvation. It has been shown that Pi is sensed by a root-localized mechanism. To sense Pi availability, two types of actions are employed by plants: 1) External Pi is sensed by a root cell plasma-membrane-localized sensor, which involves local signaling, and 2) Pi status of the whole plant (internal Pi) is sensed by an intracellular sensor, which involves systemic signaling (Nagarajan & Smith, 2011; Thibaud et al., 2010).

Regardless of the internal Pi status of the shoot, physical contact of the primary root tip with a region of low Pi can initiate adjustments in root system architecture (RSA) to enhance Pi acquisition (Stefanovic et al., 2007). In Arabidopsis, a reduction of primary root growth, which has been widely observed in response to low Pi, indicates that the primary root tip may be the site where low Pi is sensed (Chevalier et al., 2003). To date, no external or internal Pi sensors have been identified. Although internal sensors in root tip cells may perceive an internal Pi deficiency, it is likely that the Pi concentration in the apoplasm of the root tip is perceived by a plasma-membrane-localized sensor. This is based on a number of studies which have established that plasma-membrane-localized nutrient transporters can function as nutrient sensors (Ho et al., 2009; Popova et al.,
2010). For example, the Pho84 high-affinity Pi transporter in yeast functions as a Pi transceptor that can sense the Pi status (Giots et al., 2003; Popova et al., 2010). It is possible that a similar function exists for one or more members of the Pht1;1 family of Pi transporters in plants, which are orthologous to yeast Pho84.

After Pi deficiency is perceived by the root sensing system, downstream systemic signaling pathways become activated to amplify the primary Pi deficiency signals (Chiou & Lin, 2011; Lucas et al., 2013). Pi, sugar, hormones, and microRNA (miRNA) have all been implicated in signaling. Generally, systemic signaling consists of xylem-mediated root to shoot, phloem-mediated shoot to root, or mature (i.e. “source”) tissue to young (i.e. “sink”) tissue (Lough & Lucas, 2006; Lucas et al., 2013).

Many studies have highlighted the importance of sugar, especially sucrose, in Pi-starvation responses. Specifically, sucrose increases the expression of Pi starvation induced (PSI) genes and impacts changes in root system architecture (RSA) (Hammond & White, 2008; Jain et al., 2007). Studies have suggested that sucrose also acts as a systemic Pi signal (Chiou & Lin, 2011). The over-accumulation of sucrose that occurs in the early stages of Pi deficiency is translocated to roots by increased phloem loading. Therefore the root to shoot mass ratio increases. In addition, Pi deficiency can activate sucrose responsive genes. Exogenous application of sugars can activate the expression of PSR (Pi starvation response) genes, and lack of sucrose impairs the induction of PSI gene expression (Chiou & Lin, 2011; Hammond & White, 2008, 2011; Karthikeyan et al., 2007; F. Liu et al., 2011; Lloyd & Zakhleniuk, 2004; Aaron P. Smith, 2013; Zakhleniuk et al., 2001).
Many hormones can act locally and systemically in Pi signaling. Microarray analyses revealed that Pi deficiency can alter hormone production, sensitivity and transport to regulate expression of PSR genes and RSA (Chiou & Lin, 2011). It has been shown that auxin, ethylene, cytokinins, strigolactones, gibberellins and abscisic acid have all been involved in the regulation of RSA and PSR genes. Auxin, ethylene and strigolactones activate PSR and gibberellins, abscisic acid and cytokinins repress PSR (Z. Zhang et al., 2014).

Deep sequencing and microarray analyses of small RNA expression profiles demonstrated that miRNAs from several plant species can serve as systemic Pi signaling components. In particular, miR399 is up-regulated under Pi-starvation conditions and plays a key role in Pi signaling (Bari et al., 2006; Chiou et al., 2006; Fujii et al., 2005; F. Liu et al., 2010). In response to Pi deficiency, movement of miR399 from the shoot serves as a systemic signal for activating Pi uptake in roots by targeting the PHO2 ubiquitin ligase, in turn stabilizing multiple Pi transporters (Buhtz et al., 2008; S.-I. Lin et al., 2008; Pant et al., 2008).

Generally, after initial local sensing of Pi deficiency, the primary systemic signals such as Pi, strigolactones (SLs), and cytokinins (CKs) are sent from root to shoot via xylem. In shoots, the primary signals are perceived by systemic sensors. Consequently, secondary systemic signals, including Pi, sucrose, and miR399, are produced in shoots and transported to roots via the phloem. In this way, shoots send signals to both shoot apices and roots to adjust developmental processes and nutrient uptake (Chiou & Lin, 2011).
Numerous split-root experiments have indicated many characteristic changes in RSA. Ethylene synthesis and response genes, stress and hormone related response genes and developmentally related genes are more likely controlled locally, whereas the genes involved in Pi uptake, sensing and signaling and recovery, and metal homeostasis genes seem to be systemically regulated by Pi deficiency (Franco-Zorrilla et al., 2005; Thibaud et al., 2010).

Figure 2. Overview of plant responses to Pi deficiency in both root and shoot. This image was taken from a previous publication (Z. Zhang et al., 2014).

**Pi starvation transcriptional responses**

A large number of genes are differentially expressed in response to Pi deficiency to coordinate Pi uptake, recycling and stress protection in both shoots and roots (Figure 3 and Figure 4). Microarray and deep-sequencing experiments on Arabidopsis and other species including rice and maize have resulted in an understanding of Pi-deficiency adaption and homeostasis, and have identified many PSR genes involved in these
processes. Due to the vital role of Pi in metabolism and the complexity of Pi homeostasis, it is valuable to perform time course experiments when investigating PSR gene expression. There are two recognizable transcriptional responses to Pi deficiency: early responses and late responses (Hammond et al., 2003; Lan et al., 2012; R. Müller et al., 2007; Secco et al., 2013b).

It has been shown that the expression of relatively few genes are changed in either root or shoot within 24 hours of Pi deficiency. In addition, early PSR genes were shown to respond quickly (within 3 hours or less, transiently and non-specifically to Pi deficiency (Hammond et al., 2004; Secco et al., 2013a). Several general stress-responsive genes such as those that encode peroxidases, cytochrome P450s and glutathione S-transferases, as well as salicylic acid-, ethylene-, and JA-mediated abiotic- and biotic-related genes, and also signal transduction-related genes are up-regulated within a few hours (Hammond et al., 2003; W.-D. Lin et al., 2011b; Misson et al., 2005b; Y.-H. Wang et al., 2002; Wu et al., 2003). The expression of various families of transcription factors, including MYB, WRKY, basic helix-loop-helix (bHLH) and NAM, ATAF and CUC transcription factor (NAC), and Pi deficiency-specific responsive genes such as Pi transporters, acid phosphatases, and SPX domain-containing proteins are also up-regulated by Pi deficiency within hours. It was found in rice that several genes involved in photosynthesis and carbon fixation were rapidly up-regulated at the very early time point (1h) of Pi deprivation in roots. In addition, several genes involved in metal uptake such as nicotinamide synthases and metal transporters are down-regulated by 1 h of Pi deficiency in roots. In the shoots genes encoding metallothioneins, which affect metal tolerance and homeostasis, are up-regulated within 1 h. It was suggested that up-
regulation of these genes could be a mechanism to overcome the high concentration of some ions, such as iron, upon Pi starvation. Several genes involved in iron homeostasis are up-regulated in both roots and shoots after 6 h of Pi starvation. The up-regulation of iron storage genes has also been reported in Arabidopsis (Secco et al., 2013a; Thibaud et al., 2010).

Changes occur in the expression of many thousands of genes during long-term Pi deprivation to regulate downstream genes for Pi uptake, remobilization, recycling, primary and secondary metabolism, photosynthesis, hormone and protein synthesis (Amtmann et al., 2005).

Figure 3. Significantly differentially expressed gene numbers under different time course experiment of Pi deprivation in roots (R) and shoots (S). Taken from (Secco et al., 2013b)

In both roots and shoots, Pi transporters, phosphatases, RNases, and enzymes involved in metabolic bypass in glycolysis and lipid metabolism are strongly up-regulated.
On the other hand, expression of regulators of protein synthesis is down-regulated (Hammond et al., 2003; Lan et al., 2012; O’Rourke et al., 2013). In rice, genes linked to lipid metabolism, phenylpropanoid metabolism, cytochrome P450s and Pi transporters are induced, whereas genes involved in photosynthesis and N metabolism are down-regulated at the late time points of Pi starvation (Secco et al., 2013a).

![Figure 4. The summary of transcriptional regulation of phosphate starvation responses. Taken from (Z. Zhang et al., 2014)](image)

**Regulation of Pi starvation responses**

Generally, plants have evolved complex responses to maintain optimal P levels and to tolerate low P conditions (Z. Zhang et al., 2014). Plants exhibit local and systemic (long distance) sensing and signaling pathways to regulate Pi starvation. These local and systemic pathways involve a set of molecular, physiological, morphological and biochemical responses such as reduced plant growth, altered root system architecture and secretion of organic acids, phosphatases, and nucleases to increase Pi acquisition.
(Secco et al., 2013b). Functions such as transcriptional and epigenetic regulation of gene expression, RNA silencing facilitated by micro RNAs, miRNA activity, Pi transporter trafficking, and protein post-translational modification (phosphorylation, sumoylation and ubiquitination) globally regulate a wide range of Pi deficiency responses (Rojas-Triana et al., 2013; Z. Zhang et al., 2014).

**Transcriptional regulation of Pi starvation**

According to many research studies, the responses of plants to Pi deficiency are highly controlled at the transcriptional level. Furthermore, P signaling pathways are largely conserved among higher plants. Three classes of transcriptional regulators have been implicated in systemic Pi signaling: 1) MYB, 2) WRKY, and 3) bHLH. PI STARVATION RESPONSE 1 (PHR1) and PHR1-Like1 (PHL1) are two homologous GARP (the maize G2, Arabidopsis ARRs, Chlamydomonas reinhardtii PSR1 and Arabidopsis PHR1 proteins) MYB proteins that have an important role in the regulation of Pi starvation stress responses. They are up-regulated by Pi starvation and most of the transcriptional responses to Pi starvation are impacted in phr1 and phr1 phl1 mutants (Bustos et al., 2010; Rubio et al., 2001; Wykoff et al., 1999). Three PHR1 orthologs in rice were identified: OsPHR1, OsPHR2 and OsPHR3 (Meina et al., 2015). Also, overexpression of Arabidopsis and rice PHR1 led to Pi accumulation in the shoot. Under Pi deficiency PHR1 was showed to regulate the level of Pi uptake by inducing the expression of the main Pi homeostasis genes such as SPX-domain proteins (SPX1 and SPX2), PI TRANSPORTER1 (PHT1), PI TRANSPORTER TRAFFIC FACILITATOR1 (PHF1), miR827, miRNA399 and AtIPS1/At4 and accordingly increase Pi uptake (Bayle et al., 2011; Duan et al., 2008). NLA and PHO2, which have E3 ubiquitin ligase and E2
conjugase activity, respectively, can degrade PHT1. PHR1 induces miR399 and miR827, which inhibit the expression of PHO2 and NLA and consequently increase the level of Pi transporters. MYB62, another MYB transcription factor was shown to regulate PSR by changing GA signaling and biosynthesis pathways. Considered together, these data indicate an important role for MYB transcription factors to regulate the expression of PSI genes in response to Pi deficiency. (Rubio et al. 2001; Bari et al. 2006; Nilsson et al. 2007; Stefanovic et al. 2007; Duan et al. 2008; Bayle et al. 2011)

Four Arabidopsis (AtWRKY75, AtWRKY45, AtWRKY6, AtWRKY42) and one rice (OsWRKY74) WRKY proteins were shown to be involved in Pi starvation responses (Y.-F. Chen et al., 2009; Xiaoyan Dai et al., 2015; Devaiah et al., 2007; Su et al., 2015; H. Wang et al., 2014). RNAi suppression of WRKY75, which functions as a positive regulator of Pi stress responses, leads to increased anthocyanin levels, decreased Pi uptake and weakened Pi stress responses in Arabidopsis under low-Pi stress (Devaiah et al., 2007). WRKY45 increased Pi uptake by inducing the expression of PHT1;1 in Arabidopsis (H. Wang et al., 2014). WRKY6 and its homologue, WRKY42, are negative regulators of Pi starvation responses by controlling the expression of PHO1 and PHT1;1 (Y.-F. Chen et al., 2009; Eulgem et al., 2000; Su et al., 2015; H. Wang et al., 2014). Moreover, OsWRKY74 has been demonstrated to be involved in the activation of PSI genes and in the regulation of RSA, and also have roles in crosstalk between P and Fe, and P and cold stress (Xiaoyan Dai et al., 2015).

Arabidopsis bHLH32 and OsPTF1, two basic helix-loop-helix domain transcription factors, have been reported to have roles in PSRs (Z.-H. Chen et al., 2007; Yi et al., 2005). Mutation of bHLH32 causes the induction of PSRs such as an anthocyanin
accumulation, root hair formation, and expression of PSI genes, therefore bHLH32 negatively regulates PSR (Z.-H. Chen et al., 2007). Contrarily OsPTF1 overexpression improved Pi-deficiency tolerance, suggesting it acts as a positive regulator of PSRs (Yi et al., 2005).

Recent studies on Pi-starvation responses uncovered the mechanism by which SPX domain proteins regulate PHR1 expression by sensing inositol polyphosphates (InsP). PHR1 is a constitutive gene and its expression is directly down-regulated by SPX1, 2 and 4 (Lv et al., 2014; Puga et al., 2014; Z. Wang et al., 2014). SPX containing proteins that have important roles in different parts of Pi homeostasis were shown to have high affinity toward InsP by both crystal structure and function analyses (Puga et al., 2014; Wild et al., 2016). The relation between InsP and Pi starvation was also confirmed in another Arabidopsis study in which incomplete loss-of-function of the AtIPK1 gene (inositol pentakisPi 2-kinase) resulted in reduction in InsP6 and InsP7 levels and induction of PSR (Kuo et al., 2014; Stevenson-Paulik et al., 2005).

Recently, PHT1 activity was shown to be negatively regulated by two important components, ALIX and CK2α2β3 kinase. ALIX is involved in Pi starvation response by transferring PHT1 to vacuoles to be degraded, while CK2α2β3 kinase phosphorylates and retains PHT1 in the ER and inhibits its interaction with PHF1 (Cardona-López et al., 2015; Y. Chen et al., 2017). New studies on SPX-MFS proteins have implicated them in Pi homeostasis by transporting Pi across the vacuole (J. Liu et al., 2015; T.-Y. Liu et al., 2016; C. Wang et al., 2015).

A mechanism of cross talk between Pi deficiency stress and immune regulation has been defined in a recent Castrillo et al. study (Castrillo et al., 2017). Genes involved
in Pi deficiency responses impact the root microbiome community structure regardless of Pi condition. PHR1, the key regulator of Pi deficiency responses, inhibited pathogen defense signaling, suggesting Pi homeostasis is prioritized over defense responses in plants. PHR1 not only represses pathogen defense responses but also induces microbiome-enhanced responses to Pi restriction. In addition, direct interaction between immune responses and PSRs was discovered by RNA and ChIP sequencing experiments. As an example, PHR1 was shown to trigger JA and SA pathway genes involved in defense responses (Castrillo et al., 2017). Cross talk between immune system and Pi starvation responses was reported in another study. Plants under Pi deficiency condition exhibited more resistance to insect herbivory and enriched synthesis of JA and induced the JA signaling pathway. PHR1 partly regulates the induction of JA signaling pathway under Pi starvation (Khan et al., 2016).

Root tip growth inhibition responses to Pi starvation are mainly regulated by iron, LPR1 and LPR2. The normal development of stem cell niche of the primary root needs SHR (SHORT ROOT). Under Pi starvation, PRD2 is disabled and causes the trafficking of LPR1 from ER to the plasma membrane. LPR1,2 have ferroxidase activity which increases the level of Fe 3+ in the apoplast, resulting in activation Reactive Oxygen Species (ROS) activation. ROS production induces callose deposition in stem cell niches (SCN) and consequently blocks SCN intracellular movement, trafficking SHR and arresting apical meristem activity (Dong et al., 2017; J. Müller et al., 2015; Petricka & Benfey, 2008; Svistoonoff et al., 2007; Ticconi et al., 2009). STOP1, the TF that activates ALMT1 is also involved in Fe3+ accumulation and activation of ROS (Balzergue et al., 2017; Mora-Macías et al., 2017).
In eukaryotic cells, genes are complexed with core histones and other chromosomal proteins in the form of chromatin. The basic repeating unit of chromatin, the nucleosome, is composed of two copies of each of the four core histones H2A, H2B, H3, and H4 wrapped by 146 bp of DNA (Luger et al., 1997; Pontvianne et al., 2010). Most of the molecular processes such as transcription, replication, DNA repair, and recombination take place in the chromatin environment. Therefore, chromatin structure has an important role in the regulation of molecular processes (Guo et al., 2010b). Various epigenetic mechanisms—such as histone modification, DNA methylation, small interfering RNA, ATP–dependent chromatin remodeling, placement of histone variants and regulation by noncoding RNA—are involved in the regulation of the structure and function of chromatin. Among the epigenetic mechanisms, post-translational covalent histone modification, DNA methylation and small interfering RNA reflect the chromatin condensation and transcriptional state of the associated DNA, and therefore play a dominant role in modulating chromatin structure and gene activity (Jenuwein & Allis, 2001; C. Liu et al., 2010).

**H3K4me3, an Important Histone Modification Mark, Regulates Diverse Biological Processes**

The amino-terminal tails of histones that protrude from nucleosome cores can be modified by acetylation, methylation, ubiquitination, phosphorylation, glycosylation, ADP-ribosylation and sumoylation. Most of the histone modifications are conserved across kingdoms. However, there exist some differences between plants and animals in the establishment and maintenance of these modifications. In fact, only four types of histone modifications, i.e. methylation, acetylation, phosphorylation and ubiquitination, have been
characterized in plants, among which methylation is recognized as an important
epigenetic histone mark (Guo et al., 2010a; Kouzarides, 2002). Chromatin structure can
be directly changed by some histone modifications such as acetylation, weakening the
histone-DNA interactions and making ‘open’ chromatin structure that facilitates
transcription. In contrast, lysine methylation on the amino-terminal tail of H3 has no effect
on chromatin structure on its own, but interacts with additional proteins such as chromatin
remodeling and assembling factors, to modulate chromatin condensation and gene
activity (Almer et al., 1986; Flanagan et al., 2005; Y. Huang et al., 2006; J. Kim et al.,
2006; Sims et al., 2005; Williams et al., 2008; Wysocka et al., 2006; F. Xu et al., 2005).

Methylation of histones is an important modification that is involved in diverse
biological processes, including transcriptional regulation and heterochromatin formation.
Methylation can occur on lysine or arginine residues with varying numbers of methyl
groups, namely mono-, di-, and trimethylation (Ng et al., 2007). Studies have shown that
approximately two-thirds of Arabidopsis genes and half of rice genes contain methylation
to some extent at histone H3 lysine 4 (H3K4), illustrating the importance of this particular
residue in plants (Guo et al., 2010b; X. Li et al., 2008; X. Zhang et al., 2009). Enzymes
that methylate H3K4 belong to a family of proteins containing a SET (Suppressor of
variegation, Enhancer of Zeste and Trithorax) domain, which is conserved among all
eukaryotes (Shilatifard, 2012).

Trimethylation of H3K4 (H3K4me3), in particular, has been widely studied, and is
recognized as an activator of transcription (Santos-Rosa et al., 2002b; Shilatifard, 2008).
The distribution of H3K4me3 appears largely conserved among species, with this mark
being enriched in genic regions and transcription start sites (TSSs) (Barski et al., 2007;
Bernstein et al., 2002; X. Li et al., 2008; C. L. Liu et al., 2005; Ram et al., 2011; Van Dijk et al., 2010; X. Zhang et al., 2009). Enrichment at TSSs of genes correlates this modification with transcriptional activation, however more research is required to elucidate a causal relationship. H3K4me3 has also been linked to many developmental and stress-related processes in plants, such as flowering, root development, and abiotic-stress responses.

**Characterization and function of H3K4 methyltransferases**

Histone lysine methyltransferases (HKMTs) that methylate H3K4 contain an evolutionarily conserved SET domain, which is responsible for the catalytic activity of the enzyme (Shilatifard, 2012). In yeast (*Saccharomyces cerevisiae*), the HKMT Set1 is solely responsible for the methylation of H3K4, while *Drosophila melanogaster* contains three Set1 orthologs: dSet1, Trithorax, and Trithorax-related (Roguev et al., 2001; Shilatifard, 2012). In plants, all HKMTs belong to the SET Domain Group (SDG). In Arabidopsis, the HKMTs are further categorized into seven classes based on function and domain structure (Gendler et al., 2008). The methyltransferases specific to H3K4 are members of Class III, which is comprised of five homologs of Trithorax, called Arabidopsis Trithorax-like (ATX) proteins (ATX1 through ATX5) and two ATX-related (ATXR) proteins (ATXR3 and ATXR7) (Avramova, 2009; Guo et al., 2010b; Tamada et al., 2009). While all Class III proteins bear the SET domain and the post-SET domain, they also contain varying combinations of other conserved domains that allow for further sub-grouping (Table 1, Figure 5).

The roles of Class III methyltransferases are not fully understood, however recent research has provided some insights into their functions. The most studied of the ATX
homologs is ATX1, which was found to methylate H3K4 at distinct genes in Arabidopsis (Alvarez-Venegas & Avramova, 2005). In yeast, a complex known as COMPASS (Complex Proteins Associated with Set1) contains the Set1 methyltransferase, which is a homolog of ATX1. The interaction of Set1 with structural subunits of COMPASS is required for the deposition of methyl groups at H3K4 (Schneider et al., 2005). The identification of a COMPASS-like complex in Arabidopsis and its interaction with ATX1 suggests a similar scenario for plants. Indeed, knockout/knockdown mutants of COMPASS-like components exhibit reduced H3K4 methylation in Arabidopsis (Ding, Ndamukong, et al., 2012; Jiang et al., 2011). COMPASS-like interaction with other Class III methyltransferases remains to be investigated. ATX2, the other Class III-1 enzyme, shows structural parallels with ATX1, but has divergent regulatory sequences that allow for differences in function. Whereas ATX1 catalyzes tri-methylation at H3K4, ATX2 functions as a di-methyltransferase for H3K4. Also, ATX1 and ATX2 appear to regulate different subsets of genes (Saleh et al., 2008). The Class III-2 subgroup contains ATX3, ATX4, and ATX5, which are redundantly necessary for genome-wide di- and tri-methylation of H3K4 (L. Chen et al., 2017). SDG2/ATXR3 is the single methyltransferase in subgroup III-3, and was found to be the predominant enzyme responsible for genome-wide deposition of tri-methylation at H3K4. The atxr3 mutant shows a severe phenotype, with stunted growth, impaired fertility, and detrimental changes to gene expression. In addition to tri-methylation activity, SDG2/ATXR3 is capable of performing di-methylation on H3K4, however, the significance of this function is limited, as the atxr3 mutant showed only a slight reduction in genome-wide H3K4me2 (Guo et al., 2010b). ATXR7, within subgroup III-4, is able to methylate H3K4, but also may be involved with the di-methylation
of H3K36. In addition, ATXR7 was shown to target specific genes, for example, the FLOWERING LOCUS C (FLC) gene which acts as a repressor for flowering (Berr et al., 2009).

Table 1. Class III histone 3 lysine 4 methyltransferases in plants. Adapted from (Ng, Wang et al. 2007).

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Arabidopsis thaliana</th>
<th>SDG Number (Arabidopsis)</th>
<th>Oryza sativa</th>
<th>Zea mays</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-1</td>
<td>ATX1</td>
<td>SDG27</td>
<td>SDG723</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ATX2</td>
<td>SDG30</td>
<td>SDG106, SDG128</td>
<td></td>
</tr>
<tr>
<td>III-2</td>
<td>ATX3</td>
<td>SDG14</td>
<td>SDG721</td>
<td>SDG115</td>
</tr>
<tr>
<td></td>
<td>ATX4</td>
<td>SDG16</td>
<td>SDG705</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ATX5</td>
<td>SDG29</td>
<td>SDG701</td>
<td>SDG108</td>
</tr>
<tr>
<td>III-3</td>
<td>ATXR3</td>
<td>SDG2</td>
<td>SDG717</td>
<td>SDG127</td>
</tr>
<tr>
<td>III-4</td>
<td>ATXR7</td>
<td>SDG25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In rice, five Class-III HKMT orthologs have been described. Two of these, SDG723 and SDG701 are orthologs of ATX1 and SDG2/ATXR3 respectively. SDG723 has H3 methyltransferase activity whereas SDG701 shows H3K4-specific methyltransferase activity (Choi et al., 2014; K. Liu et al., 2017). In maize, five HKMTs have been reported, however, the characterization of these proteins remains limited (Springer et al., 2003).

Figure 5. Domain structure of Arabidopsis H3K4 methyltransferases. Adapted from (Ng et al., 2007)
H3K4me3 and transcriptional regulation

H3K4 and gene expression

The distribution of H3K4me3 is generally conserved among species, predominately being found within gene-rich euchromatin regions while lacking in repeat sequences of heterochromatin (X. Zhang et al., 2009). Also, H3K4me3 deposition is prevalent near the TSS of most genes (Benayoun et al., 2014; Du et al., 2013). These observations are consistent with H3K4me3 playing an important role in transcription.

Two important ways transcriptional control can be achieved are the successful formation of the Pre-Initiation Complex (PIC) and the conversion of RNA pol II phosphorylated at serine-5 (Ser5P) to the elongating form, Ser2P. A study by Ding et al. showed that RNAi knockdown of AtCOMPASS reduced the level of TATA Binding Protein (TBP) and Pol II accumulation at the promoter regions of ATX1-regulated genes. Similar results were observed when using ATX1 mutant lines, however, when ATX1 was mutated in only the SET catalytic domain, there was no observable decrease in the formation of the PIC. These results indicate that the deposition of H3K4me3 by ATX1 is not required for establishment of the PIC, but rather the formation of the ATX1/AtCOMPASS complex is responsible for regulating initiation. SET domain mutants of ATX1 also exhibit low levels of H3K4me3 at the 5′ end of ATX1-regulated genes, and show wild-type levels of Ser5P RNA Pol II but a considerable reduction in the level of Ser2P RNA Pol II. This indicates that the rate of RNA pol II exit from initiation to elongation is reduced in the mutant (Ding et al., 2011b). Therefore, the level of H3K4me3 in promoter regions is not required for the formation of PIC, but a high-level presence of this mark at +300nt downstream of the TSS is apparently required for normal levels of transcriptional elongation.
Although H3K4me3 marks in the promoter regions of ATX1-regulated genes are unnecessary for the formation of the PIC, the promoter levels of H3K4me3 can influence expression of some genes. Two important transcription factors, bZIP28 and bZIP60, regulate the expression of downstream genes induced during ER stress (i.e. accumulation of unfolded or misfolded proteins in the ER) by directly binding to their promoters, and was shown that localization of these TFs was correlated with H3K4me3. During the adaptive phase of ER stress, the level of H3K4me3 in the promoter of induced genes was increased. However, induction in the level of H3K4me3 at the promoter of ER stress-induced genes was depleted in zip28/zip60 double mutants. Further, mutation of the COMPASS-like components, ARABIDOPSIS Ash2 RELATIVE (ASH2R) and WDR5a, showed a marked decrease in the expression of these ER stress-induced genes, and interaction among bZIP28, bZIP60 and the COMPASS-like complex has been suggested. It is likely that during PIC assembly, bZIP28, bZIP60 and other general transcription factors (TFs) are recruited to the promoters of ER stress-induced genes. Once at the promoter, TFs interact with COMPASS-like components to facilitate the activity of histone methyltransferases, which deposit H3K4me3 in the promoter regions of these genes (Z.-T. Song et al., 2015).

The distribution and characteristic of this histone mark has linked H3K4me3 with transcriptional activation in rice (Booth et al., 2016; Churchman & Weissman, 2011; Guillemette et al., 2011). Studies in rice have shown that highly expressed genes contain significantly more H3K4me3 marks when compared to lower expressed genes (Du et al., 2013). The same relationship was found in Arabidopsis, furthering the correlation between H3K4me3 and gene transcription (X. Zhang et al., 2009). However, while a
correlation exists between this mark and transcriptional activation, the exact details of this relationship are not fully understood. Loss of SDG2/ATXR3 led to a substantial decrease in H3K4me3 genome-wide. Interestingly, despite the reduction in H3K4me3 levels at 25 genes analyzed, some were up-regulated, down-regulated or unaffected in sdg2 mutants, indicating that loss of H3K4me3 is not strongly correlated with gene expression (Guo et al., 2010b). Similarly, RNA-seq and ChIP-seq data showed that atx3 atx4 atx5 triple mutants displayed changes in the expression of almost 2,000 genes compared to wild-type plants, but with no strong correlation between differential expression and change in H3K4me3 (L. Chen et al., 2017).

**H3K4me3 involvement with RNA processing**

Effective mRNA splicing requires a 5′-cap which is recognized and protected by a nuclear mRNA cap-binding complex (CBC) (Calero et al., 2002). It was shown in Arabidopsis that CBC can selectively interact with both the COMPASS-like complex (which contains the H3K4me3 methyltransferase) and histone 3 lysine 36 methyltransferase, suggesting active histone modifications (H3K4me3 and H3K36me3) are involved in RNA processing. COMPASS-like requires association with CBP20, a subunit of CBC, to activate the expression of its target gene, FLC. In a CBP20 mutant, levels of H3K4me3, ASH2R enrichment, and expression of FLC were greatly reduced, suggesting CBP20 is required for binding of ASH2R to FLC chromatin, deposition of H3K4me3, and activation of gene expression. This result from the CBP20 mutant was not only found at the FLC gene; 304 loci had a similar reduction in the level of H3K4me3, signifying the role of CBP20 across a larger number of genes. CBC is needed for efficient function of COMPASS-like, however, COMPASS-like is also required for proper function
of CBC as well. CBC has two main functions, protection of the 5’cap and facilitation of efficient pre-mRNA splicing. COMPASS-like mediates the protection of the FLC pre-mRNA 5’-cap by associating with CBC and promoting CBC’s cap binding function. This conclusion arises from evidence that the level of uncapped FLC pre-mRNA was increased in both an ash2r mutant and an atx1/atx2 double mutant. In addition, CBC required COMPASS-like for sufficient splicing of introns near the cap of FLC pre-mRNA. The level of the first-intron unspliced to spliced FLC mRNA was increased in both cpb20 and ash2r mutants (Z. Li et al., 2016). This evidence taken together suggests H3K4me3 is actively involved in the regulation of RNA processing.

Role of H3K4me3 in developmental processes

Flowering time and development

Flowering is one of the most important and controlled developmental processes in higher plants. Plants have evolved complex signaling pathways to ensure appropriate transition to the reproductive phase in accordance with developmental and environmental cues. A number of molecular components, including MADS-box transcription factors such as the key floral repressor, FLOWERING LOCUS C (FLC), are involved in these pathways. Many reports have implicated H3K4 trimethylation in the regulation of floral transition, as well as in flower development.

The winter-annual habit of Arabidopsis requires vernalization to prevent flowering in fall to guarantee flowering in spring, and it was shown that active FRIGIDA (FRI) and the flowering repressor FLC are required for this process (Lee et al., 1994). FRI expression and deposition of H3K4me3 are both necessary for the efficient level of FLC transcription, which subsequently represses flowering. Investigations of the molecular
mechanism of flowering have revealed the involvement of many H3K4 trimethyltransferases, largely through their role in modulating FLC expression. Indeed, mutation of either ATX1, ATX2, SDG2, or ATXR7 results in early flowering and reduced FLC transcripts (Pien et al., 2008; Tamada et al., 2009; Yun et al., 2012). An atx1 atx2 double mutation led to a more severe reduction in FLC expression and lack of late flowering suppression, suggesting a redundant role of ATX1 and ATX2 in flowering repression. The same result was observed in the atx1 and atx1 atx2 double mutant in the presence of FRI, showing the atx1 mutant overcomes the late flowering effect of FRI (Pien et al., 2008). Rapid flowering and reduction in the level of FLC transcripts in atr7 mutants was not as severe as other mutants, but it was revealed that ATX1 and ATXR7 are both necessary for a suitable level of H3K4me3 at FLC. Furthermore, both ATX1 and ATX7 showed the direct interaction with the FLC locus. In addition, ATX1 and ATX7 not only causes the activation of FLC by H3K4me3 methylation, but also prevents repression by inhibiting H3K27me3 deposition. Since the atx1 atx7 double mutant showed more rapid flowering compared to the single mutant, and both FLC expression and H3K4me3 level were significantly lower, this suggests both classes of H3K4 trimethyltransferase are required for FLC regulation (Pien et al., 2008; Tamada et al., 2009). Mutation in ATX1, ATX2, and ATXR7 was shown to be unsuccessful for complete repression of FLC expression, so the involvement of another H3K4 methyltransferase, SDG2, was suggested. sdg2 mutants displayed a decrease of FLC expression, however, the lack of active FRI in the majority of Arabidopsis lab lines could result in a similar repression of FLC (Guo et al., 2010b). In a study by Yun et al. (Yun et al., 2012), the sdg2 mutant with an active FRI showed a low level of FLC transcripts and earlier flowering compared to the
FRI-Col, signifying that SDG2 was responsible for \textit{FLC} activation. In FRI-Col, H3K4me3 is enriched at the TSS of \textit{FLC}, while the FRI \textit{sdg2} mutant exhibited a drastic reduction of H3K4me3 around the TSS and across the gene body. Therefore, regardless of the presence of FRI, the \textit{sdg2} mutant will lead to lower expression of \textit{FLC} and subsequent early flowering. Interestingly, Yun et al. showed the \textit{atx1 atx2 atxr7} triplet mutants exhibited only partial suppression of \textit{FLC} as well, showing H3K4me3 deposition at TSS of \textit{FLC} via SDG2 has a significant role in winter-annual habit of Arabidopsis (Yun et al., 2012).

COMPASS-like complexes also affect \textit{FLC} expression through deposition of H3K4me3. Mutation of \textit{ASH2R} or the \textit{RbBP5} homolog exhibited early-flowering phenotypes and suppression of \textit{FLC} and \textit{FLC} homologs. In addition, mutation of \textit{ASH2R} led to drastic reduction in H3K4me3 genome-wide, and also at both \textit{FLC} and the \textit{FLC} homologs. Direct interaction of ASH2R at the TSS of the \textit{FLC} locus was observed, indicating ASH2R-mediated H3K4me3 deposition is involved in \textit{FLC} regulation. Taken together, it was concluded that Arabidopsis COMPASS-like core components and their methyltransferase activity are required for \textit{FLC} expression and late flowering (Jiang et al., 2011).

Genome-wide and cell specific studies show that changes in the ratio of H3K4me3 and H3K27me3 were required for regulation of gene expression during transition from shoot apical meristem (SAM) to an inflorescence meristem (IM) in rice. Genome-wide, a positive correlation between H3K4me3 and H3K27me3 in both meristems suggests the presence of both modifications in many genes. Moreover, changes in the level of H3K4me3 and H3K27me3 were increased in many IM genes compared to SAM, but with
both modifications exhibiting a mutually exclusive relationship. The H3K4me3/H3K27me3 ratio and gene expression levels displayed a correlation during the transition from SAM to IM. For example, up-regulated genes were mostly shown to have a higher H3K4me3/H3K27me3 ratio and down-regulated genes had a lower H3K4me3/H3K27me3 ratio. Also, mutations in SDG711 (H3K27 methyltransferase) or JMJD3 (H3K4 demethylase) were found to suppress levels of H3K27me3 at many genes in IM. H3K4me3 and H3K27me3 levels were increased and decreased respectively in both the jmj703 and sdg711 single mutants at the CKX2/Gn1a locus, a gene expressed in IM relating to panicle size. The sdg711 jmj703 double mutant showed similar results, however, the increase of H3K4me3 and decrease of H3K27me3 was more drastic. These results indicate the two enzymes both achieve the same increase and decrease of H3K4me3 and H3K27me3 respectively, however, the way in which this ratio is achieved is possibly through opposing functions (X. Liu et al., 2015b). A recent study utilizing cell-type specific analysis with stem cells and non-stem cells found within the SAM, H3K4me3 and H3K27me3 levels change during floral transition, with an increase in H3K4me3 corresponding to an increase in gene expression. However, the results of this study suggest that the level of H3K27me3 does not predict levels of gene expression in SAM as a result of flowering (You et al., 2017).

The rice ortholog of ATX1, OsTrx1, displays a role in regulating flowering time under long-day conditions through inhibition of Ghd7 (grain number, plant height, and heading date) transcription. OsTrx1 binds to Early heading date 3 (Ehd3), which regulates flowering upstream of Ghd7, suggesting OsTrx1 controls flowering time through this interaction. Although results indicate OsTrx1 is actively involved in flowering time
regulation, the level of H3K4me3 was not measured at the Ghd7 gene, therefore the role of H3K4me3 in this process remains unclear (Choi et al., 2014). SDG701, an ortholog of Arabidopsis SDG2, also demonstrated promotion of flowering in rice. Heading date 3a (Hd3a) and rice flowering locus T1 (RFT1), which control flowering by encoding florigens, are controlled by Early heading date 1 (Ehd1). The expression of all three of these genes was down-regulated in sdg701 mutants, suggesting SDG701 is necessary for flowering transition. Interestingly, H3K4me3 levels were high in both RFT1 and Hd3a, but not in the Ehd1 gene in wild-type. Furthermore, mutation of SDG701 caused a reduction in the level of H3K4me3 at RFT1 and Hd3a genes, but not Ehd1. Notably, the interaction between SDG701 with RFT1 and Hd3a suggests SDG701 regulates their expression by directly binding to chromatin and depositing H3K4me3 (K. Liu et al., 2017).

Involvement of H3K4 trimethyltransferases in flower development was demonstrated in sdg2 mutants in Arabidopsis. sdg2 exhibited floral defects including severe male and female reproductive organ abnormalities, as well as an inability to maintain indeterminate growth of the inflorescence meristem, resulting in a terminal flower (Guo, Yu et al. 2010). Dynamic changes of H3K4me3 were found to be a strong indicator for changes in gene expression over flower development, more than H3K27me3, although several genes are found to be marked differentially by these two modifications throughout development. Interestingly, most of the differentially marked genes (DMGs) in varying floral developmental stages showed an opposing relationship between the level of these two marks. During flower initiation, changes in H3K4me3 occurred sooner and exhibited a stronger correlation with expression relative to H3K27me3 (Engelhorn et al., 2017).
**Root and shoot development**

Evidence suggests H3K4me3 has an important role in regulating root and shoot developmental processes. For example, *sdg2* mutant plants exhibited pleiotropic phenotypes; curly leaves, short roots, and smaller organs (Guo et al., 2010b). SDG2 and H3K4me3 were shown to be necessary for proper root postembryonic development by establishing and maintaining root stem cell fate. Mutants of *SDG2* displayed hindrance of primary root growth and defective lateral root growth and development. Loss of *SDG2* function led to small root apical meristem (RAM), disorganized root stem cell niche (SCN), and as a consequence, halted meristem activity in both primary root and lateral root. Besides lack of cell identity and stem cell function, the auxin gradient in the root quiescent center vanished in *sdg2* mutants. The level of H3K4me3 genome-wide decreased in root cells, specifically in SCN cells. Generally, regulation of stem cell pluripotency and root growth development was controlled by SDG2-mediated deposition of H3K4me3, which maintains chromatin organization (Yao et al., 2013). Also, ATX1 was required for cell cycle regulation, the production of cells, and the conversion to cell elongation during root growth. *atx1* mutants displayed delayed primary root growth, decreased RAM size and activity, similar to the *sdg2* mutant. Interestingly, in the *atx1* mutant, the length and number of cells were only reduced in the proliferation domain. Additionally, longer cell cycles in the *atx1* mutant impacted cell production. ATX1, similar to SDG2, showed involvement in different steps of lateral root formation and development; for example, lateral root emergence was impeded in the *atx1* mutant. Different characteristics with respect to root development and root cell proliferation were found in the *atx1* and *sdg2*
mutants, indicating both HMTases are important for proper root growth and development (Napsucialy-Mendivil et al., 2014).

H3K4me3 is involved in early shoot development by increasing expression levels of some shoot development-related genes. In addition, H3K4me3 enrichment at these genes was shown to be assisted by the presence of H2A.Z. Transition from juvenile to adult vegetative phases is regulated by the induction in the level of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) expression. The expression level of SPL was repressed by expression of MIR156A and MIR156C, two genes producing the majority of miR156. The expression levels of MIR156A and MIR156C were reduced in both SWR1 complex (which facilitates H2A.Z deposition) component mutants (i.e. arp6 and setf) and H2A.Z-encoding gene mutants, with all mutants showing expedited transition of vegetative phases. In addition, levels of both H3K4me3 and H3K27me3 were measured in wild-type plants at MIR156A and MIR156C genes, revealing that H3K4me3 decreased gradually during shoot development at these genes and H3K27me3 increased. Mutation in ARP6 reduced the level of H3K4me3 at both genes, and interestingly, nucleosome occupancy and H3K27me3 increased at MIR156A. These results imply shoot development in Arabidopsis involves regulation of genes through H2A.Z-assisted deposition of H3K4me3 (M. Xu et al., 2018).

**Sporophyte and gametophyte development**

Berr et al. (Berr et al., 2010) reported on the involvement of SDG2 in sporophyte and gametophyte development, indicating that reduced leaf size during the vegetative phase came from impaired cell division and cell expansion in the sdg2 mutant. Sterility of the sdg2 mutants was caused by defective anther and pollen development along with
short stamens and misregulated microgametogenesis. In addition, megagametogenesis halted in most of \textit{sdg2} mutant ovules, leading to degenerate megaspore nuclei and defective embryo sacs. The expression level of \textit{SPOROCYTELESS/NOZZLE (SPL/NZZ)} and \textit{MALE STERILITY1 (MS1)}, two essential genes for male and/or female gametophyte development, were drastically reduced in \textit{sdg2} mutants. In wild-type plants, \textit{SPL/NZZ} and \textit{MS1} are known to contain bivalent domains of chromatin, marked by both H3K4me3 and H3K27me3. Therefore, SDG2-mediated H3K4me3 deposition regulates expression of two essential developmental genes, and in this way, controls molecular mechanisms for gametophyte development. Due to DNA degradation in the \textit{sdg2} mutant, it was suggested that H3K4me3 deposition plays a critical role in chromosome structure and reprogramming during meiosis as well (Berr et al., 2010).

The detailed function of SDG2 in Arabidopsis male gametophyte development was demonstrated in a recent study by Pinon (Pinon et al., 2017). SDG2 was found to play an important role in Arabidopsis postmeiotic gametogenesis via provoking chromatin decondensation, mitotic cell division, and activation of the retrotransposon ATLANTYS1. In tricellular stage pollen, \textit{sdg2} pollen grains contained only two sperm nuclei per pollen. During pollen development in the wild type plant, chromatin of vegetative cell nuclei and microspores becomes decondensed, however, chromatin was more condensed in the \textit{sdg2} mutant. In addition, loss of SDG2 function resulted in defective ATLANTYS1 activity, pollen germination, and pollen tube elongation, which is ultimately the result of the condensed chromatin state (Pinon et al., 2017).
Hormone responses

H3K4me3 was shown to have a regulatory role at several hormone-responsive genes. A reduction of transcript levels at a number of early hormone-responsive genes was observed in sdg2 mutants under hormone treatment. Consistent with reduction in expression, the H3K4me3 levels at three hormone responsive loci, SAUR27, GASA6, and KIN1 diminished in the sdg2 mutant compared to wild-type. Global status of H3 methylation was measured in WT and sdg2, showing a reduction of H3K4me3 levels in floral tissue of sdg2 mutants. In both gene-specific and global analysis, the level of H3K4me2 was increased in sdg2 mutants compared to WT, suggesting SDG2 is responsible for the transition from H3K4me2 to H3K4me3. Collectively, SDG2 plays an important role in regulating hormone-responsive gene expression by deposition of H3K4me3 (S. Kim et al., 2013).

Circadian clock

H3K4me3 has a role in the regulated expression of oscillator genes responsible for circadian clockwork. Both the expression of these genes and the levels of H3K4me3 were reduced in heterozygous sdg2 mutant plants. In wild-type Arabidopsis, the enrichment of H3K4me3 and H3K56ac was higher at evening-expressed oscillator genes at Zeitgeber Time 15 (ZT15), while a reduction in the level of both histone modifications was observed at morning-expressed oscillatory genes at ZT15. The time course analysis of both histone modifications showed H3K4me3 has a delayed accumulation when compared to H3K56ac, implying a difference in accrual could be specific to time-related activating functions. In addition, H3K4me2 levels were high at the time that H3K56ac was low and vice versa, while also, the coexistence of H3K4me3 and H3K4me2 was detected.
This suggests there is a sequential accumulation beginning with H3K56ac, then H3K4me3, and finally H3K4me2, that is temporally associated with the expression of oscillator genes through the circadian cycle. A decrease in H3K4me3 levels by nicotinamide treatment induced in the binding of a clock transcriptional repressor, PRR5, to its target genes which caused repression. Taken together, H3K4me3 plays a role in regulating the correct rhythmic transition from activation to repression in a temporal manner (Malapeira et al., 2012).

**Leaf senescence**

The epigenetic framework of the Arabidopsis genome was investigated in two studies to understand the regulatory mechanisms of gene expression during leaf senescence. The first study monitored the genome-wide distribution of H3K4me3 and H3K27me3 while the second study focused on H3K4me3 and H3K9ac during the transition into leaf senescence. The first study found that most of the genes enriched and reduced in H3K4me3 were up-regulated and down-regulated respectively in older leaves, while H3K27me3 reduced genes were also up-regulated in older leaves. However, a limited portion (5%) of differentially expressed genes showed changes in H3K4me3 and H3K27me3, therefore it is hard to draw conclusions about the importance of these two modifications in developmental transition of senescing leaves (Brusslan et al., 2012). In the second study, a stronger positive correlation between H3K4me3 levels and gene expression was shown during leaf senescence. Around 56% of increased H3K4me3 genes and 63% of decreased H3K4me3 genes were up-regulated and down-regulated during leaf senescence respectively, while a moderate positive correlation was observed between H3K9ac and gene expression. 22% of up-regulated genes showed a correlation
with an increase in H3K4me3 levels, and these genes were found to be regulatory genes that perform primary responses (Brusslan et al., 2012; Brusslan et al., 2015).

**Role in stress response**

**Drought stress**

Among the types of H3K4 methylations, tri-methylation shows the most significant change at dehydration stress-responsive genes in Arabidopsis under drought stress treatment. Furthermore, changes in the level of H3K4me3 are positively associated with transcript levels of stress response genes. In addition, the H3K4me3 distribution profile at dehydration stress-responsive genes is broader than unresponsive genes (Van Dijk et al., 2010). In contrast to Arabidopsis, a genome-wide study in rice under drought stress showed minimal expression changes as a result of differential H3K4me3. Only 13% of differential H3K4me3 genes exhibit changes in transcription level during the stress treatment. Genes that show an increase in the level of H3K4me3 displayed a corresponding increase in their expression level, whereas genes showing a decrease in H3K4me3 levels had a reduction in expression. Stress up-regulated genes with a modified level of H3K4me3 under drought stress are involved in stress-related metabolite and hormone signaling pathways. Stress down-regulated genes with a differential level of H3K4me3 are involved in photosynthesis and glycolysis (Zong et al., 2013).

Evidence from mutations in H3K4 tri-methyltransferases also indicates the involvement of H3K4me3 in drought stress response. *atx1* mutants were found to have similar transcript levels genome-wide to wild-type Arabidopsis under drought stress, however, the expression of ATX1-regulated genes is affected by drought stress. Interestingly, the levels of ATX1 transcripts were not changed, concluding that the activity
of ATX1 was post-transcriptionally inhibited by drought stress (Ding et al., 2009). Furthermore, it was demonstrated that ATX1 is required for dehydration stress signaling in both an ABA-dependent and independent manner. Mutation in ATX1 genes lead to sensitivity of dehydration stress, rapid transpiration and larger stomatal apertures due to a reduction in the level of ABA. atx1 mutants exhibited down-regulation of the ABA biosynthetic genes, ABA3 and NCED3, where NCED3 was shown to be a direct target of ATX1. Additionally, during dehydration stress treatment, Pol II occupancy and the level of H3K4me3 and ATX1 are increased at the NCED3 gene, and inductions in the level of Pol II occupancy and H3K4me3 are higher in WT plants compared to the atx1 mutant. Therefore, dehydration stress increased the level of H3K4me3 and NCED3 expression through the involvement of ATX1 (Ding et al., 2011a).

Both H3K4me3 and histone acetylation (H3K23ac, H3K9ac, and H3K27ac) are involved in regulating Arabidopsis stress-inducible genes (RD29A, RD29B, RD20, and RAP2.4) under drought stress. The levels of H3K4me3 and H3K9ac are induced under drought stress at all four genes, indicating a positive correlation between levels of histone modifications and the expression of these genes by drought stress. However, H3K23ac and H3K27ac levels increased only in RD29B, RD20, and RAP2.4. Interestingly, the level of H3K4me3 is established after RNA Pol II accumulation at RD29A and RAP2.4, showing full activation of these genes is needed for establishing H3K4me3. This result is consistent with findings regarding the role of H3K4me3 in transcriptional elongation (Kim, To et al. 2008).

Transcriptional stress memory was investigated by Ding et al (Ding, Fromm, et al., 2012) during repeated dehydration stress in Arabidopsis. Four dehydration inducible
marker genes, *RD29A*, *RD29B*, *RAB18*, and *COR15A*, were divided into two categories, non-trainable and trainable genes, based on their expression pattern during reoccurring dehydration stress. The expression levels of trainable genes (*RD29B*, *RAB18*) were increased relative to the initial stress response, however, the expression levels of non-trainable genes (*RD29A*, *COR15A*) were increased at the same level during each stress. In spite of the difference in the rate of transcription between non-trainable and the trainable genes, in the recovery states, transcript levels of both returned to initial levels. For non-trainable genes, the degree of H3K4me3 enrichment was similar after each stress treatment, but in trainable genes, H3K4me3 was significantly increased in repetitive stress treatments. In the recovery states, the level of H3K4me3 and Ser5P Pol II remained high in trainable genes even with the low level of transcription, showing Ser5P Pol II was stalled. As previously stated, H3K4me3 is necessary for the conversion of Ser5P Pol II to the elongating form, Ser2P Pol II, to promote transition from initiation to elongation. These data seems to suggest that the presence of H3K4me3 may be necessary for elongation but does not guarantee the transition. H3K4me3 behaves as an adjustable chromatin mark which is correlated with transcription levels in non-trainable genes, however, acting as an epigenetic memory mark, which is not removed during recovery states in trainable genes (Ding, Fromm, et al., 2012). Another study from Fromm and Avramova (N. Liu et al., 2014) showed that H3K4me3 serves as a memory mark, however, H3K27me3 did not serve as a memory mark in the following genes, *LTP3*, *LTP4*, *HIPP2.2*, *RD29B*, and *RAB18*. These genes were found to have different transcriptional patterns and were controlled by varying molecular mechanisms. In drought stress treatments, the levels of H3K4me3 and expression of the five genes were increased.
During the recovery state, the expression of LTP3, LTP4, and HIPP2.2 increased compared to the stressed state, however RD29B and RAB18 decreased, similar to the Ding study. Additionally, the level of H3K4me3 in LTP3, LTP4, and HIPP2.2 continued to increase from stressed to recovery, whereas RD29B and RAB18 remained constant. In the second stress treatment, H3K4me3 and expression levels increased at all genes compared to the first stress treatment. Therefore, H3K4me3 serves as a memory mark for RD29B, and RAB18 but not for LTP3, LTP4, and HIPP2.2. In contrast to H3K4me3, the level of H3K27me3 showed no change at the memory or non-memory genes during the first, second stress and recovery phases, regardless of transcriptional rate (N. Liu et al., 2014). Levels of Pol II, H3K4me3, and H3K9ac at three drought stress-inducible genes (Rd20, RD29A, and AtGOLS2) were also shown to increase in response to drought stress treatment. During the recovery phase, the expression of these genes began to decrease, and the level of H3K9ac is quickly eliminated while H3K4me3 gradually decreased, however, was not fully removed. Again, these results further suggest that H3K4me3 has a potential role in stress memory marking of certain genes (J.-M. Kim et al., 2012).

**Submergence stress**

Submergence stress causes changes in the level of tri- and dimethylation of H3K4 at both the 5′- and 3′-coding regions of rice submergence-inducible ADH1 and PDC1 genes. During submergence, the level of H3K4me3 increased, whereas there was a reduction in the relative abundance of H3K4me2. Therefore, submergence stress results in conversion of H3K4me2 to H3K4me3. This conversion happened simultaneously with an increase in RNA Pol II and transcript levels. Therefore, induction in the level of
H3K4me3 at submergence-inducible genes was positively correlated with the level of submergence-inducible gene transcripts under submerged conditions. In addition, during submergence stress, the level of H3 acetylation showed a gradual increase that occurred after the levels of RNA pol II and transcripts increased. This evidence suggests that dynamic changes in H3K4me3 and H3 acetylation leads to transcriptional regulation in response to submergence stress (Tsuji et al., 2006).

**Heat stress**

The enrichment of H3K4me3 and H3K4me2 was monitored at memory-related loci under heat stress conditions, finding that high levels of this modification were present during recovery states. Furthermore, induction of the expression levels of heat stress-inducible genes was higher during the second stress treatment. Therefore, H3K4me2 and H3K4me3 were defined as memory marks for heat stress. HSFA2, a heat stress transcription factor, was shown to bind to memory loci and assist with the maintenance of H3K4me2 and H3K4me3 levels (Lämke et al., 2016).

**Salt stress**

ABA and salt stress (125 mM NaCl) treatment leads to the enrichment of the activating marks H3K4me3 and H3K9K14 acetylation, while reducing the repressive H3K9me2 mark at some ABA and abiotic stress-responsive genes in Arabidopsis. Mutation in HDA6, Arabidopsis RPD3-type histone deacetylase, was found to cause oversensitive responses to ABA and salt stress. Under ABA and salt stress treatments, the level of H3K4me3 increased at ABA and salt-inducible genes in wild type but not in an HDA6 mutant. This suggests that initiation of H3K4me3 happens through ABA and salt stress, and HDA6 is essential for this induction (L.-T. Chen et al., 2010).
In soybean plants, three salt-stress inducible genes encoding transcription factors also showed induction in H3K4me3 and reduction in both H3K9me2 and DNA methylation in response to salinity stress (Y. Song et al., 2012). Under control conditions, all of these TFs genes were shown to be hyper-methylated but became demethylated during salt stress treatment (150 mM NaCl). Therefore, dynamic changes of histone modifications along with DNA methylation contribute to salt stress responses in soybean (Y. Song et al., 2012).

Figure 6. Summary of H3K4me3 roles in plants. H3K4me3 is one of the important histone modification in plants involved in different aspect of plant development and stress responses.

The distribution pattern of H3K4me3 and H3K9me2 was also investigated in salt-stressed Brassica napus plants (Fang et al., 2017). It was demonstrated that low concentrations of NaCl (25 mM) caused an increase in the level of H3K4me3 and a decrease in the level
of H3K9me2 but high NaCl concentrations (50 and 100 mM) resulted in the opposite effect. These results indicate that Arabidopsis and *Brassica napus* demonstrate different epigenetic changes under salt stress (Fang et al., 2017). A summary of the role of H3K4me3 is shown in (Figure 6).

**H3K4me3 relationship with other epigenetic modifications**

H3K4me3 has been shown to interact with other histone modifications, as well as other histone variants. Among them, H2A.Z and H3K27me3 are the most studied and show distinct correlations with H3K4me3. H2A.Z is a conserved histone variant among eukaryotes, having multiple roles in chromatin structure and gene expression (Jackson et al., 1996). Genome-wide, it was shown that H2A.Z was preferentially associated with H3K4me3 and has partial association with H3K27me3 in both rice callus and seedlings. In addition, the combination of H2A.Z with either H3K4me3 or H3K27me3 caused a difference in gene expression (K. Zhang et al., 2017). H2A.Z was found to be enriched in the promoter and 5' UTR regions of genes in Arabidopsis. At the promoter regions there is preferential enrichment of H3K4me3 as well, indicating the co-localization of these two marks. In contrast to promoter regions, there was no significant correlation between H2A.Z and H3K4me3 in enhancer regions. However, the activity of enhancers was demonstrated to be repressed by H2A.Z through stimulating H3K27me3 and inhibiting H3K4me3 deposition at enhancer regions. Interestingly, it was shown that mutation of *ARP6* leads to reduction of H3K4me3 around the TSS, suggesting H2A.Z deposition may regulate the level of H3K4me at the TSS. The 864 up-regulated genes with reduced H2A.Z in *arp6* were shown to have increased H3K4me3 and high nucleosome occupancy at the -1 nucleosome and decreased level of H3K27me3 at the +1 nucleosome. For that
reason it was suggested that H2A.Z may repress genes by regulating the H3K4me3 and H3K27me3 levels at the TSS (Xiaozhuan Dai et al., 2017). Recently, Cai et al. (2019) proposed that H2A.Z and H3K4me3 have antagonistic roles in the modulation of anthocyanin biosynthetic genes. The mutation of H2A.Z-encoded genes (hta9 hta11) or SWR1 complex (arp6 and pie1) results in activating anthocyanin biosynthetic gene expression by reduction and induction of H2A.Z and H3K4me3 levels, respectively (Hanyang Cai et al., 2019). In addition, ARP6 mutation reduced the expression of two vegetative phase change genes, MIR156A and MIR156C, and H3K4me3 level at these two genes, suggesting H2A.Z promotes H3K4me3 at these genes, inducing their expression (M. Xu et al., 2018). These findings are in contrast with the Cai et al. study (Hanyang Cai et al., 2019) which reported an antagonistic relationship between H2A.Z and H3K4me3 at anthocyanin biosynthetic genes.

In plants, Polycomb Repressive Complex 2 (PRC2) is a methyltransferase responsible for the deposition of H3K27me3, which represses transcription and acts antagonistically to H3K4me3. Although the distribution and relationship between these two modifications have been investigated, a conclusive pattern remains elusive. Genome-wide analysis of both marks within Arabidopsis seedlings seems to indicate a mutually exclusive relationship (X. Zhang et al., 2009). However, when considering specific stress-induced “memory genes,” H3K27me3 and H3K4me3 are found to be independent of one another, showing the capability of occurring at the same gene at the same time (N. Liu et al., 2014). Yet other evidence suggests that co-localization of these two modifications is the most likely scenario. Within the woody species, Eucalyptus grandis, it was found that the co-occupancy of these modifications is directly linked to genes involved in the
development of xylem tissue (Hussey et al., 2015). In another study by Moliter et al., (2014) it was shown that seed germination in Arabidopsis is controlled by the transition from H3K4me3 to H3K27me3 at certain seed dormancy related genes that require silencing (Molitor et al., 2014). This evidence gives rise to the notion of “bivalent histones,” suggesting co-regulation of modifications, whereas certain ratios between marks dictate transcriptional control.

In Arabidopsis, a chromatin state study using nine histone modifications, three histone variants, nucleosome density, genomic CG, and CG methylation showed a coinciding presence of H3K4me3 with H3K27me3 (Sequeira-Mendes et al., 2014). Nine chromatin states (4 types of gene enriched chromatin states, 2 heterochromatin states types, and 3 H3K27me3 enriched chromatin states) have been described to understand chromatin complexity. The highest level of H3K4me3 and its coexistence with other active histone modification marks (H3K4me2, H3 acetylation, H3K36me3, and H2Bub) and histone variants (H3.3 and H2A.Z) was observed in chromatin state 1. Chromatin state 1 was shown to be enriched at genic regions, especially TSSs with low nucleosome density. The unexpected combination of H3K4me3 with H3K27me3 was detected in chromatin state 2. Interestingly, chromatin state 2 shows a similar pattern of activating marks as chromatin state 1, however at a significantly reduced level, in addition to low expression levels. A sequential ChIP experiment performed on state 2 chromatin once more demonstrated the co-occurrence of H3K4me3 and H3K27me3 in some regions of chromatin. The moderate level of H3K4me3 with H3K4me1, H2Bub, and H3K36me3 was observed in chromatin state 3, which is associated with transcriptional elongation.
Additionally, RNA analysis showed that states 1 and 3 had the highest transcript levels, whereas state 2 was among the lowest (Sequeira-Mendes et al., 2014).

The overlap between H3K4me3 and H3K9 acetylation has been reported in many studies (Brusslan et al., 2015; Ding, Fromm, et al., 2012; N. Liu et al., 2014; Sequeira-Mendes et al., 2014). More specifically, a positive correlation of these two euchromatic marks at the same loci was observed in Arabidopsis. Although co-localization was discovered, H3K4me3 peaks demonstrated a more pronounced enrichment when compared to H3K9ac (Brusslan et al., 2015). Another study showed that sharp reduction of H3K4me3, H3K9ac, and H3 100 bp upstream of the TSS corresponds to nucleosome depletion in that region to assist binding of the transcriptional machinery. Also, H3K4me3 and H3K9ac peaks at the TSS showed a considerable correlation with transcript levels. Gene ontology (GO) enrichment analysis revealed that genes with narrow and dense distribution of both modifications are constitutively expressed genes associated with translation. However, broad distribution of these two modifications near coding regions is associated with adjustable expression of genes involved in photosynthesis, carbohydrate metabolism, and defense responses (Ha et al., 2011).

The rice genome-wide distribution of H3K4me2/3, H3K9ac, and H3K27ac are shown to be enriched in the genic region (exon, intron, and UTR regions) as well as non-transposable-element genes. In addition, a positive correlation between gene expression and these modifications was observed. Together, simultaneous presence of these histone modifications within transcribed regions of the rice genome was exhibited (Du et al., 2013).
The depletion of CHG and CHH methylation was observed in H3K4 mono/di/trimethylated genomic regions, whereas the enrichment of CG methylation was observed in H3K4me1 regions. Mutation in METHYLTRANSFERASE 1 (MET1) showed higher levels of H3K4 mono/di/trimethylation in pericentromeric heterochromatic regions, but there was no impact in the level of H3K4me in hyper-H3K4me or DNA-hypomethylated regions. Also, at nongenic silent regions, all three types of H3K4 methylation were mutually exclusive with DNA methylation. H3K4me2/3 was negatively correlated with DNA methylation, however, H3K4me1 was positively correlated with CG DNA methylation in genic regions (X. Zhang et al., 2009).

**Epigenetic Regulation of Pi Starvation Responses**

It was previously reported that Pi starvation responses can be regulated by various epigenetic mechanism such as the H2AZ variant, histone acetylation and DNA methylation (C.-Y. Chen et al., 2015; Kuo et al., 2014; Secco et al., 2015; Aaron P Smith et al., 2010; Yong-Villalobos et al., 2016; Yong-Villalobos et al., 2015).

Few studies have shown the relationship between DNA methylation and Pi starvation. In response to Pi deficiency, the transposable elements (TEs) adjacent to PSI genes are deactivated by changing the DNA methylation in the rice genome. Hypermethylation of TEs in response to Pi starvation possibly will prevent the damaging effect of TEs near PSI genes and in this way induce PSI gene expression (Secco et al., 2015). Changes in DNA methylation patterns were also demonstrated in the Arabidopsis genome (Yong-Villalobos et al., 2016; Yong-Villalobos et al., 2015). Global DNA methylation percentage of shoot and root samples were increased under Pi starvation condition. Interestingly, gene expression of MET1, DOMAINS REARRANGED
METHYLASE 1 (DRM1), DOMAINS REARRANGED METHYLASE 2 (DRM2), REPRESSOR OF SILENCING 1 (ROS1), DEMETER LIKE 2 (DML2), and DEMETER LIKE 3 (DML3) were up-regulated by Pi deficiency suggesting that DNA methylation activities are altered by Pi starvation. Furthermore, mutation in DNA methyltransferase genes impairs the expression of Pi starvation responsive genes. In addition, genome-wide changes of DNA methylation observed under Pi starvation conditions in Arabidopsis was mostly associated with transcription changes. They also showed that PHR1 directly targets DNA methyltransferase gene expression in Pi starvation (Yong-Villalobos et al., 2016; Yong-Villalobos et al., 2015).

It was validated in Smith et al. (2010) that under Pi starvation treatment the expression of a number of PSR genes are regulated by H2A.Z deposition. Mutation of ARP6 (Arabidopsis nuclear actin-related protein 6 ), a key component of SWR1, increases PSR gene transcription by reducing the level of H2A.Z (Aaron P Smith et al., 2010). As mentioned earlier, a number of PSR genes were induced in the AtIPK1 mutant, and induction of this gene was associated with a reduction in H2A.Z (Kuo et al., 2014).

Recently, the involvement of H2A.Z in modulating Pi starvation responses has been supported by investigating the global distribution of H2A.Z deposition in a rice wild-type and an ARP6-RNAi line under Pi starvation (Zahraeifard et al., 2018). Interestingly, distribution of H2A.Z was similarly changed in both Pi deficiency and ARP6-RNAi plants compared to wild type. The role of H2A.Z in gene expression was shown to be dependent on the location of its deposition. Deposition of H2A.Z in the gene body results in repressing PSR genes while H2A.Z at the TSS is positively or negatively correlated with gene expression (Zahraeifard et al., 2018).
A study of Arabidopsis histone deacetylase 19 (HD19) implicated the involvement of histone acetylation in Pi starvation responses. HDA19 regulates a number of PSRs such as a lipid remodeling by affecting SPX genes and root hair density by controlling the epidermal cell length (C.-Y. Chen et al., 2015). Another study reported a connection between H3K4me3 and Pi starvation by investigating the role of the ALFIN-LIKE 6 (ALF6) H3K4me3 reader in PSRs. The alf6 mutant showed a root hair elongation defect under Pi starvation. It was suggested that ALF6 regulates root hair elongation under Pi-deficient conditions through modulating the stability of ETC1 transcript (Chandrika et al., 2013). This model suggests that the expression of ETC1 was up-regulated by binding ALF6 to H3K4me3 deposited at the ETC1 promoter. Since the ALF6 transcript level was not altered by Pi deficiency, how Pi starvation can cause these processes is still unclear. Although the significant role of H3K4me3 in eukaryotic gene expression has been determined, studies on the genome-wide distribution pattern of H3K4me3, and its interaction with other histone modifications involved in regulation of P starvation responses are lacking.
Chapter 2. Genome-wide Distribution of H3K4 tri-methylation, H2A.Z and RNAseq under Control and P deficiency Conditions

Introduction

The H3K4me3 mark is mainly associated with genic regions and its distribution is conserved among organisms including yeast, plants and animals. H3K4me3 enrichment around gene TSSs is the most noticeable landscape of this mark, which positively correlates with transcription (Barski et al., 2007; Bernstein et al., 2002; X. Li et al., 2008; C. L. Liu et al., 2005; Ram et al., 2011; Van Dijk et al., 2010; X. Zhang et al., 2009). The powerful approach to understand the histone modification role in transcription is to investigate the relationship between different histone modifications. The coexistence of H3Kme3 and H3K27me3 with opposite function in mammalian stem cells lead to them being called bivalent histone modifications (Bernstein et al., 2006; F. Li et al., 2018). Conversely, genome-wide co-localization of H3K4me3 and H3K27me3 does not occur in the Arabidopsis genome and a mutually exclusive relationship exists between H3K4me3 and H3K27me3 at many genes (X. Zhang et al., 2009). However, a role for co-localization of these two modifications is found in the development of xylem tissue, vegetative to reproductive transition and seed germination genes (Hussey et al., 2015; X. Liu et al., 2015a; Molitor et al., 2014). In addition, an independent role of H3K4me3 and H3K27me3 and not a mutually exclusive relationship between them occurs in the Arabidopsis genome at dehydration stress responding memory genes (N. Liu et al., 2014).

Studies in yeast, mammals and plants show a strong correlation between H3K4me3 and H2A.Z. Colocalization of H2A.Z and H3K4me3 is seen at both promoters and enhancers in embryonic stem cells (Hu et al., 2013; Ku et al., 2012). However in plants, H3K4me3 only co-localizes with H2AZ at promoters but not at enhancer regions.
(Xiaozhuan Dai et al., 2017). In mammalian and yeast cells, H3K4me3 and H2A.Z marks have peaks around the TSS of genes, whereas in plant genomes they are found at both TSS and gene bodies, suggesting additional roles of these marks in plants (To & Kim, 2014). The H3K4me3 and H2A.Z colocalization has been investigated in plant genomes, but the role of H3K4me3/H2A.Z ratio and gene expression levels are lacking.

It has been proposed in many organisms including rice, soybean and Arabidopsis that H3K4me3 modification is involved in the regulation of responses to a variety of stresses including drought, heat, submergence and salt stress. However, whether this mark is involved in response to Pi deficiency is unknown.

Materials and Methods:

Plant material and growth conditions

Sterilization and pre-germination (1 day at 37°C followed by 2 days at 28°C) were carried out on rice cultivar Nipponbare (Oryza sativa ssp. japonica) seeds (Zahraeifard et al., 2018). Seeds were transferred to 12-h light/12-h dark, at 30 °C/22 °C condition to germinate for two weeks. Hydroponic systems with Yoshida Rice culture modified were performed to grow the seedlings solution as described (Secco et al., 2013b; Yoshida et al., 1971). The solution was changed every 7 days. After 21 days, seedlings were used for a 24-hour Pi deficiency treatment (solution without NaH2PO4).

Chromatin Immunoprecipitation sequencing

For each replicate, 4 g of frozen shoot tissue from 24-hours Pi deficiency and control treatment was used to perform ChIP experiments as described earlier (Aaron P Smith et al., 2010; Widiez et al., 2014) with the antibody (Millipore; lot number 2648189) against H3K4me3 and input genomic DNA as a control. Two biological replicates were
used for both input and antibody treatments. Although ChIP has been applied to a wide range of model organisms, it remains a challenging technique for plants. Because of structural anatomical differences between plant and animal cells, such as rigid cell walls, high level of cellulose and lignin, and large vacuoles in plant cells, a plant specific approach was needed to establish efficient ChIP protocols for plant systems. Therefore, we examined available protocols and developed an efficient ChIP protocol for studying the H3K4me3 mark. It was suggested in most ChIP protocols that it is better to start with fresh tissue compared to frozen tissue. By comparison, in our hands, the use of fresh tissue did not affect the DNA template yield (Table 2).

Table 2. DNA quantification in fresh and frozen samples

<table>
<thead>
<tr>
<th></th>
<th>Fresh Rice shoot</th>
<th>Frozen Rice Shoot</th>
</tr>
</thead>
<tbody>
<tr>
<td>+2μl H3K4me3-AB</td>
<td>0.210 ng/μl</td>
<td>0.160 ng/μl</td>
</tr>
<tr>
<td>+1μl H3-AB</td>
<td>0.122 ng/μl</td>
<td>0.156 ng/μl</td>
</tr>
<tr>
<td>-AB</td>
<td>&lt;0.0005 ng/μl</td>
<td>0.170 ng/μl</td>
</tr>
<tr>
<td>Input</td>
<td>3.42 ng/μl</td>
<td>7.42 ng/μl</td>
</tr>
</tbody>
</table>

First nuclei were extracted and cross-linked. Frozen plant tissues were weighed in 1 g aliquots, ground in liquid nitrogen, and transferred into Nuclear Isolation Buffer (10mM HEPES pH 7.6, 1M sucrose, 5m KCL, 5m Mgcl2, 5mM EDTA) with 1% Formaldehyde, 14 mM β-ME, 0.6% Triton X-100 and 1/2 of a large protease inhibitor tablet Before cross-linking to cross-link at room temperature for 10 minutes. 1.7 mL of filter-sterilized 2M glycine was added to terminate cross-linking. After incubation at room temperature for 5
minutes, the solution was filtered through Miracloth into fresh tubes on ice. The samples were centrifuged at 3000 g for 20 min at 4°C. The white pellets were resuspended in 300 uL Nuclear Isolation Buffer without Formaldehyde, Triton X-100, β-ME, and the large tablet, and transferred into 15% Percoll Solution (Percoll, 10 mM HEPES, 1 M sucrose, 5 mM KCl, 5 mM MgCl2, 5 mM EDTA). The resuspensions were centrifuged at 3000 g for 5 minutes at 4°C. Percoll separates the nuclei from the carried over chloroplast. The pellets were resuspended in the 600 uL of cold Nuclear Lysis Buffer (50 mM Tris-HCl, 1% SDS, 10 mM EDTA, 1/2 of a mini protease inhibitor tablet). The chromatin solutions were sonicated with a Branson Digital Sonifier 250 at power 5 in 11 cycles of on/off within 10 seconds follow spinning at 12,000 g at 4 °C for 10 minutes to remove debris. The supernatant was transferred to new tubes, and aliquots of each sample were set aside to serve as "input DNA controls". Samples were diluted in 1000 µL ChIP Dilution Buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, 167 mM NaCl), and divided equally among three tubes. The nuclear lysate should be diluted at least 10-fold to deactivate the SDS. The SDS concentration has to be less than 0.1% to prevent the antibody denaturation. 50 µL Protein A-Agarose beads was used to reduce background caused by non-specific absorption of irrelevant proteins. Samples with A-Agarose beads were incubated at 4 °C with gentle rotation, after 2 hour beads were removed by centrifugation at 12,000 g at 4 °C for 30 seconds. Two to five µl of commercially-available antibody recognizing H3K4me3 and IgG (control) was added to two tubes, and the other tube served as the "no antibody control". All tubes were incubated at 4°C overnight with gentle agitation. 50 µL Protein A-Agarose beads were added to each tubes to collect immune complexes by incubation at 4°C for 2 h with gentle agitation.
Immunocomplexes were washed several times with sequence of protein wash buffers. Immunocomplexes were centrifuged at 3,800 g at 4 °C for 30 seconds, supernatants were removed and pellets were re-suspend in 1 ml Wash Buffer 1 (1% Triton X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris-HCl, 150 mM NaCl) following a 30-second spin beads at 3,800 g at 4 °C. For a second time, supernatants were carefully removed and pellets were re-suspend in 1 ml Wash Buffer 1 and agitate at 4 °C for 5 minutes. The immunocomplexes were collected by centrifugation at 3,800 g at 4 °C for 30 seconds. These washing steps were repeated with Wash Buffer 2 (1% Triton X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris-HCl and 500 mM NaCl), Wash Buffer 3 (1% Sodium Deoxycholate, 1% IGEPAL CA-630, 1 mM EDTA, 10 mM Tris-HCl, 250 mM LiCl), and Wash Buffer 4 (1 mM EDTA, 10 mM Tris-HCl).

After washing the beads, supernatants were carefully removed and 250 µL Elution Buffer (1% SDS, 0.1 M NaHCO) were added to pellets to separate the beads. The samples were incubated at 4 °C with gentle agitation for 15 minutes, and centrifuged at 3,800 g at room temperature for 2 minutes. Supernatants were then transferred to new tubes. Another 250 µL Elution Buffer were added to the beads and the supernatant were added to the previous ones. To make the input DNA control, 490 µL Elution Buffer was added to 10 µL ‘input DNA control’. To reverse crosslinking and digest the proteins, NaCl and Proteinase K were used. At first 20 µL 5 M NaCl was added to the samples and incubated at 65 °C for 4 hours. Next Proteinase K solution (10 µL 0.5 M EDTA pH 8, 20 µL 1 M Tris-HCl pH 6.5 and 20 µg Proteinase K) was added to each sample and incubated at 50 °C for 2 hours.
520 µL phenol:chloroform solution was added to each sample to remove lipids and proteins, and incubated on shaker at room temperature for 5 minutes, following a spin at 12,000 g at 4 °C for 15 minutes. 400 µL of the aqueous layer was then moved to a new tube and 50 µL 3 M NaOAc pH 5.2, 2 µL glycogen, and 1 mL cold 100% ethanol were added. The combination was incubated at -20 °C overnight. Each sample was spun at 12,000 g at 4 °C for 20 minutes and supernatants were removed. 75% ethanol was added to the pellets and centrifuged at 12,000 g at 4 °C for 10 minutes, then the supernatants removed. The dried pellets were dissolved in 30 µL EB containing 60 ug/mL RNaseA and incubated at 37 °C for 30 minutes.

DNA was precipitated and cleaned by NaOAc, glycogen, and EtOH. Purification of ChIP DNA was carried out via Zymo clean and Concentrator kit. Libraries were constructed using 1:20 diluted adaptor from Kapa Biosystems Hyper Library Construction Kit and 10 cycle DNA amplification. Libraries were quantitated (qPCR) and multiplex together and single-end sequencing was completed in HiSeq2500 (Illumina) using a HiSeq SBS sequencing kit (version 4) for 101 cycles.

**RNA sequencing**

0.1g of frozen shoots from 24-hours Pi deficiency and control treatment were homogenized in liquid nitrogen. Total RNA was extracted via the RNeasy Plant Mini kit (Qiagen) from two replicates of each Pi deficiency and control treatment samples. The RNA-libraries were constructed and quantitated using Illumina's TruSeq Stranded mRNAseq Sample Prep kit (Illumina) and qPCR respectively. The samples were single-end sequenced with HiSeq2500 (Illumina) using a HiSeq SBS sequencing kit version 4. for 101 cycles at University of Illinois at Urbana-Champaign’s DNA service.
RNA-Seq analysis

A minimum of 58 million high-quality RNA-seq reads (100-bp single end) per sample were mapped to MSU Rice Genome Annotation Release 7.1 (MSU7.1) using Bowtie2 tools (Langmead & Salzberg, 2012). Fragments per kilobase of transcript per million mapped reads (FPKM) were calculated with Cuffdiff tool (Trapnell et al., 2012). DESeq2 was applied to identify differentially expressed gene (DEGs) (Love et al., 2014). The stringent cutoff (adjusted P-value < 0.001) recommended for a small-sample RNA-seq experiment was used (Soneson & Delorenzi, 2013). Gene ontology (GO) terms enriched in DEGs were analyzed with BiNGO and visualized with Cytoscape (Maere et al., 2005).

ChIP-Seq analysis

Around 147 million total ChIP-seq reads for H3K4me3 were checked and cleaned using FastQC and Trimmomatic-0.33 (Andrews, 2010; Bolger et al., 2014). The cleaned data were aligned using Bowtie to MSU Rice Genome Annotation Release 7.1 (MSU7.1) with one mismatch allowed to retain uniquely mapped read. The SICER software package (Zang et al., 2009) was used to define the H3K4me3 enrichment regions with the following parameters (W = 200, G = 200, FDR < 1.00E-02). The input genomic DNA was used as a background control. Differential H3K4me3 enrichment peaks between control and Pi deficiency samples were determined using SICER-df.sh shell script (W = 200, G = 200, FDR < 1.00E-02). We defined the existence of peaks with PCGs if 50% of peaks overlapped with PCGs (including 250 bp upstream and downstream) using BEDTOOLS intersect (Quinlan & Hall, 2010). Genome-wide distribution patterns of H3K4me3 and the published profiles of two marks, H2A.Z and H3K27me3, (Zahraefard et al., 2018; L.
Zhang et al., 2012) were visualized using ngs.plot (Shen et al., 2014). The K-means program in ngs.plot was used to find different pattern of H3K4me3 in the PCGs. Gene ontology (GO) terms enriched in each clusters were analyzed with BiNGO and visualized with Cytoscape (Maere, Heymans et al. 2005).

**Results**

**H3K4me3 is prominent at the 5’ end of rice protein-coding genes and co-localizes with the H2A.Z histone variant**

At first, to study the whole genome distribution of H3K4me3 under control conditions, we performed ChIP-seq on 36-day rice seedlings. Two biological replicates were used and in total, 52.7 million uniquely aligned reads were generated ($r = 0.72$; Pearson correlation coefficient; Table 3).

Table 3. Summary of ChIP-seq and RNA-seq libraries (short reads). The number of total and uniquely mapped reads for shoots from 36-day-old rice seedlings under control conditions (Ctrl) or following a 24-hour P-deficiency treatment (-P). Each sample contains two replicates (rep).

<table>
<thead>
<tr>
<th></th>
<th>Total reads</th>
<th>Uniquely mapped reads (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctrl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me3-Rep1</td>
<td>41066109</td>
<td>78.5</td>
</tr>
<tr>
<td>H3K4me3-Rep2</td>
<td>32132970</td>
<td>63.7</td>
</tr>
<tr>
<td>RNA-seq-rep1</td>
<td>58992471</td>
<td>67.57</td>
</tr>
<tr>
<td>RNA-seq-rep2</td>
<td>65082473</td>
<td>68.36</td>
</tr>
<tr>
<td>-P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3K4me3-Rep1</td>
<td>37099052</td>
<td>77.9</td>
</tr>
<tr>
<td>H3K4me3-Rep2</td>
<td>36468290</td>
<td>78.5</td>
</tr>
<tr>
<td>RNA-seq-rep1</td>
<td>64365952</td>
<td>68.17</td>
</tr>
<tr>
<td>RNA-seq-rep2</td>
<td>62848509</td>
<td>68.48</td>
</tr>
</tbody>
</table>

Rice genes were categorized into four groups based on the MSU7 genome annotation: protein-coding genes (PCG), transposable elements (TE), transposable
element-related genes (TEG), and ‘pseudogenes’ (PSG, i.e. annotated genes that are neither expressed nor transposable elements) (Kawahara et al., 2013; Q. Zhang et al., 2018). Four gene types with the distinct H3K4me3 deposition pattern was shown in Figure 7A. A prominent peak of H3K4me3 was present immediately downstream of the transcription start sites (TSS) of PCG, whereas H3K4me3 abundance was relatively low at PSG, TEG, and TE (Figure 7A).

Clustering analysis on H3K4me3 abundance at all PCGs normalized to length showed 4 apparently distinct groups of genes with differences in H3K4me3 peak width (Figure 8A). However, sorting all PCG according to gene length revealed a strong correlation between H3K4me3 pattern and gene size (Figure 8B), indicating that the general pattern of H3K4me3 at PCG is relatively consistent among all PCG. Next, PCGs were clustered based on H3K4me3 abundance at the TSS (Figure 9). Gene Ontology (GO) term enrichment analysis showed that the clusters with high and moderate H3K4me3 abundance (cluster A (1 and 2)) were enriched with housekeeping terms, including intracellular, translation, cytoplasm, and organelle (FDR<0.05), whereas clusters with relatively low H3K4me3 abundance (cluster B (3 and 4)) were enriched in stress-responsive GO terms, including response to stimulus, transferase activity, kinase activity, and cell wall (Figure 10, Supplemental Dataset 1).
Figure 7. H3K4me3 abundance is predominantly associated with the transcription start site (TSS) and co-localizes with H2A.Z. Distribution of H3K4me3 (A) and H2A.Z (B) among four gene types in shoots from rice seedlings grown under control conditions. Control input reads were used for ChIP-Seq read normalization. PCG, protein coding genes; PG, pseudogenes; TE, non-expressed transposable element-related genes; TEG, expressed transposable element-related genes. (C) Scatter plot of read counts from H3K4me3 and H2A.Z samples (Pearson correlation = 0.77). (D) Venn diagram showing the number of H3K4me3- and H2A.Z-enrichment peaks and the overlap.

The H3K4me3 patterns at different gene types (Figure 7A) are similar to those we recently demonstrated for the H2A.Z histone variant (Zahraeifard et al. 2018; Figure 7B). A key difference is that the abundance of H3K4me3 is relatively higher than that of H2A.Z at TEG and TE (Figure 7). This may reflect the presence of DNA methylation at TEG and TE that antagonizes H2A.Z deposition (Conerly et al., 2010; Zahraeifard et al., 2018;
Zemach et al., 2010; Zilberman et al., 2008). To examine the association of H3K4me3 and H2A.Z in more detail, we first computed a correlation coefficient analysis using deepTools (Ramírez et al., 2016). As shown in Figure 7C, H3K4me3 and H2A.Z were strongly correlated in the rice genome ($r = 0.77$; Pearson correlation coefficient). Next we identified and compared distinct H3K4me3 and H2A.Z peaks using SICER (Zang et al., 2009) and BEDTools (Quinlan & Hall, 2010). This analysis identified 32,886 H3K4me3 peaks and 44,804 H2A.Z peaks, of which 30813 (93% of H3K4me3 peaks) overlapped, showing substantial co-localization of these chromatin marks in the rice genome (Figure 7D).

**Figure 8.** The pattern of H3K4me3 at gene body of PCG is relatively consistent among all PCG. (A) Heat map of K-means H3K4me3 clusters across gene body. (B) H3K4me3 heat map and average plot based on the gene length in gene body of protein-coding genes (PCG) from 500 bp upstream of the transcription start site (TSS) to 500 bp downstream of the transcription termination site (TES). Five quintiles were ordered by gene length (Q1-Q5). Control input reads were used for ChIP-Seq read normalization.
Figure 9. H3K4me3 abundance varies at the transcription start site (TSS). Heat map and average plot of k-means H3K4me3 clusters around the TSS (50 bp upstream and downstream of the TSS)

H3K4me3 and H2A.Z abundance have distinct correlations with gene expression in rice

To compare H3K4me3 abundance with gene expression, we carried out RNA-seq experiments on two replicates of shoot tissues from 36-day rice seedlings (Table 3). PCG were ranked according to FPKM and divided into five expression quintiles, as well as a sixth group of genes that were not expressed (i.e. FPKM = 0). We found a relationship between transcript abundance and H3K4me3 localization around the TSS (Figure 11A,B), consistent with a positive role for H3K4me3 in gene expression (Barski et al., 2007; Bernstein et al., 2002; Santos-Rosa et al., 2002a; Van Dijk et al., 2010; X. Zhang et al., 2009). In contrast, transcript abundance was negatively correlated with TES and gene body-localized H3K4me3 (Figure 11A,B). This trend differs from a study in Arabidopsis in
which H3K4me3 abundance at both the TSS and TES positively correlated with gene expression (Van Dijk et al., 2010)

Figure 10. Over-represented functional categories of Cluster A and cluster B genes. Bingo and Cytoscape were used to identify and visualize enriched GO terms. Circle color shows p-value of enrichment.

Associations of H3K4me3 with other chromatin modifications, including H2A.Z and H3K27me3, have been observed in rice and Arabidopsis (Xiaozhuan Dai et al., 2017; K. Zhang et al., 2017). To further explore how H3K4me3 in combination with these two modifications relate to gene expression, we plotted H3K4me3 abundance along with the
H2A.Z profile from our previous work (Zahraeifard et al., 2018), and a published profile of H3K27me3 carried out on leaves from 3-week-old rice seedlings (L. Zhang et al., 2012) (Figure 11C). The H3K27me3 profile exhibited a consistent trend in which H3K27me3 is present as moderate peaks at the 5’ and 3’ flanking regions of genes with low or no expression, but is progressively depleted as expression increases. In contrast, H2A.Z and H3K4me3 exhibited more prominent peaks and more complex associations with transcript abundance. H3K4me3 abundance near the TSS was negatively correlated with gene expression, whereas gene-body H3K4me3 showed a positive correlation. On the other hand, H2A.Z exhibited a general negative correlation, except for genes not expressed, in which H2A.Z abundance was relatively low.

**Characterizing the H3K4me3 distribution changes through 24 hours of Pi deficiency**

To profile the effect of short-term Pi deficiency on the H3K4me3 distribution in the rice genome, we also carried out the ChIP-seq method on 24-hours Pi deficiency treated rice seedlings. Similar to control samples, two biological replicates were used and in total, 57.5 million uniquely aligned reads are generated (r = 0.97; Pearson correlation coefficient; Table 3). The average distribution of H3K4me3 in both control and 24- hour Pi deficiency samples were plotted across the gene body of PCG. We assigned two separate regions to the gene structure, TSS proximal region (“TSS”, 250 bp upstream to 500 bp downstream of TSS) and gene body region (“GB”, 500 bp downstream of TSS to 250 bp downstream of TES). We found the reduction in the H3K4me3 abundance around the TSS in response to Pi deficiency, whereas there was increased and decreased H3K4me3 deposition in GB (Figure 12A). We also plotted H3K4me3 abundance for housekeeping (cluster A) and stress-related (cluster B) genes in both control and Pi
deficiency. As shown in Figure 12B the height of broad H3K4me3 peaks inside of stress-related gene body was decreased, while there was a reduction and increase in H3K4me3 level at TSS and GB in housekeeping genes respectively.

Figure 11. Correlations between H3K4me3 and H2A.Z distribution and gene expression in rice shoots. Heat map (A) and distribution of H3K4me3 (B) across the gene body in control samples for six gene groups ordered based on transcript abundance level (FPKM), defined as 1st (highest) to 5th (lowest) and no expression (zero). (caption cont’d)
(C) Distribution of H3K4me3, H3K27me3 and H2AZ across the gene body from 500 bp upstream of the TSS to 500 bp downstream of the TES. Control input reads were used for ChIP-Seq read normalization.

Figure 12. Difference of H3K4me3 enrichment pattern across rice protein coding genes (PCG) under 24-hours of Pi deficiency. (A) Average plot of H3K4me3 for all PCG and (B) cluster A and B in control (Ctrl) or Pi deficiency (–Pi) samples.

To further investigate the H3K4me3 distribution changes in response to Pi deficiency in PCG, regions that were significantly enriched with H3K4me3 were identified by using SICER software in both control and Pi deficiency samples. In total 32886 and 27319 enriched regions (peaks) were recognized in control and Pi deficiency samples respectively. In addition, 26243 decreased and 770 increased regions were found in the whole rice genome. We annotated the differential regions. 18407 and 617 PCGs overlapped with at least one decreased and increased H3K4me3 regions respectively. Among 617 PCGs, 547 of them (89%) has increased regions at their GB, however, only 70 PCGs (11%) have increased regions at TSS. 11185 (61%) and 7222 (39%) PCGs
contained decreased H3K4me3 regions at their GB and TSS respectively (Table 4). Overall, these data showed that genome-wide H3K4me3 profile was changed in response to 24 hours of Pi deficiency and in PCG, these changes differ for GB and TSS regions.

Table 4. Summary of differentially H3K4me3 peak and PCG number under Pi deficiency

<table>
<thead>
<tr>
<th></th>
<th>Number of peaks</th>
<th>Number of PCG</th>
<th>PCG % in GB and TSS regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decreased</td>
<td>26243</td>
<td>18407</td>
<td>39% TSS 61% GB</td>
</tr>
<tr>
<td>Increased</td>
<td>770</td>
<td>617</td>
<td>11% TSS 89% GB</td>
</tr>
</tbody>
</table>

**Transcriptional changes under Pi deficiency**

To examine the impact of 24 hours Pi deficiency, we conducted an RNA-seq experiment by collecting shoot tissues from Pi deficiency and control samples using two replicates. Mapping of RNA-seq single-end reads to the MSU Rice Genome Annotation Release 7.1 (MSU7.1) provided on average, 68% uniquely mapped reads, while 11.6% reads were aligned multiple times (Table 3). Differential expression analysis with DESeq2 revealed that 1385 genes were significantly differentially expressed in response to 24 hours of Pi deficiency. We identified 694 up-regulated genes and 691 down-regulated genes (adjusted P-value < 0.001) (Figure 13A,B).

We next identified and visualized GO terms enriched among DEGs used BiNGO (Figure 14A, and B, Table 5). GO term enrichment networks of DEGs, created based on an ancestor-child relationship, are shown in Figure 14.
Figure 13. Identification of differentially expressed genes in response to 24 hours of Pi deficiency. (A) Differentially expressed gene numbers under P deficiency in rice. (B) MA plots of RNA-seq data by DESeq2 comparing P deficiency samples to control samples. The x-axis represents mean expression for each gene; the y-axis represents log2 fold change. Red dots indicate differentially expressed genes (Padj<0.001)

Overall, up- and down-regulated genes were shown to have functional divergences. Response to stress, lipid metabolic process, regulation of cellular biology and signal transduction GO terms were enriched in up-regulated genes, whereas growth, cell-cell signaling and lipid, carbohydrate and secondary metabolic process were enriched in down-regulated genes. Lipid metabolism was overrepresented in both up- and down-regulated gene networks. Comparison between the lipid metabolism network for up- and down-regulated genes showed that different pathways are involved in this network. Genes involved in cutin, suberin, and wax biosynthesis are among down-regulated lipid metabolism network whereas carotenoid biosynthesis genes and alpha-Linolenic acid metabolism are in up-regulated lipid metabolism network. Genes encoding phosphoethanolamine/phosphocholine phosphatase (Hur et al., 2007) and monogalactosyldiacylglycerol synthase (Mehra et al., 2016) that are involved in Pi homeostasis by phospholipid degradation and lipid remodeling, respectively, were also
up-regulated in our data. In addition, phosphoenolpyruvate carboxylase, which has a role in forming an intermediate carbon source for organic acid biosynthesis in response to Pi deficiency, was detected among up-regulated genes. OsAOS1, which encodes allene oxide synthase, the key jasmonic acid (JA) biosynthesis enzyme, (Hibara et al., 2016) allene oxide cyclase (OsAOC) (Riemann et al., 2013) and two lipoxygenase (LOX) genes (L. Liu et al., 2017) which are involved in jasmonic acid biosynthesis were up-regulated by 24 hours Pi deficiency. OsMYC2 which is known to be involved in the regulation of JA signaling is also up-regulated (Kazan & Manners, 2013; Uji et al., 2016). The induction of JA biosynthesis by Pi deficiency was reported in both rice shoot and root (Khan et al., 2016). It was observed that Pi starvation results in plant sensitivity to ethylene. Changes in the expression of the ethylene signaling pathway genes were shown to be involved in the higher ethylene sensitivity in Pi-starved plants (Lei et al., 2011; L. Song & Liu, 2015). In our RNAseq data three ETHYLENE INSENSITIVE 3-like (EIL) genes, important transcription factors of the ethylene signaling pathway, were up-regulated. In addition, the expression of Arabidopsis AP2/ERF transcription factors, which bind to the promoters of many ethylene-responsive genes, were altered in response to Pi starvation (Thibaud et al., 2010). Six rice AP2/ERF transcription factors were also up-regulated in response to short Pi deficiency in our data. Therefore, EIL genes and AP2/ERF transcription factor genes may be involved in ethylene sensitivity in Pi-starved plants. OsMAPK2, Mitogen-activated protein kinase 5, which was shown to improve low Pi tolerance in rice, was also up-regulated in our data (Hur & Kim, 2014).

Two transcription factors AUXIN RESPONSE FACTOR 7 (ARF7) and ARF19 were reported to be involved in the regulation of PHR1 expression in Arabidopsis (K.-L. Huang
et al., 2018). In addition, it was reported that Pi starvation altered lateral root development via ARF19 (Pérez-Torres et al., 2008). In our data, OsARF19 was up-regulated. In addition, auxin response repressor AUX/IAA genes were also down-regulated in our data. It was reported in Pérez-Torres study that low level of AUX/IAA repressor results in activation of auxin response factors (ARF) and alter the expression of auxin response genes and consequently induced lateral root development (Pérez-Torres et al., 2008; Z. Zhang et al., 2014). Two inorganic Pi transporters were up-regulated in our data, OsPT19 and OsPT13. OsPT13 is a Pi transporter involved in symbiotic Pi uptake and development of arbuscular mycorrhiza (AM) symbioses (Güimil et al., 2005; Yang et al., 2012). OsPT19 is one of the 26 known P transporters (F. Liu et al., 2011) which also participating in relocation of P from flag leaves to emerging grains (Jeong et al., 2017). Replacement of phospholipid with digalactosyl diacylglycerol is one of the established responses to Pi deficiency. Phosphatidic acid phosphatase (PAP), which is involved in this replacement by dephosphorylating phosphatidic acid to produce diacylglycerol, was up-regulated after 24 hours of Pi deficiency (Tjellström et al., 2008). IPs (inositol pyroPi) are known as one of the secondary messengers that has a role in Pi sensing and signaling (Chiou & Lin, 2011). AtIPK1, Inositol polyPi kinase, that produce IP5 and IP6 by phosphorylating IP4 and IP5 to was indicated to be essential for Pi homeostasis (Stevenson-Paulik et al., 2005). OsIPK2 which converting IP3 into IP4, and IP4 to IP5 and play an important role in the phytic acid biosynthesis (Suzuki et al., 2007) was up-regulated in my data. It was indicated that phosphoenolpyruvate carboxylase (PEPC) expression and also the level of the enzyme was induced in response to Pi deprivation (Gregory et al., 2009; Peñaloza et al., 2004; Plaxton & Podestá, 2006). PEPC was up-regulated in our data. One of the PSI
in the shoot is increasing the Pi recycling. Two ent-kaurene synthase genes and phosphatidic acid phosphatase, which produce inorganic Pi, were up-regulated in response to Pi deficiency.

Figure 14. Networks representing Gene Ontology (GO) terms in the Biological Process (B), cellular component (C) and molecular functions (M) category enriched in DEGs that are down-regulated (A) and up-regulated (B) by Pi deficiency. Bingo and Cytoscape were used to identify and visualize enriched GO terms. Circle color shows p-value of enrichment.
Table 5. Summary of gene ontology (GO) analysis of up-regulated DEGs (U) and down-regulated DEGs (D) of rice under P deficiency.

<table>
<thead>
<tr>
<th>Category</th>
<th>P deficiency response</th>
<th>Network</th>
<th>GO terms</th>
<th>Number (%) in DEGS</th>
<th>Number (%) in background</th>
<th>Log (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological process (B)</strong></td>
<td>Up</td>
<td>BU1</td>
<td>response to stimulus&lt;br&gt;-response to biotic stimulus&lt;br&gt;-response to stress&lt;br&gt;-response to endogenous stimulus&lt;br&gt;-response to abiotic stimulus</td>
<td>186(30.7)&lt;br&gt;63(10.4)&lt;br&gt;130(21.5)&lt;br&gt;59(9.7)&lt;br&gt;79(13)</td>
<td>5265(20.7)&lt;br&gt;10764(4.2)&lt;br&gt;3618(14.2)&lt;br&gt;1491(5.9)&lt;br&gt;2194(8.6)</td>
<td>-8.6&lt;br&gt;-10.3&lt;br&gt;-6.2&lt;br&gt;-4&lt;br&gt;-3.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BU2</td>
<td>metabolic process&lt;br&gt;-lipid metabolic process</td>
<td>381(62.9)&lt;br&gt;42(6.9)</td>
<td>14183(55.7)&lt;br&gt;991(3.9)</td>
<td>-3.8&lt;br&gt;-3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BU3</td>
<td>regulation of biological process&lt;br&gt;-regulation of cellular process</td>
<td>58(9.6)&lt;br&gt;58(9.6)</td>
<td>1588(6.2)&lt;br&gt;1464(5.8)</td>
<td>-3.1&lt;br&gt;-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BU4</td>
<td>signaling process&lt;br&gt;-signal transduction</td>
<td>58(9.6)&lt;br&gt;58(9.6)</td>
<td>1481(5.8)&lt;br&gt;1464(5.8)</td>
<td>-3.8&lt;br&gt;-4</td>
</tr>
<tr>
<td>Down</td>
<td></td>
<td>BD1</td>
<td>metabolic process&lt;br&gt;-lipid metabolic process&lt;br&gt;-carbohydrate metabolic process&lt;br&gt;-secondary metabolic process</td>
<td>383(61.5)&lt;br&gt;70(11.2)&lt;br&gt;50(8)&lt;br&gt;28(4.5)</td>
<td>14183(55.7)&lt;br&gt;991(3.9)&lt;br&gt;972(3.8)&lt;br&gt;448(1.9)</td>
<td>-2.7&lt;br&gt;-14.8&lt;br&gt;-6.1&lt;br&gt;-4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BD2</td>
<td>growth&lt;br&gt;regulation of cell size</td>
<td>26(4.2)&lt;br&gt;22(3.5)</td>
<td>581(2.3)&lt;br&gt;465(1.8)</td>
<td>-2.6&lt;br&gt;-2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BD3</td>
<td>cell-cell signaling</td>
<td>4(0.6)</td>
<td>31(0.1)</td>
<td>-2.2</td>
</tr>
<tr>
<td>Cellular component(C)</td>
<td>Down</td>
<td>CD1</td>
<td>extracellular region</td>
<td>47(7.5)</td>
<td>559(2.2)</td>
<td>-12.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CD2</td>
<td>external encapsulating structure&lt;br&gt;-cell wall</td>
<td>50(8)&lt;br&gt;50(8)</td>
<td>953(3.7)&lt;br&gt;943(3.7)</td>
<td>-6.4&lt;br&gt;-6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CD3</td>
<td>plasma membrane&lt;br&gt;Golgi apparatus&lt;br&gt;endoplasmic reticulum</td>
<td>106(17)&lt;br&gt;17(2.7)&lt;br&gt;26(4.2)</td>
<td>3116(12.2)&lt;br&gt;337(1.3)&lt;br&gt;562(2.2)</td>
<td>-3.6&lt;br&gt;-2.4&lt;br&gt;-2.8</td>
</tr>
<tr>
<td>Molecular function (M)</td>
<td>Up</td>
<td>MU1</td>
<td>transcription factor activity</td>
<td>66(10.9)</td>
<td>1773(7)</td>
<td>-3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MU2</td>
<td>catalytic activity</td>
<td>287(47.4)</td>
<td>9920(39)</td>
<td>-4.9</td>
</tr>
<tr>
<td></td>
<td>Down</td>
<td>MD1</td>
<td>catalytic activity&lt;br&gt;-hydrolase activity&lt;br&gt;-transferase activity</td>
<td>328(52.6)&lt;br&gt;110(17.7)&lt;br&gt;119(19.1)</td>
<td>9920(39)&lt;br&gt;3112(12.2)&lt;br&gt;3913(15.4)</td>
<td>-11.8&lt;br&gt;-4.4&lt;br&gt;-2.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MD2</td>
<td>lipid binding</td>
<td>20(3.2)</td>
<td>271(1.1)</td>
<td>-4.9</td>
</tr>
</tbody>
</table>
Discussion

In the present study, specific experimental approaches include ChIP-seq to map the distribution of H3K4me3, and RNA-seq analyses to investigate global gene expression were used. We defined the H3K4me3 pattern and its co-localization with H2A.Z in the rice genome. The relationship among H3K4me3, H2A.Z and H3K4me3 abundance, gene types and gene expression was also determined. In addition, analysis of the early transcriptional responses to Pi deficiency by RNA seq showed genes involved in response to stress, lipid metabolic process, regulation of cellular biology, signal transduction were induced while growth, cell-cell signaling and lipid, carbohydrate and secondary metabolic process genes were repressed.

H3K4me3 abundance has been investigated in yeast, plants and animal. Association of H3K4me3 with genes and a striking peak at TSS was conserved among them (Barski et al., 2007; Bernstein et al., 2002; Du et al., 2013; C. L. Liu et al., 2005; Santos-Rosa et al., 2002a; Van Dijk et al., 2010; X. Zhang et al., 2009; Zong et al., 2013). Our data is in agreement with this view by showing higher level of H3K4me3 at TSS of PCG compared to PSG, TEG, and TE (Figure 7A). In addition clustering analysis indicated the unvarying pattern of H3K4me3 among all PCG (Figure 8 and Figure 9) and the importance of H3K4me3 abundance at the TSS.

The difference in the level of H3K4me3 at TEG and TE compared to H2AZ was observed in our data (Figure 7A and B) and may reflect the presence of DNA methylation at TEG and TE that antagonizes H2A.Z deposition (Zilberman, Coleman-Derr et al. 2008, Conerly, Teves et al. 2010, Zemach, McDaniel et al. 2010, Zahraeifard, Foroozani et al. 2018). We previously reported that H2A.Z deposition and nucleosome patterns were generally different according to gene function. Housekeeping genes have sharp H2AZ
peaks around the TSS with have wide 5’ nucleosome-depleted region while stress-responsive genes have broad H2AZ peaks around the TSS and nucleosomes positioned on either side of the TSS (Zahraeifard et al., 2018; Q. Zhang et al., 2018). Consistently our data on H3K4me3 abundance indicate that housekeeping genes have higher levels of H3K4me3 compared to stress-responsive genes (Figure 9 and Figure 10).

Genome-wide analysis of H3K4me3 and H2AZ showed co-localization of these two marks in promoters of Arabidopsis genes and the rice genome (Xiaozhuan Dai et al., 2017; K. Zhang et al., 2017). We also found the strong association of H3K4me3 with H2AZ in the rice genome (Figure 7C and D). A strong correlation between H3K4me3 peaks at the TSS and transcription has been reported in many organisms (Barski et al., 2007; Bernstein et al., 2002; Santos-Rosa et al., 2002a; Van Dijk et al., 2010; X. Zhang et al., 2009). Here we showed that there are positive and negative relationships between the H3K4me levels and gene expression at TSS and TES respectively (Figure 11A, B). This trend differs from a study in Arabidopsis in which H3K4me3 abundance at both the TSS and TES positively correlated with gene expression (Van Dijk et al., 2010). It was reported that H3K4me at 3’ end of the genes correlate positively with antisense transcript (Cui et al., 2012), while the abundance of antisense transcript often results in the suppression of gene expression (Ponting et al., 2009), therefore 3’-H3K4me3 negatively correlates with sense transcripts in my data. The strong correlations between H3K4me3 and both expression and length were observed in my data, consistent with the observation that gene expression and length are positively correlated in plant systems, but are negatively correlated in animals (Caldwell et al., 2015; Ren et al., 2006).
The antagonistic and synergistic relationships between H3K4me3 and H2A.Z regarding gene expression were reported in Arabidopsis (Hanyang Cai et al., 2019; M. Xu et al., 2018). The expression of anthocyanin biosynthesis genes were shown to be negatively regulated by H2A.Z deposition. Mutation in ARP6 and PIE1 (members of the SWR1 complex) and HTA9 and HTA11 genes (which encode H2A.Z) result in reduced and enriched levels of H2AZ and H3K4me3 respectively, and repress the expression of anthocyanin biosynthesis genes (Hanyang Cai et al., 2019). Conversely, it was shown that H2A.Z positively regulates the expression of MIR156A/MIR156C genes involved in vegetative phase change by promoting H3K4me3 deposition. Mutation in genes involved in H2A.Z exchanging and encoding lessen the expression of MIR156A/MIR156C genes and facilitate vegetative phase transition (M. Xu et al., 2018). In our data, we found that H3K4me3 at the TSS has a strong positive correlation with gene expression, whereas H2A.Z in the gene body has a strong negative correlation with expression. At the TSS, when H3K4me3 increases with expression, H2A.Z decreases, but at the gene body, both marks increase with decreasing expression. This suggests antagonism at the TSS but not the gene body (Figure 11C).

The dynamic changes of H3K4me3 and H3K27me3 were shown to be involved in gene regulation during flower morphogenesis, xylem development, stable transmission of rice epigenetic variation and inflorescence meristem (X. Chen et al., 2015; Engelhorn et al., 2017; Hussey et al., 2017; X. Liu et al., 2015a). Through early flower development, alterations in H3K4me3 abundance occurred prior to changes in H3K27me3 and could predict changes in gene expression better. In addition, the correlation among H3K4me3, H2AZ and H3K27me3 and gene expression was previously demonstrated in rice seedling
and callus tissue. Genes with H3K4me3 and H2AZ showed high levels of expression while genes with H3K27me3 and H2AZ have low expression levels (K. Zhang et al., 2017). Here we showed the ratio between H3K4me3 and H2AZ and existing of H3K27me3 could predict the level of expression, high expression level was associated with higher level of H3K4me3 compared to H2AZ and vice versa (Figure 11C). The reduction and increase in the level of H2A.Z deposition at the TSS and GB was detected previously (Zahraeifard et al., 2018). Consistent with these data, we detected reduction of H3K4me3 deposition at both TSS and the part of GB in response to Pi deficiency (Figure 12A). Early transcription responses to Pi deficiency showed that the functional categories of our DEGs (Figure 14 and Table 5) were similar to those from previous transcriptome studies of Pi-deficient plants (Hongmei Cai et al., 2013; Secco et al., 2013b; Thibaud et al., 2010).
Chapter 3. Chromatin States Associated with Early Pi Deficiency Responses

Introduction

Phosphorus (P) is among the most limiting essential nutrients for plants. This is because plant-available forms of P (inorganic Pi (Pi)) have low solubility in soils (Holford, 1997). P fertilization of soils is required for crop plants to cope with its deficiency. However, P fertilization can result in serious environmental concerns due to depletion of P resources (Vance et al., 2003). In addition, application of P fertilizer to soils deficient in Fe is likely to increase this deficiency (Sánchez-Rodríguez et al., 2014). It is, therefore, necessary to investigate the underlying mechanisms involved in regulating P homeostasis, so as to increase the efficiency of plants to absorb and recycle P. Different kinds of Pi deficiency responses such as molecular, physiological, morphological and biochemical responses exist in plants and these responses are regulated at transcriptional and epigenetic levels (Gräff et al., 2011; Secco et al., 2013b).

Histone variants, histone posttranslational modifications and nucleosome remodeling are among major epigenetic mechanisms involved in defining the chromatin landscape. The role of histone modification, histone variants, nucleosome occupancy in eukaryotic gene expression under various stress has been investigated extensively over the last decades (Asensi-Fabado et al., 2017). However, the study of a single component of the chromatin landscape cannot dictate the epigenetic states (Margueron & Reinberg, 2010). Analyzing chromatin states, chromatin landscape patterns defined by combining multiple marks in their spatial context, is more informative for understanding the transcriptional responses under stress (Ernst & Kellis, 2012; Pan et al., 2017). For
example, it has been shown that specific chromatin states are correlated with genes differentially expressed in response to ionizing radiation (Pan et al., 2017).

Both genome-wide analyses and a study focused on a subset of key PSR (Pi starvation response) genes demonstrated the role of H2A.Z in the regulation of Pi starvation responses (Aaron P Smith et al., 2010; Zahraeifard et al., 2018). The role of H2A.Z in gene expression was shown to depend on the location of its deposition. Deposition of H2A.Z in the gene body results in repressing PSR genes while H2A.Z at the TSS is positively or negatively correlated with gene expression (Zahraeifard et al., 2018). Also, the alteration of whole genome nucleosome patterns by Pi starvation was established recently in the rice genome (Q. Zhang et al., 2018). Differentially expressed genes in response to low Pi showed both nucleosome position shifts and changes in nucleosome occupancy (Q. Zhang et al., 2018). In this chapter I examined H2A.Z, H3K4me3 and nucleosome occupancy simultaneously to identify chromatin structure changes under Pi deficiency.

Materials and Methods

Chromatin states analysis

ChromHMM (Ernst & Kellis, 2012) was used with default parameters to characterize the chromatin state maps for control and Pi deficiency samples. We used the published profile of H2A.Z ChIP-seq data (Zahraeifard et al., 2018) and nucleosome occupancy (Q. Zhang et al., 2018) (MNase-seq) of rice genome in control and under Pi deficiency conditions. In addition, we also used the H3K4me3 profile generated in Chapter 2. All input data were binarized with ChromHMM’s BinarizedBam method and input genomic DNA was used to adjust binarization thresholds locally. Concisely,
ChromHMM by default used 200 bp interval to divide the genome. The common model of chromatin states in both control and Pi deficiency samples was developed by concatenating the marks using a hidden Markov model (ChromHMM). Five chromatin states were generated based on the learned model parameters as defined in ChromHMM (Ernst & Kellis, 2012).

We analyzed the chromatin state changes using a previously described method (Fiziev et al., 2017). Briefly, 200 bp bins in the control and –Pi genome was occupied by one chromatin state. We overlapped the chromatin state annotations of control and –Pi genomes. 25 possible chromatin state transition could happen. The number of bins in each possible chromatin state transitions was quantified and called as the observed number. The expected number was calculated by multiplying the number of bins in the two chromatin states involved in each transition and then dividing by total bins in the genome. We divided the observed number by the expected number to calculate enrichment scores. States similarity between each pair of chromatin states was controlled by dividing the enrichment scores of each state transition to the enrichment scores of reverse state transition. The distribution of each chromatin states were identified using CEAS software (Shin et al., 2009).

Every PCG was assigned to one chromatin states based on the chromatin states at the TSS or the most frequent states call on 200 bp upstream and 800 bp downstream of genes TSS in control and –Pi samples. Next, we counted number of PCG in each pair of chromatin states transition. Similarly, DEGs chromatin states were determined. For bootstrap analysis, we used the script written in FORTRAN to obtain the same number of randomly selected genes and estimate the percentage of overlap between these genes.
and each group of state transition (1000 times). Binomial distribution test was done by R (pbinom, P-value < 1.00E-03).

**Circos plot**

Each rice chromosome was partitioned into bins of 5kbp. Chromatin states were merged from 200bp bins to 5kbp bins in both control and phosphorus deficiency samples. The most dominant chromatin state in each merged bin, or the chromatin state of the previous bin if most dominant chromatin state cannot be determined, was selected as the chromatin state for that bin. For gene type partition, the most dominant gene type, in base pair, was used as the bin type for each bin. Chromatin states, differential expression status, bin types for the merged bins were determined using customized scripts and visualized with a R package circlize (Gu et al., 2014).

**Chromatin state transition plot**

Chromatin states in control samples were differentially color coded. Genes in each control chromatin state were sorted based on their positions within each chromosome. Chromatin transitions for each gene were connected with lines of colors matching those used for control chromatin states. Genes in each chromatin transition were positioned according to their expression changes upon Pi deficiency treatment, with up-regulated genes on the top and down-regulated on the bottom. These transition connections were plotted with ggplot2 (Wickham, 2016).

**Total mRNA level comparison**

Read counts for each gene were first converted to RPKM in each replicate. The RPKM from all replicates was averaged for each gene and used for expression level
comparison between control and treated samples. The comparison was performed using Wilcoxon signed rank test in R.

**Results**

**H3K4me3, H2A.Z, and nucleosome occupancy define 5 chromatin states in the rice genome**

As mentioned in chapter 2, to evaluate a potential role for H3K4me3 in modulating Pi-deficiency responses, we carried out H3K4me3 ChIP-seq on shoots from plants subjected to a 24-hour Pi-deficiency treatment. As shown in Figure 12, Pi-deficiency altered H3K4me3 distribution at PCG, such that the prominent 5' peak was reduced. These data along with our prior studies (Zahraeifard et al., 2018; Q. Zhang et al., 2018) indicate that nucleosome occupancy, H2A.Z, and H3K4me3 each exhibit distinct changes in response to Pi deficiency. It is becoming increasingly clear that examining multiple chromatin marks simultaneously provides a more robust picture of the dynamic chromatin structure linked to various developmental processes and responses to stimuli (Pan et al., 2017; Yan et al., 2019). Therefore, we integrated our H3K4me3 ChIP-Seq, H2A.Z ChIP-Seq, and MNase-Seq data sets to define distinct chromatin states using ChromHMM (Ernst & Kellis, 2012). ChromHMM employs a multivariate Hidden Markov Model that scores the presence or absence of each chromatin mark to determine the major recurring combinatorial and spatial patterns of marks, i.e. chromatin states. ChromHMM identified 5 chromatin states (CS), each distinguishable from the others by differential enrichment of one or more of the marks tested (Figure 15). CS1 and CS2 were each deficient in both H2A.Z and H3K4me3, CS3 was enriched in only H2A.Z, CS4 was enriched in both H2A.Z and H3K4me3, and CS5 was enriched in only H3K4me3. Regarding nucleosome density, CS2 and CS3 had moderately higher nucleosome enrichment compared to the other 3
states (Figure 15). All the states were most likely to remain in their own states except state 5 which had high possibility of shift to state 1 (Figure 15).

Figure 15. Chromatin states prediction for control (Ctrl) and Pi deficiency (–Pi) samples defined by H3K4me3, H2A.Z and nucleosome occupancy. (A) The heat map of emission parameters for five chromatin states. Each row corresponds to a different state and each column correspond to a different mark. The darker blue color corresponds to a greater probability of observing the mark in the state. (B) The heat map of transition parameters of five chromatin states.

Next we mapped the distribution of the 5 chromatin states across the genome, which revealed biases with a number of genomic features (Figure 16, Figure 17). CS1 was the major chromatin state, accounting for 63% of the rice genome, and was enriched at TE and TEG. TE and TEG were also enriched in CS2 and CS5. This means that the transposable element-related loci were either deficient in both H2A.Z and H3K4me3 or contained H3K4me3 only. In contrast, PSG were enriched in CS2 and CS3, consistent with depletion of both H2A.Z and H3K4me3 or enrichment of only H2A.Z. Finally, PCG
were enriched in CS4, consistent with enrichment of both H3K4me3 and H2A.Z (Figure 16, Figure 17).

To more specifically characterize PCG, we calculated enrichments at the TSS, TES, and 1kb regions that encompass the TSS or TES (TSS 1kb region, 200 bp upstream to 800 bp downstream of the TSS; TES 1kb region, 800 bp upstream to 200 bp downstream of the TES). Compared to all bins within PCG, the TSS was more enriched in CS4, CS5, and CS3, whereas the TES was more enriched in CS3 and less enriched in CS4. These results indicate an overall high occupancy of H2A.Z and/or H3K4me3 at the TSS, but an enrichment of only H2A.Z at the TES (Figure 16, Figure 17).

![Figure 16](image)

**Figure 16.** Overlap fold enrichment of various genomic regions with 5 chromatin states in Ctrl (B) and –Pi (C) samples. PCG, protein coding genes; PG, pseudogenes; TE, non-expressed transposable element-related genes; TEG, expressed transposable element-related genes; TSS, transcription start site; TES, transcription termination site; TSS1Kb, 200 bp upstream to 800 bp downstream of the TSS; TES1Kb, 800 bp upstream to 200 bp downstream of the TES.
Figure 17. Circos plot showing the chromatin states (in 5kbp bins) of the whole genome. The first and second rings show the chromatin state in -Pi and Ctrl condition respectively. The third ring shows fold change of differentially expressed genes and the last ring represents four gene types. Each segment in circus plot showed one chromosome (Chr) in the rice genome.

**Pi deficiency has a dramatic impact on chromatin structure**

To characterize the impact of Pi deficiency on chromatin structure we compared the chromatin state distribution between control and Pi deficiency conditions. First we measured the genome-wide coverage changes for each chromatin state by calculating the fold change in the total number of genomic bins in the –Pi sample relative to the control. As shown in Figure 18, the –Pi sample had 2.1-fold more CS3 and 1.4-fold less CS4 compared to control. Next we quantified the chromatin state transitions between control and –Pi samples by calculating enrichment scores of state transition and also analyzing the pairwise state transition enrichment as described earlier by (Fiziev et al.,...
To control state similarity, the enrichment scores of each state transition was divided to the enrichment scores of state transition in the reverse direction (Figure 19B). This analysis revealed a general gain of H2A.Z and/or H3K4me3 marks in response to –Pi at regions deficient in both marks in control. For example, the transitions from CS2 (i.e. deficient in H2A.Z and H3K4me3) in control to CS3 (i.e. enriched in H2A.Z), CS5 (i.e. enriched in H3K4me3), and CS4 (i.e. enriched in both marks) were enriched 8.4-, 2.5-, and 3.5-fold respectively. Similarly, the CS1 to CS3 transition (i.e. gain of H2A.Z) was enriched 2.8-fold. The other major transition was from CS4 to CS3, which is consistent with a loss of H3K4me3 and/or a gain of H2A.Z.

Next we analyzed the enrichment of each chromatin state within the four gene types. In response to Pi deficiency, TE and TEG increased in CS1, but decreased in CS2 and CS5, consistent with a loss of H3K4me3. On the other hand, PSG and PCG did not exhibit any major shifts overall in response to Pi deficiency, but at the TSS of PCG, CS4 decreased and CS5 increased. Also, at the TES proximal region of PCG, CS4 decreased and CS3 increased. Together this reveals an overall trend whereby, at PCG, Pi deficiency leads to decreased H2A.Z at the TSS and decreased H3K4me3 near the TES (Figure 20).

To further examine the chromatin states changes in response to Pi deficiency, we analyzed the distribution of each chromatin states across sub-genomic regions (Promoter, Downstream of a gene, 3'UTRs, 5'UTRs, coding exons, introns, and distal intergenic) in both control and Pi deficiency samples by using CEAS software. It was observed that CS1 and CS2 distribution was not changed under Pi deficiency. Smaller percentage of CS3, CS4, and CS5 were found in the promoter of rice genome under Pi
deficiency, while a higher percentage of CS3, CS4, and CS5 distribution were found in the coding exon, intron and 5’ UTR (Figure 21).

![Log2 fold change (FC) of genomic bins occupied by each chromatin state in control compared to Pi deficiency.](image)

**Figure 18.** Log2 fold change (FC) of genomic bins occupied by each chromatin state in control compared to Pi deficiency.

<table>
<thead>
<tr>
<th>Ctrl-States</th>
<th>-Pi states</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS1</td>
<td>1.37457</td>
</tr>
<tr>
<td>CS2</td>
<td>0.46541</td>
</tr>
<tr>
<td>CS3</td>
<td>0.15201</td>
</tr>
<tr>
<td>CS4</td>
<td>0.07017</td>
</tr>
<tr>
<td>CS5</td>
<td>0.74701</td>
</tr>
</tbody>
</table>

**Figure 19.** (A) Heat map showing state transition in Ctrl and –Pi samples with raw enrichment score. (B) Heat map showing fold enrichment of transitions of chromatin states in Ctrl to –Pi.
Figure 20. Overlap fold enrichment of various genomic regions with five chromatin states in -Pi samples.

Figure 21. Genomic distribution of each chromatin states in control (Ctrl) and Pi deficiency (-Pi) samples.
To examine the chromatin state transitions of PCG in more detail, we next compared the chromatin state of each PCG at its TSS in control and Pi-deficiency samples (Figure 22). Over 40 % of PCG exhibited a chromatin state transition at their TSS in response to Pi deficiency. The largest groups of transitions were CS4 to CS3, CS4 to CS5, and CS5 to CS1 (Figure 22A). Gene Ontology (GO) enrichment analysis showed significantly enriched GO terms (FDR < 0.05) for 8 of the transition groups (Figure 22B, and Figure 23). CS4-CS3 genes were enriched in terms related to stress-responses, including oxygen binding, transcription factor activity, cell wall, and response to stimulus. In contrast, CS4-CS5 genes were enriched in terms related to translation, gene expression, RNA binding, and embryonic development. Interestingly, CS5-CS1 genes shared essentially the same enriched GO terms as CS4-CS5 genes, suggesting the presence of two sub-groups of translation-related genes that are modulated by different chromatin dynamics.
Figure 22. Chromatin state (CS) transitions of all protein coding genes (PCG) from control (Ctrl) to Pi deficiency (-Pi) conditions. (A) The size of the segment represents the number of (PCG) in each CS and the width of the ribbons represent the number of genes with a transition to another CS. (B) Networks representing Gene Ontology (GO) terms enriched in CS5-CS1, CS4-CS3 and CS4-CS5 groups. Bingo and Cytoscape were used to identify and visualize enriched GO terms. The color of the circles show the p-value of enrichment.
Figure 23. Networks representing Gene Ontology (GO) terms enriched in CS1-CS2, CS1-CS3, CS1-CS5, CS4-CS1 and CS2-CS3 groups.
Chromatin state transitions correlate with differential expression of Pi deficiency-responsive genes

We next used the results of our RNA-seq experiments to investigate the relationship between gene expression and chromatin state transitions in response to Pi deficiency. As mentioned before, we conducted RNA-seq experiments on two biological replicates of shoot tissues from control plants and those subjected to a 24-hour Pi deficiency treatment (Zahraeifard et al., 2018).

To determine whether any chromatin state transitions were significantly over- or under-represented among the DEGs, we determined the overlap between each transition and the DEGs via bootstrapping analyses (1000 iterations; binomial test, p-value < 0.001; Figure 24, Figure 25). These analyses revealed several significant biases between CS transitions and DEGs. First, down-regulation of gene expression correlated with a gain of H2A.Z, as indicated by enrichment of down-regulated genes with CS1-CS3 and CS2-CS3 transitions, under-representation of up-regulated DEGs with a CS2-CS3 transition, and under-representation of down-regulated DEGs with a CS4-CS5 transition. Second, up-regulation of gene expression correlated with a loss of H2A.Z and/or H3K4me3, as indicated by enrichment of up-regulated genes with CS3-CS1 and CS4-CS1 transitions, as well as under-representation of down-regulated DEGs with a CS5-CS1 transition. Interestingly, up- and down-regulated DEGs were both enriched among CS4-CS3 transition genes (i.e. those with a decrease in H3K4me3), suggesting a possible dual role of H3K4me3 in Pi-responsive gene modulation. In addition to the biases between DEGs and chromatin state transitions, there were also biases to groups of genes that did not transition. Both up- and down-regulated DEGs were significantly enriched among genes with CS3 that did not transition (i.e. CS3-CS3), and were under-represented among CS1-
CS1 and CS5-CS5 genes. Furthermore, up-regulated DEGs were enriched among CS4-CS4 genes and under-represented among CS2-CS2 genes. These results show that responsive genes are likely to contain H2A.Z. Taken together, the biases described above demonstrate that the chromatin state of the TSS is associated with differential expression of genes in response to Pi deficiency.

Figure 24. Chromatin state (CS) transitions are associated with differentially-expressed genes under Pi deficiency. Heat map showing fold difference between the particular number of DEGs and random genes within specific CS transitions, whereas the shading represents the significant one, red represents enriched and blue depleted.
Figure 25. Bootstrapping analysis showing the overlap between genes exhibiting chromatin state transitions and down-regulated or up-regulated genes at TSS in response to Pi deficiency. All data are means (±SD) for 1000 iterations.
To determine whether a relatively broad region of chromatin encompassing the TSS exhibits modification in accordance with differential gene expression, we repeated our bootstrapping analyses, but this time at a 1kb region (five 200-bp bins) encompassing the TSS (-800 to 200 relative to the TSS). A particular chromatin state was assigned to a gene if at least three of the five bins were of the same CS. As with the TSS-only bootstrapping, these analyses revealed a number of biases between particular CS transitions and DEGs, some of which were similar to the TSS-only, but others that were different. For example, genes with a CS1-CS3 or CS2-CS3 transition across the TSS region were again enriched among down-regulated DEGs (Figure 24; Figure S7). This suggests that a gain of H2A.Z at a relatively broad region around the TSS is linked to down-regulation of Pi-starvation response genes. On the other hand, DEGs exhibited different biases with CS4-CS3 transition genes when assaying a range across the TSS relative to the TSS alone. Rather than being over-represented, CS4-CS3 transitions across the TSS were similar to random for down-regulated DEGs, and were under-enriched among up-regulated DEGs. This indicates that a relatively narrow chromatin state transition from CS4 to CS3 at the TSS is correlated with differential expression of Pi-starvation response genes.
Figure 26. Bootstrapping analysis showing the overlap between genes exhibiting chromatin state transitions and down-regulated or up-regulated genes at region adjacent to the TSS in response to Pi deficiency. All data are means (±SD) for 1000 iterations.
Differentially-expressed genes exhibiting a CS4 to CS3 chromatin transition suggest a coordinated Pi-deficiency network targeting the apoplast

As shown above, the largest group of genes exhibiting a chromatin state shift in response to Pi deficiency was the CS4-CS3 group (Figure 22). These genes were also significantly enriched among both up- and down-regulated DEGs (Figure 24), and GO term enrichment analysis of those CS4-CS3 DEGs suggests functions linked to the cell wall, responses to biotic stress, and catalytic activity (Figure S8).

Figure 27. Gene Ontology (GO) terms enriched in differentially expressed genes (DEGs) that have a chromatin state (CS) transition of CS4-CS3. Bingo and Cytoscape were used to identify and visualize enriched GO terms. Circle color shows p-value of enrichment.

Due to the relatively limited GO term assignments for rice loci, we carried out extensive data mining on the CS4-CS3 DEGs, which allowed us to assign putative functional and subcellular localization information to 178 (91%) of the 196 DEGs (Table 6).
Table 6. Functional and subcellular localization information of 178 differentially expressed genes with CS4-CS3 transition.

<table>
<thead>
<tr>
<th>Locus</th>
<th>RNAseq fold change (P/ctrl)</th>
<th>Subcellular compartment</th>
<th>Functional category</th>
<th>Annotation</th>
<th>Abbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC_D02g00860</td>
<td>3.39</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>AIR (auxin responsive; DOMAIN-3 containing)</td>
<td>AIR</td>
</tr>
<tr>
<td>LOC_D02g00870</td>
<td>0.38</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>AIR (auxin responsive; DOMAIN-3 containing)</td>
<td>AIR</td>
</tr>
<tr>
<td>LOC_D02g00880</td>
<td>0.14</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>beta-galactosidase</td>
<td>B-gal</td>
</tr>
<tr>
<td>LOC_D02g00890</td>
<td>0.02</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>beta-galactosidase</td>
<td>B-gal</td>
</tr>
<tr>
<td>LOC_D02g00900</td>
<td>0.39</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>eno-beta-6-glucuronidase</td>
<td>B-guc</td>
</tr>
<tr>
<td>LOC_D02g00910</td>
<td>0.18</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>lipase</td>
<td>EXP</td>
</tr>
<tr>
<td>LOC_D02g00920</td>
<td>0.01</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>expansin</td>
<td>EXP</td>
</tr>
<tr>
<td>LOC_D02g00930</td>
<td>0.03</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>dirigent (involved in lignin/lignin biosynthesis)</td>
<td>LG</td>
</tr>
<tr>
<td>LOC_D02g00940</td>
<td>0.80</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>pectinesterase</td>
<td>PEC</td>
</tr>
<tr>
<td>LOC_D02g00950</td>
<td>0.34</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>pectin hydrolase</td>
<td>PEC</td>
</tr>
<tr>
<td>LOC_D02g00960</td>
<td>0.11</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>pectin hydrolase</td>
<td>PEC</td>
</tr>
<tr>
<td>LOC_D02g00970</td>
<td>0.15</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>pectidic acid esterase</td>
<td>PEC</td>
</tr>
<tr>
<td>LOC_D02g00980</td>
<td>0.14</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>pectidic acid esterase</td>
<td>PEC</td>
</tr>
<tr>
<td>LOC_D02g00990</td>
<td>0.12</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>pectidic acid esterase</td>
<td>PEC</td>
</tr>
<tr>
<td>LOC_D02g01000</td>
<td>0.03</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>pectidic acid esterase</td>
<td>PEC</td>
</tr>
<tr>
<td>LOC_D02g01010</td>
<td>0.30</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>pectidic acid esterase</td>
<td>PEC</td>
</tr>
<tr>
<td>LOC_D02g01020</td>
<td>0.53</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>suberic anhydride protease</td>
<td>PEX</td>
</tr>
<tr>
<td>LOC_D02g01030</td>
<td>7.28</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>stress induced cell wall structure protein</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01040</td>
<td>0.20</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>coprin domain containing; germin-like</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01050</td>
<td>0.04</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>peptidase H4 asparagine amino peptide A</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01060</td>
<td>0.04</td>
<td>apoplast</td>
<td>cell wall structure</td>
<td>systein cysteine proteinase-3</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01070</td>
<td>0.02</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>A/B barrel domain protein</td>
<td>a/b</td>
</tr>
<tr>
<td>LOC_D02g01080</td>
<td>0.11</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>GASRI - Gibberellin-regulated GAAS/GAST 1 (branched protein)</td>
<td>GASR</td>
</tr>
<tr>
<td>LOC_D02g01090</td>
<td>0.03</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>glycosidase</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01100</td>
<td>0.03</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>lucine-rich repeat, exoribonuclease (LRX)</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01110</td>
<td>0.27</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>subtilisin/C10 glycosidase</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01120</td>
<td>0.98</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>RALF1 - Rapid Acidification Factor</td>
<td>RALF</td>
</tr>
<tr>
<td>LOC_D02g01130</td>
<td>0.19</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>RALF1 - Rapid Acidification Factor</td>
<td>RALF</td>
</tr>
<tr>
<td>LOC_D02g01140</td>
<td>0.06</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>RALF1 - Rapid Acidification Factor</td>
<td>RALF</td>
</tr>
<tr>
<td>LOC_D02g01150</td>
<td>0.01</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>RALF1 - Rapid Acidification Factor</td>
<td>RALF</td>
</tr>
<tr>
<td>LOC_D02g01160</td>
<td>0.01</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>RALF1 - Rapid Acidification Factor</td>
<td>RALF</td>
</tr>
<tr>
<td>LOC_D02g01170</td>
<td>0.16</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>RALF1 - Rapid Acidification Factor</td>
<td>RALF</td>
</tr>
<tr>
<td>LOC_D02g01180</td>
<td>0.97</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>UF55038 domain protein</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01190</td>
<td>0.01</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>UF55038 domain protein</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01200</td>
<td>13.83</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>fiber expressed protein</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01210</td>
<td>0.08</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>UF55038 domain protein</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01220</td>
<td>0.91</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>UF55038 domain protein</td>
<td></td>
</tr>
<tr>
<td>LOC_D02g01230</td>
<td>0.15</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>glycerol-3-PH acyltransferase</td>
<td>G3P</td>
</tr>
<tr>
<td>LOC_D02g01240</td>
<td>0.03</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>glycerol-3-PH acyltransferase</td>
<td>G3P</td>
</tr>
<tr>
<td>LOC_D02g01250</td>
<td>0.00</td>
<td>apoplast</td>
<td>signal transduction</td>
<td>GDSL-like lipase/acylhydrolase</td>
<td>GDSL</td>
</tr>
</tbody>
</table>

(table cont’d)
<table>
<thead>
<tr>
<th>Locus</th>
<th>RNAseq fold change (P/ctrl)</th>
<th>Subcellular compartment</th>
<th>Functional category</th>
<th>Annotation</th>
<th>Abbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC_Du02g08710</td>
<td>0.03</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>UDOS-like lipase/phytase lipase</td>
<td>UDOSL</td>
</tr>
<tr>
<td>LOC_Du02g08610</td>
<td>3.18</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>LIP11 - Proteinase-inhibitor seed storage LTP protein</td>
<td>LIP1L</td>
</tr>
<tr>
<td>LOC_Du01g06010</td>
<td>0.20</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>LIP144 - Proteinase-inhibitor seed storage LTP protein</td>
<td>LIP1L</td>
</tr>
<tr>
<td>LOC_Du01g05810</td>
<td>0.05</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>LIP20 - Proteinase-inhibitor seed storage LTP protein</td>
<td>LIP1L</td>
</tr>
<tr>
<td>LOC_Du01g05610</td>
<td>0.13</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>LIP1L - Proteinase-inhibitor seed storage LTP protein</td>
<td>LIP1L</td>
</tr>
<tr>
<td>LOC_Du01g05410</td>
<td>0.08</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>non-specific phospholipase A</td>
<td>NPC</td>
</tr>
<tr>
<td>LOC_Du01g05310</td>
<td>0.15</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>cytochrome P450</td>
<td>CYP</td>
</tr>
<tr>
<td>LOC_Du01g05210</td>
<td>0.05</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>phospholipase A</td>
<td>PLA</td>
</tr>
<tr>
<td>LOC_Du01g05010</td>
<td>0.24</td>
<td>apoplastic</td>
<td>lipid metabolism</td>
<td>WAX2</td>
<td>WAX</td>
</tr>
<tr>
<td>LOC_Du01g04810</td>
<td>20.91</td>
<td>apoplastic</td>
<td>secondary metabolism</td>
<td>cytochrome P450</td>
<td>CYP</td>
</tr>
<tr>
<td>LOC_Du01g04710</td>
<td>2.33</td>
<td>apoplastic</td>
<td>secondary metabolism</td>
<td>cytochrome P450</td>
<td>CYP</td>
</tr>
<tr>
<td>LOC_Du01g04610</td>
<td>2.55</td>
<td>apoplastic</td>
<td>secondary metabolism</td>
<td>peroxidase</td>
<td>P45O</td>
</tr>
<tr>
<td>LOC_Du01g04510</td>
<td>0.33</td>
<td>apoplastic</td>
<td>secondary metabolism</td>
<td>xylooligodehydrogenase synthase</td>
<td>XOD</td>
</tr>
<tr>
<td>LOC_Du01g04410</td>
<td>3.09</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>heparen-alpha-glucosaminidase N-acetyl transferase</td>
<td>HGL</td>
</tr>
<tr>
<td>LOC_Du01g04310</td>
<td>2.64</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>arabinogalactan protein type 1 (AGP)</td>
<td>AGP</td>
</tr>
<tr>
<td>LOC_Du01g04210</td>
<td>0.51</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>arabinogalactan protein type 1 (AGP)</td>
<td>AGP</td>
</tr>
<tr>
<td>LOC_Du01g04110</td>
<td>0.12</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>arabinogalactan protein type 1 (AGP)</td>
<td>AGP</td>
</tr>
<tr>
<td>LOC_Du01g04010</td>
<td>0.15</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>arabinogalactan protein type 1 (AGP)</td>
<td>AGP</td>
</tr>
<tr>
<td>LOC_Du01g03910</td>
<td>0.14</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>arabinogalactan protein type 1 (AGP)</td>
<td>AGP</td>
</tr>
<tr>
<td>LOC_Du01g03810</td>
<td>0.16</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>MCDMNA-like</td>
<td>MCDMNA</td>
</tr>
<tr>
<td>LOC_Du01g03710</td>
<td>0.44</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>RIOODD type acyltransferase</td>
<td>RIOODD</td>
</tr>
<tr>
<td>LOC_Du01g03610</td>
<td>4.48</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>RIOODD type acyltransferase</td>
<td>RIOODD</td>
</tr>
<tr>
<td>LOC_Du01g03510</td>
<td>0.23</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>RIOODD type acyltransferase</td>
<td>RIOODD</td>
</tr>
<tr>
<td>LOC_Du01g03410</td>
<td>0.20</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>RIOODD type acyltransferase</td>
<td>RIOODD</td>
</tr>
<tr>
<td>LOC_Du01g03310</td>
<td>0.13</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>RIOODD type acyltransferase</td>
<td>RIOODD</td>
</tr>
<tr>
<td>LOC_Du01g03210</td>
<td>0.05</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>RIOODD type acyltransferase</td>
<td>RIOODD</td>
</tr>
<tr>
<td>LOC_Du01g03110</td>
<td>0.17</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>RIOODD type acyltransferase family protein</td>
<td>RIOODD</td>
</tr>
<tr>
<td>LOC_Du01g03010</td>
<td>2.86</td>
<td>plasma membrane</td>
<td>cell wall structure</td>
<td>type II R2 domain-containing GPI-anchored protein</td>
<td>LYM</td>
</tr>
<tr>
<td>LOC_Du01g02910</td>
<td>0.76</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>calcium translocon ATPase</td>
<td>Ca ATPase</td>
</tr>
<tr>
<td>LOC_Du01g02810</td>
<td>0.02</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>GH28M16 - Calmodulin-related calcium sensor</td>
<td>CAM</td>
</tr>
<tr>
<td>LOC_Du01g02710</td>
<td>0.91</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>E3 ligase</td>
<td>E3 ligase</td>
</tr>
<tr>
<td>LOC_Du01g02610</td>
<td>0.00</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>TRIP H2 finger E3 ligase</td>
<td>E3 ligase</td>
</tr>
<tr>
<td>LOC_Du01g02510</td>
<td>0.08</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>TRIP H2 finger E3 ligase</td>
<td>E3 ligase</td>
</tr>
<tr>
<td>LOC_Du01g02410</td>
<td>0.13</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>TRK, IRAK, CIREX K1.5 - PERONIA ortholog</td>
<td>FER</td>
</tr>
<tr>
<td>LOC_Du01g02310</td>
<td>0.14</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>TD calmodulin-binding motif protein</td>
<td>TD CAM</td>
</tr>
<tr>
<td>LOC_Du01g02210</td>
<td>0.78</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>leucine-rich receptor kinase</td>
<td>LRR</td>
</tr>
<tr>
<td>LOC_Du01g02110</td>
<td>7.18</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>leucine-rich receptor kinase 6</td>
<td>LRR</td>
</tr>
<tr>
<td>LOC_Du01g02010</td>
<td>0.05</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>G-box leucine protein receptor kinase</td>
<td>LRR</td>
</tr>
<tr>
<td>LOC_Du01g01910</td>
<td>0.01</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>TRK, IRAK, CIREX K1.16</td>
<td>FER</td>
</tr>
<tr>
<td>LOC_Du01g01810</td>
<td>0.21</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>tyrosine protein kinase</td>
<td>tyrosine protein kinase</td>
</tr>
</tbody>
</table>

(table cont’d)
<table>
<thead>
<tr>
<th>Locus</th>
<th>RNAseq fold change</th>
<th>Subcellular compartment</th>
<th>Functional category</th>
<th>Annotation</th>
<th>AB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC_Os02g43870</td>
<td>0.33</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>receptor kinase</td>
<td>RLK</td>
</tr>
<tr>
<td>LOC_Os02g58520</td>
<td>0.20</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>receptor kinase</td>
<td>RLK</td>
</tr>
<tr>
<td>LOC_Os03g03060</td>
<td>0.98</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>receptor-like kinase</td>
<td>RLK</td>
</tr>
<tr>
<td>LOC_Os03g49570</td>
<td>0.25</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>TF, TRAF, CUL7, kinase</td>
<td>RLK</td>
</tr>
<tr>
<td>LOC_Os04g30010</td>
<td>1.74</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>UWA-KXX2 - WAK receptor-like kinase</td>
<td>WAK</td>
</tr>
<tr>
<td>LOC_Os05g26280</td>
<td>16.04</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>DOF/PHD domain protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g77110</td>
<td>7.18</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>Hsp70 protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os05g03600</td>
<td>2.93</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>Thaumatin</td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g74114</td>
<td>3.49</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>wound-induced protein IM12</td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g42430</td>
<td>0.36</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>xyloglucan glucose/sorbitose dehydrogenase</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g25360</td>
<td>0.15</td>
<td>plasma membrane</td>
<td>signal transduction</td>
<td>transferase family protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g13140</td>
<td>0.14</td>
<td>plasma membrane</td>
<td>lipid metabolism</td>
<td>TORKED 1-like</td>
<td>FOPK</td>
</tr>
<tr>
<td>LOC_Os01g02230</td>
<td>2.82</td>
<td>plasma membrane</td>
<td>lipid metabolism</td>
<td>RAD</td>
<td>HAD</td>
</tr>
<tr>
<td>LOC_Os05g10210</td>
<td>0.07</td>
<td>plasma membrane</td>
<td>lipid metabolism</td>
<td>RAD</td>
<td>HAD</td>
</tr>
<tr>
<td>LOC_Os01g11100</td>
<td>0.03</td>
<td>plasma membrane</td>
<td>secondary metabolism</td>
<td>UMP-glycerocephospho-1'glycerol transferase</td>
<td>UDP</td>
</tr>
<tr>
<td>LOC_Os04g71510</td>
<td>0.10</td>
<td>plasma membrane</td>
<td>primary metabolism</td>
<td>aristolochic acid transporter</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g25050</td>
<td>3.95</td>
<td>plasma membrane</td>
<td>primary metabolism</td>
<td>ATP/ATP transferase</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g42204</td>
<td>7.14</td>
<td>plasma membrane</td>
<td>primary metabolism</td>
<td>GABA permease</td>
<td></td>
</tr>
<tr>
<td>LOC_Os02g74420</td>
<td>0.19</td>
<td>plasma membrane</td>
<td>cell growth</td>
<td>ARFPD/DEP/1ROPD/2F1</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g37170</td>
<td>0.29</td>
<td>plasma membrane</td>
<td>cell growth</td>
<td>LIPase-activating protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os04g40100</td>
<td>2.57</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>BTB/1-Tino-b, Brca-1, brac, Tintrack, Binso, Complex protein</td>
<td>BTB</td>
</tr>
<tr>
<td>LOC_Os03g04446</td>
<td>1.25</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>UBA1-2-like</td>
<td>E7-like</td>
</tr>
<tr>
<td>LOC_Os05g25396</td>
<td>6.15</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>ZW domain-containing 35G protein</td>
<td>ZAD</td>
</tr>
<tr>
<td>LOC_Os05g28180</td>
<td>4.11</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>ZW domain-containing 35G protein</td>
<td>ZAD</td>
</tr>
<tr>
<td>LOC_Os02g74850</td>
<td>7.16</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>PDG/DEK/XR2 nucleasase superfamily protein</td>
<td>FOU</td>
</tr>
<tr>
<td>LOC_Os03g06550</td>
<td>4.48</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>PDG/DEK/XR2 nucleasase superfamily protein</td>
<td>FOU</td>
</tr>
<tr>
<td>LOC_Os03g12085</td>
<td>4.41</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>OSB1R1</td>
<td>DII</td>
</tr>
<tr>
<td>LOC_Os04g71480</td>
<td>4.18</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>circadian clock coupling factor ZOT</td>
<td>ZOT</td>
</tr>
<tr>
<td>LOC_Os03g20070</td>
<td>3.83</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>DORFY156 domain protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g92020</td>
<td>3.42</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>DORFY156 domain protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g07170</td>
<td>3.26</td>
<td>nucleus</td>
<td>signal transduction</td>
<td>pHM1/PMC2-like copper</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g17420</td>
<td>0.77</td>
<td>nucleus</td>
<td>secondary metabolism</td>
<td>ATP-dependent caspase(s) (Cyp) protease-orthomerase protein</td>
<td>CYP</td>
</tr>
<tr>
<td>LOC_Os03g37830</td>
<td>0.25</td>
<td>nucleus</td>
<td>secondary metabolism</td>
<td>maternal effect embryo arrest 59</td>
<td>MEA</td>
</tr>
<tr>
<td>LOC_Os01g06010</td>
<td>0.79</td>
<td>nucleus</td>
<td>cell growth</td>
<td>late embryogenesis abundant protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os03g85420</td>
<td>36.36</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>AP2 domain containing protein, expressed</td>
<td>AP2</td>
</tr>
<tr>
<td>LOC_Os03g40760</td>
<td>8.34</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>AP2 domain containing protein, expressed</td>
<td>AP2</td>
</tr>
<tr>
<td>LOC_Os03g50370</td>
<td>0.95</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>DREB ERF/AP2 transcription regulation factor</td>
<td>AP2</td>
</tr>
<tr>
<td>LOC_Os03g48220</td>
<td>0.11</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>DREB ERF/AP2 transcription regulation factor</td>
<td>AP2</td>
</tr>
<tr>
<td>LOC_Os03g98446</td>
<td>1.43</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>ZU222, AP2 domain containing transcription regulation factor</td>
<td>AP2</td>
</tr>
<tr>
<td>LOC_Os03g43170</td>
<td>4.76</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>UTP transcription regulation factor</td>
<td>UTP</td>
</tr>
</tbody>
</table>

(table cont’d)
<table>
<thead>
<tr>
<th>Locus</th>
<th>RNAseq fold change (P/ctrl)</th>
<th>Subcellular compartment</th>
<th>Functional category</th>
<th>Annotation</th>
<th>Abbr</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC_Os03g46040</td>
<td>4.07</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>hexin-kopf-hex transcription regulation factor</td>
<td>FKH</td>
</tr>
<tr>
<td>LOC_Os03g46140</td>
<td>3.33</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>hexin-kopf-hex transcription regulation factor</td>
<td>FKH</td>
</tr>
<tr>
<td>LOC_Os03g59200</td>
<td>3.73</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>heat shock factor</td>
<td>PFD</td>
</tr>
<tr>
<td>LOC_Os10g58164</td>
<td>0.22</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>homedomain protein rickets 4</td>
<td>H鲜明</td>
</tr>
<tr>
<td>LOC_Os11g44840</td>
<td>4.48</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>MTHB transcription regulation factor</td>
<td>MTHB</td>
</tr>
<tr>
<td>LOC_Os12g46010</td>
<td>0.89</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>one cut domain transcription regulation factor</td>
<td></td>
</tr>
<tr>
<td>LOC_Os12g46430</td>
<td>0.16</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>FCA12 family transcription regulation factor</td>
<td></td>
</tr>
<tr>
<td>LOC_Os13g26010</td>
<td>3.08</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>chloroplast V-ATPase subunit 2</td>
<td>P2K</td>
</tr>
<tr>
<td>LOC_Os10g58440</td>
<td>0.91</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>RNA polymerase II subunit A</td>
<td>PA</td>
</tr>
<tr>
<td>LOC_Os12g54080</td>
<td>7.23</td>
<td>nucleus</td>
<td>transcription regulation</td>
<td>MTHK1YE</td>
<td></td>
</tr>
<tr>
<td>LOC_Os13g26070</td>
<td>0.85</td>
<td>cytosol</td>
<td>signal transduction</td>
<td>MTHK1YE</td>
<td></td>
</tr>
<tr>
<td>LOC_Os15g11090</td>
<td>3.84</td>
<td>cytosol</td>
<td>signal transduction</td>
<td>giberelin receptor</td>
<td>GO</td>
</tr>
<tr>
<td>LOC_Os15g03080</td>
<td>5.74</td>
<td>cytosol</td>
<td>signal transduction</td>
<td>serine/threonine protein kinase</td>
<td>Kinase</td>
</tr>
<tr>
<td>LOC_Os14g48140</td>
<td>5.70</td>
<td>cytosol</td>
<td>signal transduction</td>
<td>STE_MRK kin1 MAP3K18 - STE kinase</td>
<td>Kinase</td>
</tr>
<tr>
<td>LOC_Os15g48140</td>
<td>4.76</td>
<td>cytosol</td>
<td>signal transduction</td>
<td>STE_MRK kin1 MAP3K18 - STE kinase</td>
<td>Kinase</td>
</tr>
<tr>
<td>LOC_Os16g04030</td>
<td>0.75</td>
<td>cytosol</td>
<td>signal transduction</td>
<td>ZRT, YTV, zin c ring do main protein</td>
<td>ZRT1</td>
</tr>
<tr>
<td>LOC_Os14g74030</td>
<td>11.33</td>
<td>cytosol</td>
<td>signal transduction</td>
<td>wound-induced protein WIR1</td>
<td></td>
</tr>
<tr>
<td>LOC_Os15g19070</td>
<td>3.38</td>
<td>cytosol</td>
<td>lipid metabolism</td>
<td>phospha thiolesamine for-methyltransferase</td>
<td>PMB</td>
</tr>
<tr>
<td>LOC_Os12g30080</td>
<td>0.34</td>
<td>cytosol</td>
<td>secondary metabolism</td>
<td>L-aminoacylpropane-1-carboxylate oxidase</td>
<td>ACO</td>
</tr>
<tr>
<td>LOC_Os16g18146</td>
<td>2.83</td>
<td>cytosol</td>
<td>secondary metabolism</td>
<td>coenzyme dehydrogenase</td>
<td>DML</td>
</tr>
<tr>
<td>LOC_Os13g27170</td>
<td>2.18</td>
<td>cytosol</td>
<td>secondary metabolism</td>
<td>L-lysine-L, methionine-dependent methyltransferases</td>
<td>DAAS</td>
</tr>
<tr>
<td>LOC_Os14g71100</td>
<td>0.29</td>
<td>cytosol</td>
<td>secondary metabolism</td>
<td>flavin reductase</td>
<td>SULF</td>
</tr>
<tr>
<td>LOC_Os15g07010</td>
<td>0.29</td>
<td>cytosol</td>
<td>secondary metabolism</td>
<td>thioredoxin reductase</td>
<td>SULF</td>
</tr>
<tr>
<td>LOC_Os14g12070</td>
<td>0.13</td>
<td>cytosol</td>
<td>secondary metabolism</td>
<td>thioredoxin reductase</td>
<td>SULF</td>
</tr>
<tr>
<td>LOC_Os13g40050</td>
<td>2.48</td>
<td>cytosol</td>
<td>primary metabolism</td>
<td>L-alo-threonine aldolase</td>
<td></td>
</tr>
<tr>
<td>LOC_Os14g38080</td>
<td>4.27</td>
<td>cytosol</td>
<td>cell growth</td>
<td>dynamin light chain type 1 domain microtubule motor</td>
<td></td>
</tr>
<tr>
<td>LOC_Os12g47020</td>
<td>1.73</td>
<td>plastid</td>
<td>signal transduction</td>
<td>protein kinase</td>
<td>NPR1</td>
</tr>
<tr>
<td>LOC_Os12g47020</td>
<td>3.17</td>
<td>plastid</td>
<td>lipid metabolism</td>
<td>alpha beta lip hydroxase</td>
<td>a/b lip</td>
</tr>
<tr>
<td>LOC_Os12g47010</td>
<td>4.21</td>
<td>plastid</td>
<td>lipid metabolism</td>
<td>chalcone flavanone isomerase</td>
<td>FA</td>
</tr>
<tr>
<td>LOC_Os12g47070</td>
<td>5.27</td>
<td>plastid</td>
<td>lipid metabolism</td>
<td>fatty acid hydroxylase</td>
<td>FA</td>
</tr>
<tr>
<td>LOC_Os12g47070</td>
<td>0.25</td>
<td>plastid</td>
<td>lipid metabolism</td>
<td>chloroplast 2-ka domain fatty acid composition</td>
<td>FA</td>
</tr>
<tr>
<td>LOC_Os12g47020</td>
<td>13.21</td>
<td>plastid</td>
<td>lipid metabolism</td>
<td>tyrosylglycine lipase</td>
<td>Lipase</td>
</tr>
<tr>
<td>LOC_Os12g50010</td>
<td>0.15</td>
<td>plastid</td>
<td>lipid metabolism</td>
<td>SWB/MOS2 domain protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os12g50080</td>
<td>4.85</td>
<td>plastid</td>
<td>secondary metabolism</td>
<td>methane oxide synthase (MOS)</td>
<td>AOS</td>
</tr>
<tr>
<td>LOC_Os12g52070</td>
<td>3.43</td>
<td>plastid</td>
<td>primary metabolism</td>
<td>Tricose-to Pi aldolase</td>
<td></td>
</tr>
<tr>
<td>LOC_Os12g52080</td>
<td>0.32</td>
<td>plastid</td>
<td>primary metabolism</td>
<td>pyruvate kinase</td>
<td></td>
</tr>
<tr>
<td>LOC_Os12g54010</td>
<td>0.06</td>
<td>plastid</td>
<td>primary metabolism</td>
<td>SPT1-related protein kinase</td>
<td></td>
</tr>
<tr>
<td>LOC_Os12g50140</td>
<td>0.33</td>
<td>mitochondria</td>
<td>signal transduction</td>
<td>alpha beta lip hydroxase</td>
<td>a/b lip</td>
</tr>
<tr>
<td>LOC_Os15g08040</td>
<td>5.22</td>
<td>mitochondria</td>
<td>signal transduction</td>
<td>FTP1, penta lecoenzyme peptide repeat protein</td>
<td>PRPI</td>
</tr>
<tr>
<td>LOC_Os16g08140</td>
<td>4.78</td>
<td>mitochondria</td>
<td>secondary metabolism</td>
<td>DOG-Fe oxygenase</td>
<td>DOG</td>
</tr>
<tr>
<td>LOC_Os16g24070</td>
<td>0.09</td>
<td>mitochondria</td>
<td>cell growth</td>
<td>UDP-glucose-4-epimerase 4A protein</td>
<td></td>
</tr>
<tr>
<td>LOC_Os15g20080</td>
<td>0.34</td>
<td>ER</td>
<td>signal transduction</td>
<td>CHSR3 domain-containing protein</td>
<td>CHSR3</td>
</tr>
</tbody>
</table>

(table cont’d)
These DEGs encode components with putative functions in signal transduction (37%), cell wall structure (23%), lipid composition (13%), transcription regulation (10%), secondary metabolism (9%), primary metabolism, or cell growth (3%), that are mostly targeted to the apoplast (31%), plasma membrane (28%), nucleus (18%), cytosol (10%), or plastid (6%) (Figure 28A). Strikingly, more than half (53%) of the DEGs are predicted to encode proteins targeted to the apoplast or plasma membrane, and have functions in signaling or cell wall and lipid composition. Among this group are a number of pectinases, arabinogalactan proteins (AGPs), and expansins that mostly are down-regulated by the 24-hour Pi deficiency treatment (Figure 28B, Table 6).

A previous study in Arabidopsis identified a similar response of cell wall hydrolytic enzyme-encoding loci in roots subjected to Pi-deficiency treatments of 1, 6, and/or 24 hours (W.-D. Lin et al., 2011a). Together this suggests that modification of the cell wall is an early and prominent response to Pi deficiency in roots and shoots across plant species. In addition to the down-regulation of cell wall-related components was a large group of signaling components, including many receptor-like kinases (RLKs), that were predominantly up-regulated (Figure 28, Table 6). One of the RLKs is a *Catharanthus roseus* RLK1-like kinase orthologous to Arabidopsis FERONIA (FER), which has been shown to regulate cell expansion in response to diverse developmental and environmental cues (Liao et al., 2017). For example, during salinity stress, FER maintains...
cell wall integrity, and is necessary for root growth recovery (Feng et al., 2018). Recently it was demonstrated that FER is one component of a signaling module that transduces cell-wall signals during salt stress (Zhao et al., 2018). In the absence of salt stress, a group of apoplastic leucine-rich repeat extensins (LRX) bind to RAPID ALKALINIZATION FACTOR (RALF) peptides. In response to salt stress, LRX and RALF dissociate, and RALF peptides bind FER. This results in FER internalization and, subsequently, inhibition of growth and initiation of stress responses. Calcium transients and SITE-1 PROTEASE (S1P) activity also play roles in RALF/FER signaling (Feng et al., 2018; Stegmann et al., 2017). Notably, our CS4-CS3 DEG list also contains genes encoding six RALF peptides, an LRX, several Ca2+ transport-related components (e.g. Ca2+ ATPase and calmodulin), and two S1P proteases (Table 6).

In addition to the signaling and cell wall components were a number of transcription factors among the CS4-CS3 DEGs, including five AP2 superfamily factors, two HLH factors, and two WRKY transcription factors. These represent families of transcription factors known to be responsive to a number of biotic and abiotic stressors. It is tempting to speculate that these regulatory genes, along with the differentially-expressed CS4-CS3 structural genes comprise a transcriptional regulatory network aimed at transducing Pi deficiency signals and initiating reduced cellular growth and tolerance to low Pi (Figure 29).

A 24-hour Pi deficiency-treatment poises translation-related genes for inactivity

The CS5-CS1 and CS4-CS5 at TSS transition groups with 2495 and 2355 number of genes are among the largest groups of transitions after CS4-CS3 group. They both have essentially the same GO terms enriched, including translation. However, they are not significantly correlated with either up or down-regulated DEGs. Pi starvation response
genes are impacted in a short- or long-term manner (Hongmei Cai et al., 2013; Secco et al., 2013b; Thibaud et al., 2010). It is possible that CS5-CS1 and CS4-CS5 genes belong to late transcriptional Pi responses. Thus, they have chromatin changes that occur in response to Pi starvation, but exhibit no corresponding transcript differences within the time-scale of our RNA-seq experiment. Consistent with this, a previous transcriptomic study of rice shoots (Secco et al., 2013b) shows a large number of translation-related genes that are down-regulated by 21 days of Pi deficiency, and many of these have a chromatin state transition. To determine whether CS5-CS1 and CS4-CS5 chromatin state transition genes were significantly over- or under-represented among the 21 day down-regulated genes (Secco et al., 2013b), we determined the overlap between each transition and the down-regulated genes via bootstrapping analyses. Indeed, this analysis indicated a bias of genes down-regulated during prolonged Pi deficiency with our CS5-CS1 and CS4-CS5 transition genes (1000 iterations; binomial test, p-value < 0.01; Figure 29).

To investigate if P deficiency resulted in low mRNA biosynthesis, we compared the expression level of control and Pi deficiency samples. The total mRNA level of Pi deficiency samples were significantly lower than the control samples (P-value < 2.2e-16, Wilcoxon signed-rank test). This further supports a decline in translation in response to Pi deficiency.
Figure 28. Predicted functions and subcellular locations of differentially-expressed genes (DEGs) having a chromatin state (CS) transition of CS4-CS3.

Figure 29. Predicted responses and functions of differentially-expressed genes having a chromatin state (CS) transition of CS4-CS3.
Figure 30. Bootstrapping analysis showing the overlap between genes exhibiting chromatin state transitions (CS5-CS1, CS4-CS5) and genes down-regulated in shoots following a 21-day Pi deficiency treatment (Secco et al., 2013b).

Figure 31. Density plot of RNAseq data from control (Ctrl) and Pi deficiency samples.
Discussion

Plants cope with environmental stress by altering chromatin features such as histone modifications, histone variants and nucleosome occupancy. The role of each feature in the regulation of plant stress responses has been studied discretely (C.-Y. Chen et al., 2015; Kuo et al., 2014; Secco et al., 2015; Aaron P Smith et al., 2010; Yong-Villalobos et al., 2016; Yong-Villalobos et al., 2015; Zahraeifard et al., 2018; Q. Zhang et al., 2018), however the combination of these epigenetic changes associate with Pi-starvation response was uncharacterized. We defined distinct chromatin states across the rice genome, by integrating multiple aspects of chromatin structure, including H2A.Z, H3K4me3, and nucleosome positioning. In addition, we found the chromatin-level mechanisms linked to the response of rice seedlings to Pi deficiency. Moreover, examination of specific chromatin states transition showed that loss of H3K4me3 at the TSS of subgroups of differentially expressed genes that are implicated in lipid biosynthesis and cell wall integrity. Our findings demonstrate that in response to Pi starvation, rice employs adaptation strategies to reduce cell wall expansion and alter lipid composition in order to reduce shoot growth and enhance recycling of endogenous Pi.

Next, we compared the profile of the CS4-CS3 DEGs to public transcriptome studies using the signature tool in Genevestigator (Hruz et al., 2008). The CS4-CS3 DEG expression profile had substantial overlap with several pairwise comparisons from a previous study on rice lamina joint development (Zhou, Xiao et al. 2017). Comparisons between older stages of development (maturation or post-maturation) with a younger stage showed similar expression profiles as our CS4-CS3 Pi deficiency DEGs. Interestingly, cell-wall thickening is a prominent feature during younger stages of lamina joint development, and this declines over time. This may suggest that Pi deficiency results
in decreased cell wall thickening, or more generally, a decrease in cell elongation. Transcriptomic profiles of several biotic and abiotic (e.g. salinity and heat) stressors also showed high similarity to our CS4-CS3 DEG profile, suggesting the apparent apoplastic signaling network overlaps with multiple stressors. Interestingly, a transcriptomic profile of the *jmj706* mutant, which lacks a H3K27me3 demethylase, is also similar to the CS4-CS3 DEG profile. This is consistent with our results showing an important role for chromatin structure in modulating gene expression of this signaling network.

The pectin network changes have an important role in early responses to Pi starvation (Ogden et al., 2018). Hoehenwarter et al (2016) showed pectin has a role in Pi mobilization. It was also shown that there is a correlation between root growth inhibition under Pi deficiency and changes in the expression of cell wall modifying genes (Hoehenwarter et al., 2016). Alteration of gene expression involved in modifying pectin was among the early responses to Pi deficiency. The expression of cell wall hydrolytic enzymes such as a pectinesterase, polygalacturonases, pectate lyases, and pectin methylesterases was down-regulated while a pectinesterase inhibitor was up-regulated under Pi deficiency. These data showed the repression of genes involved in modifying pectin (W.-D. Lin et al., 2011a). It was well established that pectin has important roles in cell growth (Cosgrove, 2018). Pectin with low levels of methyl groups cause reduction in cell wall extensibility and inhibit growth (Derbyshire et al., 2007). The role of pectin acetylesterase as a structural regulator in planta was demonstrated. This enzyme modulates the status of pectin acetylation to alter the physiochemical properties of the cell wall's polysaccharides, affecting cell extensibility (Gou et al., 2012). In our data, in CS4-CS3 DEG, pectin hydrolase (2), pectinesterase (4) and pectinacetylesterase (1)
genes were down-regulated. It means in early responses to Pi deficiency in shoot the pectin modifying enzyme was down-regulated to decrease cell wall extensibility and inhibit shoot growth.

In response to long term Pi deficiency cellulose synthase-like protein were found to be down-regulated (Misson et al., 2005a). On the other hand CELLULOSE SYNTHASE-LIKE5 (CSLB5), was up-regulated under short term Pi deficiency and cslb5 mutants showed shorter root hairs under Pi-deficient conditions. In our data CSLF8 - cellulose synthase-like was down-regulated.

Mutation of the cellulose synthase gene, \textit{AtCesA8\textbackslash IRX1}, causes Arabidopsis plants to become more tolerant to drought stress, and to become dwarfed. ABA, proline and soluble sugars over-accumulate in this mutant (Z. Chen et al., 2005). It was known that the main sink for soluble sugars is the cell wall cellulose (Babb & Haigler, 2001). Drought stress results in induction of the content of some sugars by inhibiting some enzyme activities involved in cellulose synthesis (Foyer et al., 1998). In our data CSLF8 - cellulose synthase-like was down-regulated. In response to Pi deficiency one of the shoot responses is increase in sugar concentration. This may happen by down-regulation of cellulose biosynthesis.

The extracellular concentrations of AGPs were shown to be induced under salt stress in a variety of plant species (Lamport et al., 2006; L Zagorchev et al., 2008). The role of AGPs as storage molecules for Ca2+ was previously reported. AGPs discharge Ca2+ in response to different stimuli to trigger Ca-dependent signal cascades (Lyuben Zagorchev et al., 2014). In our data AGPs were up and down-regulated. In response to Pi deficiency, AGPs may discharge Ca2+ to turn on the signaling cascade.
Pectin cross-linking was altered in response to salt stress. It was shown that high salinity causes defects in pectin cross-linking, leading to weakening of the cell wall. Changes in cell wall softness was sensed by the extracellular domain of FER which interacts directly with pectin. These interactions induce cell-specific [Ca2+] transients that recover root growth and also maintain cell-wall integrity. Our data support the existence of a similar mechanism that occurs in response to Pi deficiency.
Chapter 4. The Role of SET DOMAIN GROUP2 in Pi Starvation Responses

Introduction

Histone methyltransferases (HMTs) are histone modifying enzymes that catalyze the addition of methyl groups donated from S-adenosyl methionine (SAM) to the histone (Triever et al., 2002). Methylation can occur on histones H3 and H4 at both lysine and arginine amino acids (Ng et al., 2007). Two types of H exist, Histone lysine N-methyltransferase (HKMTs) and histone–arginine N-methyltransferase (HRMTs) (Kouzarides, 2002), but the majority of Arabidopsis HMTs are HKMTs (X. Wang et al., 2007). Five lysine methylation sites have been described in plants, lysines 4, 9, 27 and 36 of H3, and lysine 20 of H4 (Pfluger & Wagner, 2007). These modifications are “written” by different HKMTs. Two families of enzymes have been identified that methylate lysine: the first group is called the SET-domain containing group (named after the Drosophila histone methyltransferases su (var) 3-9, Enhancer-of –zeste, and Trithorax) and the second group are the DOT1-like proteins (non-SET domain containing lysine specific) (Guo et al., 2010b; Nguyen & Zhang, 2011). All known HKMTs in plants belong to the SET domain group (SDG). The SDG proteins with HKMT activity can be divided into five classes in plants, based on their domain architecture (the characteristic SET domain, cysteine rich region and additional conserved domain) and/or differences in enzyme activity (C. Liu et al., 2010; Ng et al., 2007; Pontvianne et al., 2010). H3K4 is one of the important histone marks in Arabidopsis and rice. More than two-thirds of Arabidopsis genes contain at least one type of H3K4me1,me2,me3 (Guo et al., 2010b). Half of the protein-coding genes in rice have di- and/or trimethylated H3K4 on the chromatin (X. Li et al., 2008). H3K4me3 preferentially located at actively transcribed genes plays
important roles in maintaining the proper expression level of numerous genes. Several Arabidopsis SET domain group proteins have been identified and shown to belong to Trithorax group (Trx G) and are associated with H3K4 trimethyltransferase activity (Gendler et al., 2008). However, among Trx G proteins, the ATXR3/SDG2 is a major H3K4 tri-methyltransferase and plays a prominent role in H3K4 deposition (Berr et al., 2010; S. Kim et al., 2013; Yao et al., 2013; Yun et al., 2012). SDG2 is a large and unique protein, which is highly conserved in plants. SDG2 proteins have the SET domain, which has been shown to have tri-methyltransferase activity (Berr et al., 2010). In addition, loss of SDG2 leads to severe and genome-wide decrease in H3K4me3. It was demonstrated in a study by Guo et al (Guo et al., 2010a) that SDG2 is responsible for high levels of H3K4me3 at the vast majority of genes in Arabidopsis. Not only is SDG2 required for H3K4me3 deposition but it is also required for normal expression of a large number of genes as shown by using Affymetrix ATH1 microarrays and real-time RT-PCR (Guo et al., 2010). It was found that 271 genes were up-regulated and 321 genes were down-regulated by greater than four-fold in an sdg2 mutant. Furthermore, they found that the genes up-regulated in sdg2 are involved in a wide range of biological processes, whereas transcription factors and DNA-binding proteins are overrepresented in down-regulated genes. Therefore it was suggested that SDG2 is required to maintain the proper expression level of many transcriptional regulators (Guo et al., 2010a). SDG2 has also been implicated in the regulation of various hormone responsive genes and many developmental processes, including development of sporophyte and gametophytes, root growth and development and flowering time. Therefore, the knockdown of its function
results in a pleiotropic phenotype, such as dwarfism, impaired male and female gametophyte development (Berr et al., 2010; Yao et al., 2013; Yun et al., 2012).

I sought to employ chromatin immunoprecipitation in this study to compare the distribution of H3K4me3 between wild type and sdg2 mutant at P homeostasis genes to understand the role of SDG2 in Pi starvation responses. Unfortunately, despite my efforts in growing sdg2 mutant plants in a variety of ways, and attempts at generating SDG2-RNAi knockdown lines, I was not able to develop a means to collect enough sdg2 tissue for the analyses.

Materials and Methods

Plant material and genotyping the T-DNA insertion mutant line, sdg2-1

The T-DNA insertion mutant line, sdg2-1 obtained from Dr. Zhang’s laboratory and Arabidopsis thaliana Col ecotype seedlings were grown in ¼ MS plates under 16-h-light/8-h dark for 2 weeks. The seedlings were transferred to hydroponic solution. After 2 weeks half of the seedlings were transferred to a solution lacking Pi (-Pi) and the rest to solution with Pi. Plant tissues (shoot and root) from wild type were harvested. To collect tissue from the sdg2 mutant it was necessary to do genotyping. The leaves from seedlings which seemed to be homozygous, were collected for genomic DNA extraction. PCR amplification with gene specific primers was done to determine the genotype for the T-DNA.

Clonal propagation of sdg2 roots by sustained root culture

For sustained root culture, plants from wild type and heterozygous sdg2-1/+ seeds were grown on sterile agar at 22 °C under continuous light for 12 days after germination. Homozygous sdg2-1 seedlings were distinguished and confirmed by PCR, transferred to
Arabidopsis root culture medium on an orbital shaker for 2-4 weeks (Czakó et al., 1993). sdg2-1 and wild type roots were excised and transferred to fresh ARC media. Two days of IAA treatment was given to roots. Roots were rinsed and grown in ARC media for 1-2 months. Arabidopsis root culture (ARC) contains MS salts, 4.3 g/L; Vitamix stock (Vitamin B1, 500 mg/100 ml; Vitamin B6, 50 mg/100ml; Glycine, 100 mg/100ml; Nicotinic acid, 50 mg/100ml; Folic acid, 25 mg/100 ml; Biotin, 50 mg/100ml) 2 ml/L; 3% Sucrose; myo-inositol, 200 mg/L.

**Generation of Arabidopsis SDG2-RNAi knockdown lines**

The pHANNIBAL system was used for the cloning of an SDG2 RNAi construct. Sense and antisense cDNA fragments corresponding to the SDG2 gene were cloned into the two multiple cloning sites to form an inverted repeat. The cassette was released by cutting at flanking NotI sites and cloned into the unique site of the A. tumefaciens binary vector pART27 (Yin et al., 2005). Vectors and cloning strategies for RNAi are shown in Figure 31. The resulting RNAi construct in pART27 was introduced into A. tumefaciens strain GV3101, which was used to transform plants by using the floral dip method. RNAi transgenic plants were characterized by kanamycin selection plates. Transgenic plants carrying the RNAi (T1, T2 and T3) were identified by selection using kanamycin.

**Results**

**Identify the correlation between the genes that are misregulated under Pi deficiency and the sdg2 mutant**

The histone methyltransferase, SDG2, was shown to be a major H3K4me3 methylase in Arabidopsis (Guo et al., 2010a). Mutation of SDG2 resulted in a decrease in the
abundance of H3K4me3, and resulted in the misregulation of numerous genes (Guo et al., 2010a). Upon visual inspection of the specific genes, we noticed that many of the

![Figure 32. Vectors and cloning strategies for RNAi](image)

genes are also related to P homeostasis pathways. We sought to examine the correlation between P signaling genes and mutation of SDG2 in more detail. Therefore, gene set enrichment (GSEA) was performed on a list of misregulated genes in the *sdg2* mutant by using the AgriGO database (Du et al., 2010). We selected those genes whose
expression were increased more than five times in the sdg2 mutant. Figure 33 shows the top 7 significantly enriched GO categories obtained from the GSEA analysis. Among these enriched GO terms is “response to Pi starvation”.

Figure 33. Significantly enriched GO categories. Genes up-regulated 5-fold or greater in the Arabidopsis sdg2 mutant (“Input list”) comprise GO categories (x-axis) that are enriched above the corresponding genes present in the entire Arabidopsis genome (“Background/Reference”).

The GO term analysis indicated that a significant proportion of the genes up-regulated by mutation of SDG2 are known to be responsive to Pi deficiency. To determine the extent of this overlap, the list of genes in the sdg2 mutant that were up-regulated 5-fold or greater were compared to those genes in a previous transcript profiling study (Bustos et al., 2010) that were up-regulated 5-fold or greater by Pi deficiency. We found that 33.3% (61/183) of the up-regulated genes in the sdg2 mutant were also up-regulated by Pi deficiency (Figure 34), confirming that a major proportion of the genes impacted by mutation of SDG2 are related to Pi starvation responses.
Compare the H3K4me3 levels at P and Fe homeostasis genes between wild type and 
sdg2 mutant/RNAi lines in Arabidopsis and rice

The expression of a large number of genes including P homeostasis genes is 
misregulated in sdg2 mutants. According to Guo et al (Guo et al., 2010a) and Gene Set 
Enrichment Analysis (Figure 33), we found out that many of the up-regulated genes in 
sdg2 show significant enrichment of P starvation related genes. The hypothesis for this 
aim was that SDG2 is responsible for changing the level of H3K4me3 at P deficiency 
genes. To characterize the role of SDG2 in H3K4me3 at P deficiency genes, we sought 
to perform chromatin immunoprecipitation (ChIP) and real-time PCR at selected genes 
and compare the chromatin levels of H3K4me3 between wild type and sdg2 mutant in 
Arabidopsis. We know that heterozygous sdg2 mutant plants are phenotypically
indistinguishable from wild type. In contrast, homozygous sdg2 mutant plants show some morphological defects at 6–8 days after germination, including less shoot growth with curly leaves and significantly shorter roots. They remain dwarfed, with smaller rosettes throughout vegetative growth, and flower significantly earlier (Figure 35A). Furthermore, sdg2 mutants are completely sterile (Guo et al., 2010a). The T-DNA insertion mutant line, sdg2-1, was obtained from Dr. Zhang’s laboratory. The leaves from 2-3 weeks old seedlings which seemed to be homozygous, were collected for genomic DNA extraction. PCR amplification with gene specific primers was done to determine the genotype for the T-DNA (Figure 35B).

Because of the severe growth defects, it was very challenging to identify adequate homozygous sdg2 mutants for the ChIP experiments. To surmount this difficulty we did clonal propagation of sdg2 roots by sustained root culture and also produced transgenic Arabidopsis seedlings containing an RNA interference (RNAi) construct targeting SDG2. Unfortunately, because of the growth and development defect in the sdg2 mutant, we could not get sufficient sdg2 root tissue with sustained root culture methods. Finally, SDG2 expression was analyzed by RT-PCR to identify transgenic plants that exhibit an appropriate reduction in SDG2 expression. Unfortunately, none of our RNAi lines showed significant reduction in the level of SDG2 transcripts.
Figure 35. Phenotypes of sdg2 (A) sdg2 seedlings are small and remain dwarfed throughout development (B) PCR result using DNA extracted from wild type (Col), sdg2 mutant.
Conclusions

Phosphorus is one of the mineral nutrients vital for plant growth, development and reproduction. Not only is it a major component of nucleic acids and phospholipids but it also plays an important role in regulation of signaling, enzymatic reactions and metabolic pathways. Due to the low solubility and mobility of Pi, its deficiency is one of the major factors that limit crop productivity. Therefore, it is essential to investigate the underlying mechanisms involved in Pi-starvation responses to improve Pi-use efficiency. Plants have developed complex responses to tolerate Pi deficiency and retain the level of Pi. The responses to Pi deficiency are highly controlled at the transcriptional level. In general, plants cope with environmental stress by altering chromatin features such as histone modifications, histone variants and nucleosome occupancy. The role of each feature in regulation of plant stress responses has been studied discretely, however the combination of these epigenetic changes associated with Pi-starvation responses has been uncharacterized. Recently, software programs such as chromHMM enables merging multiple chromatin marks together and facilitates the discovery of chromatin structure and dynamics related to responses to internal and external stimuli. Therefore, understanding the chromatin-level mechanisms that regulate responses to stimuli are important for upcoming efforts to develop crops with high abilities to adapt to stress conditions. Here we sought to understand how expression of Pi-starvation response genes are regulated at the transcriptional and chromatin levels in rice genome.

In the first part of this research study, the genome-wide distribution pattern of H3K4me3 was examined in the rice genome at four different gene types. We established that there is a prominent H3K4me3 peak at the 5’ end of protein-coding genes but not at
other gene types (Figure 7). We also observed that there is a strong correlation between the pattern of H3K4me3 localization and gene length (Figure 8). This finding is consistent with studies that showed the conserved pattern of H3K4me3 at the TSS of all organisms (Howe et al., 2017). Clustering analysis showed H3K4me3 abundance varies at the TSS of PCG, and PCG with a high level of H3K4me3 at the TSS are enriched in housekeeping genes, whereas low-H3K4me3 PCG are enriched in stress-responsive genes (Figure 9). In both Arabidopsis and rice the correlation between H2A.Z at the 5’ end of PCG and housekeeping genes and H2A.Z pattern across gene bodies with gene responsiveness were established (Coleman-Derr & Zilberman, 2012; Zahraeifard et al., 2018). The pattern of H3K4me3 and H2A.Z are similar in PCG, but there is a higher level of H3K4me3 at TE and TEG compared to H2A.Z (Figure 7). The presence of DNA methylation at transposable loci and its antagonistic relationship with H2A.Z may cause the low abundance of H2A.Z at these regions (Conerly et al., 2010; Zahraeifard et al., 2018; Zemach et al., 2010; Zilberman et al., 2008). The correlation coefficient and peak calling analysis demonstrated the co-localization of H3K4me3 and H2A.Z in the rice genome (Figure 7), which is in agreement with previous studies (Xiaozhuan Dai et al., 2017; K. Zhang et al., 2017). Next, we investigated the correlation between H2A.Z and H3K4me3 abundance and gene expression. The result reflected the positive and negative correlation between the abundance of H3K4me3 at TSS and TES respectively, while H2A.Z at both TSS and TES were negatively correlated with gene expression (Figure 11). H3K4me3 at the TSS was shown to be strongly associated with gene expression (X. Zhang et al., 2009). Finally, the effect of Pi deficiency at H3K4me3 distribution was
examined. It was shown that Pi deficiency altered H3K4me3 distribution of PCG, especially at TSS (Figure 12).

In the second part of the study, I identified and characterized five chromatin states in the rice genome (Figure 15). We found that the distribution of these chromatin states were different at transposable-related loci, PSG and PCG. In addition, there were some biases toward the TSS and TES in the rice genome (Figure 16). A comparison between the chromatin state distribution of control and Pi deficiency conditions revealed a global increase of H2A.Z and/or decrease of H3K4me3 (Figure 17, Figure 18, Figure 19). Moreover, the chromatin state distribution of transposable-related loci and TSS and TES of PCG were changed in response to Pi deficiency. Pi deficiency results in H2A.Z and H3K4me3 reduction at TSS and TES (Figure 20). To further investigate PCG chromatin state transitions, we examined the chromatin state of each PCG at control and Pi deficiency conditions. In response to Pi deficiency, 40% of all protein-coding genes show a chromatin state transition (Figure 22). In response to Pi deficiency, the largest groups of genes displaying chromatin state transition were CS4-CS3, CS5-CS1 and CS4-CS5. GO term enrichment analysis showed similar and translation-related terms for both CS5-CS1 and CS4-CS5 while CS4-CS3 had unique terms related to stress-responses (Figure 22). Analysis of overlap between chromatin state shifts and differentially-expressed genes revealed several of these transitions are significantly enriched in subsets of genes differentially-expressed by Pi deficiency (Figure 24, Figure 25). Interestingly, CS4-CS3 transition was enriched in both up- and down-regulated genes. This finding suggests a complex role of H3K4me3 in regulation of gene expression under Pi deficiency. Furthermore, we found up-regulation of gene expression associated with H2A.Z and/or
H3K4me3 loss however, down-regulation of gene expression are correlated with gain of H2A.Z (Figure 24).

In the last part of this study, we performed a data mining analysis to find the putative function and subcellular localization of the most prominent subgroup, DEGs exhibiting a CS4-CS3 transition. This showed that around 53% of CS4-CS3 DEGs encode components with signaling, cell wall modification and lipid metabolism functions, and are targeted to the apoplast or cell membrane (Figure 27 Figure 28). Therefore, examination of specific chromatin state transitions showed that loss of H3K4me3 at the TSS of subgroups of DEGs that are implicated in lipid biosynthesis and cell wall integrity. The presence of genes encoding FER, six RALF peptides, and LRX among CS4-CS3 DEGs suggest that RALF and FER are involved in a signaling module that transduces cell-wall signals during Pi deficiency similar to their role during salt stress (Figure 29). Together my findings suggest that in response to Pi starvation, rice employs adaptation strategies to reduce cell wall expansion and alter lipid composition in order to reduce shoot growth and enhance recycling of endogenous Pi.
References


natural variation between Arabidopsis accessions. *Plant, cell & environment, 26*(11), 1839-1850.


Lan, P., Li, W., & Schmidt, W. (2012). Complementary Proteome and Transcriptome Profiling in Phosphate-deficient Arabidopsis Roots Reveals Multiple Levels of


patterns in response to dehydration stress in Arabidopsis thaliana. *BMC plant biology, 10*(1), 238.


Starvation but Promotes Expression of Key Housekeeping Genes. *Journal of experimental botany*.


Vita

Maryam Foroozani was born in 1981, in Iran. She received her bachelor’s degree in plant biology in 2003 from University of Isfahan, Isfahan, Iran and her master’s degree in Plant Cell & Developmental Biology in 2007 from Shiraz University, Shiraz, Iran. Then, she had taught at school for 5 years before she joined Dr. Aaron Smith’s lab in the fall of 2012. Her main doctoral work focused on on understanding how expression of phosphate-starvation response genes are regulated at the transcriptional and epigenetic levels. She will receive her PhD’s degree in December 2019 and plan to pursue her post-doctoral training on plant stress adaptation and epigenetic silencing.