FUNCTIONS OF ARABIDOPSIS C-TERMINAL DOMAIN PHOSPHATASE-LIKE 4 IN GLOBAL TRANSCRIPTIONAL REGULATION

A Dissertation

by

AKIHITO FUKUDOME

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Hisashi Koiwa
Committee Members,	Timothy Devarenne
	Libo Shan
	Xiuren Zhang
	Keyan Zhu-Salzman
Head of Department,	Dirk B. Hays

May 2017

Major Subject: Molecular and Environmental Plant Sciences

ABSTRACT

Phosphoregulation of the carboxyl-terminal domain of RNA polymerase II largest subunit (pol II-CTD) couples transcription and co-transcriptional modification of nascent RNA. Although the molecular mechanisms have been extensively studied in vertebrates, understanding of that in plants is still in its infancy. Through genetics, biochemical and transcriptomic approaches, this dissertation work characterizes functions of a pol II-CTD phosphatase-like protein from *Arabidopsis thaliana*, CPL4, in the phosphoregulation of pol II-CTD during protein-coding and non-coding RNA transcriptions.

CPL4 interacts with and dephosphorylates pol II-CTD both *in vitro* and *in vivo*, showing that CPL4 regulates pol II-CTD phosphorylation status in Arabidopsis. An amino acid substitution in the catalytic motif abolished the phosphatase activity of CPL4. The catalytically inactive protein strongly inhibits transcription in transient assays, likely due to a dominant negative effect. Deletion of Breast cancer C-terminal (BRCT) domain alleviates the inhibitory effect of the catalytically inactive CPL4, suggesting that BRCT domain is necessary for CPL4's function. A suite of xenobiotic stress responsive genes shows constitutive up-regulation in CPL4 knockdown transgenic ($CPL4_{RNAi}$) lines, indicating that CPL4 negatively regulates the toxic chemical detoxification pathway.

The $CPL4_{RNAi}$ plants accumulate aberrant 3'-extended transcripts from many pol II-dependent small nuclear RNA (snRNA) loci. The snRNA 3'-extension gives rise to a

snRNA transcript fused with a downstream protein-coding gene (DPG) if present. Such snRNA-DPG fusion transcripts can be found in other plant species. A short, unstable non-coding RNA produced from a protein-coding locus driven by a transposable-embedded snRNA promoter can yield the full-length product in the $CPL4_{RNAi}$ plants. These snRNA-DPGs can be induced in wild type by salt stress, which affects pol II-CTD phosphorylation status. These results indicate a potential stress-inducible conversion of non-coding RNA transcription into protein-coding transcription mediated by pol II-CTD phosphoregulation.

$CPL4_{RNAi}$ root explants exhibit enhanced capability of *de novo* shoot organogenesis due to cytokinin hypersensitivity and earlier induction of shoot apical meristem regulatory genes. A potential involvement of an operon-like cluster of thalianol biosynthesis genes in the $CPL4_{RNAi}$ organogenesis phenotype is implicated. Taken together, Arabidopsis CPL4 is an essential pol II-CTD phosphatase, regulating stress-responsive and organogenesis pathways through protein-coding and non-coding RNA transcriptions.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my committee chair, Dr. Hisashi Koiwa, for his patience, guidance and support throughout the course of this research. I also thank my committee members, Dr. Timothy Devarenne, Dr. Libo Shan, Dr. Keyan Zhu-Salzman and Dr. Xiuren Zhang for fruitful discussion, encouragement and support. My thanks also go to Dr. Alan Pepper for his guidance and support in the experiments related to hypocotyl growth measurements in Chapter IV; Dr. Misato Ohtani and Dr. Munetaka Sugiyama for *srd2-1* mutant lines. I also thank my friends and colleagues for assisting, encouraging and inspiring me.

I truly appreciate various funding sources supporting my Ph.D. work. Japan Student Services Organization (JASSO) supported my first three years of Ph.D. study. Department of Horticultural Sciences, Dr. Fred Davis and Dr. David Reed provided me with invaluable opportunities as a Teaching Assistant. Texas A&M University Office of Graduate and Professional Studies has supported my last year through Dissertation Fellowship.

Finally, I'd like to thank my parents for their understanding, encouragement and continuing support.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Hisashi Koiwa of MEPS/the Department of Horticultural Sciences, Professor Timothy Devarenne and Professor Xiuren Zhang of MEPS/the Department of Biochemistry and Biophysics, Professor Libo Shan of MEPS/the Department of Plant Pathology and Microbiology, and Professor Keyan Zhu-Salzman of MEPS/the Department of Entomology.

Raw data of mass spectrometry in Chapter II (Figure 2.1) was provided by former Post-doctoral researcher Xiaoqiang Wu of the Koiwa lab, Ph. D. student Krishna Kumar of MEPS/the Department of Biology, Kimberly May and Dr. William K. Russell of the Department of Chemistry. Raw microarray data I analyzed in Chapter II was provided by former Ph. D. student Emre Aksoy and Dr. Hisashi Koiwa of the Koiwa lab. From RNA samples I submitted, raw RNA-seq data analyzed in Chapter III and Chapter IV were generated by Otogenetics Corporation (Norcross, GA) and Texas A&M University Genomics and Bioinformatics Service, respectively. Raw data for im-ncRNAs RNA-seq in Chapter III was provided by Dr. Hisashi Koiwa of the Koiwa lab. U12-LUC2 transgenic materials and the associated data I analyzed in Chapter IV were provided by undergraduate student Ruby Trejo and Dr. Hisashi Koiwa of the Koiwa lab.

All other work conducted for the dissertation was completed by the student independently. Graduate study was supported by Student Exchange Support Program (Scholarship and Tuition Fee for Long-Term Study Abroad) from Japan Student

Services Organization (Fall 2011 – Summer 2014), Teaching and Research Assistantships from the Department of Horticultural Sciences (Fall 2014 – Summer 2016), and Dissertation Fellowship from Texas A&M University Office of Graduate and Professional Studies (Fall 2016 – Spring 2017).

This dissertation work was supported by the National Science Foundation (MCB0950459), USDA-CSREES (2010-34402-20875) "Designing Food for Health", National Center for Research Resources (1 S10 RR022378-01), and the Office of Vice-President for Research, Texas A&M University.

NOMENCLATURE

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
Pol	DNA-dependent RNA polymerase
RPB	RNA polymerase B (II) subunits
CTD	Carboxyl-terminal domain
Ser2/5/7-PO_4	Phosphorylated serine at position 2, 5 or 7 of a pol II-CTD repeat
αSer2/5/7P	Antibodies against phosphoserine Ser2-, Ser5-, or Ser7-PO_4
TF	Transcription factor
CPL	RNA polymerase II CTD phosphatase-like
RNAi	RNA interference
DNSO	*De novo* shoot organogenesis
SDS	Sodium dodecyl sulfate
PAGE	Polyacrylamide gel electrophoresis
RT	Reverse transcription
PCR	polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
GUS	β-glucuronidase
Col	Colombia (Arabidopsis ecotype)
Ler	Landsberg erecta (Arabidopsis ecotype)
LUC	Luciferase

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Carboxyl-terminal domain of RNA polymerase II (Pol II-CTD) in eukaryotic gene expression

DNA-dependent RNA polymerases in eukaryotes

DNA-dependent RNA polymerases are essential protein complexes that convert genetic information stored in DNA into RNA through a process called transcription. The first RNA polymerizing activity was detected *in vitro* in rat liver nuclei homogenate almost six decades ago (Weiss and Gladstone 1959). A decade later, three types of essential DNA-dependent RNA polymerases conserved in eukaryotes (pol I, pol II and pol III) were first identified in sea urchin (Roeder and Rutter 1969). Additionally, plants have evolved two more pol II derivatives identified as pol IV and pol V (Onodera et al. 2005; Ream et al. 2009). Each RNA polymerase activity exerts different degree of sensitivity to a mushroom-derived toxin α-Amanitin which blocks translocation of DNA and RNA if properly bound; pol I, IV and V are insensitive; pol II is highly sensitive; pol III is slightly sensitive (Lindell et al. 1970; Cramer et al. 2001; Haag et al. 2012). These traits have allowed researchers to dissect polymerase specificity of transcription *in vitro* (Waibel and Filipowicz 1990; Yukawa et al. 1997; Haag et al. 2012). Among the five DNA-dependent RNA polymerases, pol II is the sole DNA-dependent RNA polymerase responsible for protein-coding messenger RNA (mRNA) transcription, as well as several classes of non-coding RNAs including small nuclear RNAs and microRNA precursors

(Myer and Young 1998; Hirose and Manley 2000; Hernandez 2001; Lee et al. 2004). Therefore, pol II activity and its regulation are pivotal for all biological processes in eukaryotic organisms. Other RNA polymerases are responsible for transcription of various types of non-coding RNAs; most ribosomal RNA (rRNA) is transcribed by pol I; transfer RNA (tRNA), 5S rRNA, U6 small nuclear RNA and majority of small nucleolar RNA (snoRNA) are transcribed by pol III (Bark et al. 1987; Paule and White 2000; McStay et al. 2002; Dieci et al. 2007). The plant-specific pol IV and pol V transcribe short precursor RNA involved in transcriptional gene silencing and non-coding RNA serving as a scaffold for heterochromatin formation machinery, respectively (Wierzbicki et al. 2008; Blevins et al. 2015).

Structure and composition of C-terminal domain of pol II

The gene expression by pol II is dynamically regulated by post-translational modifications, particularly phosphorylation, on the carboxyl-terminal domain of the largest subunit of pol II, RPB1 (Corden 1990; Phatnani and Greenleaf 2006). I refer to the domain as "pol II-CTD" throughout this dissertation. Pol II-CTD consists of tandemly repeated heptapeptide motifs and is conserved in animals and plants (Tyr_1-Ser_2—Pro_3-Thr_4-Ser_5-Pro_6-Ser_7; Y-S-P-T-S-P-S) (Allison et al. 1988; Martin and Medina 1991). The number of repetitions varies depending on organisms: 52 repeats in human, 26 repeats in the budding yeast *Saccharomyces cerevisiae*, and 32-39 repeats in the model plant *Arabidopsis thaliana* (Dietrich et al. 1990; Koiwa et al. 2002; Pikaard et al. 2008). Among the seven amino acid residues in each repeat, Tyr1, Ser2, Thr4, Ser5 and Ser7 are subjected to phosphorylation and dephosphorylation (Phatnani and

Greenleaf 2006; Hsin et al. 2011; Mayer et al. 2012; Descostes et al. 2014). The remaining Pro3 and Pro6 are not phosphorylated, but can be *cis*- or *trans*- through a phosphorylation-specific peptidyl-prolyl isomerase (PPIase) Pin1; the different isomerization state can impact different pol II-CTD phosphatase's recruitment and activities (Xu et al. 2003; Zhang et al. 2012b). The numerous combinations of the phosphorylation status of all residues in the heptapeptide tandem repeats are sometimes referred to as "CTD-code" (Buratowski 2003; Egloff and Murphy 2008).

Pol II-CTD couples transcription and co-transcriptional processing

Pol II-CTD kinases and pol II-CTD phosphatases catalyze the phosphorylation and dephosphorylation of the residues, respectively, throughout each transcription cycle (Dahmus 1996; Archambault et al. 1997; Archambault et al. 1998; Hengartner et al. 1998; Cho et al. 1999). A specific phosphorylation pattern on the pol II-CTD at a specific transcriptional stage would recruit a particular set of co-transcriptional processing machinery such as elongation factors, capping enzymes, spliceosomes and polyadenylation complexes to the transcribing pol II (discussed in later sections). As such, the pol II-CTD phosphoregulation couples the transcriptional activity of pol II and the processing of the nascent transcript RNAs. In accordance with this view, pol II-CTD is located proximal to where nascent RNAs exit, although the entire structure of pol II-CTD itself is not resolved presumably due to its extensive structural flexibility (Cramer et al. 2001; Bushnell and Kornberg 2003; Spahr et al. 2009; Suh et al. 2013). Interestingly, the CTD fused to RPB4 or RPB6 which are positioned proximal to RPB1 in the pol II complex is still functional, while the CTD fused to RPB9 located distal to

RPB1 is not. These observations showcase the importance of the CTD position in the complex (Suh et al. 2013).

1.2 Phosphoregulation of pol II-CTD during transcription of protein-coding and non-coding genes

Pol II-CTD phosphoregulation during protein-coding RNA transcription

Truncation of pol II-CTD does not affect the polymerase activity and accuracy of transcription initiation *in vitro*, while it is detrimental to pol II function *in vivo* (Zehring et al. 1988; Buratowski and Sharp 1990), highlighting the importance of pol II-CTD mediated regulation of transcription in living cells. At the initial stage of transcription, the phosphorylation level of the pol II-CTD is very low (hypo-phosphorylated form; pol II_A) when the pol II binds to a promoter region of a protein-coding gene (Lu et al. 1991). Then, Ser5 positions in the pol II-CTD become phosphorylated by a pol II-CTD kinase associated with the general transcription factor TFIIH (cyclin-dependent kinase 7; CDK7 in mammals). The kinase activity is a part of the pre-initiation complex, consisting of general transcription factors (TFIIA, TFIIB, TFIID, TFIIF and TFIIH) that unwind the double-stranded DNA by a helicase activity and guide the pol II complex to a transcription start site (Holstege et al. 1996). The Ser5-PO_4 marks on the pol II-CTD recruit a capping enzyme complex for the protective modification on the 5'-end of a nascent RNA transcript (Cho et al. 1997; Ghosh et al. 2011) and a histone H3 lysine 4 (H3K4) methyltransferase for the activation of transcription (Ding et al. 2011).

After pol II starts transcriptional elongation with the Ser5-PO_4 marks on the pol II-CTD (hyper-phosphorylated form; pol II_O), the Ser5-PO_4 is typically dephosphorylated by a Ser5/7-P-specific pol II-CTD phosphatase SSU72 (Ganem et al. 2003; Krishnamurthy et al. 2004; Reyes-Reyes and Hampsey 2007; Zhang et al. 2012a). At the same time, the serine residues at position 2 become gradually phosphorylated (Ser2-PO_4) by the action of a positive transcription elongation factor b (p-TEFb/CDK9) (Bonnet et al. 1999; Cho et al. 2001). The Ser2-PO_4 signals peak at the 3'-end of the gene (Tietjen et al. 2010; Hajheidari et al. 2013). Therefore, Ser2-PO_4 marks on pol II-CTD represent elongating/terminating pol II complex, whereas pol II enriched with Ser5-PO_4 is considered as at its initiating stage. The Ser2-PO_4 signals are essential for recruiting spliceosome and Cleavage and polyadenylation specificity factor (CPSF) subunits for mRNA splicing and polyadenylation, respectively (Licatalosi et al. 2002; Gu et al. 2013).

Following the transcriptional termination, the pol II-CTD has to be extensively dephosphorylated for the next round of transcription, because any residual phosphorylation on pol II-CTD may prevent pol II from associating with a promoter region of a gene (Lu et al. 1991). A Ser2-PO_4 specific pol II-CTD phosphatase TFIIF-interacting phosphatase 1 (FCP1) plays an essential role in dephosphorylation of Ser2-PO_4 and the recycling process (Archambault et al. 1998; Cho et al. 1999; Kobor et al. 1999; Cho et al. 2001; Fuda et al. 2012).

Non-coding RNA transcription by pol II; small nuclear RNA as a model

In addition to the protein-coding genes, pol II is responsible for transcription of several classes of biologically important non-coding RNAs, including microRNA precursors and small nuclear RNAs (snRNA) (Myer and Young 1998; Hirose and Manley 2000; Hernandez 2001; Lee et al. 2004). The snRNA is a class of non-coding RNAs that are essential for splicing of the pre-mRNAs in the nucleus (Lerner et al. 1980; Ohshima et al. 1981; Sontheimer and Steitz 1993). Typically, a eukaryotic organism has five classes of major snRNAs that are abundant (U1, U2, U4, U5 and U6 snRNA) and a few minor snRNAs that are less abundant (U11, U12, U4atac and U6atac snRNAs). Except for U6 snRNAs, which are transcribed by pol III, all these snRNAs are transcribed by pol II (Hernandez 2001).

The RNA polymerase specificity is dictated by DNA sequence elements embedded in the snRNA promoter region, and the requirements are slightly different among species (Hernandez 2001). In humans, for instance, if a TATA box is present near the proximal sequence element (PSE), pol III transcribes the downstream snRNA gene; if a TATA box is absent in the core promoter region, then pol II transcribes the snRNA (Lobo and Hernandez 1989). In the fruit fly *Drosophila melanogaster*, the polymerase specificity is determined by differences in one of two PSE elements, not by the presence of TATA-box (Jensen et al. 1998). In higher plants, it has been determined that two distinct sequence elements, the upstream sequence element (USE) and the TATA box, constitute the core snRNA promoter. The spacing between the two elements in the promoter region determines which RNA polymerase will transcribe the snRNA

gene in plants (Waibel and Filipowicz 1990). If the spacing is 32-34 nucleotides in length, then the pol II is responsible for the transcription of the snRNA; the shorter spacing such as 22-24 nucleotides dictates the transcription by pol III. The approximately ten nucleotides difference of the USE-TATA box spacing requirement between pol II and pol III in plants apparently corresponds to one DNA helical turn (Waibel and Filipowicz 1990). These core snRNA promoter elements are recognized and bound by a snRNA activator protein complex (SNAPc) as well as general transcription factors TFIIA, TFIIB, TFIIF and TFIIE (Kuhlman et al. 1999; Ohtani and Sugiyama 2005).

Similarly to protein-coding genes that typically have a polyadenylation signal for the 3'-end processing, the pol II-dependent snRNA genes also contain a snRNA-specific 3'-end signal sequence called the 3'-box (Hernandez 1985; Yuo et al. 1985). Although consensus sequences in the 3'-box are more resistant to mutations, altering several core sequences in the 3'-box results in defective termination and read-through transcription of the snRNA in a transient expression system using tobacco (*Nicotiana plumbaginifolia*) protoplasts (Connelly and Filipowicz 1993). In animals, transcription initiation driven by the snRNA promoter elements and 3'-end processing dictated by the 3'-box signal are coupled, whereas the 3'-end processing of snRNA in plants does not require snRNA-promoter elements (Hernandez and Weiner 1986; Connelly and Filipowicz 1993).

Pol II-CTD and its phosphoregulation play important roles in the pol II-dependent snRNA transcription. For example, truncation of pol II-CTD or inhibition of CTD kinase activities by chemical inhibitors such as DRB (5,6-dichloro-1-β-D-

ribofuranosyl-1H-benzimidazole) is shown to reduce the proper 3'-end processing of U2 snRNA in human cells (Medlin et al. 2003; Uguen and Murphy 2003). Also, knockdown of a Ser5/7-specific pol II-CTD phosphatase SSU72 results in defective 3'-end formation of U2 and U4 snRNAs in chicken cells (Wani et al. 2014). Phosphorylation at Ser2 and Ser7 positions are known to be especially important for the snRNA transcription because the dual mark recruits a snRNA 3'-end processing complex known as integrator complex and other pol II-CTD phosphatases to the snRNA-transcribing pol II in humans (Egloff 2012). A plant counterpart of the integrator complex in *Arabidopsis thaliana* containing CPSF73 has recently been identified (Liu et al. 2016).

The abovementioned sequencing elements important for snRNA transcription and maturation (USE, TATA box, the spacing rule and the 3'-box), seem to be widely conserved in the plant kingdom, as they can be seen both in dicots (*Arabidopsis thaliana*, pea - *Pisum sativum*, potato - *Solanum tuberosum*, tomato - *Lycopersicon esculentum*, soybean – *Glycine max*, common bean – *Phaseolus vulgaris*) and monocots (maize – *Zea mays*) (Solymosy and Pollak 1993). Thus, it is possible that higher plants share a common regulatory mechanism for snRNA transcription regulation through pol II-CTD phosphoregulation. The findings in *Arabidopsis thaliana* described in this dissertation will be useful for further characterizations of similar mechanisms in other plant species.

1.3 Pol II-dependent, functional long non-coding RNAs

In addition to small non-coding RNAs such as snRNA, tRNA and rRNA, advances in sequencing technologies have revealed unprecedented numbers of long non-

coding RNAs (lncRNA) in transcriptomes of many organisms; 15,767 in human (GENCODE 24 (Harrow et al. 2012), 10,481 in mouse (GENECODE M12) (Mudge and Harrow 2015), 4,761 in *Arabidopsis thaliana*, 8,594 in *Oryza sativa*, 4,403 in *Zea mays* and several thousands in other plants (Szczesniak et al. 2016). Although biological significances of many of lncRNAs remain elusive, several examples of pol II-dependent lncRNA with essential functions have been described.

Pol II-dependent lncRNA functioning as scaffold for protein complex recruitment

Many lncRNAs are known to function as a scaffold for protein complexes. For instance, X Inactive Specific Transcript (XIST) is a lncRNA longer than 17 kbp in mice and human, and functions in X-chromosome inactivation by physically coating one of the X-chromosomes in female mammal (Brown et al. 1992); It is transcribed by pol II and achieves the X-chromosome inactivation by recruiting Polycomb Repressive Complex 2 (PRC2) which deposits repressive histone modification H3K27me3 on the associated X-chromosome (Navarro et al. 2005; Zhao et al. 2008). Hox antisense intergenic RNA (HOTAIR) lncRNAs in human are transcribed from the HOXC locus and represses expression of other HOX loci on different chromosome in-trans; a 5'-region of HOTAIR transcript binds the PRC2 complex, whereas its 3'-end region binds LSD1 complex which demethylates H3K4me2 associated with active transcription; through such collective recruitment of the histone modification machinery on select target loci, the lncRNA HOTAIR negatively regulates expression of developmentally important genes such as HOXs encoding homeodomain transcription factors (Rinn et al. 2007; Tsai et al. 2010). Another lncRNA, COLD ASSISTED INTRONIC

NONCODING RNA (COLDAIR) in *Arabidopsis thaliana*, plays a role similar to HOTAIR in human but acts in cis; COLDAIR lncRNA is transcribed by pol II from the first intronic region of *FLOWERING LOCUS C* (*FLC*) locus in response to cold exposure; the COLDAIR lncRNA binds and recruits PRC2 to the FLC locus to repress the expression of FLC, resulting in promotion of flowering; this FLC repression by the cold-inducible COLDAIR lncRNA is the molecular basis of vernalization, a horticultural and agricultural practice to stimulate flowering (Heo and Sung 2011).

Pol II-dependent lncRNA functioning as structural components of nuclear bodies

Several pol II-dependent lncRNAs play an architectural role for subcellular compartments/bodies. Nuclear Enriched Abundant Transcript 1 (NEAT1) is a structural component and required for paraspeckle formation in nuclei (Clemson et al. 2009). Satellite-III (SatIII) lncRNA constitutes nuclear stress bodies (nSB) whose formation can be induced by heat shock, UV, heavy metal and oxidative stresses in human cells (Jolly et al. 2004; Valgardsdottir et al. 2005; Valgardsdottir et al. 2008); SatIII lncRNA also serves as a scaffold to recruit transcription factor CREBBP and splicing factor SRSF1 to nSB (Goenka et al. 2016). Another intriguing example of stress-inducible lncRNA derives from protein-coding transcription; Downstream of Genes (DoG) transcripts are osmotic stress responsive long transcripts derived from 3'-extended transcription from more than 10 % protein-coding genes in human; DoG transcripts associate with chromatin, reinforcing nuclear scaffold to protect it from osmotic stress (Vilborg et al. 2015). The widespread DoG production in response to osmotic stress showcases that the switching from protein-coding transcription to functional lncRNA

production is possible. This opens up possibilities for other types of transcriptional conversions in response to environmental cues.

1.4 Stress-triggered changes in the pol II-CTD phosphoregulation

In addition to the dynamic phosphoregulation on pol II-CTD during each transcription cycle, various types of biotic and abiotic stresses can influence the phosphorylation status of pol II-CTD. Severe heat stress causes hyper-phosphorylation of pol II-CTD in vertebrates and yeast, whereas *Drosophila melanogaster* (insect) cells show dephosphorylation trends (Patturajan et al. 1998; Lavoie et al. 1999). If the heat stress is mild, human cells also exhibit dephosphorylation due to inhibition of CTD kinases (Dubois et al. 1994; Venetianer et al. 1995). The Ser2-PO_4 specific CTD phosphatase FCP1 in *Drosophila melanogaster* is required for the efficient expression of heat shock genes; when DmFCP1 is knocked-down, pol II cannot be efficiently loaded to the heat shock genes due to defects in pol II-recycle step, resulting in accumulation of hyper-phosphorylated pol II in non-chromatin-bound fraction (Fuda et al. 2012). Pol II-CTD in fission yeast *Schizosaccharomyces pombe* responds to nitrogen starvation by elevating Ser2-PO_4 level via a CTD kinase CTDK-I activated by a stress-inducible MAP-kinase Sty1 (Sukegawa et al. 2011). Mouse cells exposed to osmotic or oxidative stress yield give rise to a pol II-CTD with extensive Ser5-PO_4, migrating faster than the hyper-phosphorylated form (II_O) in SDSPAG gel; this unique phosphorylated form, called II_m, is dependent on the ERK-type MAP kinases (Bonnet et al. 1999).

In plants, multiple pol II-CTD phosphatases have been isolated from forward genetics targeting abiotic and biotic stress response regulators, which will be discussed in section 1.6 (Koiwa et al. 2002; Xiong et al. 2002; Guan et al. 2014; Li et al. 2014). All of these examples clearly show that pol II-CTD phosphoregulation can respond to environmental factors.

1.5 Factors involved in pol II-CTD phosphoregulation in the model plant, *Arabidopsis thaliana*

Although still in its infancy, pol II-CTD phosphoregulation and the regulatory factors in plants have been identified and characterized in the model plant *Arabidopsis thaliana*. The Arabidopsis pol II-CTD on the largest subunit NRPB1 consists of 15 repeats of a heptapeptide showing perfect match to the consensus sequence (Y-S-P-T-S-P-S), 11 repeats with one substitution at the seventh residue (Y-S-P-T-S-P-X), 5 repeats with two substitutions and 2 repeats with three substitutions (between 1544aa - 1813aa of amino-acid sequence NP_195305.2).

Pol II-CTD kinase homologs in Arabidopsis thaliana

As for pol II-CTD kinases, Arabidopsis genome contains three homologs of TFIIH/CDK7 (CDKD;1, CDKD;2 and CDKD;3) (Shimotohno et al. 2003), and two homologs of p-TEFb/CDK9 (CDKC;1 and CDKC;2) (Cui et al. 2007). Additionally, a CDK8 homolog CDKE;1 and a plant-specific CDKF;1 have been identified (Umeda et al. 1998; Joubes et al. 2000; Shimotohno et al. 2004; Wang and Chen 2004; Hajheidari et al. 2012). Although *in vitro* CTD kinase activities have been shown for all of these

CTD kinases, their positional preference and involvement in pol II-CTD phosphoregulation *in vivo* are not fully established. The *cdkd123** triple mutant seedlings show more reduction in Ser5-PO_4 levels than other phosphoserines, although *in vitro* assays show the CDKDs can phosphorylate Ser2-PO_4, Ser5-PO_4 and Ser7-PO_4 (Hajheidari et al. 2012).

The plant-specific CDKF;1 can phosphorylate other CDKDs and exhibit strong preference toward Ser7-PO_4 residues *in vitro*, and *cdkf;1* mutant seedlings show reduced Ser7-PO_4 as well as Ser2-PO_4 and Ser5-PO_4 depending on developmental stages, reflecting CDKF;1's function in activating other CDKDs *in vivo* (Hajheidari et al. 2012). The capping of microRNA and trans-acting siRNA precursor transcripts is affected by *cdkd123**, indicating that the Ser5-PO_4 is important for capping of these transcripts in plants (Hajheidari et al. 2012). Interestingly, the defect in the Ser7-PO_4 level in *cdkf;1* mutant apparently does not result in a reduction in stable snRNA level detected by RT-PCR, although it is pivotal for snRNA 3'-end maturation in animals (Egloff et al. 2007; Egloff et al. 2010; Hajheidari et al. 2012). This may indicate different functions of Ser7-PO_4 in plants, yet the possibility of the snRNA 3'-end processing defects in the *cdkf;1* mutant has not been excluded.

Although phosphoserine specificity of Arabidopsis CDKCs has not been demonstrated, multiple lines of evidence indicate the functionality of Arabidopsis CDKCs. CDKC;1-CYCT;1 complex purified from Arabidopsis protoplast was able to facilitate the *in vitro* transcription activity of CDK9-depleted HeLa cell nuclear extract, suggesting that the Arabidopsis CKDC;1 can be a positive factor of transcription (Fulop

et al. 2005). In the same study, it was demonstrated that immunoaffinity purified CDKC;1 from *Medicago sativa* preferentially phosphorylate Ser5 residues of recombinant CTD substrate (Fulop et al. 2005). In line with an expected function of a Ser2-PO_4 CTD kinase homolog, Arabidopsis CDKC;2 co-localizes with a spliceosome component SRp34 in nuclei (Kitsios et al. 2008). Another spliceosomal SR protein interacting protein AtCyp59 was shown to be co-immunopurified with pol II; the ectopic expression of AtCyp59 in Arabidopsis callus results in a reduced pol II-CTD phosphorylation level in Ser5-PO_4 and Ser2-PO_4 (Gullerova et al. 2006). AtCyp59, as a peptidyl-prolyl cis–trans isomerase (PPIase) domain containing protein, might play a role similar to the Pin1 PPIase in animals which modulates the pol II-CTD phosphorylation status by isomerization of Proline residues in the heptapeptide repeats (Zhang et al. 2012b).

Other types of CDKs in plants, CDKA and CDKB, have not been shown to participate in pol II-CTD phosphoregulation. Rather, CDKAs and CDKBs play major roles in cell cycle and shoot apical meristem regulation (Francis 2007; Nowack et al. 2012). Actively dividing or proliferating cells, which are mainly found on apical meristems, go through a well-defined cell cycle comprising of four distinct phases; Gap1 (G1), Synthesis (S), Gap 2 (G2) and Mitosis (M) phases (Francis 2007). Cells in G1 phase are preparing for the subsequent S phase where DNA replication occurs. A peak in CDKA1 activity can be observed at the G1-S phase (Joubes et al. 2000). After completion of DNA replication in S phase, G2 cells grow and synthesize proteins necessary for the mitotic cell division and cytokinesis in the following M phase to

produce two daughter cells. Both CDKA1 and CDKB1;1 show the highest activity at the G2-M phase (Joubes et al. 2000). After M phase, the daughter cells can go into G1 phase to be ready for another cycle, or become a quiescent state called G0. In plants, the phytohormone cytokinin (CK) can stimulate G0-G1 re-entry through CK-responsive D-type cycling and CDKDs activating CDKAs (Riou-Khamlichi et al. 1999; Oakenfull et al. 2002; Francis 2007). Reduced CDKAs activity in shoot apical meristem (SAM) resulted from SAM-specific expression of dominant-negative CDKA;1 driven by the promoter of *SHOOT MERISTEMLESS* (*STM*) gene, an essential homeodomain transcription factor involved in SAM regulation and shoot organogenesis, causes premature differentiation of meristem cells and a morphological disorder (Gaamouche et al. 2010).

Involvement of transcriptional and co-transcriptional regulatory factors in de novo organogenesis

In *Arabidopsis thaliana*, forward genetic screenings searching for genes involved in *de novo* organogenesis of shoots and roots have identified multiple genes related to transcriptional and co-transcriptional regulation such as pre-rRNA processing (*ROOT INITIATION DEFECTIVE*, *RID2* and *RID3*), spliceosome activity (*RID1*) and a snRNA transcription activator complex subunit (*SHOOT REDIFFERENTIATION DEFECTIVE2*, *SRD2*) (Yasutani et al. 1994; Sugiyama 2003; Ohtani and Sugiyama 2005; Ohtani et al. 2013). Mutations in a TFIID subunit *TATA-BOX BINDING PROTEIN ASSOCIATED FACTOR 12* (*TAF12*) and a chromatin remodeling factor *PICKLE* have been isolated as *CYTOKININ HYPERSENSITIVE* mutants, resulting in the

cytokinin-hypersensitivity-mediated greening of calli (Kubo and Kakimoto 2000; Furuta et al. 2011; Kubo et al. 2011).

De novo organogenesis of roots and shoots can be artificially induced by incubating explants on auxin-rich media and cytokinin-rich media, respectively (SKOOG 1950; Skoog 1957). *De novo* root organogenesis can be induced solely by auxin-rich media, whereas *de novo* shoot organogenesis (DNSO) requires incubation on CK-rich shoot induction media (SIM) followed by pre-incubation on auxin-rich callus induction media (CIM) to gain competence for shoot organogenesis (Che et al. 2007). In the process, exogenous CK is perceived by *ARABIDOPSIS HISTIDINE KINASE* (*AHK*) family receptor kinases known as *AHK2*, *AHK3* and *AHK4/CRE1* (Nishimura et al. 2004). The signals from the CK receptor are transduced through the multistep His-Asp phosphorelay system in the cytoplasm and activate *ARABIDOPSIS RESPONSE REGULATOR* (*ARR*) family transcription factors in the nucleus (Hwang and Sheen 2001; Howell et al. 2003). ARRs can be divided into type-A and type-B genes based on the domain structure; Type-A ARRs such as *ARR5*, *ARR7* and *ARR15* are transcriptionally responsive to exogenous CK, whereas type-B ARRs are not (D'Agostino et al. 2000). Because *ARR5* expression level correlates with the endogenous CK level, it is used as an indicator of CK content in various tissues (Aloni et al. 2004; Aloni et al. 2005). Type-B *ARR*s also up-regulate type-A *ARR* expression, and type-A ARRs partly function as negative regulators of type-B ARRs and CK response (To et al. 2004). In the DNSO process, activation of some type-A ARRs such as *ARR15* on SIM

requires a CIM-pre-incubation period, while another type-A *ARR5* can be induced on SIM without CIM-pre-incubation (Che et al. 2007).

APETALA2/Ethylene Responsive Factor (AP2/ERF) family *ENHANCER OF SHOOT REGENERATION 1* (*ESR1*) and *ESR2* are genes that show the earliest response to SIM containing high CK. *ESR1* expression gradually increases until 5-day post transfer to SIM, while *ESR2* expression begins a few days later than *ESR1* (Matsuo et al. 2009). *ESR1* was originally isolated from screening for cDNAs whose overexpression confers enhanced shoot formation without CK (Banno et al. 2001). Both expressions activate an NAC (an acronym for No Apical Meristem/Arabidopsis thaliana activating factor/Cup-shaped cotyledon 2) transcription factor *CUP-SHAPED COTYLEDON 1* (*CUC1*) expression (Ikeda et al. 2006; Matsuo et al. 2009). The expression of CUC transcription factors is necessary for activating a homeodomain transcription factor *STM*. Constitutive overexpression of *CUC1* or *CUC2* results in a higher frequency of shoot formation on calli incubated at a lower concentration of CK, without bypassing the CK requirement (Daimon et al. 2003). The *CUC2* expression does require CIM pre-incubation, and has been suggested as an early marker for the acquisition of competence for root and shoots organogenesis (Che et al. 2007; Motte et al. 2011). Another essential homeobox transcription factor required for shoot apical meristem (SAM) regulation and shoot organogenesis, *WUSCHEL* (*WUS*), also requires CIM-pre-incubation to be activated during incubation on SIM (Che et al. 2007). The expression of these key SAM regulators *CUCs*, *STM* and *WUS* is a hallmark of competence and initiation of shoot meristem for DNSO on SIM.

Except for CDKDs, there are no pol II-CTD kinases or pol II-CTD phosphatases in plants that have been implicated in the *de novo* root/shoot organogenesis pathway. Because the FCP1 pol II-CTD phosphatase in human has been shown to be involved in mitotic exit by dephosphorylating Wee1 kinase (Visconti et al. 2012), it is possible that FCP1 homologs in plants might be involved in the cell cycle regulation.

1.6 Pol II-CTD phosphatase-like (CPL) proteins in *Arabidopsis thaliana*

Although pol II-CTD kinases and phosphatases have been extensively studied in vertebrates, the understanding of those in plants remains elusive. Through forward genetic screening approaches using an osmotic stress-responsive reporter gene system in *Arabidopsis thaliana*, two pol II-CTD phosphatase-like (CPL) family genes, CPL1 and CPL3, were identified as negative regulators of responses to abiotic and biotic stresses such as cold, drought, osmotic, salt and bacterial elicitors (Koiwa et al. 2002; Xiong et al. 2002; Jiang et al. 2013; Guan et al. 2014). Subsequent amino-acid similarity-based searches identified two additional homologs: CPL2 and CPL4 (Koiwa et al. 2004; Bang et al. 2006). These CPLs all contain Asp-based metal-dependent phosphatase-like FCP1-homology (FCPH) domain with a DXDX(T/V) signature catalytic motif. CPL3 and CPL4 contain a Breast Cancer C-terminus (BRCT) domain at their C-terminus, which is conserved in animal and fungal pol II-CTD phosphatases. On the other hand, CPL1 and CPL2 belong to the plant-specific class of pol II-CTD phosphatases due to a lack of the BRCT domain and the presence of double-stranded RNA-binding domains. Also, Arabidopsis genome contains 19 Small CTD-phosphatase-like CTD phosphatase (SSP)

family genes and one SSU72-like gene, some of which have been confirmed to dephosphorylate pol II-CTD substrates with distinct substrate specificities *in vitro* (Feng et al. 2010).

Biochemical studies have revealed that phosphatase activity of bacterial recombinant CPL1 and CPL2 proteins prefer Ser5-PO_4 over Ser2-PO_4 *in vitro*, and function redundantly in an essential manner because the double mutant is lethal while the individual single mutant lines do not show strong phenotype; the *cpl2* mutant show reduced fertility, early flowering and increased sensitivity to salt (Koiwa et al. 2002; Koiwa et al. 2004; Ueda et al. 2008). Recombinant CPL3 and CPL4 proteins interact with a TFIIF subunit RAP74 *in vitro*, indicating their involvement in transcriptional regulation (Bang et al. 2006). CPL3 shows preference toward Ser2-PO_4 *in vitro* and is also involved in dephosphorylation of pol II-CTD hyper-phosphorylated by CDKCs activated by MAPKs upon a bacterial elicitor flagellin-22 treatment (Li et al. 2014).

In addition to pol II-CTD, CPL1 has been implicated in phosphoregulation of other CPL1-interacting proteins. Recently, it has been shown that CPL1 facilitates Nonsense Mediated Decay of a certain set of transcripts by dephosphorylating eukaryotic translation initiation factor 4A3 (eIF4III); the phosphoregulation on eIF4III impacts the protein's localization and accumulation of the NMD target RNAs (Cui et al. 2016). CPL1 interacts with SERRATE (SE) and a double-stranded RNA binding protein HYPONASTIC LEAVES 1 (HYL1) involved in microRNA maturation; CPL1 dephosphorylates HYL1, facilitating its activity and loading into a microRNA processing machinery (Manavella et al. 2012). The CPL1's involvement in HYL1-

phosphoregulation is consistent with the reduced accumulation of miRNA observed in *cpl1* mutants (Jeong et al. 2013a). Also, CPL1 interacts with K-homology domain containing RNA-binding protein (RCF3, Regulator of C-Repeat Binding Factor Gene Expression 3) which is dephosphorylated by CPL1 (Chen et al. 2015) and facilitates CPL1/2-mediated dephosphorylation of HYL1 (Karlsson et al. 2015). The interactions between CPL1 and RCF3/HYL1/SE are mediated through the dsRNA-binding domain of CPL1, a non-canonical functional motif for a pol II-CTD phosphatase (Jeong et al. 2013a; Jeong et al. 2013b; Chen et al. 2015; Karlsson et al. 2015). The involvement of Ser5-PO_4 specific pol II-CTD phosphatases in microRNA biogenesis is consistent with the involvement of Ser5-PO_4 kinase TFIIH homologs CDKDs in the same pathway (Hajheidari et al. 2012). The acquisition of dsRNA-binding domains might have conferred CPL1 and CPL2 unique functions as pol II-CTD phosphatases participating in the phosphoregulation of microRNA biogenesis machinery, which by itself exists in animals (Paroo et al. 2009).

Among the four CPLs, CPL4 is the smallest and closest homolog to FCP1, an essential pol II-CTD phosphatase in the budding yeast *Saccharomyces cerevisiae* (Archambault et al. 1997). Consistently, knockdown of *CPL4* by RNA interference (RNAi) results in severe growth and developmental defects (Bang et al. 2006), suggesting that *CPL4* is an essential gene in *Arabidopsis thaliana*. Yet, a definitive evidence of CPL4 as a genuine pol II-CTD phosphatase had been missing. The remaining CPL family genes seem to play roles in more specific pathways rather than in general biological processes because their single loss-of-function mutant lines did not

show strong phenotypes in normal growth condition while each of them results in altered stress responsive gene expressions (Koiwa et al. 2002; Bang et al. 2006; Ueda et al. 2008; Li et al. 2014). Given the importance of CPL4 for the survival of *Arabidopsis thaliana*, my dissertation focuses on the detailed characterization of CPL4 as an essential pol II-CTD phosphatase and its functions in pol II-CTD phosphoregulation, protein-coding and non-coding RNA transcriptions in *Arabidopsis thaliana*.

1.7 Scope of this dissertation

The major focus of my dissertation is to dissect functions of the essential pol II-CTD phosphatase-like protein CPL4 in the phosphoregulation of pol II-CTD for both global and stress-triggered transcriptional regulation in *Arabidopsis thaliana*. The choice of the model plant *Arabidopsis thaliana* as a tool for this study was based on the availability of various genetic materials (relevant mutants and transgenic plants), genome sequence information resources and means of detailed biochemical characterizations. Although a previous study found that the knockdown of *CPL4* (AT5G58003) in Arabidopsis by RNAi causes severe growth and developmental defects (Bang et al. 2006), the molecular basis of the phenotype and the function of the protein was uncharacterized.

In Chapter II, I establish CPL4 as an essential and major pol II-CTD phosphatase in Arabidopsis by showing the physical interaction between CPL4 and pol II and CPL4's CTD-phosphatase activity toward phosphorylated CTD substrates *in vitro*. Furthermore, the overexpression of CPL4 results in a reduction of the global pol II-CTD

phosphorylation level, while the knockdown of *CPL4* enhances accumulation of hyper-phosphorylated pol II *in vivo*, indicating that CPL4 is responsible for adjusting the global pol II-CTD phosphorylation level. Then, I examine the impact of the pol II-CTD phosphorylation changes on the global protein-coding gene expression profile, using *CPL4* knockdown lines that over-accumulate hyper-phosphorylated pol II due to the reduced activity of CPL4. Co-expression network analysis on the differentially expressed genes identified by a microarray analysis reveals that a suite of the xenobiotic stress-responsive genes is particularly up-regulated in the CPL4 knockdown lines, even without the stress. Accordingly, CPL4 knockdown lines are found to be less sensitive to a general herbicide, chlorsulfuron. After the chemical treatment, the response level of the xenobiotic stress responsive genes are comparable to that of wild-type plants, indicating that the function of CPL4 is likely to suppress the basal expression of the stress responsive genes.

Chapter III presents a novel concept of gene expression regulation that is a conversion of non-coding RNAs into protein-coding transcripts by altering 3'-end processing/termination of upstream non-coding RNA transcription through pol II-CTD phosphorylation status changes, which is potentially inducible by environmental factors such as salt stress. This chapter mostly focuses on one class of major non-coding RNAs, snRNA. Transcriptome analysis by RNA-seq identifies many pol II-dependent snRNA loci showing abnormal 3'-extended transcripts in *CPL4* knockdown plants. Such 3'-extension of snRNA gives rise to chimeric snRNA-mRNA transcripts if a protein-coding locus is present in the proximal downstream region. Furthermore, *CPL4* knockdown

specifically facilitates expression of a snRNA-mRNA chimeric isoform without changing a non-chimeric isoform expression from a locus having both annotations (AT3G55850). Additionally, a promoter-based search identifies a non-snRNA locus (AT1G20320) driven by a pol II-dependent snRNA promoter that produces short-unstable transcripts only detectable in a 3'-5' exosome mutant, hen2. This locus accumulates longer transcripts spanning the entire coding sequence in the *CPL4* knockdown plants. By analyzing the expression pattern of multiple protein-coding genes located near snRNAs, salt-stress is found to upregulate most of the snRNA-proximal downstream genes in wild-type plants. Consistently, salt-stressed wild-type plants show 3'-extension of snRNAs, specific up-regulation of chimeric snRNA-mRNA isoform in AT3G55850 and accumulation of AT1G20320 long transcripts. These observations in wild-type plants are reliant on a snRNA transcription factor SRD2, suggesting that the 3'-extension and chimeric snRNA-mRNA transcript production are dependent on the snRNA-transcription. Salt stress is shown to alter global pol II-CTD phosphorylation status *in vivo*. Chromatin-immunoprecipitation (ChIP) analysis assesses the pol II-CTD phosphorylation status on the snRNA-extension regions in the CPL4 knockdown plants and wild-type plants exposed to salt stress.

Finally, Chapter IV discusses the functions of CPL4 in *de novo* shoot organogenesis, which is known to require snRNA and SRD2 (Ohtani and Sugiyama 2005; Ohtani et al. 2015). $CPL4_{RNAi}$ lines show compromised lateral root formation and lateral root primordia development, similarly to *srd2-1* mutant. $CPL4_{RNAi}$ root explants exhibit enhanced shoot regeneration. Transcriptome analysis reveals that cytokinin (CK)

responsive gene expression in $CPL4_{RNAi}$ is similar to that in WT, while multiple major stem cell regulator genes are up-regulated regardless of CK in $CPL4_{RNAi}$ root explants during shoot regeneration. Additionally, an operon-like cluster of tricyclic triterpene compound biosynthesis genes is found to be hyper-induced $CPL4_{RNAi}$ roots tissue incubated on CK-containing shoot induction media. Genetic analysis indicates that the enhanced *de novo* shoot organogenesis phenotype is reliant on SRD2.

Pol II-CTD phosphoregulation is one of the fundamental gene expression regulatory mechanisms conserved in all eukaryotes. However, the significance and mechanistic understanding of this gene expression regulatory mechanism in plants are still in their infancy. Here, my dissertation describes a detailed characterization of the essential pol II-CTD phosphatase in *Arabidopsis thaliana* and the outcome of misregulated pol II-CTD phosphorylation in a global and stress-response expression of both protein-coding and non-coding genes. It also describes a novel type of gene expression regulation by adjusting 3'-end processing/termination of non-coding RNAs through pol II-CTD phosphoregulation. Because the snRNA and its transcription mechanism are widely conserved in the plant kingdom, this can be a potential target of genetic engineering for a new type of stress-inducible gene expression strategies in economically important crops.

CHAPTER II

ARABIDOPSIS CPL4 IS AN ESSENTIAL C-TERMINAL DOMAIN PHOSPHATASE THAT SUPPRESSES XENOBIOTIC STRESS RESPONSES*

2.1 Introduction

DNA-dependent RNA polymerase II (pol II) catalyzes the production of all mRNAs and miRNAs and many non-coding RNAs (Hirose and Manley 2000; Hsin et al. 2014). In eukaryotes, pol II is highly conserved and consists of 12 subunits. The C-terminal domain of the largest catalytic subunit of pol II (RPB1 CTD) is a major regulatory domain and its phosphorylation status determines the activity of the pol II complex and various pol II-associated proteins (Hsin and Manley 2012; Hajheidari et al. 2013). The CTD contains tandemly repeated heptads with a consensus sequence of $Y_1S_2P_3T_4S_5P_6S_7$ (Nawrath et al. 1990). The number of repeats and the degree of deviation from the consensus sequence correlates roughly with evolutionary complexity; thus, mammals have 52 repeats (Corden et al. 1985), Drosophila has 42 repeats (Greenleaf et al. 1987), fission yeast *Schizosaccharomyces pombe* has 29 repeats (Azuma et al. 1991), and the microsporidian *Encephalitozoon cuniculi* has 15 repeats (Hausmann et al. 2004).

---

*Used with permission from Fukudome A, Aksoy E, Wu X, Kumar K, Jeong IS, May K, Russell WK, Koiwa H (2014) Arabidopsis CPL4 is an essential C-terminal domain phosphatase that suppresses xenobiotic stress responses. The Plant Journal 80(1):27-39. doi: 10.1111/tpj.12612. Copyright 2014 by the authors.

In humans, the first 26 CTD repeats contain 18 consensus sequences and 8 variants, and this configuration is able to support cell viability. In contrast, the second half of the CTD contains only 3 consensus and 24 variant sequences and is unable to support cell viability (Hsin et al. 2014). The CTD of Arabidopsis pol II contains 16 consensus repeats and 18 variant sequences (Dietrich et al. 1990; Hajheidari et al. 2013). The CTD undergoes waves of phosphorylation and dephosphorylation during the transcription cycle, under the control of various position-specific CTD kinases and phosphatases (Dahmus 1996). Studies in animals and fungi indicate that all residues in the CTD heptads except for prolines, i.e., Tyr1, Ser2, Thr4, Ser5, and Ser7, are targets of regulatory phosphorylations (Hsin and Manley 2012). Phosphorylation marks placed on different residues in each repeat generate CTD codes that are uniquely associated with the pol II transcription status (Eick and Geyer 2013; Hajheidari et al. 2013). Phosphorylated CTD is a target of CTD-associating proteins that recognize various CTD codes and differentially regulate transcription and cotranscriptional mRNA processing (Hsin and Manley 2012).

Among the various CTD phosphorylation sites, those at Ser2 and Ser5 have been analyzed most extensively (Hirose and Manley 2000; Buratowski 2009; Hsin and Manley 2012), and recent studies showed a role for Ser7 phosphorylation in mRNA and snRNA transcription by pol II (Egloff et al. 2007; Akhtar et al. 2009; Egloff 2012; Egloff et al. 2012). The pol II in the transcription initiation complex is phosphorylated at CTD Ser5 and Ser7 by the TFIIH-associated kinase CDK7 (cyclin-dependent kinase 7) (Akhtar et al. 2009; Egloff et al. 2010). CTD-Ser5-PO_4 is essential for mRNA capping

enzyme recruitment and activation (Ho and Shuman 1999). The Ser5-PO_4 marks are partially removed during transcription elongation, and CDK9 increases the Ser2-PO_4 level (Peterlin and Price 2006). When transcription is terminated, the CTD is dephosphorylated and pol II is recycled. Several classes of protein phosphatases are responsible for CTD dephosphorylation (Hsin and Manley 2012). The major CTD phosphatase, FCP1 (TFIIF-interacting CTD phosphatase 1), and its homologs belong to a family of Asp-based metal-dependent phosphatases (AMPs) (Thaller et al. 1998). The signature motif of the AMPs is DXDX(T/V), with the first aspartic acid residue functioning as a phosphoryl acceptor and the second stabilizing the leaving phosphate group (Kobor et al. 1999; Wang et al. 2002). FCP1 contains a phosphatase catalytic domain and a BRCT (*Br*east *C*ancer 1 *C-t*erminal) domain, which is highly conserved among eukaryotes. FCP1 dephosphorylates both Ser2-PO_4 and Ser5-PO_4, but preferentially dephosphorylates Ser2-PO_4 (Archambault et al. 1997; Archambault et al. 1998; Kobor et al. 1999). Mutations in the DXDX(T/V) motif abolish the catalytic activity of FCP1 (Kobor et al. 1999). Yeast FCP1 is essential for survival (Kobor et al. 1999), whereas genetic mutations in human *FCP1* are associated with congenital cataracts facial dysmorphism neuropathy syndrome (Varon et al. 2003). In addition to FCP1 family proteins, SCP1 (Small CTD Phosphatase 1) family proteins with only an FCP1-like catalytic domain dephosphorylate Ser2- and Ser5-PO_4 and negatively regulate neural (Yeo et al. 2003; Yeo et al. 2005; Wrighton et al. 2006; Visvanathan et al. 2007) and bone (Knockaert et al. 2006) development.

The domain structure of Arabidopsis CTD-phosphatase-like AMP can be classified into three groups, CPL1-like, CPL3-like, and SSPs (Bang et al. 2006; Koiwa 2006). The CPL1-like proteins contain C-terminal double-stranded-RNA binding domains, which are unique to plants (Koiwa et al. 2002; Koiwa et al. 2004). The CPL3-like-proteins resemble FCP1 and contain a C-terminal BRCT domain (Koiwa et al. 2002; Bang et al. 2006). SSP (*S*CP1-like *s*mall *p*hosphatase) proteins contain only a phosphatase catalytic domain and thus resemble animal SCP1 (Feng et al. 2010). CPL1-like proteins are able to dephosphorylate CTD-Ser5-PO_4 specifically (Koiwa et al. 2004), whereas SSPs contain both Ser5-PO_4-specific and Ser2/Ser5-PO_4-specific members (Feng et al. 2010). By contrast, only limited information is available for CPL3-like proteins, i.e., CPL3 and CPL4. Genetic studies have shown that CPL3 is involved in ABA signaling and that CPL4 is required for normal growth and development (Bang et al. 2006). CPL3 could hydrolyze CDP-star, a general phosphatase substrate (Koiwa et al. 2002), but no specific CTD phosphatase activity has been shown for any CPL3-like phosphatase.

Here I present evidence that Arabidopsis CPL4 functions as a pol II CTD phosphatase. CPL4 interacted with pol II *in vivo*, and dephosphorylated pol II CTD-PO_4 in *in vitro* assays using peptide and intact CTD substrates. Plants with altered CPL4 levels exhibited changes in pol II CTD phosphorylation levels, consistent with the CTD phosphatase activity observed *in vitro*. Expression of a dominant negative form of CPL4 strongly impaired general gene expression and cell viability of *Nicotiana benthamiana* host cells, and Arabidopsis plants silenced for *CPL4* exhibited upregulation of a specific

set of xenobiotic stress-regulated genes. These results not only establish that CPL4 is an essential plant CTD phosphatase, equivalent to animal and fungal FCP1, but also suggest a novel function for CPL4 in the xenobiotic stress response.

2.2 Materials and methods

Plant transformation and induction of cell cultures

Transformed cell cultures were prepared according to the previously described procedure (Jeong et al. 2013b). Briefly, transgenes for overexpression cassettes, namely CPL4-FSG, CPL4-9myc, or mCherry-FSG (mC-FSG), were introduced into Arabidopsis Col-0 *rdr6-11* plants using *Agrobacterium tumefaciens* and the floral spray method (Chung et al. 2000), and T1 transformants were selected on media containing 1/4x MS salts, 0.7% agar, 25 μg/mg hygromycin B, and 100 μg/ml Clavamox. Hygromycin-resistant seedlings were screened by immunoblot using anti-FLAG-HRP conjugate. Positive plants were then cut into small pieces and cultured on callus induction/growth medium (1x MS salts, 2% sucrose, 0.2 g/L KH_2PO_4, 0.1 g/L myo-inositol, 1x B5 vitamins, 2 mg/L 2,4-D, 0.05 mg/L kinetin and 0.9% agar). The resulting CPL4-FSG, CPL4-9myc, and mC-FSG cell lines were maintained by transferring cells to fresh medium every week (Doelling and Pikaard 1993).

In vitro phosphatase assays

For *in vitro* activity assays, CPL4 was purified from Arabidopsis CPL4-FSG cells following the TAP procedure (Methods S1), but replacing the TAP extraction buffer with protein extraction buffer (50 mM Tris-HCl, pH 7.5; 100 mM NaCl; 4 mM

$MgCl_2$; 5 mM DTT) supplemented with 1 mM PMSF and 1x proteinase inhibitor cocktail for plant cell extracts (Sigma). The $MgCl_2$ was omitted from the protein extraction buffer in the final elution step of the TAP procedure so that $MgCl_2$ carry-over in the *in vitro* reaction was minimized. Phosphatase assays were performed essentially as described previously (Feng et al. 2010). For the phosphatase assay with synthetic phosphopeptide substrates, reaction mixture (25 µl) containing 50 mM Tris-acetate, pH 6.5; 10 mM $MgCl_2$; 25 µM CTD-phosphopeptides; and 1 µg of CPL4 was incubated for 2 h at 37^{o}C. The reactions were quenched by adding 3x vol. of malachite green reagent (BIOMOL Research Laboratories, PA). Phosphate release was determined by measuring ABS_{650} and the values were interpolated to generate a phosphate standard curve. For assays with recombinant CTD protein substrates, phosphatase reaction mixtures (5 µl) containing 50 mM Tris-acetate, pH 6.5; 10 mM $MgCl_2$; 1 µg of phosphorylated GST-AtCTD; and 200 ng of CPL4 were incubated for 0, 30, 60, or 120 min and the Ser5, Ser2, and Ser7 phosphorylation status was analyzed by immunoblot analysis.

Transient reporter gene assay using Nicotiana benthamiana leaves

pBSVirEL2LUC and pBSVirUOiLUC, containing an enhanced 35S promoter or a ubiquitin (UBQ1) promoter, respectively, driving the luciferase reporter gene were prepared by recombining pEnEL2LUC (GenBank accession number KF545095) or pEnUOiLUCThsp (GenBank accession number KF545094) with pBSVirHygGW using LR clonase, respectively. The resulting plasmids were introduced into GV3101(pMP90RK). pBin35S-p19 was provided by Dr. Baulcomb and introduced into

GV3101. For a negative control, pBSVirGFP containing a synthetic $(Aocs)_3A_{mas}P_{mas}$-GFP cassette derived from pEnSOTG was used (Koiwa *et al.,* 2004).

Agrobacterium-mediated transient expression was performed as described previously (Jeong et al. 2013a). Induced Agrobacterium suspensions were mixed at a ratio of 4:4:1.5 (LUC:effecter:P19). Approximately 300 µl of each mixture was infiltrated into the leaves of 4- to 7-week-old *N. benthamiana* plants. Luminescence images were taken 48 to 60 h after infiltration by infiltrating leaves with luciferin solution (10 mM MES/KOH, pH 5.6; 10 mM $MgCl_2$; and 100 µM luciferin) and exposing them to an EMCCD camera. After image acquisition, 7-mm leaf discs were excised from each infiltrated spot and proteins were extracted and analyzed with anti-c-myc and anti-GFP antibodies.

Microarray analysis

Total RNAs were extracted from 10-day-old Col-0 *gl-1* wild type and $CPL4_{RNAi}$ line 7 using an RNeasy Plant Mini Kit (Qiagen). Microarray analysis using Affymetrix ATH1 GeneChips was performed as previously described (Aksoy et al. 2013). The microarray data were deposited into the NASC Database under accession number 658.

Bioinformatic analyses

After validating the microarray results of select genes by qPCR (Table 2.3), functional Gene Ontology (GO) annotation was performed using the GO Functional Categorization tool in TAIR (Appendix A2.1). Co-expressed gene clusters among the C4UTs were identified using the Network Drawer tool in the ATTEDII database (http://atted.jp/), using 227 C4UTs as input (Clusters 1- 3; Figure 2.8 and Table 2.5).

Using the Over-/Under-Representation Analysis tool in Genetrail (http://genetrail.bioinf.uni-sb.de/) (Backes et al. 2007), functional GOs significantly over-represented in each of the clusters were retrieved (*p*-values<0.01, calculated by the Hypergeometric Distribution method followed by False Discovery Rate adjustment; Appendix A2.2). Perturbations that up-regulate the 23 Cluster 1 genes were identified and retrieved using the Hierarchical Clustering tool in Genevestigator (https://www.genevestigator.com/gv/index.jsp).

RNA extraction and RT-qPCR

Total RNAs were extracted using TRIzol reagent and treated with DNase I to eliminate genomic DNA contamination (Aksoy et al. 2013). Total RNA samples (2 µg) were reverse-transcribed using random hexamers and Superscript™ III Reverse Transcriptase (Life Technologies) or GoScript™ Reverse Transcriptase (Promega). The reverse transcription products were analyzed using a LightCycler 480 (Roche Diagnostics) and EvaGreen qPCR Master Mix (Bullseye). Two housekeeping genes, *UBQ10* (AT4G05320) and *GAPDH* (AT1G13440), were used as internal controls for normalization, by taking the geometric mean of their Ct values (Czechowski et al. 2005). In the chlorosulfuron response experiment, *GAPDH* (AT1G13440) alone was used as the internal control, since *UBQ10* (AT4G05320) showed responsiveness to the treatment (up to 3.79-fold up-regulation in the Col-0 *gl-1* wild type).

Chlorosulfuron treatment

For the germination assays, plants were grown on medium containing 1/2 x MS salts, 1% sucrose, and 0.7% agar with or without 40 nM chlorosulfuron. For root growth

measurement, seeds were germinated on medium containing 1/2 x MS salts, 1% sucrose, and 0.7% agar. Four-day-old vertically-grown seedlings were transferred to the same media with or without 10 nM chlorosulfuron. Primary root growth was measured on the third day post transfer. For RT q-PCR analyses, 10-day-old seedlings grown on medium containing 1/2 x MS salts, 1% sucrose, and 0.7% agar were sprayed with 0.03% Silwet containing 0, 40, or 600 nM chlorosulfuron. Total RNA was extracted 24 h after treatment.

Preparation of CPL4 expression vectors

To express tagged CPL4 in Arabidopsis, the *CPL4* coding sequence was placed between the SalI-ScaI sites of pEnEOimCherryFSGThsp (GenBank accession number KF537341) upstream of the 3xFLAG tag and SG-tag (Van Leene et al. 2008), resulting in pEnEOCPL4F3SGThsp, and between the SalI-EcoRI sites of pEnEOi9mycThsp (GenBank accession number KF537343), resulting in pEnEOiCPL4-9mycThsp. The resulting plasmids and negative control (pEnEOmCherryF3SG; GenBank accession number KF537342) were recombined with pMDC99 (Curtis and Grossniklaus 2003) or pBSVirHygGW (GenBank accession number KF537344) using LR Clonase (Life Technologies). The D128A substitution was introduced into the pEnEOi-CPL4-9mycThsp plasmid using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) and specific primers (Appendix P), to produce the $CPL4^{D128A}$ variant. For the *N. benthamiana* transient assay, cDNA fragments encoding *CPL4* and its variants were cloned between the SalI-EcoRI sites of pEnEOi9mycThsp and recombined into pBSVirHygGW. To express GFP-CPL4 fusion proteins, CPL4 and its variant cDNAs

containing 5'- and 3'- *attB* sequences were amplified by PCR and cloned into pDonrZeo (Life Technologies) by the BP reaction according to the manufacturer's protocol. The resulting pDon-CPL4 clones were recombined with pMDC43. The resulting pMDC plasmids were introduced into *Agrobacterium tumefaciens* GV3101, and the pBSVir plasmids into GV3101 (pMP90RK).

Tandem affinity purification and mass spectrometry

Tandem affinity purification (TAP) was performed essentially as described previously, with slight modification (Van Leene et al. 2008). Stably transformed CPL4-FSG cells were harvested 7-10 days after subculturing, and were homogenized in 2 ml/g TAP extraction buffer (50 mM Tris-HCl, pH 9.0; 100 mM NaCl; 12.5% glycerol; 2.5 mM EDTA; 10 mM β-mercaptoethanol; and 20 mM sodium fluoride) supplemented with 1 mM PMSF and 1x Proteinase Inhibitor Cocktail for Plant Cell and Tissue Extracts (Sigma). The crude extract was filtered through two layers of Miracloth (Millipore) and centrifuged at 13,000 rpm for 5 min to remove debris. One-fortieth volume of IgG Sepharose6 Fast Flow (GE) was added to the supernatant, and the samples were incubated for 4-6 h at 4^{o}C with rotation. Precipitates were washed three times with TST buffer (50 mM Tris-HCl, pH 7.5; 150 mM NaCl; and 0.05% Tween 20), and then one time with TEV cleavage buffer (25 mM Tris-HCl, pH 8.0; 150 mM NaCl; 0.5% NP-40; 0.5 mM EDTA; and 1 mM DTT). TAP-tagged proteins were eluted from the beads by cleavage with tobacco etch virus (TEV) protease (Blommel and Fox 2007) in TEV cleavage buffer overnight at 4^{o}C. One-twentieth volume bed of Streptavidin Sepharose (GE) was added to the eluate and the samples were incubated for 1-3 h at 4°C.

The beads were then washed three times with TAP extraction buffer, and proteins were eluted with TAP extraction buffer supplemented with 10 mM desthiobiotin. The eluate was separated by SDS-PAGE and visualized by CBB staining. For protein identification by mass spectrometry (MS), the protein bands of interest were manually excised (approximately 2 mm strips) and processed and analyzed as described previously (Jeong et al. 2013b).

Co-immunoprecipitation

Stably transformed CPL4-9myc cells were harvested 7-10 days after subculturing, and were homogenized in 2 ml/g CelLytic™ P (Sigma) buffer supplemented with 10% glycerol, 1 mM PMSF, and 1x Proteinase Inhibitor Cocktail for Plant Cell and Tissue Extracts (Sigma). Debris and the insoluble fraction were removed firstly by centrifugation, and then by filtering through a glass bead column. The concentration of proteins in the supernatant was adjusted to 2 µg/µl with CelLytic buffer. The supernatant (1 ml) was pre-cleared by incubating with 10 µl of protein G Sepharose (GE healthcare) for 45 min at 4°C with rotation. Next, the pre-cleared supernatant was incubated with 20 µg of anti-RPB1 or control (anti-COBRA) antibody and 10 µl of protein G Sepharose for 2 h at 4°C with rotation. The beads were then washed three times with CelLytic buffer, each for 15 min with rotation. Proteins were eluted by boiling the beads in 25 µl SDS-PAGE loading buffer, and then 10 µl of supernatant was analyzed by immunoblotting.

Protein extraction for pol II detection

Cells or seedlings were homogenized in nuclear extraction buffer (see TAP procedure described above). The crude extract was filtered through a glass bead column to remove debris. The filtered extract represents the total protein fraction (designated as total extract). The protein concentration was adjusted to 3 µg/µl. For tobacco RPB1 detection, the harvested leaves were homogenized in nuclear extraction buffer. The crude extract was filtered through two layers of Miracloth (Millipore) to remove the debris. The protein concentration was adjusted to 3 µg/µl, and 0.4% Triton X-100 was then added to the extract. The insoluble fractions were precipitated by centrifugation at 4,300 g for 10 min, resuspended in SDS-PAGE loading buffer, and then analyzed by immunoblotting.

Detection of pol II phosphorylation status by immunoblotting

For detection of RPB1 and determination of the phosphorylation status, proteins were separated by 5% SDS-PAGE and transferred to PVDF membrane by semi-dry blotting with transfer buffer containing 25 mM Tris-base, 192 mM glycine, and 10% SDS. The membrane was washed twice with TBST buffer (20 mM Tris-HCl, pH 7.4; 137 mM NaCl; and 0.1% Tween 20), and blocked with TBST buffer containing 6% skim milk for 1 h. The antibodies used were described below. Chemiluminescence was developed with SuperSignal West Femto Chemiluminescent Substrate (Pierce), and images were acquired using an EMCCD camera and processed by WinView software. For Figure 2.1b, 2.1e, 2.2b, 2.4c, 2.4d, and 2.6, the same membrane was reprobed with different position-specific phosphor-serine antibodies. After imaging, the membrane was

incubated twice with stripping buffer (200 mM glycine, pH 2.2; 1% Tween 20; and 1% SDS), twice with TBS buffer, and then twice with TBST buffer. Each washing step was performed for 10 min at room temperature. The stripped membrane was then blocked again with 6% skim milk and subjected to reprobing with another antibody (see below).

Antibodies

For detection of RPB1 and determination of its phosphorylation status *in vivo,* anti-RPB1 (Kang et al. 2009), 8WG16 (αCTD, Abcam), H5 (anti-Ser2-PO_4, Covagen), 3E10 (anti-Ser2-PO_4, Millipore), H14 (anti-Ser5-PO_4, Covagen), and 4E12 (anti-Ser7-PO_4, Active Motif) were used. Anti-c-myc antibodies (GenScript) were used for Co-IP and detection of transiently expressed proteins in tobacco leaves. Anti-FLAG antibody (Sigma) was used to detect mCherry and CPL4 in the TAP-related experiments. Anti-GST-HRP antibody (GE Healthcare) was used to detect GST-AtCTD substrates.

Recombinant protein preparation using E. coli

cDNA fragments encoding wild-type and the D128A variant of CPL4 with an N-terminal TEV protease cleavage site and C-terminal Strep-tag II were cloned into pCold-TF-GW by the Gateway LR reaction. The resulting pCold-TF-TEV-CPL4 and pCold-TF-TEV-CPL4D128A vectors were transformed into the *E. coli* DH10B strain. Recombinant proteins were expressed by inducing bacterial cultures with 0.5 mM IPTG and cold shock (15^{o}C) for 24 h. After harvesting, cells were resuspended in protein extraction buffer (50 mM Tris-HCl, pH 7.5; 100 mM NaCl; 4 mM $MgCl_2$; and 5 mM DTT) supplemented with 1 mM PMSF and lysed by sonication. Triton X-100 (1%) was added to the lysate, and the soluble fraction was recovered by

centrifugation. Fusion proteins were purified by incubating the supernatant with StrepTactin Sepharose (GE Healthcare). The N-terminal TF-tag on the fusion proteins was removed by on-column cleavage with TEV protease, and the recombinant CPL4 proteins were eluted from the beads with elution buffer (50 mM Tris-HCl, pH 7.5; 100 mM NaCl; and 5 mM DTT) supplemented with 10 mM desthiobiotin, as described above.

Genetic analysis

A heterozygous T-DNA insertion line (*cpl4-2*, T3 generation) was identified in the GABI_926A04 seed pool. After self-pollination, T4 plants were characterized using PCR to determine the segregation ratio. For reciprocal cross analysis, a T3 *cpl4-2*/ + plant and a Col-0 plant were used as parents, and F1 seedlings were characterized to determine the segregation ratio. Genomic DNA was extracted as described (Tsugama et al. 2011). The genotype of CPL4 locus was determined by PCR using primer pairs CPL4for/84mm_CPL4_R or T-DNA_8474/84mm_CPL4_R to detect WT and *cpl4-2* alleles, respectively.

2.3 Results

CPL4 associates with RNA polymerase II in vivo

CPL4 is thought to be an ortholog of animal and fungal FCP1, based on their high degrees of amino acid sequence similarity and the essential role of CPL4 in plant growth and development. However, direct evidence that functionally links CPL4 to pol II has yet to be obtained. To address this gap, I sought to identify and characterize proteins that associate with CPL4 *in vivo*. To purify the CPL4 complex, a CPL4 transgene C-terminally fused to an FSG-tag (3xFLAG-streptavidin-binding peptide-2xprotein G domains) was introduced into Arabidopsis Col-0 *rdr6-11* (Figure 2.1a). The *rdr6-11* host was used because it is impaired in post-transcriptional gene silencing, and thus supports a high level of CPL4-FSG transgene expression (Butaye et al. 2004). T1 plants overproducing CPL4-FSG were identified and used to establish cell cultures. Similarly, an mCherry-FSG (mC-FSG) transgenic cell line was prepared as a negative control. Immunoblot analyses of CPL4-FSG and mC-FSG cell extracts detected approximately 80-kDa and 60-kDa immunopositive bands corresponding to tagged CPL4 and mCherry, respectively (Figure 2.1b). To confirm that CPL4-FSG has CTD phosphatase activity *in vivo*, the CTD phosphorylation status of pol II in CPL4-FSG cells was determined (and mC-FSG) (Figure 2.1b). In the control (mC-FSG) cells, immunoblot of total pol II using anti-RPB1 detected two bands of ca. 220 kDa and 260 kDa, corresponding to the hypophosphorylated pol II_A and hyperphosphorylated pol II_O forms, respectively. The pol II_O form was strongly recognized by antibodies specific to the CTD Ser2-PO_4 (H5), Ser5-PO_4 (H14), and Ser7-PO_4 (4E12) marks. CPL4-FSG cells

had substantially less pol II_O than did pol II_A, suggesting that *CPL4-FSG* expression decreases pol II CTD phosphorylation *in vivo* (Figure 2.1b). Phosphorylation at Ser2 showed the greatest decrease in CPL4-FSG cells, whereas phosphorylation at Ser5 and Ser7 decreased only moderately. These data suggest that *CPL4-FSG* expression gives rise to a functional CTD phosphatase with a substrate preference for CTD Ser2-PO_4.

Using tandem affinity purification of CPL4-FSG, I recovered an immunopositive band of approximately 62 kDa, corresponding to the CPL4 fusion protein after cleavage of Protein G domains. SDS-PAGE analysis of the final eluate revealed several peptides that were co-purified with CPL4 (Figure 2.1c). TOF-MS/MS analysis of these peptides identified the three largest subunits of pol II, i.e., RPB1, RPB2, and RPB3, in addition to CPL4 and its fragments, indicating that CPL4 binds to the pol II complex (Table 2.1). This was confirmed by a bi-directional pull-down assay. Since the protein G domain of the FSG tag interferes with specific pol II immunoprecipitation, the pol II complex was immunoprecipitated using a cell line expressing CPL4-9xmyc (CPL4-myc). The immunocomplex precipitated with anti-RPB1 but not with control IgG contained both RPB1 and CPL4 (Figure 1d). Similarly, the CPL4-FSG complex precipitated with IgG sepharose contained pol II (RPB1) (Figure 2.1e). Together, these results establish that CPL4 and pol II form a complex, suggesting that pol II is a target of CPL4. Interestingly, this analysis did not identify additional pol II-associated proteins or expected CPL4-interacting proteins, such as TFIIF. Since a substantial proportion of CPL4 remained in the pellet fraction, other CPL4 complexes that are more difficult to extract from the nuclei may exist.

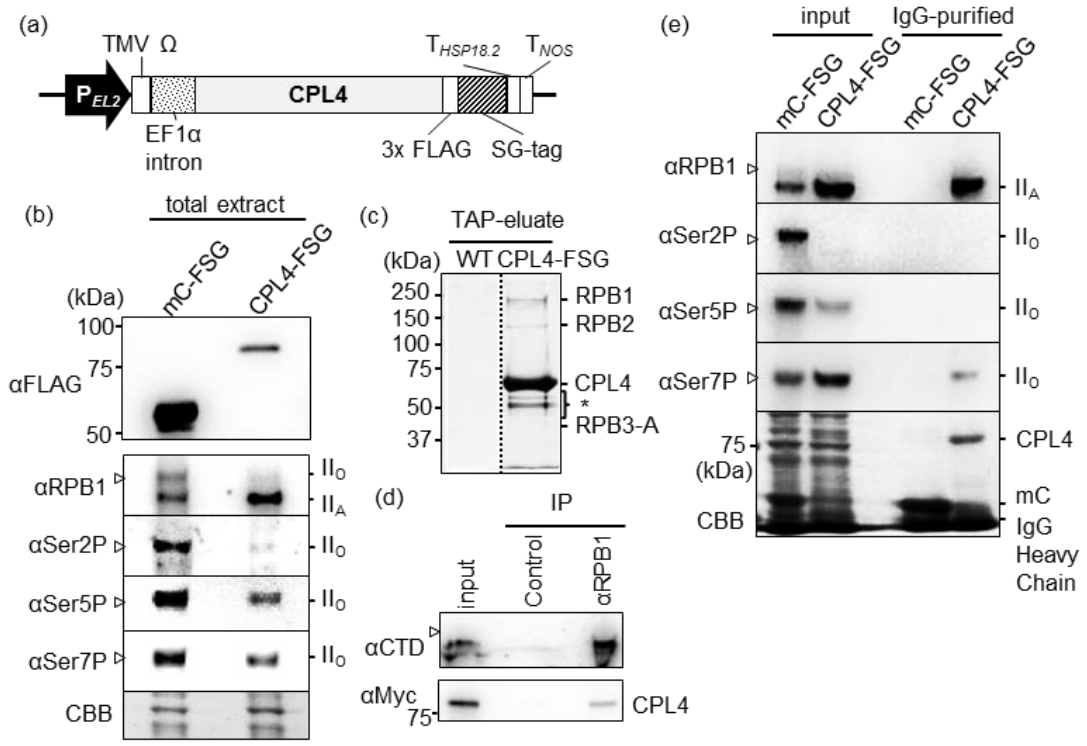


Figure 2.1 Arabidopsis CPL4 interacts with pol II.

(a) Schematic representation of the CPL4 overexpression cassette. Black line represents noncoding region. (b) Overexpression of CPL4 in Arabidopsis and an analysis of the pol II phosphorylation status. Total proteins were extracted from Arabidopsis calli overexpressing mCherry-FSG (mC-FSG) or CPL4-FSG (CPL4-FSG) cells. Total extracts containing 13 μg protein were separated on a 5% SDS-PAGE gel and analyzed by immunoblot with anti-FLAG (αFLAG), anti-RPB1 (αRPB1), H5 (anti-Ser2-PO_4, αSer2P), H14 (anti-Ser5-PO_4, αSer5P), and 4E12 (anti-Ser7-PO_4, αSer7P) antibodies. The bottom panel represents CBB-stained low molecular weight proteins on the same gel, and serves as a loading control. (c) SDS-PAGE analysis of the CPL4 complex. The identity of each band determined by TOF-MS/MS analysis is shown on the right. An asterisk indicates CPL4 fragments. (d) CPL4 co-immunoprecipitates with pol II. The immunocomplex was precipitated from CPL4-myc callus extract using αRPB1 or control IgG. RPB1 and CPL4-myc in the immunoprecipitates were detected by immunoblotting using anti-CTD (8WG16, αCTD) and anti-c-Myc (αMyc) antibody, respectively. Forty percent of total immunoprecipitate was loaded. (e) Phosphorylation status of RPB1 bound to CPL4. Pol II (RPB1) in the CPL4-FSG complex precipitated by IgG sepharose was analyzed by immunoblot. Thirty micrograms of input proteins and 10% of total immunoprecipitates were loaded. The bottom panel (loading control) shows CBB staining of low molecular weight proteins on the same gel. Open arrowheads indicate the position of the 250-kDa marker.

Table 2.1 CPL4-interacting peptides identified by TOF-MS/MS

Protein	Protein Score	Observed Peptides	Observed Mass	Mass error (ppm)	Mascot Score (Ion)	C.I.%
RPB1	545	HLQDGDFVLFNR	1460.71	-12	68	99.0
		VVQFGILSPDEIR	1472.79	-15	73	99.7
		VVISFDGSYVNYR	1518.74	-13	62	96.7
		EAFEWVIGEIESR	1564.74	-14	85	100.0
RPB2	405	LFDQSDAYR	1114.49	-22	69	99.1
		TEIPIIIVFR	1200.71	-22	62	95.8
CPL4	1190	LEETGVSFR	1037.49	-39	66	98.5
		AKPEDHPLWK	1220.59	-44	69	99.4
		HKDNLIVIER	1236.66	-40	82	100.0
		DLKPEEEYLK	1263.59	-48	64	97.6
		GWIDAANYLWMK	1467.65	-42	82	100.0
		QPEENFGLEQLKK	1559.73	-46	112	100.0
		QAHALFFENVDEGISNR	1946.87	-32	152	100.0
		LYLVLDLDHTLLNTTILR	2126.15	-33	108	100.0
		SLSELKSDESEPDGALATVLK	2189.03	-41	69	99.4
		QAHALFFENVDEGISNRDVR	2317.07	-26	86	100.0
RPB3	161	DDSFIFTVESTGAVK	1615.80	6	85	100.0

Proteins identified by TOF-MS/MSAll TOF-MS/MS data were searched against the UniProt protein sequence database using the GPS Explorer (Applied Biosystems) software (Jeong et al. 2013b).

CPL4 dephosphorylates both Ser2 and Ser5 of pol II CTD

Immunoblot analysis showed that the RPB1 that was copurified with CPL4 was predominantly a hypophosphorylated form (pol II_A) (Figure 2.1e, top), lacking CTD Ser2- and Ser5-PO_4 marks, and with reduced levels of Ser7-PO_4 marks. In the CPL4-FSG cell extract, the Ser2- and Ser5-PO_4 marks were greatly decreased, but Ser7-PO_4 was increased. This increase in Ser7-PO_4 could be due to a compensatory effect of host cells to maintain CTD phosphorylation levels in response to low Ser2- and Ser5-PO_4 levels. Because CPL4 does not dephosphorylate Ser7-PO_4 as effectively as it does Ser2-PO_4 or Ser5-PO_4, the Ser7-PO_4 level could increase in CPL4-FSG cells.

Affinity-purified CPL4 was used to confirm the Ser-specific CTD phosphatase activity of CPL4. Two substrate systems were used, i.e., the synthetic $(YSPTSPS)_4$ phosphopeptides $(CTD\text{-}Ser5\text{-}PO_4)_4$ and $(CTD\text{-}Ser2\text{-}PO_4)_4$, together with intact pol II CTD after AtMPK3 phosphorylation. As shown in Figure 2.2a, CPL4 dephosphorylates both $(CTD\text{-}Ser5\text{-}PO_4)_4$ and $(CTD\text{-}Ser2\text{-}PO_4)_4$. Similar to animal and fungal CTD phosphatases (Kobor et al. 1999; Hausmann and Shuman 2002), the phosphatase activity of CPL4 is Mg^{2+}-dependent. In contrast to the phosphatase activity toward peptide substrates, the activity toward full-length CTD substrate (GST-AtCTD-PO_4) showed a strong preference for Ser2-PO_4 over Ser5-PO_4 and Ser7-PO_4 (Figure 2.2b).

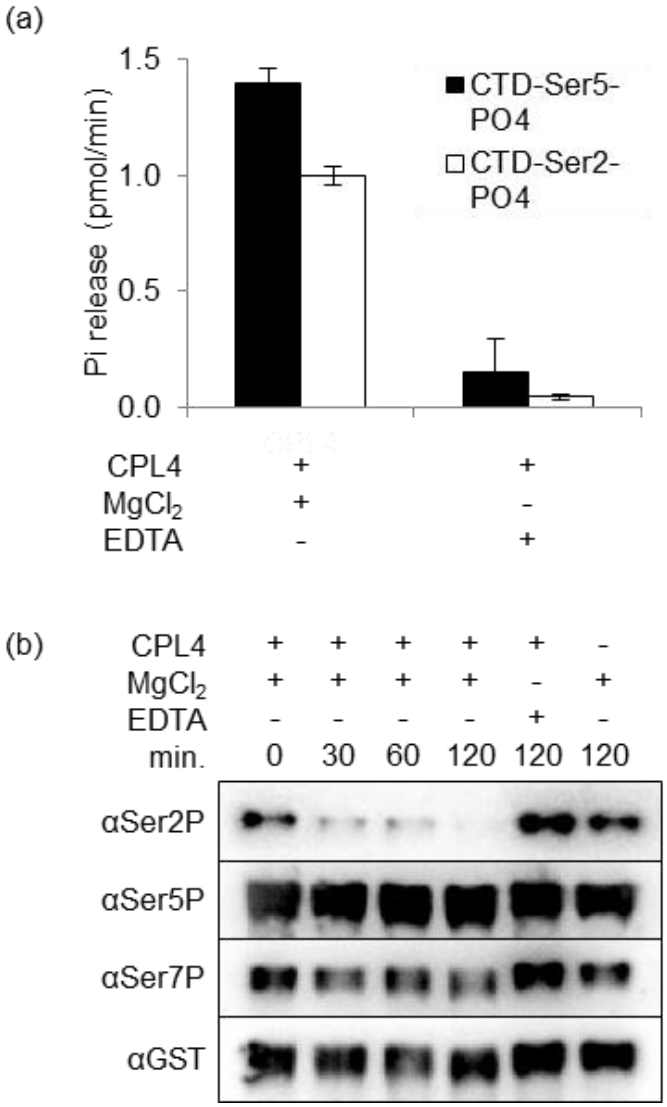


Figure 2.2 In vitro CTD phosphatase activity of TAP-CPL4.
(a) CTD phosphatase assay using synthetic phosphopeptide substrate (CTD-Ser5-PO_4 or CTD-Ser2-PO_4). I determined the phosphatase activity by quantifying the released phosphates using malachite green reagent and measuring ABS_{650}. The bars represent standard errors of triplicates. (b) The CTD phosphatase assay used GST-AtCTD-PO_4 substrates. Samples containing 100 ng of substrate were loaded in each lane. The phosphatase activity toward Ser2-, Ser5-, and Ser7-PO_4 was analyzed by immunoblotting using H5 (αSer2P), H14 (αSer5P), and 4E12 (αSer7P), respectively. The reactions were performed in the presence or absence of $MgCl_2$. αGST was used as the loading control.

To confirm that CPL4, but not low-abundance CTD phosphatases co-purified with CPL4, was responsible for the observed CTD phosphatase activity, recombinant CPL4 was expressed in *E. coli* (Figure 2.3a). Cleavage of the N-terminal TF-tag by TEV and affinity purification by immobilized Strep-Tactin yielded highly purified recombinant CPL4 (Figure 2.3b). As a negative control, $CPL4^{D128A}$ with a point mutation in the catalytic motif was prepared. Recombinant CPL4 but not $CPL4^{D128A}$ was able to dephosphorylate a general phosphatase substrate (*p*-nitrophenyl phosphate) in a Mg-dependent manner (Figure 2.3c). Furthermore, CPL4 could dephosphorylate both CTD-Ser2-PO_4 and CTD-Ser5-PO_4 with similar efficiency (Figure 2.3d). These results establish that CPL4 is a functional CTD phosphatase. The substrate preference of CPL4 toward different Ser-PO_4 marks was affected by the CTD substrate preparation method. This indicates that the context of the CTD codes can impact the CTD phosphatase activity and the specificity of CPL4. Factors that affect CTD codes include deviation of the CTD sequence from the consensus heptad and CTD kinases that generate CTD phosphorylation patterns (Lin et al. 2002).

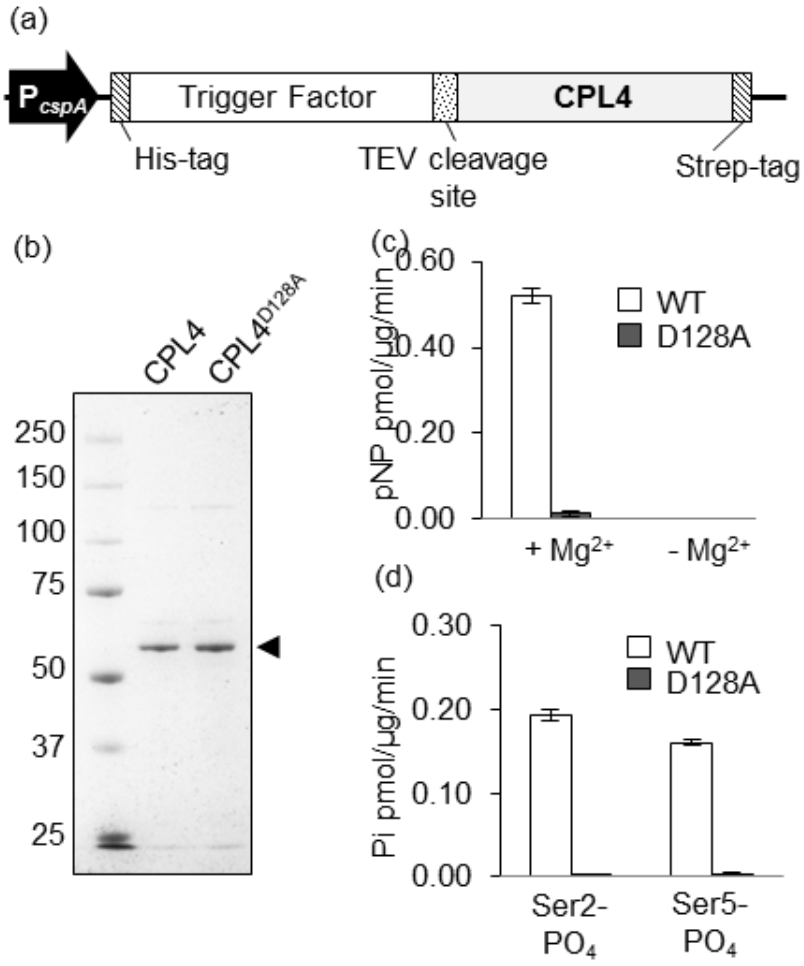


Figure 2.3 Phosphatase activity of bacterial recombinant CPL4 proteins.
(a) Schematic representation of the recombinant CPL4 expression cassette. (b) SDS-PAGE analysis of affinity-purified recombinant CPL4 and CPL4^{D128A} proteins. (c) Phosphatase activity of the purified CPL4 proteins toward *p*-nitrophenyl phosphate (pNPP). Thirty micrograms of CPL4s was incubated with 40 mM pNPP in the presence or absence of 10 mM $MgCl_2$. The reaction was stopped by adding 1 M sodium carbonate, and the released *p*-nitrophenol (pNP) was quantified by measuring ABS_{405}. (d) A CTD-phosphatase assay was performed using synthetic phosphopeptides. Twenty micrograms of CPL4 proteins was used in a 25-µl reaction as described in Figure 2.2a. Error bars represent the standard deviation derived from triplicate experiments.

Transient co-expression of CPL4 negatively affects gene expression in vivo

Dephosphorylation of pol II CTD can occur at multiple phases in the transcription cycle, and has the potential to both promote and inhibit the pol II transcriptional activity. To test if pol II hypophosphorylation induced by CPL4 overexpression affects gene expression in general, Agrobacterium-mediated transient assays were performed using a firefly luciferase (*LUC*) reporter gene and an *N. benthamiana* host. A set of expression constructs encoding wild-type CPL4, $CPL4_{1\text{-}338}$ lacking the essential BRCT domain (negative control), $CPL4^{D128A}$, and $CPL4^{D128A}{}_{1\text{-}338}$ were prepared (Figure 2.4a). These fragments contain a predicted N-terminal nuclear localization signal (amino acid 44-56), and localized to nuclei when expressed *in planta* as GFP-fusion proteins (Figure 2.4e), similar to intact CPL4 (Bang et al. 2006).

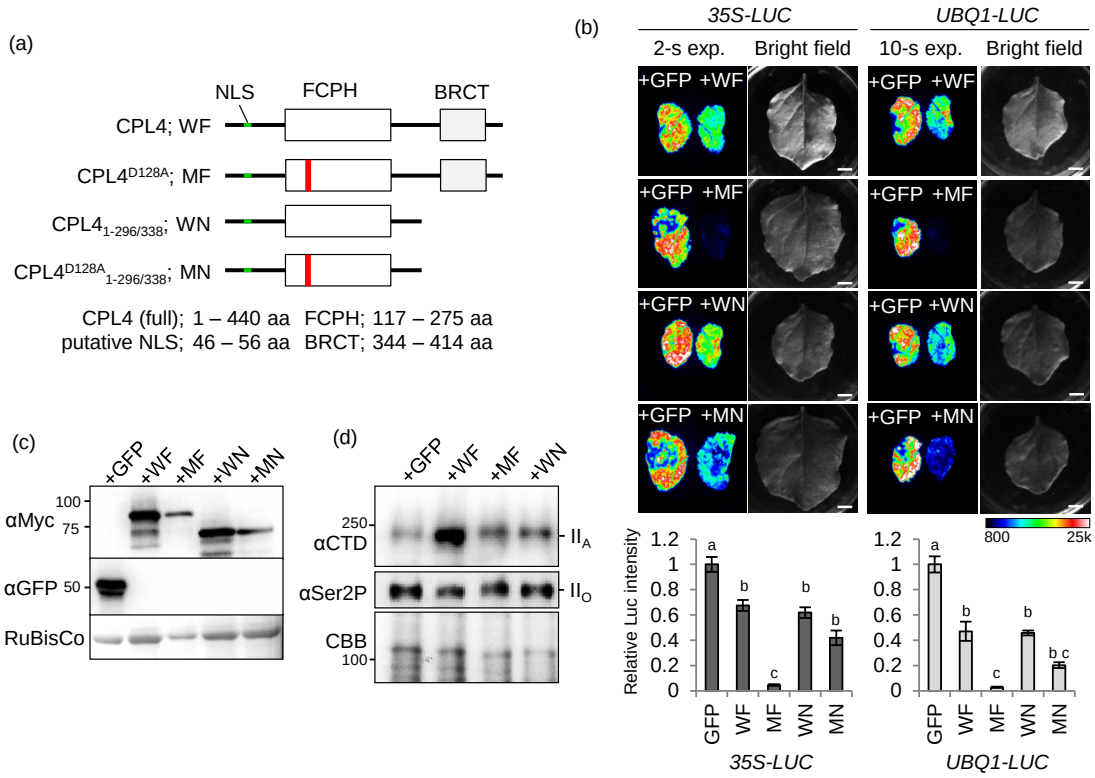


Figure 2.4 Transient reporter gene expression assay using *N. benthamiana* leaves.

(a) Schematic diagram of CPL4 and its variant proteins. FCPH, FCP1 homology domain; BRCT, Breast Cancer 1 C-terminal domain; and NLS, nuclear localization signal. (b) Co-expression of CPL4 and its variants with CaMV *35S* promoter-*LUC* and *UBQ1* promoter-*LUC*. (Top) Luminescence and bright-field images of leaves infiltrated with Agrobacterium suspension. LUC luminescence was documented 48 h post infiltration. White bars represent 10 mm. exp., exposure. (bottom) Quantitative analysis of luminescence levels. The presented luminescence levels are mean values of three biological replicates. Bars indicate the SEM of biological replicates ($n \geq 4$). Different letters show significant differences between constructs ($p<0.05$, one-way ANOVA followed by Tukey's HSD post hoc test). (c) αMyc and αGFP immunoblots showing the successful co-expression of effecter proteins. Total proteins were extracted from 7-mm leaf discs by boiling in 100 µl SDS-extraction buffer (Tsugama *et al.* 2011), and 10 µl of total extracts was loaded onto a SDS-PAGE gel. (d) Pol II phosphorylation status of infiltrated leaves. Pellet fractions of leaf extracts (30 µg protein) were subjected to immunoblot analysis. For detection of unphosphorylated CTD and Ser2-PO_4, 8WG16 and H5 antibodies were used. (e) The full-length and N-terminal 296-amino-acid fragments of CPL4 were targeted to the nuclei. GFP (top) and bright-field (bottom) images of *N. benthamiana* leaves infiltrated with Agrobacteria carrying different GFP-CPL4 constructs. Images were taken 2 days post-infiltration. Bars indicate 25 µm. (f) $CPL4^{D128A}$ expression is detrimental to *N. benthamiana.* Agrobacteria containing CPL4 variant constructs were infiltrated into whole *N. benthamiana* leaves and photographed 8 days post-infiltration.

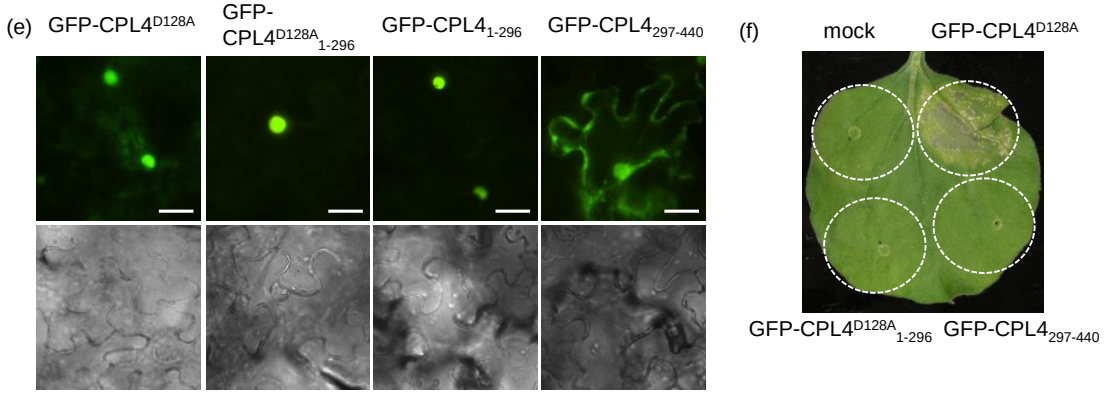


Figure 2.4 Continued.

The *LUC* reporter genes were co-expressed with myc-tagged effecter genes (Figure 2.4b). Both CaMV 35S promoter-*LUC* and Arabidopsis *UBQ1* promoter-*LUC* produced similar results. The production of effecter proteins was confirmed by immunoblot (Figure 2.4c). Coexpression of *35S-LUC* with a neutral effecter (*GFP*) in *N. benthamiana* produced strong luminescence (Figure 2.4c). The *CPL4* effecter decreased *LUC* expression; however, this was likely unrelated to the specific CPL4 function, because a similar decrease in LUC activity was observed with $CPL4_{1-338}$ lacking an essential BRCT domain. Unexpectedly, the $CPL4^{D128A}$ effecter strongly inhibited LUC expression, and at a later stage, caused necrosis of the leaf tissue (Figure 2.4f). The nature of this toxicity is not clear. The onset of necrosis (7-10 days) was slower than observed in typical hypersensitive responses caused by avirulence gene expression (within 1 day) (Espinosa et al. 2003); therefore, it is unlikely that $CPL4^{D128A}$ triggers typical programmed cell death. The inhibition and cellular toxicity was attenuated in $CPL4^{D128A}_{1-338}$ lacking the BRCT domain, indicating the effects of

CPL4^{D128A} were caused by alteration of an innate function of CPL4. Since FCP1 and phosphorylated pol II form a stoichiometric complex via the BRCT domain of FCP1 (Archambault et al. 1998; Yu et al. 2003), by analogy, catalytically inactive CPL4^{D128A} may compete with the native CPL4 homolog for pol II CTD-PO_4 and cause a dominant-negative effect. However, transient CPL4^{D128A} expression did not change the overall level of the Ser2-PO_4 mark. Therefore, the direct mechanism by which CPL4 ^{D128A} represses gene expression is not clear. CPL4 ^{D128A} may inhibit pol II dephosphorylation in only a small sub-population of pol II, or only at a specific step in transcription. In addition, the possibility that gene repression is the secondary effect of CPL4^{D128A} toxicity should not be excluded.

The overexpression of wild-type CPL4 strongly decreased the pol II phosphorylation level, caused the pol II_A form to accumulate, and the Ser2-PO_4 mark to decline. This was not observed with $CPL4_{1-338}$, establishing that the BRCT domain is important for CTD phosphatase activity *in vivo* (Figure 2.4d). The impact of wild-type CPL4 overexpression in *N. benthamiana* on erasing the CTD Ser2-PO_4 mark was weaker than that observed in stably transformed Arabidopsis CPL4-FSG cells, perhaps because of the differences in expression system. In contrast to the stably transformed CPL4-FSG cells, the transient expression samples were harvested only 48 h after Agrobacterium infiltration. Also, not all cells in the infiltrated *N. benthamiana* tissues likely overexpress *CPL4*, and the presence of untransformed cells or those with low levels of CPL4 would dilute the impact of CPL4 on Ser2-PO_4, decreasing the chance of Ser2 dephosphorylation.

Reduction in CPL4 activity promotes pol II CTD hyperphosphorylation

The above results suggest that inhibition of CPL4 function reduces gene expression and cell viability, implying that CPL4 and its homologs exhibit major CTD phosphatase activities essential for survival. However, the effect of inhibiting CPL4 on the pol II phosphorylation status was not obvious in the transient expression system. Unfortunately, I was unable to recover viable transgenic plants expressing $CPL4^{D128A}$, perhaps due to its toxicity. Furthermore, a T-DNA insertion mutant line (*cpl4-2*, GABI_926A04) containing an insertion in the 5th exon of At5g58003 (at base 23481044 of Chromosome 5) did not produce any homozygous mutant progeny. The segregation ratio of the selfed progeny of heterozygous *cpl4-2/+* was severely skewed, i.e., +/+:*cpl4-2/+*: *cpl4-2/cpl4-2*=68:53:0, suggesting the gamete lethality of *cpl4-2* (Table 2.2). To determine whether male or female gametes could transmit *cpl4-2,* I performed a reciprocal cross analysis using *cpl4-2/+* and wild type Col-0 as parents. When Col-0 pollen was used to pollinate *cpl4-2/+*, the progeny segregated at a +/+:*cpl4-2/+*ratio of 32:16, consistent with the lack of aborted ovules in the siliques of heterozygous plants (Figure 2.5). This indicates that the *cpl4-2* mutation could be transmitted through the female gamete relatively normally. However, when *cpl4-2* pollen was used to pollinate Col-0, the progeny segregated at a +/+:*cpl4-2/+* ratio of 27:2. A substantial decrease in pollen transmission of the *cpl4-2* mutation and our inability to recover homozygous *cpl4-2* mutants further indicate the essential function of CPL4 in plant growth and development.

Table 2.2 Genetic analysis of the *cpl4-2* line

Parental genotype		Progeny	Progeny genotype		
Male	**Female**	**total**	**+/+**	***cpl4-2*/+**	***cpl4-2/cpl4-2***
cpl4-2/+	*cpl4-2*/+	121	68	53	0
cpl4-2/+	Col-0	29	27	2	n.a.
Col-0	*cpl4-2*/+	48	32	16	n.a.

The CPL4 locus of seedlings from self- or cross-pollinated progeny was genotyped. +, wild-type CPL4; cpl4-2, T-DNA insertion (GK-926A04); n.a., not applicable.

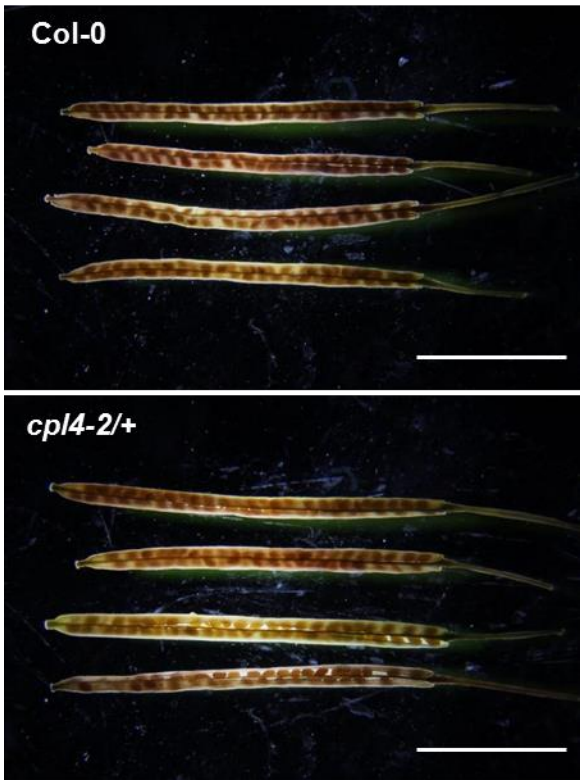


Figure 2.5 Siliques from Col-0 and a *cpl4-2* heterozygous (cpl4-2/+) plant.
Siliques were harvested from seven to eight-week old self-pollinated Col-0 and T3 *cpl4-2* heterozygous (*cpl4-2*/+) plants, and then observed under a dissecting microscope. Bars indicate 5 mm.

To determine the effect of compromised CPL4, I used a *CPL4*$_{RNAi}$ line in which CPL4 expression was constitutively decreased. *CPL4*$_{RNAi}$-6 and 7 are two independent lines showing a moderate growth inhibition phenotype, as reported previously (Bang et al. 2006). In these lines, the abundance of pol II$_O$ in young seedlings was substantially increased (Figure 2.6). Among the three CTD-PO$_4$ epitopes, the increase in Ser2-PO$_4$ signal was the greatest. This demonstrates that CPL4 accounts for the major CTD phosphatase activity in Arabidopsis, and that CPL4 mainly targets CTD Ser2-PO$_4$ of pol II *in vivo*.

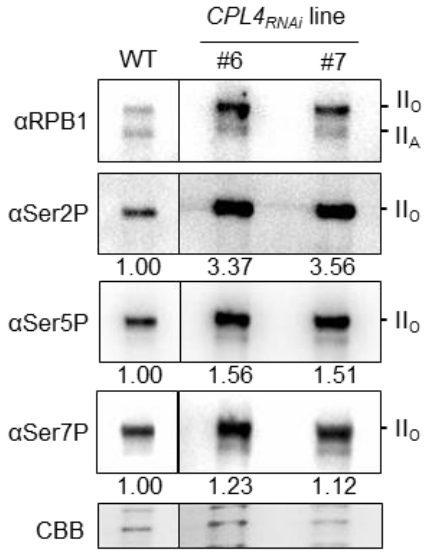


Figure 2.6 Pol II phosphorylation status in 10-day-old *CPL4*$_{RNAi}$ seedlings.
Proteins (15 μg) in total extracts were separated on a 5% SDS-PAGE gel and analyzed by immunoblotting using αRPB1 and phosphoserine-specific antibodies, as indicated. The bottom panel represents CBB-stained low molecular weight proteins on the same gel (loading control).

Knockdown of CPL4 activates expression of DTX3 and its co-expressed genes

I next analyzed the gene expression in the $CPL4_{RNAi}$ line using an Affymetrix ATH1 Gene Chip. cRNA probes were prepared from triplicate total RNA samples extracted from unstressed young (10-day-old) seedlings of the wild type or $CPL4_{RNAi}$-7 line, and used for hybridization and data collection. Hybridization data were processed as described previously (Aksoy et al. 2013). A total of 232 and 224 genes were significantly up- and down-regulated by more than 2-fold, respectively, in the $CPL4_{RNAi}$-7 datasets relative to those in the control. This number appears to be relatively low, considering the global function of FCP1 suggested in other organisms (Kobor et al. 1999). This is likely due to the relatively mild knock-down of *CPL4* by RNAi. Among the T1 plants produced after the introduction of $CPL4_{RNAi}$, viable seeds were recovered only from plants with a mild phenotype (Bang et al. 2006). To confirm the microarray data, the expression levels of 30 upregulated transcripts were determined by RT-qPCR analysis. The up-regulation of 25 genes was confirmed, indicating that the microarray results were generally reliable (Table 2.3). Genes that were not confirmed to be differentially expressed were excluded from subsequent analyses. I also confirmed that *CPL4,* but not other CPL family genes, was specifically knocked down in $CPL4_{RNAi}$ lines (Table 2.4). The estimated expression level of *CPL4* in $CPL4_{RNAi}$ differed in the microarray and RT-qPCR data (48% and 13% of wild type levels, respectively). This is likely due to the fact that the region used to produce CPL4 hairpin RNA overlaps with the ATH1 probe sequences, but not with the amplicon used for qPCR.

Table 2.3 Validation of microarray results by RT-qPCR on select C4UTs

AGI	Gene	Array Fc	RT-qPCR Fc (±SEM)
AT1G61280	*Phosphatidylinositol N-acetylglucosaminyltransferase, GPI19/PIG-P subunit*	51.68	39.77 (±2.97)**
AT2G04050	*DTX3; MATE family protein*	13.70	471.80 (±107.42)*
AT5G42800	*DIHYDROFLAVONOL 4-REDUCTASE; DFR*	11.44	6.85 (±1.14)*
AT5G51440	*HEAT SHOCK PROTEIN 20-LIKE; HSP20L*	10.44	12.27 (±0.07)*
AT1G05680	*UDP-GLYCOSYL TRANSFERASE 74E2*	9.60	18.91 (±5.06)*
AT2G04070	*DTX4; MATE family protein*	8.82	301.37 (±36.46)*
AT1G56650	*PRODUCTION OF ANTHOCYANIN PIGMENT 1*	6.42	11.81 (±1.63)*
AT5G13170	*SENESCENCE-ASSOCIATED GENE 29*	6.15	4.62 (±0.58)*
AT2G18190	*AAA-type ATPase family protein*	6.00	91.31 (±5.95)**
AT3G61630	*CK RESPONSE FACTOR 6*	5.88	3.33 (±0.37)*
AT5G03190	*CONSERVED PEPTIDE UPSTREAM OPEN READING FRAME 47*	5.65	1.65 (±0.18)*
AT4G14630	*GERMIN-LIKE PROTEIN 9*	5.49	1.68 (±0.13)*
AT5G60250	*C3HC4-type RING finger family protein*	5.40	104.38 (±17.11)*
AT2G41730	*Unknown protein*	4.92	4.31 (±0.37)*
AT2G21640	*UP-REGULATED BY OXIDATIVE STRESS*	4.90	12.50 (±1.47)*
AT1G32870	*NAC domain protein 13*	4.23	4.52 (±0.10)*
AT5G43450	*ACC oxidase*	4.12	4.72 (±0.33)*
AT2G36800	*UDP-GLUCOSYL TRANSFERASE 73C5*	3.84	1.98 (±0.17)*
AT4G37370	*CYTOCHROME P450, FAMILY 81, SUBFAMILY D, POLYPEPTIDE 8*	3.46	3.35 (±0.25)**
AT2G03760	*SULFOTRANSFERASE 1*	3.37	1.60 (±0.08)**
AT3G22370	*ALTERNATIVE OXIDASE 1A*	2.70	3.28 (±0.65)*
AT2G04040	*DTX1; MATE family protein*	2.42	15.22 (±4.81)*
AT2G36750	*UDP-GLUCOSYL TRANSFERASE 73C1*	2.37	14.72 (±3.24)*
AT3G13080	*MULTIDRUG RESISTANCE PROTEIN 3*	2.33	16.43 (±2.79)*
AT3G22840	*EARLY LIGHT-INDUCABLE PROTEIN 1*	2.26	2.43 (±0.26)*

For RT-qPCR, values are fold-change (Fc) and standard error of mean (SEM) calculated from three biological replicates. * and **; $p<0.05$ and 0.01, Student's t-test between mean values of $CPL4_{RNAi}$ and wild type. CPL4-Up Transcripts; C4UTs.

Table 2.4 Expression levels of CPL family genes in $CPL4_{RNAi}$ line

Probe Set	AGI	Gene	Microarray Fc (*p*-value)	RT-qPCR Fc (±SEM)
n.a.	AT4G21670	*CPL1*	n.a.	1.30 (±0.13)
251134_at	AT5G01270	*CPL2*	1.34 (5.97×10^{-5})	1.05 (±0.10)
255843_at	AT2G33540	*CPL3*	1.23 (4.81×10^{-3})	1.25 (±0.17)
247894_at	AT5G58003	*CPL4*	0.48 (5.62×10^{-9})	0.13 (±0.01)**

P-values in microarray analysis were calculated by one-way ANOVA with asymptotic *p*-value computation followed by Tukey HSD, using GeneSpringGX 11.0.2 software (Agilent, CA) (Aksoy *et al.* 2013). For RT-qPCR, values are fold-change (Fc) and standard error of mean (SEM) calculated from three biological replicates. * and **; $p<0.05$ and 0.01, Student's *t*-test between mean values of $CPL4_{RNAi}$ and wild type.

Therefore, the *CPL4* transcript level was likely reduced to 13% of the wild type, as estimated by the RT-qPCR analysis. Gene ontology analysis using the GA annotation tool in TAIR revealed that $CPL4_{RNAi}$ affected the expression of genes associated with diverse cellular functions (Appendix A2.1). Of these, I collectively refer to 227 genes upregulated in $CPL4_{RNAi}$ as *C4UTs* (*CPL4-Up T*ranscripts).

The gene expression networks affected by $CPL4_{RNAi}$ were analyzed using a gene co-expression analysis tool, ATTED-II (Obayashi et al. 2011), and 227 C4UTs as query. Three C4UT clusters, C1, C2, and C3, that exhibit a high degree of co-regulation were detected (Figure 2.7a, Figure 2.8, Table 2.5). The C1 cluster represented 23 *C4UT*s that were strongly upregulated in $CPL4_{RNAi}$, including 12 genes that showed up-regulation of more than 4 fold.

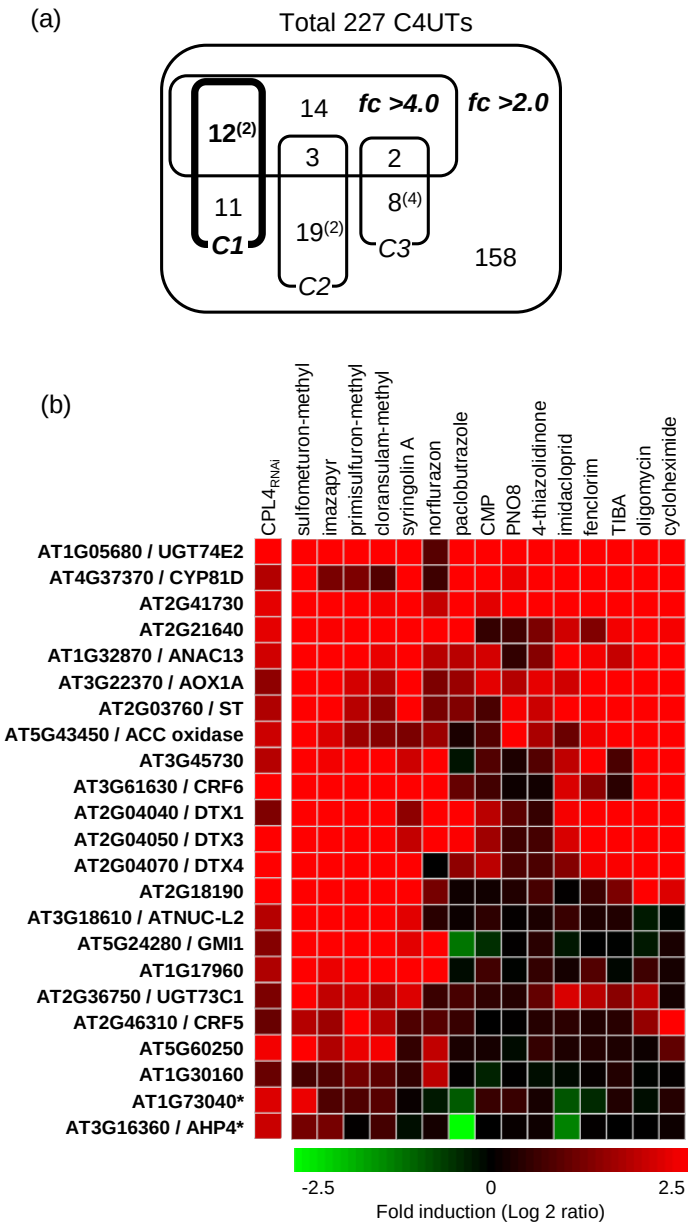


Figure 2.7 Identification of a xenobiotic stress-responsive gene cluster up-regulated by *CPL4*$_{RNAi}$.

(a) Venn diagram representing the induction levels of C4UTs and three major co-expression clusters identified by NetworkDrawer in the ATTED-II database (see Table 2.3 and Figure 2.8 for all cluster information.) Cluster 1 (C1) is enriched for highly up-regulated genes compared with Cluster 2 (C2) and 3 (C3). Numbers in parentheses indicate connections by protein-protein interaction. fc, fold-change. (b) Herbicides and various chemicals induce the expression of cluster 1 genes. The expression levels of C1 genes in plants exposed to xenobiotic stress. Microarray data were retrieved using Genevestigator (https://www.genevestigator.com).

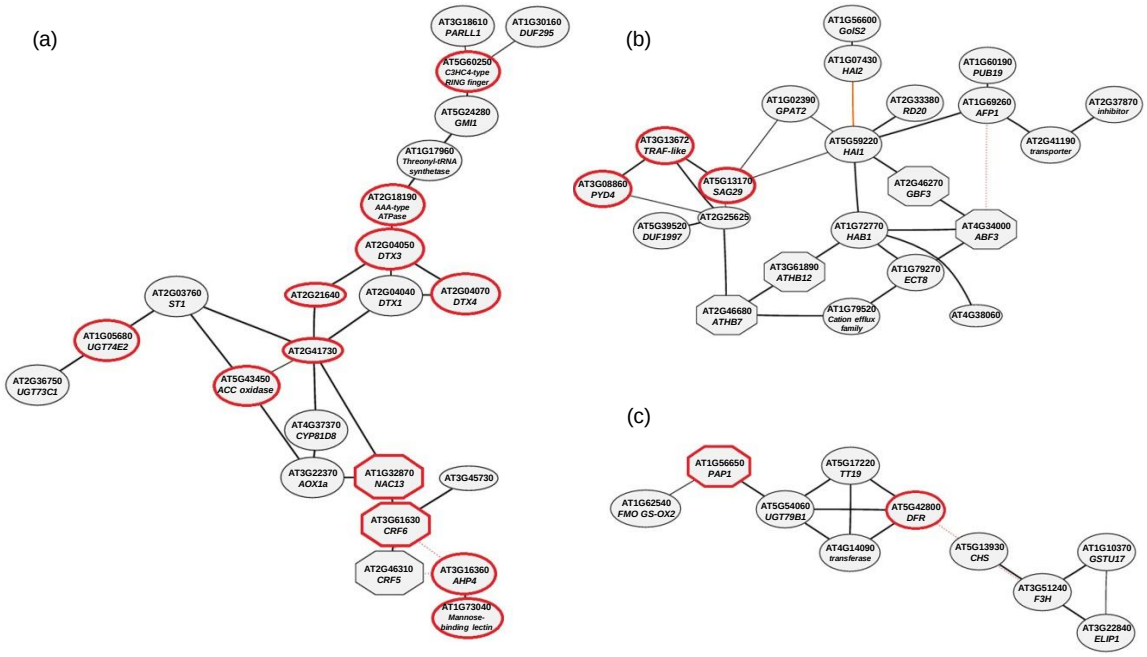


Figure 2.8 Co-expression clusters of C4UTs.
Gene co-expression networks were drawn using the NetworkDrawing tool in ATTED-II (http://atted.jp/). Genes upregulated by more than 4 fold in $CPL4_{RNAi}$ plants are encircled in red. Three clusters containing more than 10 genes were termed Cluster 1 (a), Cluster 2 (b), and Cluster 3 (c). A detailed list of genes in each cluster is presented in Table 2.

The degree of upregulation was weaker for genes in C2 and C3, with the majority of C4UTs in these clusters being upregulated by less than 4 fold. Gene ontology analysis indicated that the C2 cluster represented ABA-inducible genes, and the C3 cluster represented flavonoid biosynthesis pathway genes (Appendix A2.2).

Table 2.5 Co-expression clusters among C4UTs

Probe Set	AGI	Fold change	Gene
		Cluster 1 (23)	
263402_at	AT2G04050	13.70	*DTX3; MATE family protein*
263231_at	AT1G05680	9.60	*UDP-GLYCOSYL TRANSFERASE 74E2*
263401_at	AT2G04070	8.82	*DTX4; MATE family protein*
263061_at	AT2G18190	6.00	*AAA-type ATPase family protein*
251282_at	AT3G61630	5.88	*CK RESPONSE FACTOR 6*
247615_at	AT5G60250	5.40	*C3HC4-type RING finger family protein*
260522_x_at	AT2G41730	4.92	*Unknown protein*
263515_at	AT2G21640	4.90	*UP-REGULATED BY OXIDATIVE STRESS*
262357_at*	AT1G73040	4.62	*Mannose-binding lectin superfamily protein*
261192_at	AT1G32870	4.23	*NAC domain protein 13*
249125_at	AT5G43450	4.12	*ACC oxidase*
259329_at*	AT3G16360	4.05	*HPT PHOSPHOTRANSMITTER 4*
252539_at	AT3G45730	3.55	*unknown protein*
256852_at	AT3G18610	3.52	*PARALLEL1-LIKE 1*
253046_at	AT4G37370	3.46	*CYTOCHROME P450, FAMILY 81, SUBFAMILY D, POLYPEPTIDE 8*
255893_at	AT1G17960	3.38	*Threonyl-tRNA synthetase*
264042_at	AT2G03760	3.37	*SULFOTRANSFERASE 1*
258452_at	AT3G22370	2.70	*ALTERNATIVE OXIDASE 1A*
249784_at	AT5G24280	2.53	*GAMMA-IRRADIATION AND MITOMYCIN C INDUCED 1*
263403_at	AT2G04040	2.42	*DTX1; MATE family protein*
265197_at	AT2G36750	2.37	*UDP-GLUCOSYL TRANSFERASE 73C1*
256188_at	AT1G30160	2.07	*Protein of unknown function (DUF295)*
266606_at	AT2G46310	2.06	*CK RESPONSE FACTOR 5*
		Cluster 2 - ABA(22)	
245982_at	AT5G13170	6.15	*SENESCENCE-ASSOCIATED GENE 29*
258983_at	AT3G08860	4.48	*PYRIMIDINE 4*
256789_at	AT3G13672	4.04	*TRAF-like superfamily protein*
245627_at*	AT1G56600	3.84	*GALACTINOL SYNTHASE 2*
267080_at	AT2G41190	3.63	*Transmembrane amino acid transporter family protein*
261077_at*	AT1G07430	3.35	*HIGHLY ABA-INDUCED PP2C GENE 2*
266327_at	AT2G46680	3.33	*ARABIDOPSIS THALIANA HOMEOBOX 7*
264217_at	AT1G60190	3.21	*PLANT U-BOX 19*
262940_at	AT1G79520	3.06	*Cation efflux family protein*

Genes in three co-expression clusters identified by ATTED-II analysis of 227 *C4UTs*. See Figure 2.8 for each co-expression network. Asterisks indicate genes connected to the cluster by protein-protein interaction.

Table 2.5 Continued

Probe Set	AGI	Fold change	Gene
		Cluster 2 - ABA(22)	
255795_at	AT2G33380	2.99	*RESPONSIVE TO DESSICATION 20*
251272_at	AT3G61890	2.93	*ARABIDOPSIS THALIANA HOMEOBOX 12*
259418_at	AT1G02390	2.76	*GLYCEROL-3-PHOSPHATE SN-2-ACYLTRANSFERASE 2*
266555_at	AT2G46270	2.61	*G-BOX BINDING FACTOR 3*
264102_at	AT1G79270	2.58	*EVOLUTIONALLY CONSERVED C-TERMINAL REGION 8*
247723_at	AT5G59220	2.52	*HIGHLY ABA-INDUCED PP2C GENE 1*
253022_at	AT4G38060	2.31	*CLAVATA COMPLEX INTERACTOR 2*
260357_at	AT1G69260	2.15	*ABI FIVE BINDING PROTEIN 1*
249454_at	AT5G39520	2.07	*Protein of unknown function (DUF1997)*
253263_at	AT4G34000	2.04	*ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR 3*
259922_at	AT1G72770	2.03	*HYPERSENSITIVE TO ABA 1*
265913_at	AT2G25625	2.02	*unknown protein*
266098_at	AT2G37870	2.01	*Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein*
		Cluster 3- flavonoid (10)	
249215_at	AT5G42800	11.44	*DIHYDROFLAVONOL 4-REDUCTASE; DFR*
245628_at	AT1G56650	6.42	*PRODUCTION OF ANTHOCYANIN PIGMENT 1*
248185_at	AT5G54060	3.38	*UDP-GLUCOSYL TRANSFERASE 79B1*
245624_at	AT4G14090	3.22	*anthocyanidin 5-O-glucosyltransferase*
250083_at	AT5G17220	3.02	*TRANSPARENT TESTA 19*
252123_at*	AT3G51240	2.79	*FLAVANONE 3-HYDROXYLASE*
250207_at*	AT5G13930	2.44	*CHALCONE SYNTHASE*
258321_at*	AT3G22840	2.26	*EARLY LIGHT-INDUCABLE PROTEIN 1*
264436_at*	AT1G10370	2.17	*GLUTATHIONE S-TRANSFERASE U17*
265122_at	AT1G62540	2.16	*FLAVIN-MONOOXYGENASE GLUCOSINOLATE S-OXYGENASE 2*

The C1 cluster is the most strongly upregulated cluster among the three; however, gene ontology analysis did not provide a clear function commonly associated with the C1 cluster. Therefore, I focused on analyzing the C1 cluster in this study. The C1 cluster includes three multidrug and toxic compound extrusion (MATE) transporters and two UDP-glucosyltransferases, and thus appears to function in defense against toxic

chemicals. A hierarchical clustering tool in Genevestigator (https://www.genevestigator.com/gv/index.jsp) (Zimmermann et al. 2004) was used to identify microarray datasets that highly upregulate 23 genes in the C1 cluster. This revealed a group of datasets that were largely based on chemically treated plants (Figure 2.7b). Most significantly, nearly all genes in the C1 cluster are upregulated in plants treated with acetolactate synthase (ALS) inhibitor herbicides, such as sulfometuron methyl, cloransulam-methyl, and imazapyr (Figure 2.7b). Among the *C4UT*s, the MATE transporters *DTX3* (AT2G04050) and *DTX4* (AT2G04070), which have been proposed to detoxify toxic chemicals (Manabe et al. 2007), were induced 16.8-86.3 and 35.5-79.5 fold, respectively, upon herbicide treatments. Therefore, I hypothesized that the co-upregulation of genes in the C1 cluster indicates the elevated herbicide/xenobiotic stress response in $CPL4_{RNAi}$. *C4UTs* contain two additional genes, MRP3 and UGT73C5, potentially involved in xenobiotic detoxification. These genes were excluded from the ATTEDII platform, because the corresponding ATH1 probes recognize two targets. Upregulation of these genes in $CPL4_{RNAi}$ plants was confirmed by RT-qPCR (Table 2.3).

Multi-drug efflux/detoxification genes are constitutively activated but not hyperactivated in $CPL4_{RNAi}$

To determine if the elevated transcriptional xenobiotic defense response of $CPL4_{RNAi}$ confers resistance to xenobiotics, I analyzed the herbicide response of $CPL4_{RNAi}$. As shown in Figure 2.9, $CPL4_{RNAi}$ lines were more tolerant than the wild type to chlorosulfuron during seed germination and seedling establishment.

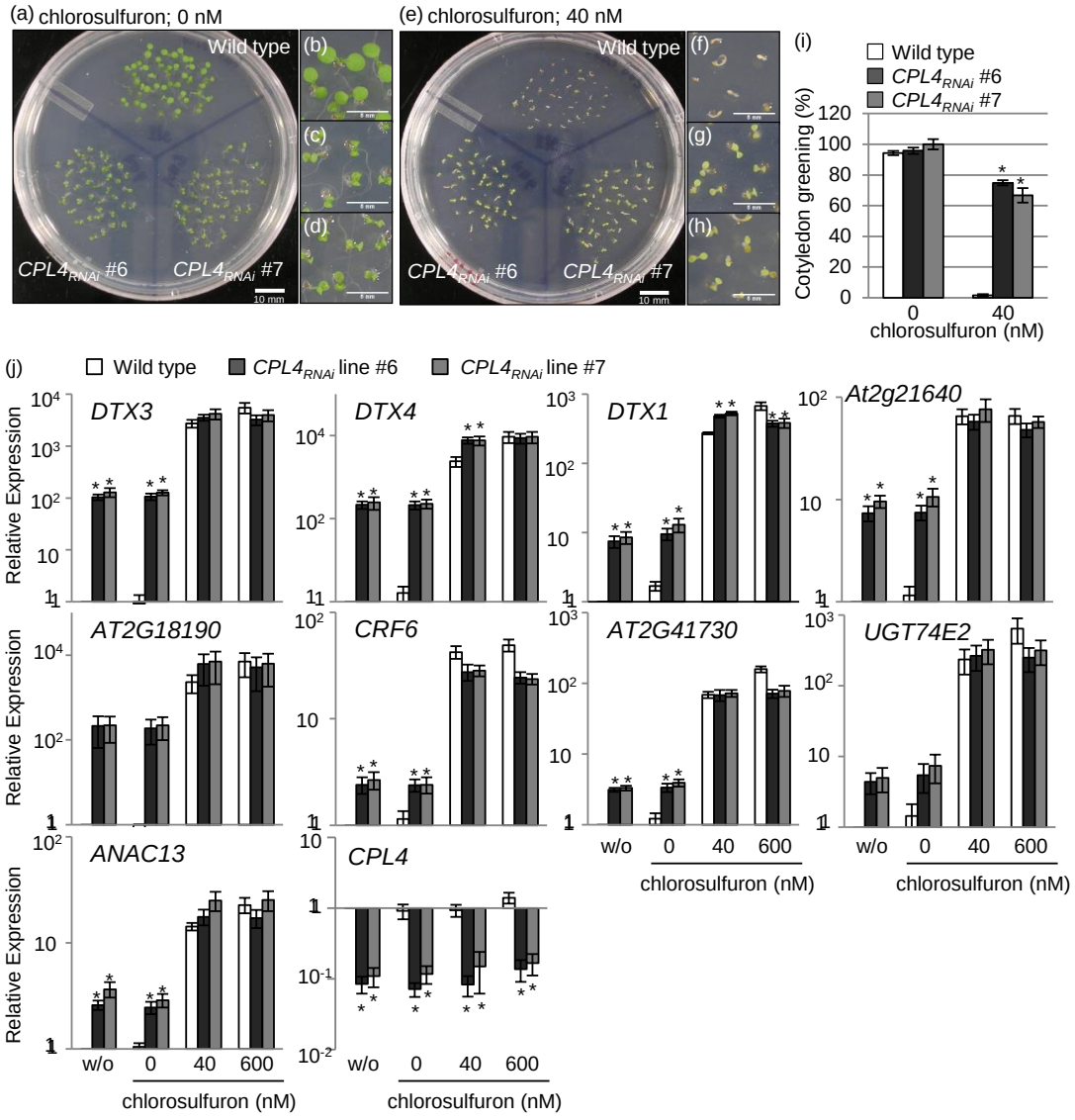


Figure 2.9 Response of *CPL4*$_{RNAi}$ lines to chlorosulfuron.

(a-h) Photographs of 5-day-old seedlings grown on medium containing 0 nM (a-d) or 40 nM (e-h) chlorosulfuron. Magnified images of each genotype (b and f, wild type; c and g, *CPL4*$_{RNAi}$ #6; and d and h, *CPL4*$_{RNAi}$ #7) are shown. (i) Rate of cotyledon greening observed among 5-day-old seedlings grown on medium containing with the indicated concentration of chlorosulfuron. Cotyledon greening was visually scored based on the opening and greening of the cotyledons (g-h). (j) Expression of representative *C4UT*s and *CPL4* in wild type and *CPL4*$_{RNAi}$ lines without treatment (w/o), with water treatment (0 nM) or with chlorosulfuron treatments. The gene expression levels relative to the untreated wild type are presented as the mean values of three biological replicates. The y-axis uses a base-10 logarithmic scale. Bars indicate the SEM of biological replicates. * p<0.05, Student's t-test between mean values of *CPL4*$_{RNAi}$ and the wild type under the same conditions.

On medium containing 40 nM chlorosulfuron, 60-70% of *CPL4*$_{RNAi}$ lines but only 1.6% of wild-type plants were able to develop green cotyledons (Figure 2.9a-i), suggesting that the up-regulation of C1 cluster genes contributes to the xenobiotic stress tolerance of *CPL4*$_{RNAi}$. A moderate tolerance was also observed at the seedling level (Figure 2.10). RT-qPCR analysis of C1 cluster *C4UT*s during herbicide treatment revealed that *CPL4*$_{RNAi}$ constitutively expresses xenobiotic-responsive transcripts (Figure 2.9j). When plants were treated with chlorosulfuron, wild-type and *CPL4*$_{RNAi}$ plants exhibited similar levels of *C4UT*s. The expression level of CPL4 itself was not affected by chlorosulfuron treatment (Figure 2.9j). Therefore, it is likely that the reduction in *CPL4* expression de-represses, but does not enhance, the xenobiotic stress response.

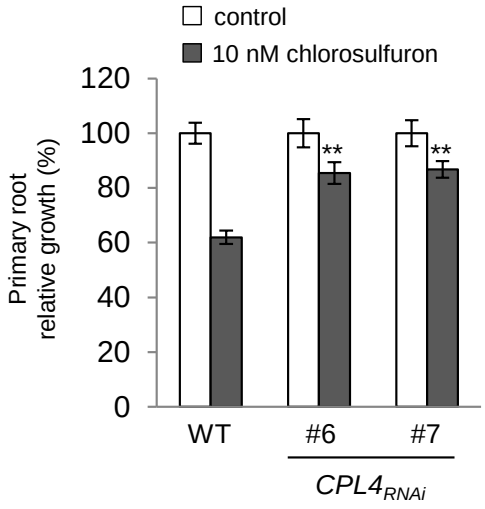


Figure 2.10 Primary root growth of wild-type and *CPL4*$_{RNAi}$ seedlings exposed to chlorosulfuron.

Four-day-old vertically-grown seedlings were transferred to control medium (½ MS medium with 1% sucrose) or to medium containing 10 nM chlorosulfuron. The primary root growth was measured on the third day post transfer, and is shown as a percentage relative to the growth observed under control conditions. ** $p<0.01$, Student's t-test between mean values of the wild type (WT) and each *CPL4*$_{RNAi}$ line under chlorosulfuron treatment. Bars indicate the SEM of 17-24 seedling measurements. The experiment was repeated twice with similar results.

2.4 Discussion

The coordinated phosphorylation and dephosphorylation of pol II-CTD represents a pivotal regulatory mechanism of eukaryotic transcription. CTD kinases have been extensively studied in a range of organisms, including plants, and the distinct substrate specificities and roles of various CTD kinases in transcription cycles have been established. In contrast, our understanding of the function of plant CTD phosphatases has been limited to a handful of studies (Koiwa et al. 2002; Xiong et al. 2002; Koiwa et al. 2004; Hausmann et al. 2005; Bang et al. 2006; Ueda et al. 2008; Jin et al. 2011; Aksoy et al. 2013; Jeong et al. 2013a). In this chapter, I provide evidence that CPL4 is a major CTD phosphatase in Arabidopsis that is involved both in general and xenobiotic stress response-specific gene expression. CPL4 exhibited various hallmark activities, such as the interaction with pol II and dephosphorylation of pol II, both *in vitro* and *in vivo*. CPL4 purified either from *E. coli* or from plants dephosphorylated synthetic CTD phosphopeptide at both Ser2 and Ser5, and transgenic plants overexpressing CPL4 showed reductions in both the Ser2- and Ser5-PO_4 marks of pol II, indicating that CPL4 dephosphorylates CTD-PO_4 at Ser2 and Ser5. However, similar to animal and fungal FCP1, the position specificity of phosphatase activity appeared to be influenced by the preparation method of CTD-PO_4 substrates, because a strong preference for Ser2-PO_4 was observed when full-length CTD phosphorylated by MPK3 but not synthetic CTD phosphopeptide was used as a substrate.

In a transient co-expression assay, overexpression of catalytically compromised $CPL4^{D128A}$ inhibited the reporter gene expression. The strong inhibitory effect of

CPL4^{D128A} was dependent on the presence of the BRCT domain, which is thought to form a CTD-binding cleft structure (Ghosh et al. 2008). This suggests that CPL4^{D128A} functions as a dominant negative form that competes with endogenous CPL4 and inhibits general gene expression. While it was not possible to stably overexpress CPL4^{D128A}, the phenotype of strong *CPL4$_{RNAi}$* lines was consistent with this hypothesis (Bang et al. 2006). The growth of *CPL4$_{RNAi}$* plants is significantly inhibited (Bang et al. 2006), and the level of pol II$_O$ is substantially increased compared to the wild type. It remains to be clarified whether dephosphorylation by CPL4 occurs during transcription termination and/or if CPL4 has a specific function in activating transcriptionally inactive, phosphorylated pol II, such as the pol IIm or pol IIe forms (Palancade et al. 2001).

A suite of genes responsive to xenobiotic stress, i.e., the C1 cluster, was upregulated in *CPL4$_{RNAi}$* plants. These genes include 3 transport-related genes (DTX1, DTX3, and DTX4) and 4 detoxification enzymes (CYP81D8, ST1, UGT74E2, and UGT73C1). Two additional C4UTs (MRP3 and UGT73C5) also likely participate in xenobiotic stress tolerance. Of these, DTX1 is able to transport a broad range of chemicals, including toxic chemicals/xenobiotics (i.e., berberine, palmatine, norfloxacin, and ethidium bromide) when heterologously expressed in *E. coli* (Li et al. 2002). Since *DTX1-DTX4* exhibit high levels of amino acid sequence identity (90%), and DTX1, 3, and 4 (together with other genes) form a gene cluster, it is likely that DTX3 and DTX4 have broad substrate specificities, similar to DTX1. Similarly, MRP3 can transport dinitrobenzene-glutathione conjugate (Tommasini et al. 1998). Several studies have

shown that UGTs are active on endogenous and xenobiotic substrates. The UGTs upregulated in *CPL4*$_{RNAi}$ use a broad range of endogenous substrates as well as xenobiotic substrates. For example, UGT74E2 can use indole butylic acid (IBA) (Tognetti et al. 2010), UGT73C1 can use zeatin (Hou et al. 2004) and 2,4,6-trinitrotoluene (TNT) derivatives (HADNT/HDNT: hydroxyaminodinitrotoluen/aminodinitrotoluen) (Gandia-Herrero et al. 2008), and UGT73C5 can use brassinosteroids (Poppenberger et al. 2005) and fungal mycotoxin deoxynivalenol (Poppenberger et al. 2003) as substrates. These transporters and UGTs may detoxify chlorosulfuron in plants directly (Christopher et al. 1991). However, functional analyses of DTX1-DTX4 as well as of UGTs are necessary to determine whether DTX and UGT proteins directly contribute to the herbicide resistance of *CPL4*$_{RNAi}$. Also, the mechanism that confers specific induction of the C1 cluster in *CPL4*$_{RNAi}$ remains to be determined. Several transcription factors, such as ERF (ETHYLENE RESPONSE FACTOR), CRF6 (CK RESPONSE FACTOR 6), CRF5, ARR7 (ARABIDOPSIS RESPONSE REGULATOR 7), ARR5, and ANAC13, were identified in our microarray analysis as being strongly co-regulated with the C1 cluster in *CPL4*$_{RNAi}$. These transcription factors are candidates for further investigation to establish the regulatory relationship among *C4UT*s.

What initially triggers the expression of cluster C1 genes in *CPL4*$_{RNAi}$ plants? While the representative chemicals capable of inducing genes of the C1 cluster are herbicides that inhibit ALS, exposure to a wide range of xenobiotics and UV-B is capable of inducing the expression of genes of the C1 cluster, indicating that up-

regulation of the C1 cluster in the $CPL4_{RNAi}$ line is not solely due to a failure in branched-chain amino acid (BCAA) biosynthesis. Indeed, several C1 cluster genes are early-response genes during herbicide treatment (i.e., induced within 6 h of exposure), being induced presumably before a deficiency in BCAAs sets in. Rather, upregulation of C1 cluster genes may represent an elevated defense response to general xenobiotic chemicals or toxic metabolites. If so, $CPL4_{RNAi}$ cells that have been predisposed to toxic metabolites exhibit elevated genotoxic stress responses, which includes up-regulation of C1 cluster genes. The finding that the expression levels of C1 cluster genes did not differ in the wild type and $CPL4_{RNAi}$ line after herbicide treatment support a model in which $CPL4_{RNAi}$ plants are preconditioned, but not hyper-responsive, to xenobiotic stress.

CHAPTER III

SALT-INDUCIBLE CONVERSION OF SMALL NUCLEAR RNA TRANSCRIPTION INTO PROTEIN-CODING EXPRESSION MEDIATED BY POL II-CTD PHOSPHOREGULATION IN ARABIDOPSIS

3.1 Introduction

Eukaryotic transcriptome consists of protein-coding polyadenylated mRNAs and non-coding RNAs (ncRNAs), the latter have no or little protein-coding potential (Liu et al. 2006). It is widely accepted that expression of not only mRNAs but also ncRNAs are regulated by various developmental and environmental stimuli, suggesting unique functions for ncRNA specific to growth, development, and stress responses (Brown et al. 1992; Valgardsdottir et al. 2005; Valgardsdottir et al. 2008; Heo and Sung 2011). ncRNA can be largely classified into canonical and non-canonical ncRNAs, the former include ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and small nuclear/nucleolar RNAs (snRNAs and snoRNAs), the latter include micro RNA, small interfering RNA, long ncRNA, and intermediate ncRNA (Wang et al. 2014). Early studies established roles of canonical ncRNAs in protein synthesis and RNA maturations (Lerner et al. 1980; Ohshima et al. 1981; Sontheimer and Steitz 1993), whereas advances in the area of gene silencing pathways revealed roles of many non-canonical ncRNAs as transcriptional and translational regulators of gene expression (Ruiz et al. 1998; Reinhart et al. 2002; Xie et al. 2005; Wierzbicki et al. 2009). In model plants, Arabidopsis, some ncRNAs and ncRNA-like sequence have more than one function. tRNA-like structures

in mRNAs can function as a systemic long-distance transport signal of mRNA (Zhang et al. 2016). Another example is that miRNA precursor (pri-miR165a) functions as mRNA to produce a regulatory peptide (Lauressergues et al. 2015). Whether or not other classes of ncRNAs can have multiple functions has been unclear. Indeed, for U snRNAs, a major class of canonical ncRNA, have only been associated with snRNP functions in RNA maturation processes (Lerner et al. 1980; Ohshima et al. 1981; Sontheimer and Steitz 1993), however, mRNAs with embedded U snRNA structure have been identified in Arabidopsis (Hare et al. 2003).

There are two classes of U snRNAs; Sm-class snRNAs including U1, U2, U4, U4atac, U5, U11 and U12 snRNAs are transcribed by RNA polymerase II (pol II), and have a 5'-trimethylguanosine (TMG) cap structure and binding sites for Sm proteins (Lerner and Steitz 1979). LSm-class snRNAs such as U6 and U6atac are transcribed by RNA polymerase III. Transcriptional activation of snRNA promoter resembles that of mRNA; it requires a conserved snRNA-specific transcriptional activator, SNAPc complex, which binds to Proximal Sequence Element (PSE) or Upstream Sequence Element (USE) (Sadowski et al. 1993; Ohtani and Sugiyama 2005), and general transcription factors, such as TATA-binding proteins, TFIIB, TFIIE, TFIIF that assemble on TATA-box sequence (Kuhlman et al. 1999). The transcription termination and 3'-end formation mechanisms of snRNA are studied more in detail in animals and resemble these of mRNA albeit it requires a snRNA-specific promoter (Hernandez 1985; Hernandez and Lucito 1988). Like pre-mRNAs, pre-snRNAs are transcribed beyond the 3′ end of mature snRNAs, and processed by the function of the integrator complex that

is recruited to pol II with specific phosphorylation patterns at the carboxyl terminal domain (CTD) of the largest subunit (Egloff et al. 2010). In plants, transcriptions of both classes of snRNAs are regulated by USE (upstream sequence element; RTCCCACATCG) and TATA box in the promoter region, but spacing between two elements distinguishes them: the pol II-dependent Sm class shows [USE-N_{32-35}-TATA box] configuration whereas pol III-dependent LSm-like class shows [USE-N_{23-24}-TATA box] (Waibel and Filipowicz 1990).

Pol II-CTD consists of multiple repeats of a conserved heptapeptide Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7 (Allison et al. 1988; Nawrath et al. 1990). All residues in heptads except prolines are targets of pol II-CTD kinases and phosphatases that regulate the activities and functions of pol II complex during the transcription cycles. Studies in animal systems have shown that the roles of pol II-CTD phosphoregulation in snRNA transcription largely mirror those in mRNA transcription (Egloff 2012; Egloff et al. 2012; Wani et al. 2014). During transcription activation and promoter escape, phosphorylation at Ser5-PO_4 is required to recruit capping enzymes (Ho and Shuman 1999; Wen and Shatkin 1999). Ser2-PO_4 is crucial for recruiting the integrator complex during the termination and 3'-end processing (Egloff et al. 2010; Gu et al. 2013; Davidson et al. 2014), but not during transcription elongation (Medlin et al. 2003), highlighting differences in the significance of pol II-CTD phosphoregulation between snRNA and protein-coding gene transcriptions. In addition, Ser7-PO_4 phosphorylation is needed for efficient recruitment of the Integrator complex (Egloff et al. 2007; Egloff 2012; Egloff et al. 2012). By contrast, phosphorylation of Ser5 appears inhibitory for

the recruitment of the integrator complex. Information for snRNA transcription in plants is limited, but studies showed that it uses components homologous to animal counterparts such as 11-bp USE (RTCCCACATCG) (Waibel and Filipowicz 1990), SNAPc complex (Ohtani and Sugiyama 2005), and the integrator complex (Liu et al. 2016). Notable differences between plant and animal snRNA transcription are that plant snRNA 3'-end processing does not require a snRNA promoter (Connelly and Filipowicz 1993). Involvement of pol II CTD phosphorylation in regulation of plant snRNA biogenesis is not clear, but CTD kinase mutants with decreased CTD phosphorylation levels did not show decreased snRNA levels in Arabidopsis (Hajheidari et al. 2012). Currently, no pol II-CTD phosphoregulation factors in plants have been characterized in co-transcriptional processing of snRNA.

Phosphorylation status of pol II-CTD responds to various developmental and environmental signals. For example, human, simian, rodent and avian cells treated with severe heat shock exhibit higher pol II-CTD phosphorylation, whereas amphibian and insect pol II apparently reduce the CTD phosphorylation level after heat shock (Lavoie et al. 1999). Similarly, heat stress has been shown to elevate Ser2-phosphorylation level in budding yeast *Saccharomyces cerevisiae* (Patturajan et al. 1998). On the other hand, mild heat stress tends to cause de-phosphorylation of pol II-CTD due to CTD-kinase inhibition in human cells (Dubois et al. 1994; Venetianer et al. 1995). In Drosophila melanogaster, pol II-CTD dephosphorylation by CTD-phosphatase FCP1 with a preference toward Ser2-PO_4 is required for efficient transcription of heat shock genes (Fuda et al. 2012). DmFCP1 depletion by RNAi was shown not to cause significant

changes in pol II-CTD phosphorylation level on heat shock genes. Instead, it caused a dramatic increase of non-chromatin-bound hyper-phosphorylated pol II. In fission yeast *Schizosaccharomyces pombe*, nitrogen starvation induces phosphorylation of Ser2 via a stress-inducible MAP-kinase Sty1 phosphorylating a pol II-CTD kinase CTDK-I (Sukegawa et al. 2011). Osmotic stress and oxidative stress have been shown to give rise to a pol "IIm" form which is phosphorylated extensively on Ser5 due to ERK/MAPK activities (Bonnet et al. 1999). In Arabidopsis, CTD phosphoregulation is an integral part of various biotic and abiotic stress signaling including microbial-associated molecular patterns (MAMPs)/innate immunity signaling (Li et al. 2014), osmotic-stress response (Koiwa et al. 2002; Xiong et al. 2002), and hormone signaling (Ueda et al. 2008; Matsuda et al. 2009). In ecological scale, an allelopathic phytochemical juglone (5-hydroxy-1,4-naphthalenedione) function by inhibiting the PPIase Pin1 and causes dephosphorylation of pol II-CTD based on the studies using animal cell models (Hennig et al. 1998; Chao et al. 2001).

CTD-phosphatase-like 4 (CPL4) is a genuine CTD phosphatase in Arabidopsis thaliana and is orthologous to human and fungal FCP1 (Fukudome et al. 2014). Knocking down CPL4 expression by RNAi ($CPL4_{RNAi}$) induces global CTD hyperphosphorylation and upregulation of more than 200 genes indicating that CPL4 functions as a major CTD phosphatase in Arabidopsis. In the present study, I show that $CPL4_{RNAi}$ also affect snRNA biogenesis, promoting 3'-extension of snRNA transcripts. Surprisingly, this resulted in production of chimeric snRNA-mRNA fusion transcripts with protein-coding potential. Expression of these fusion transcripts, which are termed as

"snR-DPG (downstream protein-coding gene)", is dependent on snRNA transcription machinery, and is also produced in wild-type background when CTD phosphorylation status is altered by salt stress and developmental signals. These results revealed a novel mechanism of plant stress-responsive gene expression that uses pol II-CTD phosphorylation and snRNA 3'-end processing as regulatory nodes.

3.2 Materials and methods

Plant materials and growth condition

Unless otherwise stated, seeds were sown on ¼ MS media containing 0.5% sucrose and 1.5% Agar and treated with cold for 2 days. Then, they were grown vertically at 25 ^{o}C and long-day (16-hour day/8-hour night) condition. 10-15-day old seedlings were used for experiments. *srd2-1* in Ler background was kindly provided by Dr. Misato Ohtani and Dr. Munetaka Sugiyama. Callus cultures were induced from excised a few-day-old seedlings incubated on callus induction medium containing 1x MS salts, 2% sucrose, 0.2 g/L KH2PO4, 0.1 g/L myo-inositol, 1x B5 vitamins and 0.9% agar). The callus cultures were maintained in dark and 25 ^{o}C, and transferred to fresh callus induction medium once in every 7-10 days.

RT-PCR and qPCR

RNA extraction and reverse transcription were done as previously described (Fukudome et al. 2014). Total RNAs were extracted using TRIzol reagent, followed by DNase I treatment to eliminate DNA contamination. One to two microgram of total RNAs were converted to first-strand cDNA by GoScript® Reverse Transcriptase

(Promega, http://www.promega.com/). The reverse transcription products were analyzed using a LightCycler 480 (Roche Diagnostics, http://www.roche.com/) and Bullseye EvaGreen qPCR MasterMix (http://www.midsci.com). GAPDH (AT1G13440) was used as an internal control for normalization in qPCR. In RT-PCR experiments of snR-DPGs, *KAKU4* (AT4G31430) was chosen as a control due to its low expression level closer to these snR-DPGs and stability. Primers used are listed in Appendix P.

RNA-seq

10-day old WT and $CPL4_{RNAi}$ #7 seedlings horizontally grown on ¼ MS media containing 0.5% sucrose were subjected to total RNA extraction using RNeasy Plant Mini kit (QIAGEN). The total RNA was then submitted to Otogenetics Corporation (Norcross, GA) for RNA-seq analysis (paired-end, 100-bp). After validating the integrity and purity of total RNA were using Agilent Bioanalyzer and OD260/280, 2 µg of total RNA was used for rRNA depletion using Ribo-Zero Magnetic Gold Kit (EpiCentre). Recovered total RNA from rRNA depletion was subjected for cDNA synthesis using SMARTer PCR cDNA Synthesis kit (Clontech Laboratories, Inc., Mountain View, CA USA, catalog# 634925). Sequencing was performed on the Illumina HiSeq 2000 (Illumina, CA USA) with chemistry v3.0 and using the 2×100bp paired-end read mode and original chemistry from Illumina according to the manufacturer's instructions.

For RNA-seq analysis of im-ncRNAs, total RNAs were purified from 7-day-old Col-0 wild-type and *hen2-4* mutant roots (Lange et al. 2014) growing on media containing 1/4 x MS salts, 0.5% sucrose 1.5% agar, using miRNA-easy kit (Qiagen). Sequencing libraries were prepared using TruSeq Stranded Total RNA with Ribo-Zero

Plant kit. To enrich im-ncRNA, fragmentation step was omitted, and library cDNA with 100-500nt were purified using Pippin (Sage Science) prior to sequencing by Illumina HiSeq (125 nt, paired-end).

Bioinformatics

For RNA-seq analysis, raw read files (fastq) from our own experiments and from Sequence Read Archive (SRA) were uploaded to Galaxy (https://usegalaxy.org/) and all analysis reported was done on the platform. After QC (FastQC) and adapter-trimming (Trim Galore!), the processed reads were mapped with Tophat (v2.0.14) against reference assembly Ensembl TAIR10 from Illumina iGenome. Because the original coordinates for each of 74 snRNAs from the ASRG database (Wang and Brendel 2004) were not consistent with TAIR10 genome, I manually updated each coordinate by running blast search. Coverage depth of each of snRNA-extension region was calculated by SAMtools; BedCov, and normalized by total number of mapped reads (Depth of Coverage per base per Million; DCPM). FPKM values for LAF3 isoforms and fold-change values of DPGs were obtained by running Cuffdiff. For coverage visualization in trackster, the mapping files (bam) were converted to bigwig format through BEDTools; Genome Coverage and Wig/BedGraph-to-bigWig converter. The coverage was normalized by total number of mapped reads.

Pol II-dependent snRNA promoter search in Arabidopsis genome

The pol II-dependent promoter (PIIsnR) motifs (RTCCCACATCGN_{32-35}TATAA) were searched against *Arabidopsis thaliana* TAIR10 genome sequence using the Find Individual Motif Occurrences (FIMO) tool in the MEME suite (Bailey et al.

2009; Grant et al. 2011). Due to a size limitation of the input sequences, the search was separately conducted on chromosome 1-3 and chromosome 4-5 (parameters; match p-value<1E-6; scanning both strands). Total 150 matched sequences with q-value<0.1 (corrected p-value for multiple testing) were considered for further analysis. For each PIIsnR, coverage (DCPM) of 200-bp region in starting from 23-bp downstream of the TATA-box was computed as described above. The 150 sequences were classified into three classes ("Repeat", "snRNA" and "novel") based on the phylogram computed and visualized by the multiple sequence alignment tool Clustal Omega and the TreeView software, respectively (Page 1996; Sievers et al. 2011).

snR-DPG search in other organisms

For identification of potential snR-DPG transcripts in other species, Blastn search against Refseq_rna database using Arabidopsis U1, U2, U4, U5 and U12 snRNA sequence as query was performed. Transcripts in NM/XM (mRNAs, predicted model included) categories with similarity to the query snRNA sequences on the same strand were selected. Then, transcripts with “hit-start” position larger than the query snRNA length, and those with the Alignment coverage less than 50% were filtered to focus on transcripts harboring snRNA sequences on 5’-end. In total, 150 Refseq_rna transcripts were obtained. Each of the 150 transcripts was then subjected to Blastn search against EST (est) data set to find if there are ESTs supporting the snRNA-DPG fusion transcript (Table 2).

U12-LUC2 reporter system

A luciferase expression vector pEnEOLUC2Thsp was derived from pEnEL2LUC (GenBank accession # KF545095.1) by excising firefly luciferase gene (LUC) and inserting LUC2 (Promega) and terminator of small heat shock protein (HSP18) (Nagaya et al. 2010). The wild-type U12-LUC2 construct was prepared using a 1.5 kbp genomic fragment (chromosome 1, 22602141- 22603607) corresponding between 1.0 kbp upstream of U12 transcription start site and 24 bp downstream of At1g61280 start codon (ATG), which was placed upstream of LUC2 to generate translational fusion of At1g61280-LUC2 driven by U12/At1g61280 promoter. The m, 3m, and m3m variants were prepared by introducing mutations at USE (chromosome 1: 22603035-22603046, GTCCCACATCG to GgCaaACATCG) and/or 3'-box (chromosome 1: 22603073-22603079, AGTAAAT to TCGCGAC) by overlap extension protocol (Ho et al. 1989). Resulting U12-LUC2 plasmids were recombined with pCB302-GW (a Gateway derivative of pCB302 (Xiang et al. 1999)) using LR clonase (Life Technologies). Resulting pCB302U12-LUC2 plasmids were introduced into *Agrobacterium tumafaciens* GV3101 and used for floral transformation (Chung et al. 2000). The T1 transformants were selected on soil by spraying 30µg/ml Liberty herbicide as described previously (Bang et al. 2006).

For luciferase assay, seeds of reporter transgenic plants were germinated on media containing 1/4 x MS salts, 0.5 % sucrose, 0.7% agar supplemented with 10 µg/ml phosphinothricine. After sowing seeds, media plates were stratified for 2 days and then kept at 25°C under the 16h light/8h dark cycle for 8 days. Luciferase activity was

measured after spraying 1 mM luciferin. Image acquisition with a charge-coupled device (CCD) system and processing with WINVIEW software (Roper Scientific, Trenton, NJ) were performed essentially as described (Koiwa et al. 2002). Briefly, signal noises were subtracted from the acquired images by using images acquired without samples. Any negative values were converted to zero. Then, total signals from all seedlings were divided by number of seedlings (Figure 3.2c) or signals from each ppt-resistant seedling were counted individually (Table 3.1).

Protein extraction for RNA polymerase II-CTD phosphorylation

Callus or seedling tissues were homogenized in protein extraction buffer (50 mM Tris-HCl, pH 9.0; 100 mM NaCl; 12.5% glycerol; 2.5 mM EDTA; 10 mM beta-mercaptoethanol; and 20 mM sodium fluoride) supplemented with 1 mM PMSF and 1x Proteinase Inhibitor Cocktail for Plant Cell and Tissue Extracts (Sigma). The crude extracts were then filtered through glass beads column to remove debris. The filtrate as total protein fraction was analyzed by western blotting (5% SDSPAGE gel).

Salt treatment

Ten to thirteen day old seedlings were placed onto two-layer of filter papers saturated with liquid ¼ MS media containing 300 mM NaCl in a petri dish for 30 – 200 min at 25 ^{o}C with a lid on. After treatments seedlings were snap frozen with liquid nitrogen and stored at -80 ^{o}C until total RNA extraction described above.

Pol II occupancy assessment by chromatin immunoprecipitation-qPCR (ChIP-qPCR)

Two-day post transfer calli were used. Prior to crosslinking, wild-type (Col) calli were immersed in liquid CIM containing 300 mM NaCl and incubated on gentle

horizontal shaking at room temperature for 200 min. The container was covered with foil to block light. After salt treatment, NaCl-treated calli were harvested along with untreated wild-type and $CPL4_{RNAi}$ calli. By gently covering the cells with paper towels, excess extracellular liquids were removed. Subsequently, Chromatin Immunoprecipitation was conducted as previously described, with some modifications (Saleh et al. 2008; Castillo-Gonzalez et al. 2015). All buffers were ice cold unless otherwise stated. The harvested calli were subjected to crosslinking using vacuum infiltration system. Calli were immersed in crosslinking buffer containing 0.4 M sucrose, 10 mM Tris-HCl pH8.0, 1 mM EDTA, 1% formaldehyde and 1 mM PMSF, and then were vacuum infiltrated for 15 min (2 min vacuum, release, 8 min vacuum, release, additional 5 min vacuum). The formaldehyde was quenched by adding final 100 mM Glycine to the solution, followed by additional 5 min vacuum infiltration. The crosslinked materials were washed 5 times with ice cold water. Then, after removal of extracellular water by paper towels, fresh weight of each samples were measured (approximately 1.5 – 3 g) and flash frozen in liquid nitrogen. The frozen cells were ground into powder using pre-chilled motar and pestle, and resuspended in 5 vol of ice-cold HONDA buffer (0.44 M sucrose, 1.25% Ficoll, 2.5% Dextran T40, 20 mM Hepes pH7.4, 10 mM MgCl2, 0.5% Triton X-100, 5 mM DTT, 1 mM PMSF and 1x Protease Inhibitor Cocktail for for plant cell and tissue extracts (Sigma, P9599)). The suspensions were incubated on a rotator in cold room for 15-30 min to complete homogenization. The homogenized slurry was filtered through two-layer of Miracloths (EMD Millipore). The debris left on the first-round of Miracloth was washed off and re-suspended in

HONDA buffer, then the suspension was filtered through new two-layer Miracloth. The filtrates were combined and centrifuged at 2000 g for 15 min of 4 ^{o}C. After removing supernatant, the pellet fraction was resuspended in 500 µl to 1 ml Nuclear Lysis buffer containing 50 mM Tris-HCl pH8.0, 10 mM EDTA pH8.0, 1% SDS, 1mM PMSF and 1x Protease Inhibitor Cocktail (Sigma, P9599). Chromatin-DNA was sheared into approximately 500-bp fragment on ice by using Sonic Dismembrator 60 (Fisher Scientific), 10 cycles of 15-sec-pulse with 1-min interval between each cycle to prevent overheating. The sonicated solution was centrifuged at 15,000 rpm for 10 min of 4 ^{o}C to precipitate debris. DNA concentration of the supernatant was measured using NanoDrop, taking Nuclear Lysis buffer as blank. The sonicated chromatin solution was then aliquoted and stored in -80 oC, or immediately proceeded to immunoprecipitation (IP). The sonicated chromatin containing 20 to 30 µg DNA was diluted 10-times with Chromatin Dilution buffer (16.7 mM Tris-HCl pH8.0, 167 mM NaCl, 1.2 mM EDTA and 1.1% Triton X) to a final volume of 500 µl in a 1.5 ml-non-stick tube. Then, the diluted chromatin solution was incubated with antibodies for overnight at 4 ^{o}C. Amount of antibody used were 8.64 µg and 9 µg for anti-RPB1 polyclonal antibody and anti-Ser2-PO_4 polyclonal antibody (Abcam; ab5095), respectively. Following the antibody incubation, 30-40 µl Dynabeads Protein G (Thermo Fisher) washed with Binding/Washing buffer (BW buffer; 20 mM Tris-HCl pH8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS and 1 mM PMSF) was added to the overnight-incubated sonicated chromatin and further incubated for 2.5 hours at 4 ^{o}C with gentle rotation. Then, the beads were washed 3 times with 1 ml BW buffer, followed by 2-time

washes with 1 ml TE buffer (10 mM Tris-HCl pH8.0 and 1 mM EDTA). Each wash was 5-min rotation at 4 ^{o}C. After the final wash, beads were transferred to 0.6 ml tube and the supernatant was completely removed. Subsequent reverse-crosslink and elution procedures follow the Chelex-based method previously described (Nelson et al. 2006). To the washed beads, 50 µl of 10% Chelex (w/v, in water) solution was added. Also, 0.5 µl of the original sonicated chromatin was added to 49.5 µl of 10% Chelex to prepare 1% input DNA solution. Chromatin-DNA complexes were reverse-crosslinked by incubating the beads-Chelex solution at 100 oC for 10 min. Then, 0.5 µl of protease K (Invitrogen) was added to each tube and incubated at 43oC for 1 hour. After the protease K treatment, the enzymes were deactivated by incubating the tube at 100 ^{o}C for 10 min. The supernatant (approximately 35 µl recoverable without taking Chelex bed) containing eluted DNA was analyzed by quantitative PCR as described above. No antibody was added to negative control samples (NoAb).

3.3 Results

Up-regulation of AT1G61280 in $CPL4_{RNAi}$ lines is coupled with 3'-extension from upstream U12 snRNA gene

Our previous microarray study of $CPL4_{RNAi}$ detected three major co-regulation clusters of genes affected by silencing CPL4 expression. Interestingly, however, the most upregulated gene in $CPL4_{RNAi}$ (AT1G61280: putative GPI19/PIG-P subunit), did not show any co-regulation pattern with the major gene clusters. The Expressed Sequence Tags (GenBank accession DR367811.1 and DR384064.1) for AT1G61280 revealed a unique transcript structure that contains a whole U12 snRNA at its 5'UTR. Apparently, transcription of AT1G61280 can originate from a transcription start site of U12 snRNA gene (AT1G61275) locating 290 bp upstream of AT1G61280. This suggested that upregulation of AT1G61280 in $CPL4_{RNAi}$ is due to alteration of U12 snRNA transcription regulation. To test if $CPL4_{RNAi}$ indeed produces extended U12 snRNA, i.e., U12-AT1G61280 fusion transcript, I conducted RT-qPCR analysis targeting three intergenic regions between U12 snRNA and AT1G61280 (Figure 3.1a-c). In wild type, the only transcript detected in this region is mature U12 snRNA (Figure 3.1b). In contrast, in $CPL4_{RNAi}$, transcripts were detected throughout the intergenic region and A1G61280. In addition, the continuous U12-AT1G61280 transcript was detected by RT-PCR. These results suggested that CPL4 regulates expression of AT1G61280 via a non-conventional regulatory mechanism, extending 3' end of upstream snRNA transcripts.

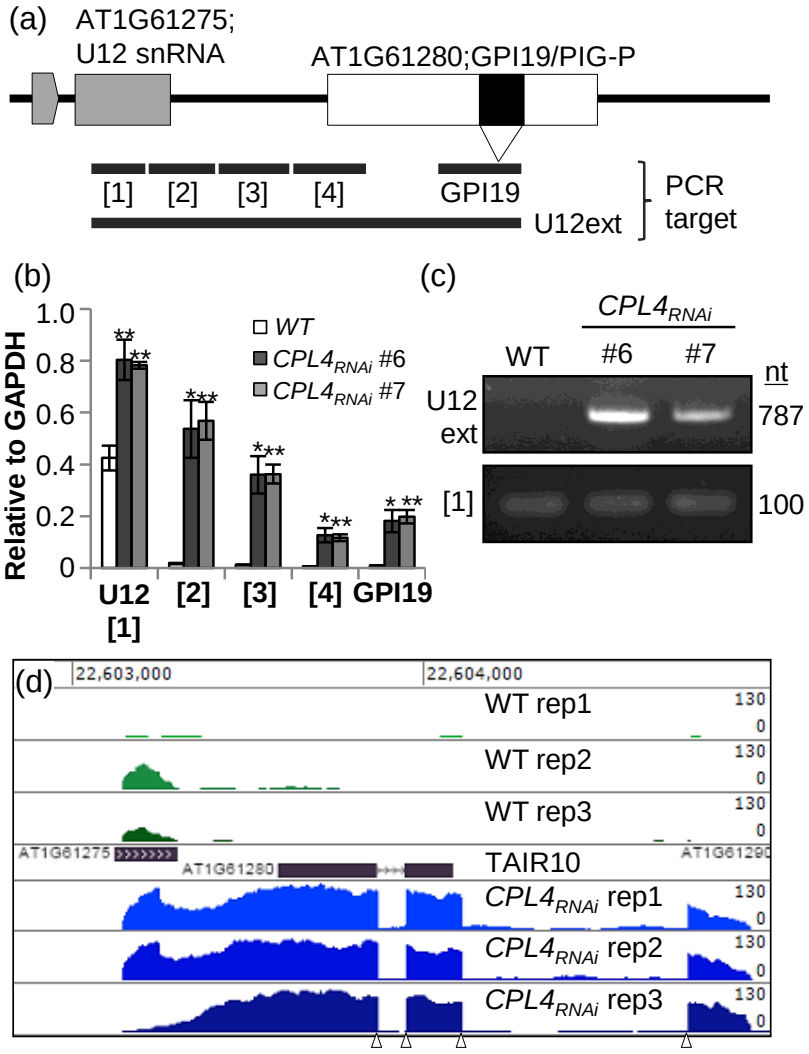


Figure 3.1 Up-regulation of AT1G61280 in *CPL4$_{RNAi}$* is coupled with 3'-extension from upstream U12 gene.

(a) RT-qPCR targets in the genomic region corresponding to AT1G61275 (U12 snRNA) and AT1G61280 (GPI19/PIG-P). Blue bars represent resulting amplicon. (b) Expression level of each fragment described in (a) relative to GAPDH, measured by RT-qPCR [$2^{(\text{Cp of GAPDH - Cp of target})}$]. Asterisks indicate significant difference from WT, one-tailed Student's t-test $p<0.05$ (*) or $p<0.01$ (**). (c) PCR analysis detecting the 3'-extended fragment from U12. (d) Read coverage over AT1G61275-AT1G61280 locus in WT and *CPL4$_{RNAi}$* lines. Coverage values normalized by library size (total amount of mapped read) and associated bigWig track formats for visualization in Trackster viewer were generated by BEDTools and Wig/BedGraph-to-bigWig converter on Galaxy server, respectively.

Removing 3'-box from U12 enhances downstream protein-coding gene expression

The U12 snRNA region of the U12-AT1G61280 fusion transcript likely form secondary structures, and may prevent translation of downstream ORF. To test if the U12-AT1G61280 fusion transcript is translatable like mRNA, we prepared a reporter gene using a 1467-bp genomic fragment, spanning from 981 bp upstream of AT1G61275 (U12) to 22 bp downstream of ATG codon of AT1G61280, fused to firefly luciferase (LUC) coding region (Figure 3.2a). Furthermore, the role of regulatory sequence in U12 USE and 3'-box were tested by introducing mutations in the reporter gene. In the Col-0 host plants, the LUC activity from WT reporter which contains both intact USE and 3'-box was detectable but weak. By contrast, the 3m reporter with compromised 3'-box showed highest LUC activities, which were abolished by additional mutations in USE (m3m, Figure 3.2b&c). These results show that the U12-LUC2 expression is largely dependent on the snRNA promoter and snRNA 3'-extension can enhance downstream protein-coding gene expression.

The impact of $CPL4_{RNAi}$ on U12-LUC2 reporter genes was tested by crossing the reporter lines to $CPL4_{RNAi}$. Interestingly, $CPL4_{RNAi}$ showed very strong induction of wild-type U12-LUC2 reporter gene (Table 3.1). The expression level was even stronger than that of the Col-0 U12-LUC2(3m) lines. Expression of the 3m reporter was also enhanced in CPL4_{RNAi} background, similar to WT U12-LUC2 (fold-induction 140.55 and 52.67, respectively), whereas the expression levels of the U12-LUC2 variants with USE mutation (m and m3m; fold induction 6.14 and 2.21, respectively) were not enhanced substantially in $CPL4_{RNAi}$ background. This indicates that $CPL4_{RNAi}$ promotes

not only read-through of snRNA 3'-box, but also the continuation of read-through transcription at the downstream of 3'-box.

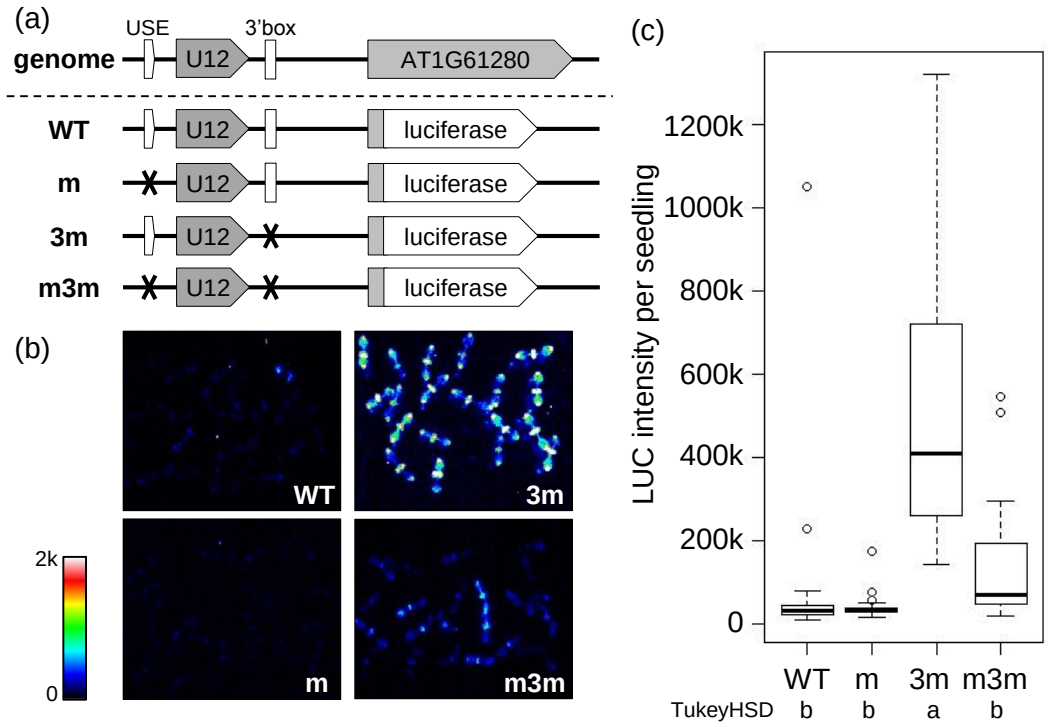


Figure 3.2 Downstream protein-coding gene expression is facilitated by 3'box mutation.

(a) Design of U12-LUC2 reporter system. AT1G61280 coding region is replaced with luciferase. Upstream sequencing element; USE. (b) Luciferase activity of 1-week old seedlings. A representative line with median luc intensity is chosen for each construct. 20-min exposure. (c) Boxplot showing mean luc intensity per seedling of independent U12-LUC2 lines. # of independent lines examined is 21, 24,17, 24 for WT, m, 3m, m3m, respectively. For each line, 18 to 32 seedlings were counted. See Materials and Methods for detail.

Table 3.1 Effects of *CPL4*$_{RNAi}$ on U12-LUC2 reporter expressions

U12-LUC2 Genotype	n	LUC activity count per seedling (million)		*p*-value (t-test)
		AVE	SE	
U12_WT	18	0.078	± 0.010	-
U12_WT x *CPL4*$_{RNAi}$ F1	13	10.908	± 0.921	6.02E-08
U12_3m	14	2.589	± 0.161	-
U12_3m x *CPL4*$_{RNAi}$ F1	10	136.375	± 30.433	1.73E-03
U12_m	14	0.021	± 0.001	-
U12_m x *CPL4*$_{RNAi}$ F1	8	0.131	± 0.011	2.50E-05
U12_m3m	17	0.086	± 0.010	-
U12_m3m x *CPL4*$_{RNAi}$ F1	17	0.191	± 0.015	2.95E-06

LUC activity in individual F1 seedlings from the cross between U12-LUC2 transgenic lines and *CPL4*$_{RNAi}$ line was quantified. See Fig. 2a for the reporter gene designs.

Pol II-dependent snRNA loci show 3'-extension in CPL4$_{RNAi}$ *lines*

To test if 3'-extension caused by *CPL4*$_{RNAi}$ is limited to U12 snRNA or shared by other snRNAs, RNA-seq analysis of *CPL4*$_{RNAi}$ transcriptome was conducted. The sequencing cDNA libraries containing cDNA from ncRNAs were prepared using the rRNA depletion protocol. As shown in Figure 3.1d, high level of RNA-seq reads were continuously mapped to the region spanning U12 and AT1G61280, consistent with the RT-PCR analyses. The reads coverage spans entire AT1G61280 coding region and 3'-UTR, supporting the idea that this transcript retains translatability. To analyze of all snRNA transcripts in Arabidopsis genome, RNA-seq reads mapped at 75 snRNA loci (Wang and Brendel 2004) were inspected individually. Among the 75 snRNA loci, 56 of them showed appreciable expression in WT (depth of coverage per base per million reads mapped (dcpm) higher than 0.1). Based on the spacing between USE and TATA-

box in their promoter, 48 and 8 snRNA loci are pol II- and pol III-dependent class, respectively. atU6-5 is not clear because of 26-nt spacing and non-canonical TATA box structure compared with other U6s. It also lacks a conserved T-stretch signal at its 3'-end (Figure 3.3a).

I determined levels of transcript mapped to downstream (50nt) of each of the 56 snRNA loci (Figure 3.3a; highlighted in gray). The analysis revealed that $CPL4_{RNAi}$ shows significantly higher coverage in the 3'-extension regions of Pol II-dependent snRNAs, but not in those of Pol III-dependent U6 snRNA loci (Figure 3.3b). The $CPL4_{RNAi}$ showed 3'-extension in all class of major snRNAs (Figure 3.3c-f; U1, U2, U4 and U5), and minor snRNAs (U12, U4atac, U4atac2, U11). The 3'-extension of select snRNA loci were independently confirmed by RT-PCR (Figure 3.4). These results showed that knockdown of CPL4 globally promoted pol II to escape transcriptional termination or 3'-end processing of snRNA, leading to accumulation of 3'-extended, read-through transcripts.

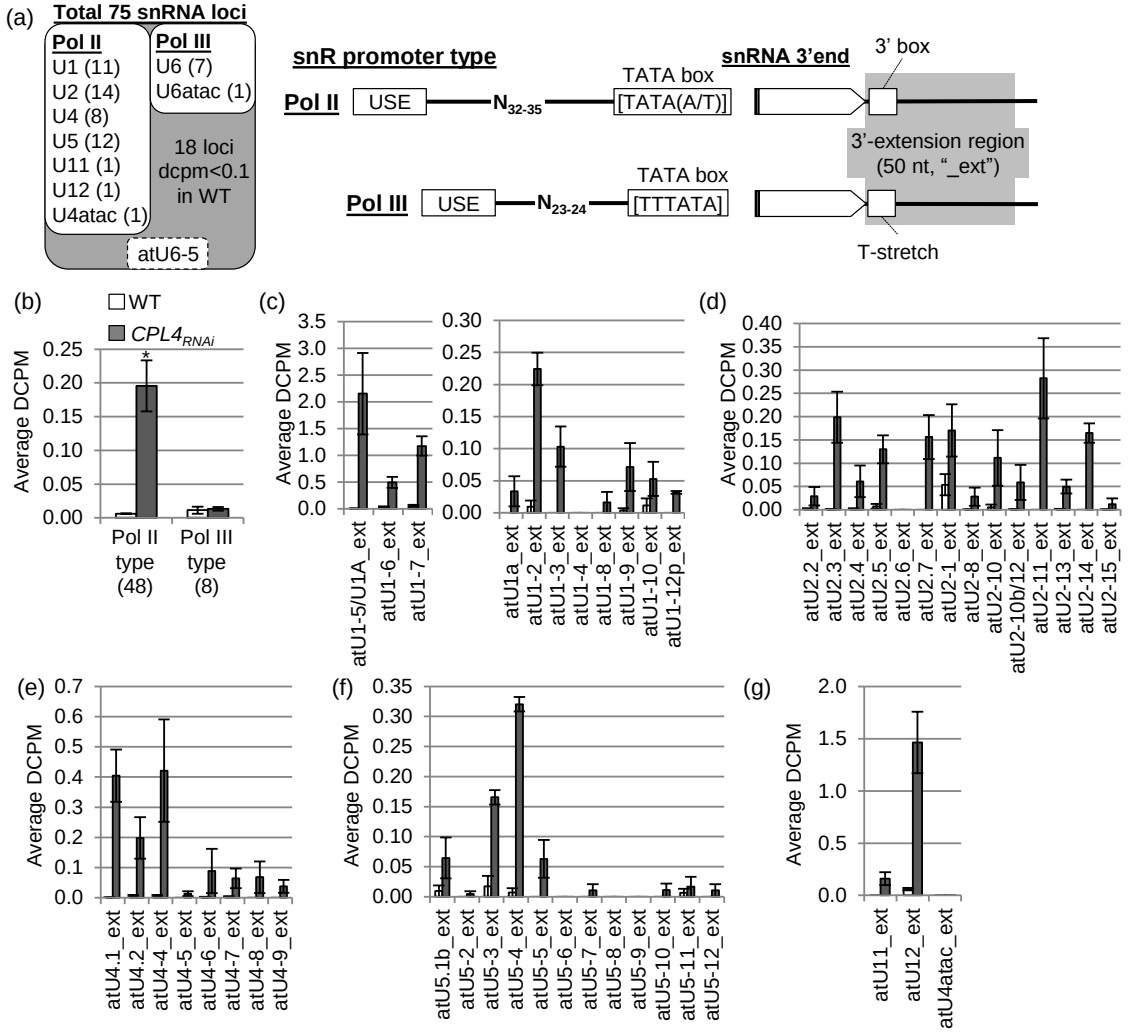


Figure 3.3 Detection of 3'-extended snRNA transcripts in *CPL4$_{RNAi}$* line.

(a) Expression level of snRNA loci measured as depth of coverage per base per billion reads mapped (DCPM) was examined for 75 snRNA loci. Among those, 57 loci with DCPM higher than 0.1 in WT were selected for further analysis. RNA polymerase specificity of each loci was determined based on the spacing between two core promoter elements known in Arabidopsis, which are USE; upstream sequence element and TATA box. The 3'box signal and poly-T stretch sequences serve as snRNA transcription termination/3'-processing signal for Pol II and Pol III, respectively. A U6 snRNA, atU6-5, has a non-canonical spacing between USE and TATA box and lacks T-stretch signal in its 3'-end, thus was removed from further analysis. (b-h) Depth of coverage of the 3'-extended regions of 56 snRNAs selected in *CPL4$_{RNAi}$*. (normalized by total number of mapped reads in million, Depth of Coverage per base per million; DCPM). (b) Average DCPM of all snRNA-extensions examined. An asterisk indicates $p<0.05$, two-tailed student's t-test. (c) U1 snRNAs, (d) U2, (e) U4, (f) U5, (g) U11, U12 and U4atac minor snRNAs. All error bars represent standard error from biological triplicate.

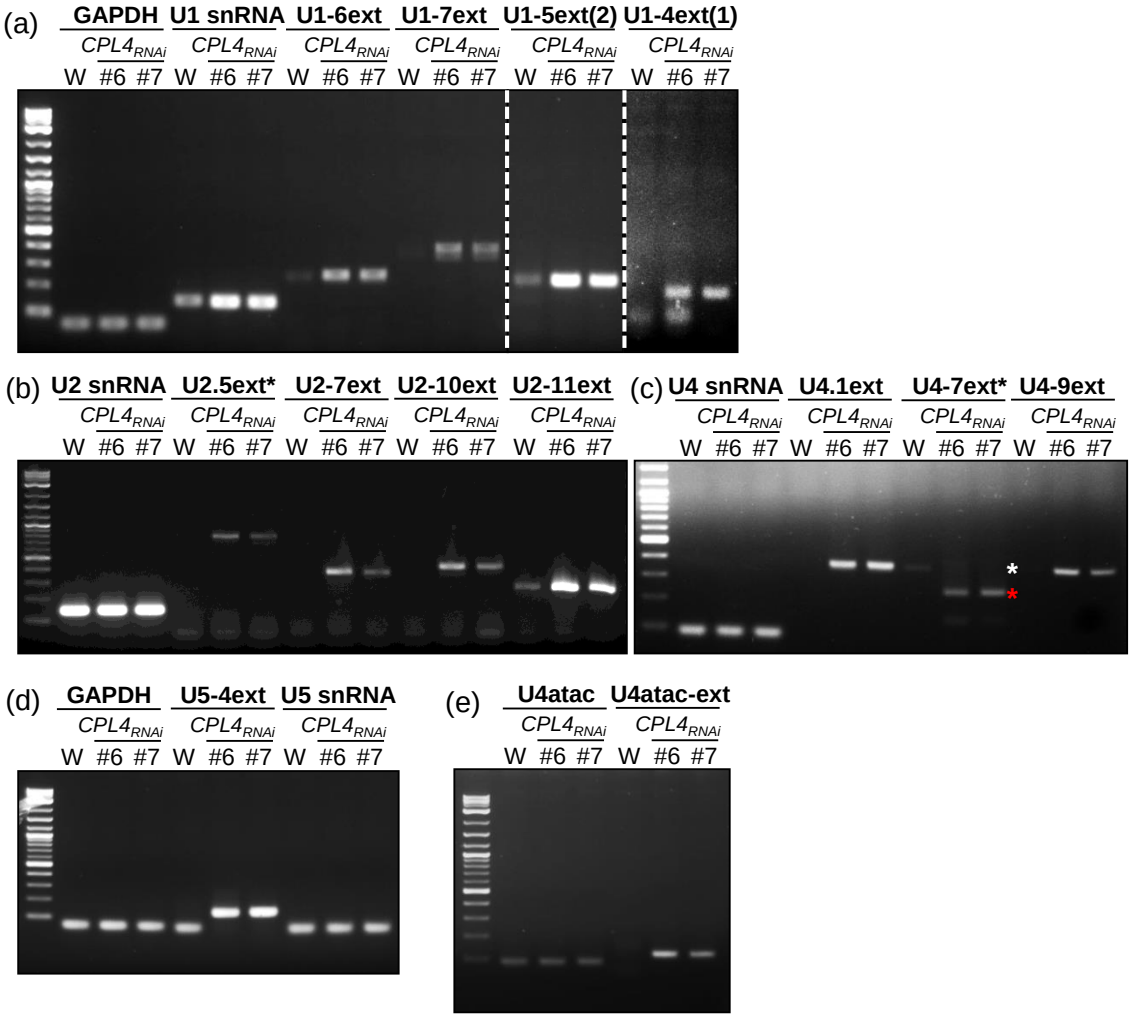


Figure 3.4 Detection of 3'-extended snRNA transcripts in *CPL4$_{RNAi}$* line by RT-PCR.

Random-primed cDNA library of total RNAs from wild type (W) and *CPL4$_{RNAi}$* lines were analyzed by PCR with primers targeting snRNA and its 3'-extended region. (a) U1, (b) U2, (c) U4, (d) U5 and (e) U4atac snRNA extensions of select loci were analyzed. White and red asterisks indicate the expected size of amplicon and actual products (c). Based on RNA-seq coverage, the discrepancy is likely due to splicing events within the amplicon. For U2.5ext detection, the reverse primer is designed to span the junction of 1st and 2nd exon of the downstream protein-coding gene.

Chimeric snR-DPG transcripts accumulate in $CPL4_{RNAi}$ lines

Among 48 pol II-dependent snRNAs examined, 18 of them have a downstream protein coding gene (DPG) within 1 kbp downstream of the snRNA annotation (Table 3.2). The majority of the identified DPGs are up-regulated in the $CPL4_{RNAi}$ line (Table 3.2). The coding sequences of DPG are properly spliced as exemplified in the U2.5-AT5G09590 chimeric transcript (Figure 3.4b), indicating that extended snRNAs enter the mRNA maturation pathway. To test if upregulation of DPG in $CPL4_{RNAi}$ is indeed due to the snRNA extension or to activation of overlapping mRNA promoter, I conducted a detailed characterization of *LONG-AFTER FAR-RED 3* (*LAF3*) locus (AT3G55850). Notably, *LAF3* locus produces two mRNA isoforms with or without embedded snRNA. Isoform1 (*LAFISF1*: GenBank accession AY295343.1, BX823543.1) sequence starts with atU5.1b snRNA, whereas isoform2 (*LAFISF2*) starts 223-bp downstream of the U5 snRNA sequence (Figure 3.5a). The isoform structures and their distinct transcription start sites indicate that expression of LAF3 isoforms are driven by two independent promoters; ISF1 expression is driven by snRNA promoter, and ISF2 by standard mRNA promoter. In $CPL4_{RNAi}$ background, the chimeric *LAF3ISF1* specifically over-accumulated while *LAF3ISF2* showed no considerable up-regulation (Figure 3.5b-c), indicating that at least in LAF3 locus, the snRNA-extension in $CPL4_{RNAi}$ affects only on the U5-snRNA-fused *LAF3ISF1*, without affecting the transcription of snRNA-independent *LAF3ISF2*.

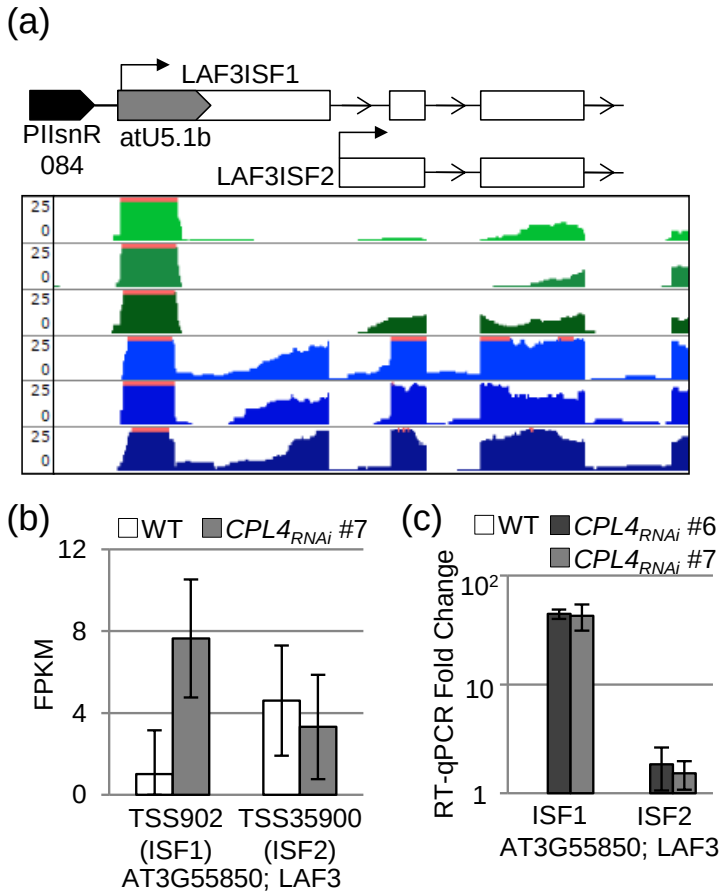


Figure 3.5 Expression pattern of U5 snRNA-fused LAF3ISF1 in *CPL4$_{RNAi}$*.
(a) A representative mapping coverage track of the AT3G55850; LAF3 genomic region from WT (upper) and *CPL4$_{RNAi}$* (lower) RNA-seq results. Red bars represent ISF1 specific features (i.e. atU5.1b snRNA and 1st intron). (b) FPKM of LAF3ISF1 (TSS902) and LAF3ISF2 (TSS35900) in WT and *CPL4$_{RNAi}$*, computed by Cuffdiff. Bars indicate confidence intervals. (c) Increase of ISF1 expression in *CPL4$_{RNAi}$* lines measured by RT-qPCR. Forward primers designed to detect specific isoform and a reverse primer spanning 2nd intron-exon junction were used.

Table 3.2 Eighteen pol II-dependent snRNA loci with proximal DPG on the same strand

snR	snR-3'ext level	snR-DPG spacing (bp)	DPG AGI	DPG description	RNA-seq FC(log2)	q_value
atU12	High	298	AT1G61280	GPI19/PIG-P subunit	6.30	0.11
atU2-11	moderate	120	AT1G09800	Pseudouridine synthase family protein	2.58	0.01
atU1-2	moderate	-170	AT4G23420	NAD(P)-binding Rossmann-fold superfamily protein	-2.42	0.00
atU2-1	moderate	272	AT1G16825	Reticulon family protein	0.72	0.10
atU5-3	moderate	689	AT1G70190	Ribosomal protein L7/L12	0.28	0.32
atU2.7	moderate	463	AT5G61450	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0.54	0.02
atU2.5	moderate	340	AT5G09590	HSC70-2	1.05	0.00
atU1-3	moderate	396	AT5G51680	hydroxyproline-rich glycoprotein family protein	n/a	1.00
atU4-6	low	129	AT1G11870	SRS; Seryl-tRNA synthetase	-0.30	0.14
atU5.1b	low	-116	AT3G55850	LAF3ISF2	0.96	0.00
atU2.4	low	61	AT3G56820	-	0.89	0.07
atU2-13	very low	60	AT2G20410	RNA-binding ASCH domain protein	0.65	0.03
atU4-9	very low	802	AT1G79970	-	-0.13	0.56
atU2-8	very low	455	AT5G67560	ATARLA1D; ADP-ribosylation factor-like A1D	0.11	0.61
atU4-5	very low	337	AT5G25770	α/β-Hydrolases superfamily protein	0.45	0.29
atU5-7	very low	476	AT4G02530	-	0.35	0.04
atU5-6	nd	1045	AT1G04470	unknown function (DUF810)	n/a	1.00
atU5-9	nd	854	AT1G79540	-	0.20	0.45

snR-3'ext level is defined as snR-ext region dcpm>1.0, High; 1.0>dcpm>0.1, moderate; 0.1>dcpm>0.05, low; dcpm<0.05, very low; no 3'-extension, nd. snR-DPG spacing column shows distance in bp between the last base of snRNA and the first base of DPG; negative values indicate the snRNA is embedded in the DPG. Fold-change (log2) of DPG and associated q-values are obtained by Cuffdiff.

A transposon-embedded PIIsnR drives AT1G20320(SSP14) locus to produce a 3'-extention-capable short unstable ncRNA

The above results suggested that CPL4 regulates a suite of protein-coding genes that are driven by pol II-dependent snRNA promoters (PIIsnR). Next, I tested if there are protein coding genes without snRNA sequence but are regulated by PIIsnR. I first located all the pol II-dependent snRNA promoters in the *Arabidopsis thaliana* genome. Using Find Individual Motif Occurrences (FIMO) algorithm in MEME suite with USE(RTCCCACATCG)-N_{32-35}-TATAA as input motifs (Bailey et al. 2009; Grant et al. 2011), 150 Pol II-dependent snRNA promoter motifs (PIIsnR) were identified in Arabidopsis TAIR10 genome sequence (Figure 3.6a). The phylogenetic clustering analysis of identified motifs revealed that 77 of them show high similarity to each other and are part of repeat sequences related to transposable elements (TE), therefore classified as "Repeat". Most of the identified TEs are helitron-type non-autonomous DNA transposons, ATREP5, or its fragments, along with ATREP1, ATREP3 and Helitrony (Figure 3.6b and Table 3.3). Indeed, PIIsnR sequences can be found in the consensus sequence of ATREP5 and Helitrony1D (Bao et al. 2015); PIIsnR motifs are also found in consensus sequences of TEs such as LTR and Gypsy from other green plants (Table 3.3). Forty-nine "snRNA" motifs are located upstream of known snRNA loci. The remaining 24 motifs not associated with either "repeat" or "snRNA" categories were classified as "novel", which included one protein-coding gene AT1G20320. Expression levels of these loci were inspected in RNA-seq profile (ribo-zero protocol) of Col-0 plants. Outside of snRNA class, most of loci that belong to repeat or novel classes

did not produce significant levels of transcripts except 3 snoRNAs in the "novel" class. This could be due to promoter defects and/or the transcriptional gene silencing. Alternatively, relatively short and/or unstable transcripts could be difficult to detect by the standard RNA-seq approach. To improve detection, the RNA-seq strategy was modified. First, to stabilize short-lived ncRNA, a *hen2-4* mutant, which is defective in nuclear exosome pathway for RNA degradation was used as a host. Second, RNA-seq protocol was modified to sequence intermediate-length ncRNA (im-ncRNA). Briefly, RNA-seq libraries were prepared using modified ribo-zero protocol without any RNA fragmentation steps, and cDNA corresponding to 100-500 nt transcripts were sequenced in a strand-specific manner using the paired-end protocol. The im-ncRNA sequencing protocol improved detection of PIIsnR-driven transcripts; in particular, the im-ncRNA profile for *hen2-4* revealed production of transcripts from nine out of 24 "novel" loci (Figure 3.6a). These loci include three monocistronic snoRNAs (Marker et al. 2002), 2 MIRNAs, a long ncRNA (AT4NC021500) and a protein-coding AT1G20320.

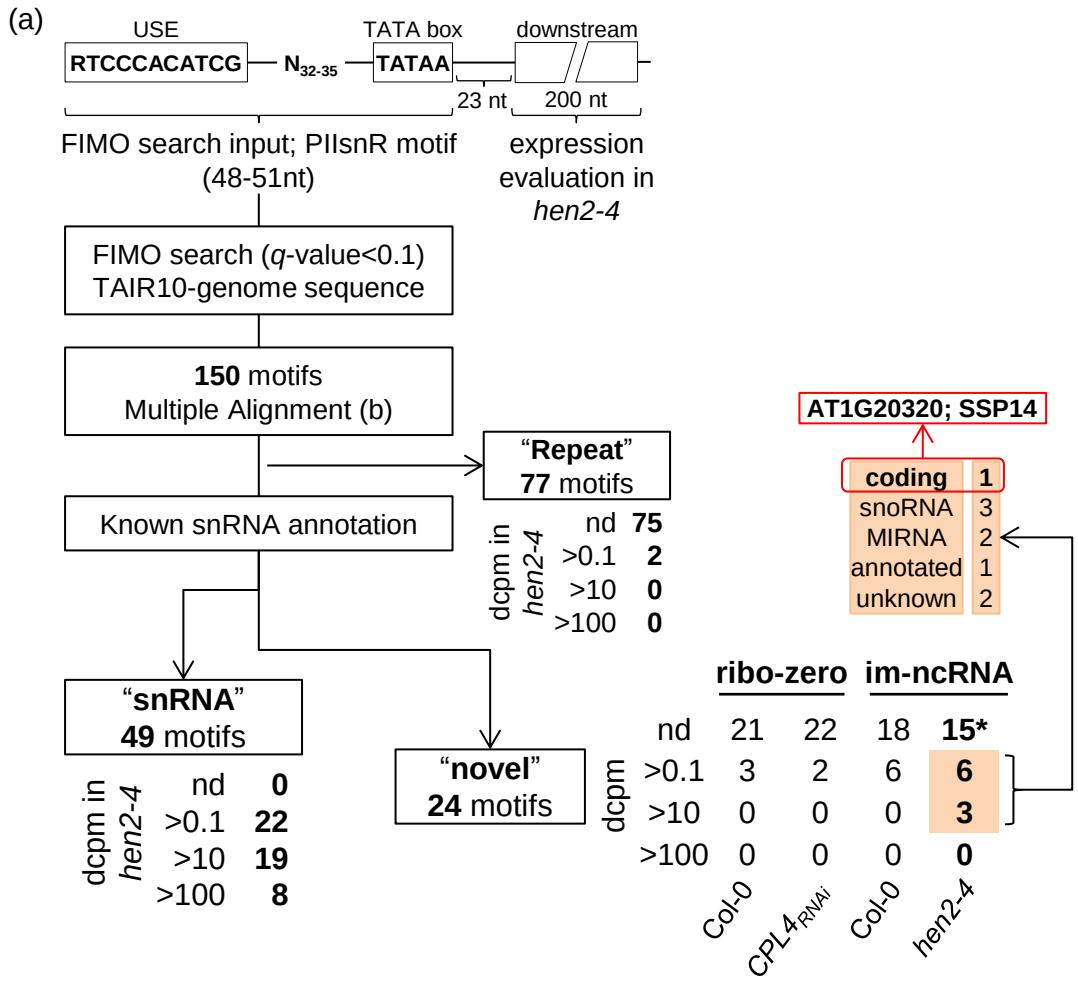


Figure 3.6 Identification of loci potentially regulated by Pol II-dependent snRNA promoter in Arabidopsis.

(a) Strategy for identifying loci potentially regulated by pol II-dependent snRNA promoter. A motif containing two core promoter elements for snRNA transcription (upstream sequence element; USE and TATA box) spaced by 32-35 nt sequence was used as an input in FIMO search against Arabidopsos TAIR10 genome. The downstream 200 nt region starting at 24 nt downstream of the TATA box was used to evaluate the expression of each loci in *hen2-4* mutant size-selection RNA-seq (Depth of Coverage per base Per Million reads mapped; DCPM, see Experimental Procedure). *PIIsnR090, 039, 070 in "Other" category show average dcpm higher than 0.1, but reads are mapped on the opposite strand therefore are considered as nd. (b) continues to next page.

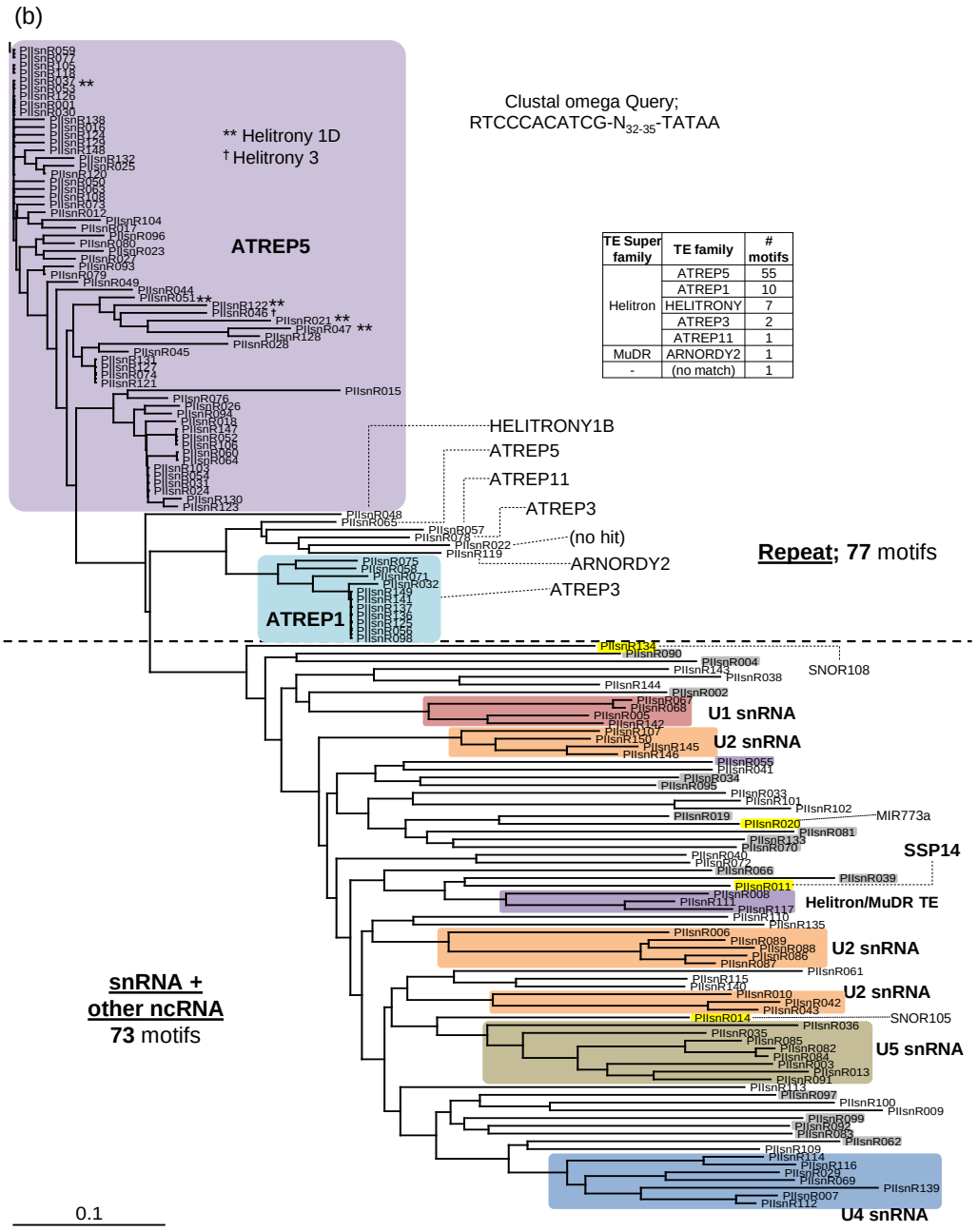


Figure 3.6 continued. (b) the 150 motifs (48-51nt long) were subsequently subjected to multiple alignment using Clustal Omega. The alignment result was presented as a phylogram by using Tree View. A group of 77 motifs shows extensive similarity to each other (above center line). Based on this nature, the motifs are categorized as "Repeat". The individual "Repeat" motifs were then used as a query for FIMO search against 31,189 transposable element annotations from TAIR10. For each of the other 73 motifs left, corresponding locus was manually examined and assigned if any. The scale bar shows 0.1 nucleotide substitution per motif. Clades consisting of more than three motifs associated with the same downstream locus category were highlighted.

Table 3.3 Consensus sequences of green plant TEs bearing PIIsnR motif

Sequence Name	Species	TE superfamily	Str	Start	End	q-value	Matched Sequence (RTCCCACATCG-N32-35-TATAA)
Gypsy-140_SBi-LTR	Sorghum bicolor	LTR	+	468	516	1.4E-03	GTCCCACATCGCCTGTCCAGAAGAGGTGGGAGGCTTTTCTAGGCTATAA
Gypsy-29_Mad-LTR	Malus domestica	LTR	+	177	224	2.7E-03	GTCCCACATCGACCGCGGACAAAAGCTGGAGACTCTCCTCAACTATAA
HARB-1N1_Mad	Malus domestica	Harbinger	+	467	515	3.1E-03	GTCCCACATCGGCTGTGGGAGAGGTTTGAGCAATCAAACATGCTTATAA
Gypsy-3_Pru-LTR	Prunus persica	LTR	+	154	204	6.2E-03	GTCCCACATCGGAAATATGAGCACAGTGCACACCTCCCAAGGCCTATATAA
Gypsy-5_Pru-LTR	Prunus persica	LTR	+	179	227	3.3E-03	GTCCCACATCGGAACTTTGTGCAAACCTCCTACTTTCACCTCCCTATAA
Gypsy-9_AIp-LTR	Arachis ipaensis (Peanut)	LTR	+	88	136	1.4E-03	GTCCCACATCGCCTAATACTCGAAGGCTCCCCCTCCCCTACTAGTATAA
Copia-38_JC-LTR	Jatropha curcas	LTR	+	19	67	3.7E-03	ATCCCACATCGAAAGAAAGGGAAGGGAATAGGGAGTTGTTTGGCTATAA
ATREP5	Arabidopsis thaliana	Helitron	+	1045	1094	1.0E-02	GTCCCACATCGCTTAAAAAAATTGGACAATGGTCAAGAGCCATACTATAA
ATREP5	Arabidopsis thaliana	Helitron	+	481	530	1.0E-02	GTCCCACATCGCTTAGAAAAATTGGACAATGGTTCAGACCCATATTATAA
HELITRONY1D	Arabidopsis thaliana	Helitron	+	430	479	1.0E-02	GTCCCACATCGCTTAAAAAAATTGGACATTGGTTCAGAGCCATACTATAA
Helitron-6_ALy	Arabidopsis lyrata	Helitron	-	10006	10056	1.8E-02	ATCCCACATCGGGACGGTTGACTAAAATAAATCACTTACGTTTAGATATAA

Interestingly, reads mapped to AT1G20320 locus in *hen2-4* were limited to the 5'-end region (Figure 3.7a). AT1G20320 is an intronless gene encoding a SCP1-like small phosphatase family protein, SSP14 (Koiwa 2006). The PIIsnR promoter of AT1G20320 overlaps with AtREP5 helitron fragments, and is flanked by another DNA helitron AtREP3 in an opposite direction (Figure 3.8), suggesting this region underwent rearrangements of TEs during evolution. The protein coding sequence shows no similarity to any of the U1, U2, U4, U5 and U12 snRNAs but possesses a 3'-box like-sequence near the end of the ncRNA reads detected in *hen2-4* (Figures 3.7, 3.8). Interestingly, expression of full-length AT1G20320 transcripts was specifically detected in $CPL4_{RNAi}$ seedling but not in Col-0, suggesting the production of AT1G20320 is regulated at the level of ncRNA extension by snRNA termination mechanism. Interestingly, in ThaleMine RNA-seq Expression (source – Araport; https://apps.araport.org/thalemine/) the expression pattern of AT1G20320 in wild-type plants was strictly limited to pollens where *CPL4* expression was the lowest and the other pol II-dependent snRNA loci also showed 3'-extensions (Figure 3.9) (Loraine et al. 2013; Krishnakumar et al. 2015). Because the AT1G20320 and snR-DPG transcripts are similarly regulated by CPL4, hereafter, I collectively refer both snR-DPG and AT1G20320 transcripts as snR-DPG transcripts unless specified. ncRNA produced from At1g20320 was termed $ncRNA_{SSP14}$.

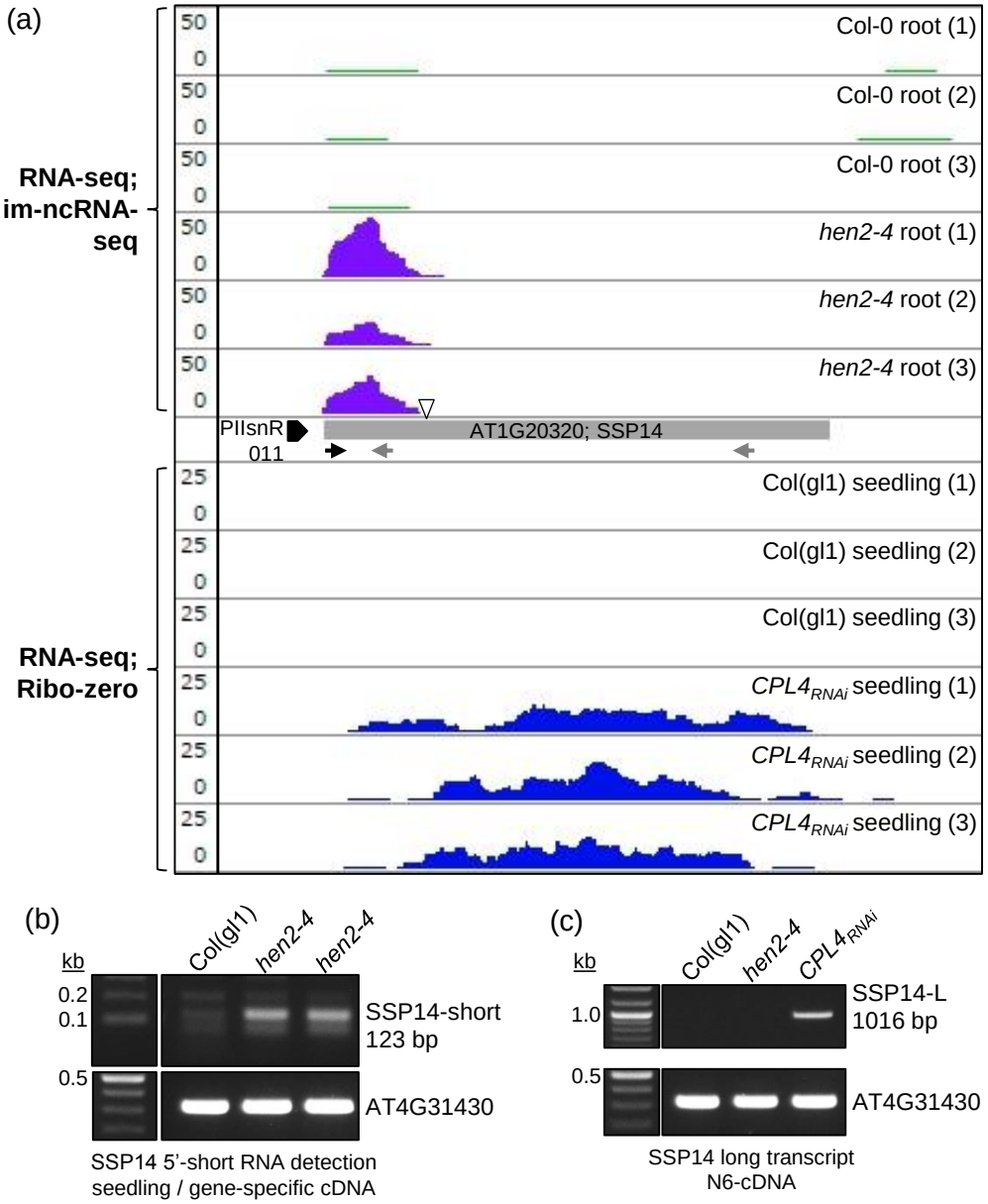


Figure 3.7 SSP14 locus produces unstable short transcripts in WT and long transcript in *CPL4$_{RNAi}$*.

(a) RNA-seq coverages. See Materials and Methods for the im-ncRNA-seq procedure. (b) 5' short RNA detection by RT-PCR from gene-specific cDNA (c) long transcript detection by RT-PCR from N6-primed cDNA. a triangle indicates presence of 3'-box-like sequence. See Figure 3.8 for detailed features on the promoter and 5'-region of this locus.

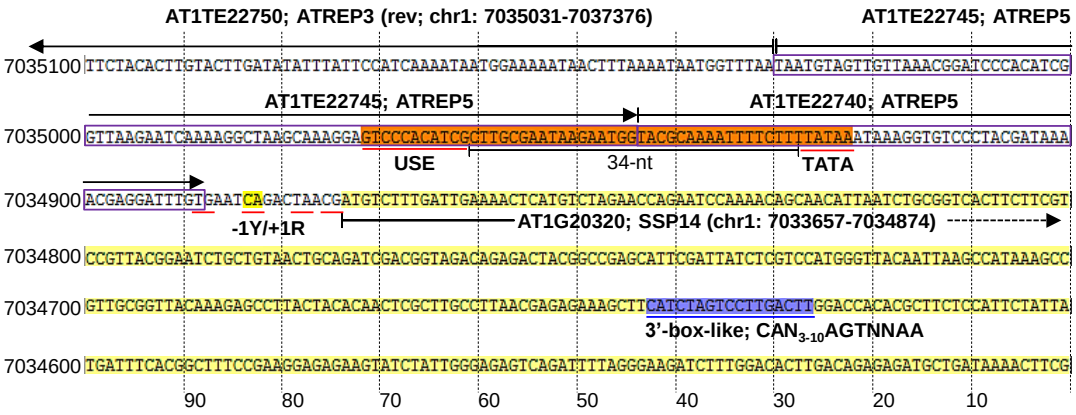


Figure 3.8 Features on TE-overlapping PIIsnR promoter region and 5'-coding sequence of AT1G20320; SSP14.

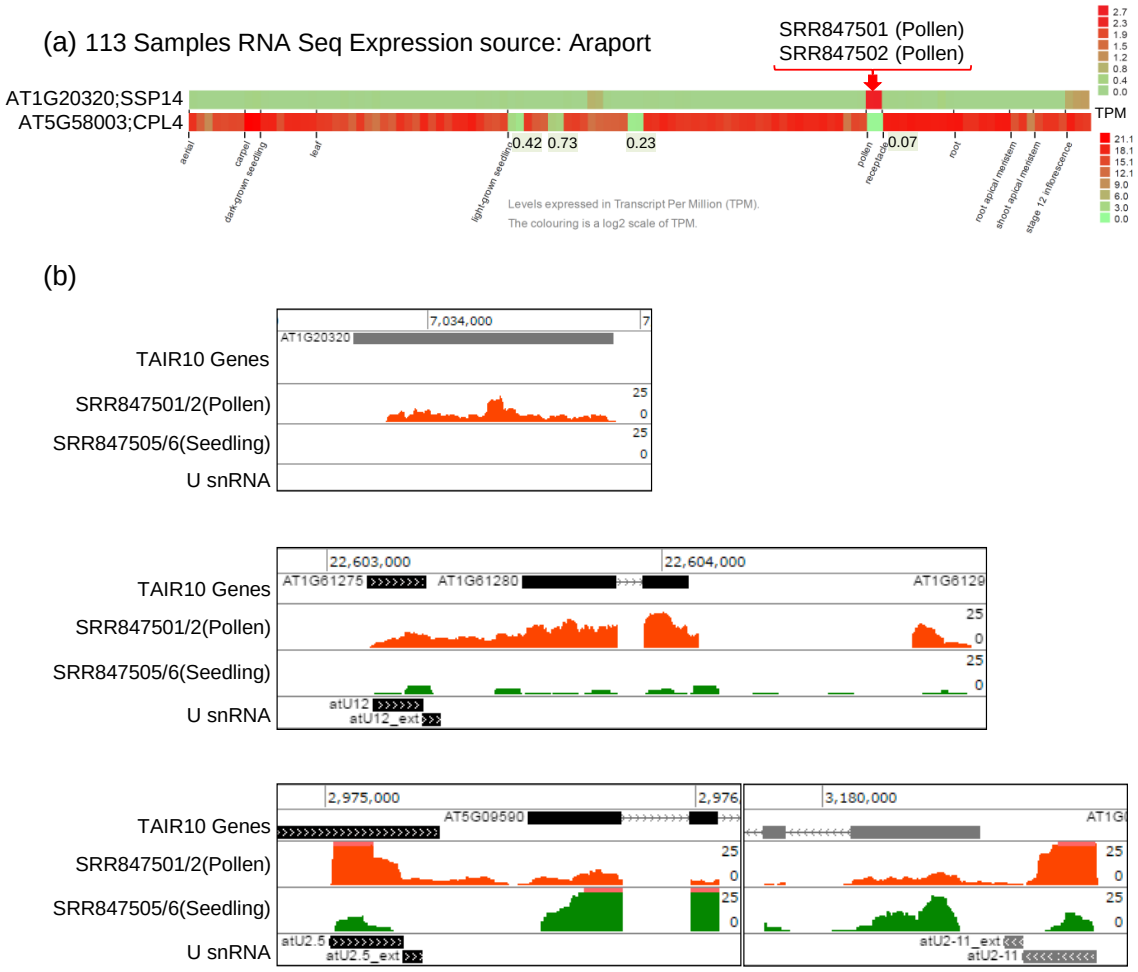


Figure 3.9 SSP14 expression and snRNA 3'-extensions in pollen.
(a) Expression pattern of SSP14 and CPL4 in 113 RNA-seq Exoression available at Araport (https://apps.araport.org/thalemine/). The heat maps show expression levels in Transcripts Per Million (TPM). Four numbers below CPL4 heat map indicate the TPM value of corresponding weak expression spot. (b) Mapping coverage tracks of select snR-DPG loci. Orange and green tracks represent pollen (SRR847501/2) and seedling (SRR847505/6) experiments from SRP022162 RNA-seq study.

snR-DPG transcripts are widely spread in other plant species

PIIsnRs are embedded in some transposons in plants, implying that snR-DPG loci with the potential to produce fusion transcripts may be widespread in plant species other than *Arabidopsis thaliana*. BLAST search using Arabidopsis snRNA sequences against RefSeq mRNA database detected 150 potential snR-DPG transcripts in wide-range of plant species including dicots and monocots, and one species of fish. EST clones spanning snRNA-sequence and downstream protein coding region were found for 17 genes (Table 3.4), establishing production of snR-DPG transcripts from these loci. Combined with the observation that PIIsnR sequences embedded in transposable elements in other green plants, these findings indicate that snR-DPG loci can frequently occur in plants. A few combinations of snR-DPG seem to be conserved, such as U12-GPI19 and U1/2-cytochrome c oxidase subunit 5b-1, while others are unique in each plant species.

Table 3.4 snRNA-DPG transcripts supported by EST spanning snRNA-DPG

Query	GeneID	Refseq (hit)	Species	EST spanning snRNAext region	downstream protein-coding gene
AtU1	828441	NM_118471	*Arabidopsis thaliana*	EG447525.1	NAD(P)-binding Rossmann-fold superfamily protein
AtU5	824752	NM_115444	*Arabidopsis thaliana*	EG419000.1	LAF3-duplicate (upstream of LAF3, flanked by TE)
AtU12	106450998	XM_013892844	*Brassica napus*	ES911942.1	GPI19
AtU2	103851809	XM_009128676	*Brassica rapa*	DY013485.1	F-box/FBD/LRR-repeat protein (AT5G56420-like)
AtU4	18055013	XM_006450628	*Citrus clementina*	DY288252.1	hypothetical protein
AtU1	103492240	XM_008452530	*Cucumis melo*	JG490974.1	N-acetyltransferase p20-like
AtU1	103499956	XM_008463125	*Cucumis melo*	JG468085.1	F-box/WD-40 repeat-containing protein (AT3G52030-like)
AtU2	103494723	XM_008456053	*Cucumis melo*	JG501029.1	cytochrome c oxidase subunit 5b-1
AtU1	4349283	XM_015757279	*Oryza sativa*	CT861291.1	arsenate reductase 2.1
AtU1	9271734	XM_015777263	*Oryza sativa*	AU173873.1	cytochrome c oxidase subunit 5b-1
AtU2	4343789	XM_015789462	*Oryza sativa*	CI285625.1	cytochrome c oxidase subunit 5b-1
AtU2	4328372	XM_015772229	*Oryza sativa*	CB682952.1	proline-, glutamic acid- and leucine-rich protein 1
AtU2	4328599	XM_015771499	*Oryza sativa*	CI711653.1	SCO1 homolog 1, mitochondrial
AtU1	100278403	NM_001329196	*Zea mays*	FL163030.1	hypothetical protein
AtU4	100502310	NM_001196788	*Zea mays*	DV511687.1	cysteine synthase
AtU1	100832877	XM_010236666 /XM_003574239	*Brachypodium distachyon*	DV482065.1	arsenate reductase 2.1
AtU4	103186729	NM_001292254	*Callorhinchus milii*	JK858370.1	serine/threonine-protein phosphatase 2A catalytic subunit beta isoform

Potential snR-DPG transcripts bearing snRNA sequences on their 5'-end are retrieved from NCBI database by BLAST search (see Materials and Methods). Transcripts with corresponding EST are shown. See Materials and Methods for the search procedure.

Accumulation of snR-DPG transcripts in CPL4$_{RNAi}$ depends on SRD2

The above efforts identified a number of snR-DPG loci that are regulated by CPL4 and snRNA transcription machinery. However, the involvement of the latter is established only for the U12-AT1G61280 locus (Figure 3.2) by reporter-gene based assays. To establish the relationship between snR-DPG loci and snRNA transcription machinery at the genome scale, I analyzed the impact of the *srd2-1* mutation on

expression of several representative snR-DPG loci (*LAF3ISF1*: atU5.1b- AT3G55850, atU2.5-AT5G09590, atU2-11-AT1G09800 and AT1G20320) in wild-type and $CPL4_{RNAi}$ background. *srd2-1* encodes a temperature-sensitive allele of SNAPc 50 subunit required for transcription activation of snRNA (Yasutani et al. 1994). The $CPL4_{RNAi}$ *srd2-1* double homozygous seedlings were isolated after the genetic cross. The $CPL4_{RNAi}$ *srd2-1* line exhibited smaller cotyledon phenotype similar to the parental $CPL4_{RNAi}$ line, and slightly shorter roots (Figure 3.10). Before analysis, I confirmed that the $CPL4_{RNAi}$ *srd2-1* maintained $CPL4_{RNAi}$ effects by testing *CPL4* knock-down level and up-regulation of *DTX3*, a marker gene for $CPL4_{RNAi}$ effect (Figure 3.11a, Chapter II; (Fukudome et al. 2014)). When I compare expression of LAF3 isoforms in *srd2-1* mutant and parental wild type (Ler-0) under standard condition (25°C), *LAF3ISF1* accumulation is specifically decreased in *srd2-1* while *LAF3ISF2* isoform is unaffected, confirming that the $CPL4_{RNAi}$ specifically affected PIIsnR-dependent transcripts (Figure 3.11a). Similarly, expression levels of select snR-DPG as well as that of extended $ncRNA_{SSP14}$ were decreased substantially in $CPL4_{RNAi}$ *srd2-1* (Figure 3.11a). Therefore, I concluded expression of snR-DPG loci is indeed regulated by snRNA transcription machinery.

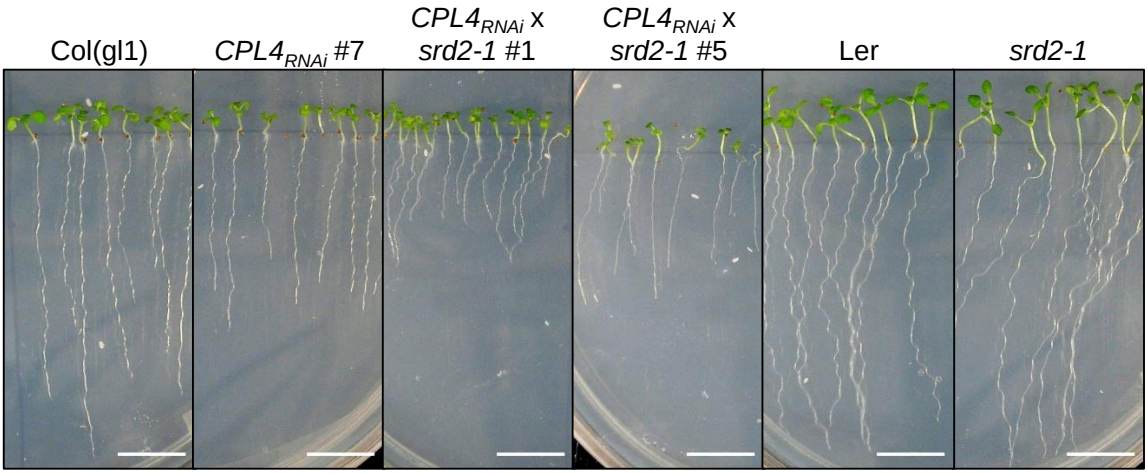


Figure 3.10 Growth of *CPL4*$_{RNAi}$ x *srd2-1* double mutants.
Vertically grown 10-day old seedlings grown on ¼ MS medium with 0.5% sucrose. White bars indicate 10 mm.

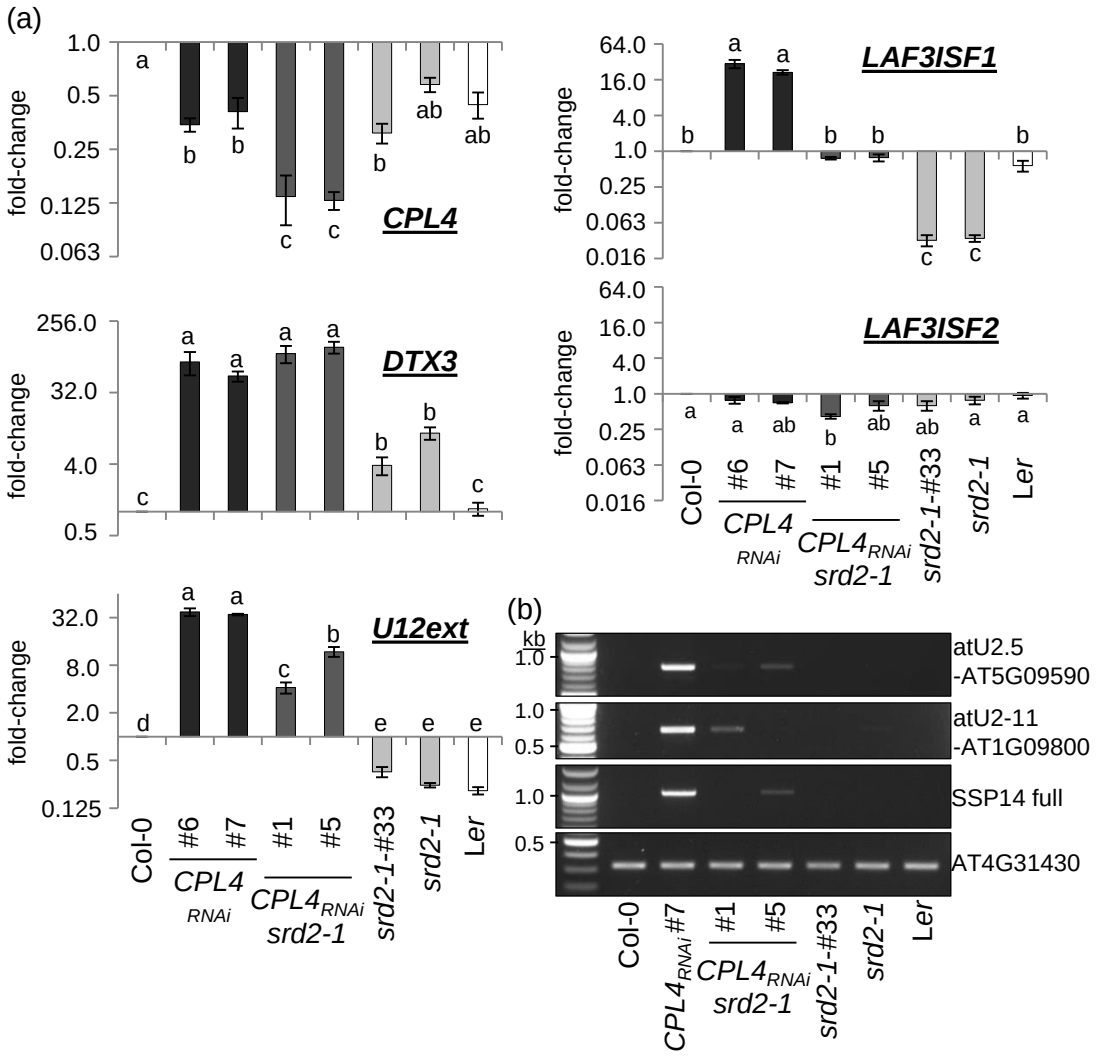


Figure 3.11 Accumulation of poly(A)-chimeric snR/mRNA transcripts in *CPL4$_{RNAi}$* depends on an snRNA activating complex subunit *SRD2*.

(a) oligo-dT-based RT-qPCR analysis on *CPL4$_{RNAi}$* x *srd2-1* double mutant lines (F4 generation) and respective parents (ecotypes; Col(gl1) for *CPL4$_{RNAi}$* (♀) and L*er* for *srd2-1* (♂)). An asterisk indicates a null segregant (*srd2-1*/*srd2-1*, without *CPL4$_{RNAi}$* transgene). Mean fold-change values relative to Col(gl1) are plotted in log(2)-scale, with bars representing standard error (n=3). Genotypes with same letters are not significantly different (One-way ANOVA followed by Tukey HSD test, $p<0.05$; analysis on log-transformed values). For LAF3, lowercase and uppercase letters are assigned to ISF1 and ISF2, respectively. (b) oligo-dT-based RT-PCR analysis on representative snR-mRNA loci.

Salt stress triggers snRNA 3'-extension and up-regulation of snR-DPG in wild type in SRD2 dependent manner

The genome-wide activation of expression of snR-DPG in $CPL4_{RNAi}$ is indicative that pol II CTD plays an active role in the expression of a specific subset of genes; however, little is known about the cues that trigger expression of snR-DPG. Based on a growing notion that the transcriptional responses to environmental stress involve pol II-CTD phosphoregulation (Lavoie et al. 1999; Fuda et al. 2012; Li et al. 2014), I hypothesized that snR-DPG expressions could be activated by environmental stress. To test this hypothesis, I searched AtGenExpress microarray expression dataset for a condition that up-regulates these DPGs. I found that salt-treatment induces many of the DPGs that locate within 500 bp from snRNA, especially in the root (Figure 3.12). I also examined an RNA-seq experiment including salt-stressed samples (Cui et al. 2016) and found reads corresponding to snRNA-extended regions in the salt-stressed seedlings (Figure 3.13).

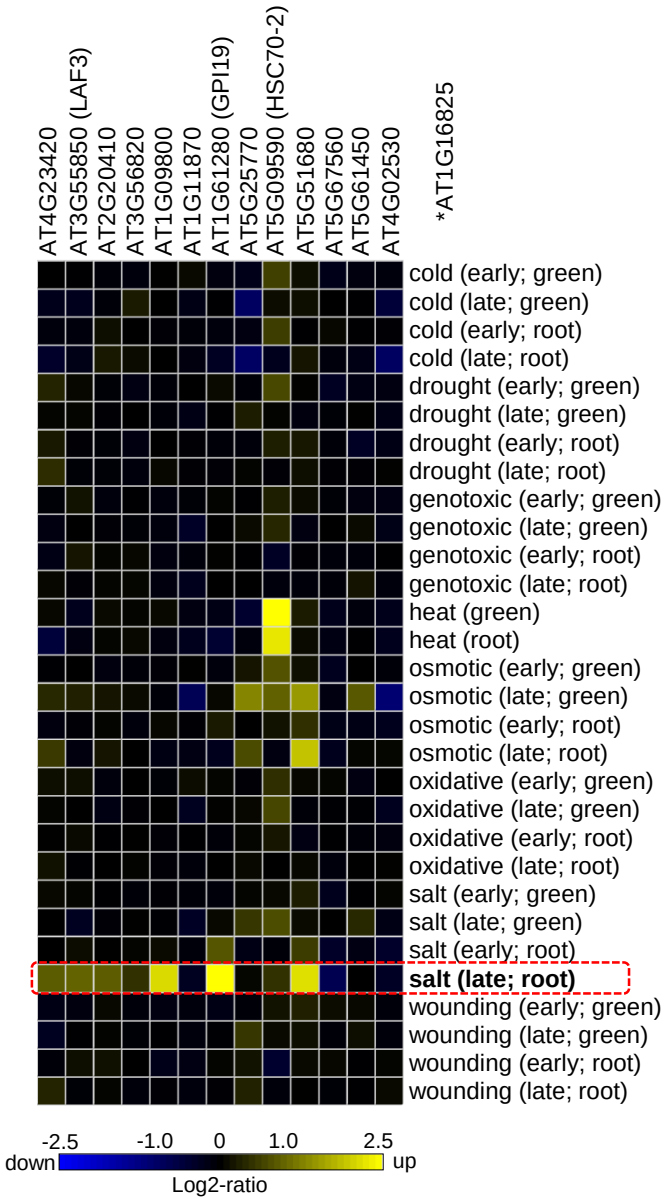


Figure 3.12 DPGs near snRNA tend to be up-regulated in roots treated with salt. Relative expression levels of 13 DPG loci located within 500-nt from pol II-dependent snRNA (Table1; no ATH1 probe is available for AT1G16825) in response to various abiotic stresses are retrieved from AtGenExpress microarrays (under Experiment AT-00120 in Genevestigator; 30 perturbations; 243 samples). DPGs are sorted by distance from upstream snRNA (the closest is on left).

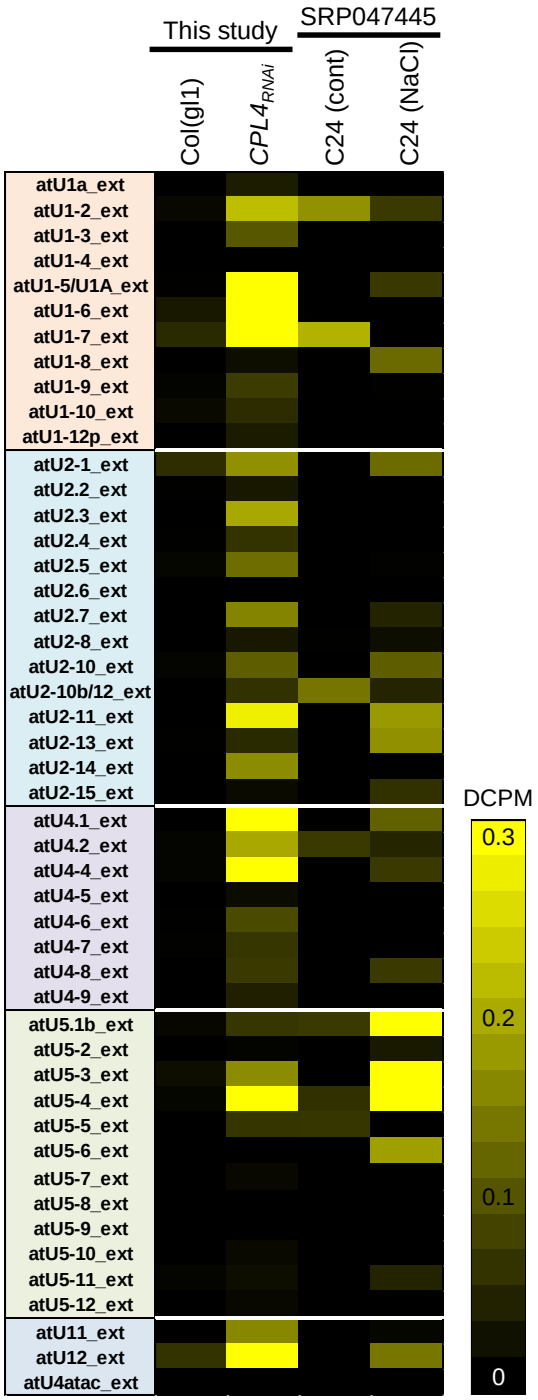


Figure 3.13 Detection of snRNA-3' extension in a salt RNA-seq study.
The heatmap shows the dcpm values from each PII-snR extension region examined in Fig. 3. A poly-(A) RNA-seq data including 300 mM NaCl treated seedlings (SRP047445) were imported to Galaxy server from EBI, and analyzed as described in Materials and Methods.

To independently validate salt-induction of snR-DPG and test if it is via the snRNA expression mechanism, I analyzed snR-DPG expression in salt-treated wild-type (Ler-0) and *srd2-1* plants. Interestingly, salt treatment induced accumulation of snR-DPGs including *LAF3ISF1* and *SSP14* transcripts, specifically in Ler-0 wild type but not in the *srd2-1* mutant (Figure 3.14). The snRNA-independent *LAF3ISF2* was not affected by the treatment. Importantly, atU4.1 snRNA locus without DPG also shows 3'-extended transcripts, indicating that the induction is independent of DPG-promoter (Figure 3.14). These results demonstrate that the snR-DPG accumulations are induced by salt-stress in Arabidopsis through snRNA transcription and 3'-extension. Similar snR-DPG accumulation in response to salt was observed in Col-0 seedlings as well, but $CPL4_{RNAi}$ did not show further induction of snR-DPG by the treatment (Figure 3.15).

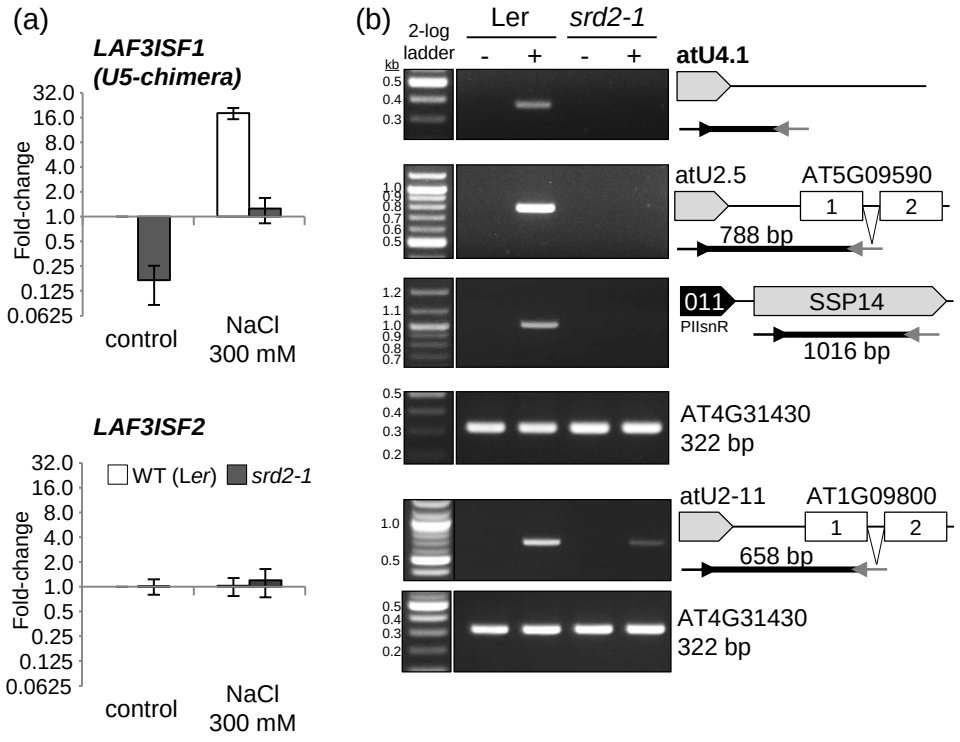


Figure 3.14 Salt-inducible accumulation of chimeric snR/mRNA transcripts in wild-type depends on *SRD2*.

(a) RT-qPCR and (b) RT-PCR of total RNA prepared from WT(ecotype Ler) and *srd2-1* seedlings treated with 300 mM NaCl for 200 min. Bars represent SEM from three independent biological replicates. + or - indicate with or without 300 mM treatment.

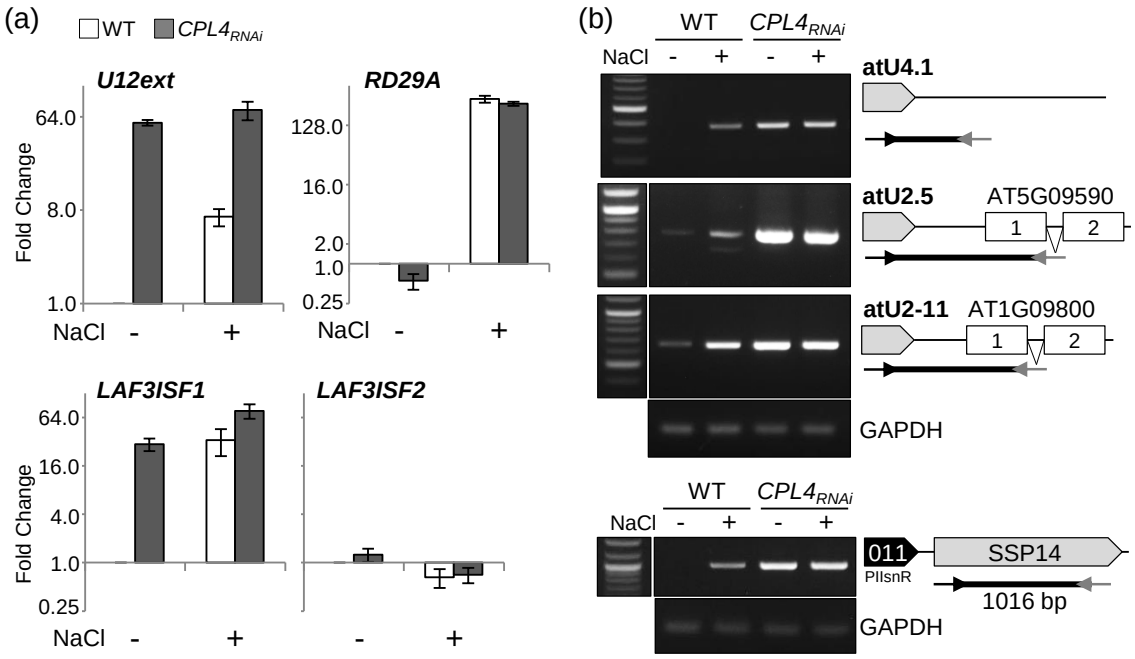


Figure 3.15 Salt-inducible accumulation of 3'-extendend snRNA and chimeric snR/mRNA transcripts.

(a) RT-qPCR and (b) RT-PCR of total RNA prepared from WT (Col) and *CPL4*$_{RNAi}$ seedlings treated with 300 mM NaCl for 200 min. RD29A is included as a marker for salt stress. + or - indicate with or without 300 mM NaCl treatment. Bars represent SEM from three independent biological replicates.

Salt-stress alters pol II-CTD phosphorylation status and pol II occupancy on snR-DPG

The similarity of snR-DPG expression patterns induced by salt-treatment to that of $CPL4_{RNAi}$ line suggested that salt stress affects pol II CTD phosphorylation status. Therefore, I proceeded to characterize relationship among salt stress, pol II CTD phosphorylation, and snR-DPG expression. Due to the slow growth and reduced fertility of $CPL4_{RNAi}$ plants, I used Arabidopsis callus system, which was established previously (Fukudome et al. 2014). RT-qPCR analysis showed that expression of snR-DPG was specific to salt-inducible in Col-0, but constitutive in $CPL4_{RNAi}$ callus, similar to the seedlings (Figure 3.16c). Immunoblot analysis of CTD phosphorylation profiles revealed the overall reduction of pol II-CTD phosphorylation levels in response to the salt treatment, visualized as the reduction of hyperphosphorylated pol II_O form and accumulation of hypophosphorylated pol II_A form (Figure 3.16a). Immunoblot analyses using position-specific anti-CTD-PO_4 antibodies revealed that salt stress promotes the reduction of Ser2-PO_4, Ser5-PO_4 and Ser7-PO_4 (Figure 3.16a-b). This was contrasting to $CPL4_{RNAi}$ pol II profile, which constitutively showed a high level of pol II_O (Fukudome et al. 2014). Salt treatment promoted the reduction of pol II_O and accumulation of pol II_A in $CPL4_{RNAi}$; however, it occurred slower than the transition in the wild-type callus consistent with the reduced CTD phosphatase activity in the RNAi line. Interestingly, the level of Ser5-PO_4 in $CPL4_{RNAi}$ did not decrease during the salt treatment (Figure 3.16a).

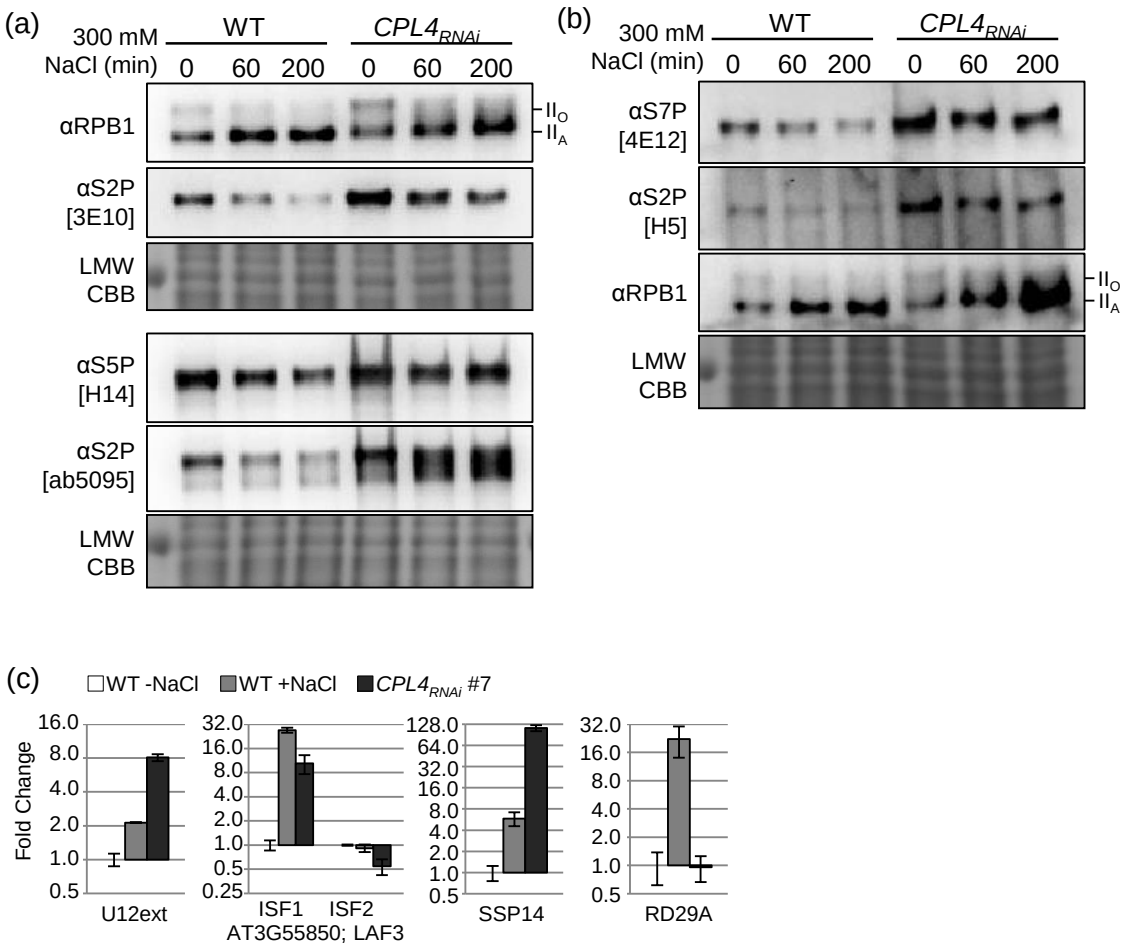


Figure 3.16 Salt-stress causes pol II-CTD dephosphorylation.
Pol II CTD phosphorylation levels detected by western blotting. Approximately 12 µg (a) and 24 µg (b) of total proteins from 2-dpt calli treated with 300 mM NaCl for indicated time. For each membrane, CBB-staining of low molecular weight area of the same gel not subjected to the blotting is shown as loading control. Data taken from the same membrane after stripping and reprobing are stacked. αSer2P, anti-Ser2-PO_4; αSer5P, anti-Ser5-PO_4; αSer7P, anti-Ser7-PO_4. (c) Relative expression levels of select 3'-extended snRNAs in salt-treated 2dpt-callus by RT-qPCR. Absolute fold-change values relative to WT(-)NaCl are plotted in log2-scale. Bars represent SEM of biological triplicates.

The immunoblot results were unexpected because salt treatment and *CPL4*$_{RNAi}$ had opposite effects on pol II CTD phosphorylation levels even though both result in the similar snRNA-extension and resulting snR-DPG accumulation in the host cells. Because immunoblots assess CTD phosphorylation status as a whole, but not at individual genes, I performed chromatin immunoprecipitation (ChIP) assays to test if pol II loading and CTD phosphorylation levels at the snRNA locus change during the salt-induced snRNA 3'-extension (Figure 3.17). I chose U12 locus as a model because (1) U12 is a single copy locus, unlike other snRNAs (2) U12 exhibits notable 3'-extension in *CPL4*$_{RNAi}$ and salt treatment. As shown in Figure 3.17, loading of pol II decreased at the 3'-box (3'-1) and the 3'-extended region (3'-2) relative to the promoter region (pro) in untreated wild type, while it stayed as high as the promoter level at the 3'-2 region in salt-stressed wild-type and in *CPL4*$_{RNAi}$ calli (Figure 3.17b, c). The occupancy of pol II with Ser2-PO4 mark was lower at the pro and the 3'-1 region in salt-stressed wild type and in *CPL4*$_{RNAi}$ than in unstressed wild type (Figure 3.17d). Similar to total pol II, unstressed wild type showed a decline of the Ser2-PO4 level at the 3'-extended region (3'-2) relative to the promoter in the unstressed wild type but not in salt-stressed wild-type and in *CPL4*$_{RNAi}$ calli (Figure 3.17e). These results indicate that salt and *CPL4*$_{RNAi}$ impact global CTD-Ser2-PO_4 levels in opposite directions, however, similarly promote retention of pol II at the snRNA 3'-extended region.

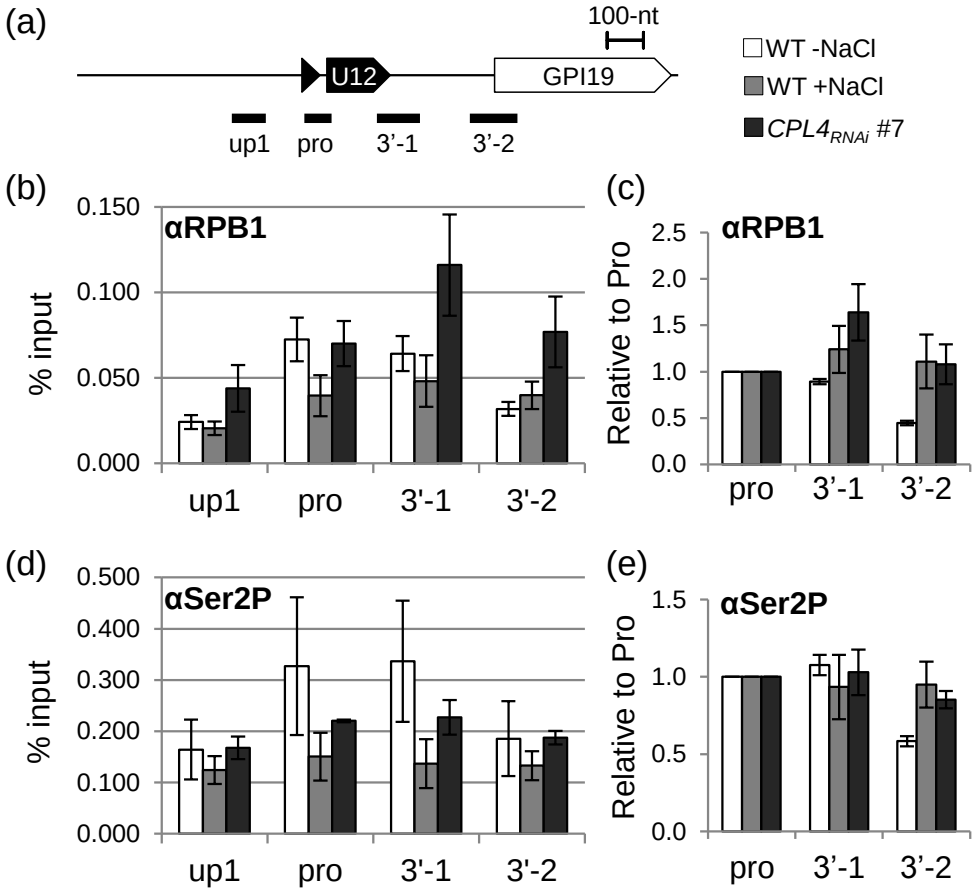


Figure 3.17 Pol II occupancy on U12 snRNA extension region in the salt treated and *CPL4$_{RNAi}$* calli.

(a) schematic diagram of the U12-GPI19 locus. Target regions for quantitative PCR analysis are indicated by black lines. A black triangle right in front of U12 snRNA locus is its promoter. (b-e) The amount of pol II on each target region were determined by ChIP-qPCR analysis using (b-c) αRPB1, (d-e) anti-Ser2-PO_4 (αSer2P) polyclonal antibody [ab5095], (b,d) and (c,e) show % input and signal levels relative to the promoter (pro) region, respectively. Error bars represent SEM of biological triplicates. No-antibody control yielded maximum %input of 0.0031.

3.4 Discussion

Pol II-CTD phosphoregulation governs transcription of both protein-coding and non-coding genes. Although plant kinases and phosphatases involved in the regulation have been identified and characterized, little is known about the significance of the phosphoregulation in non-coding RNA transcription in plants. In this study, I show that a CTD phosphatase CPL4 plays a pivotal role in choosing the fate of pol II transcribing U snRNA either for "normal transcription termination/3'-processing" or "continuation of transcription to produce 3'-extended snRNA". I demonstrated that the latter is not a mere transcriptional abnormality but is a part of transcriptional regulatory strategies in plants. Furthermore, I found that plants may use other ncRNAs as extension substrates to generate protein-coding mRNA. These snRNA/ncRNA-to-mRNA conversions occur not only in $CPL4_{RNAi}$ plants but also observed in wild type when plants are environmentally challenged (salt stress) and perhaps at specific tissues like pollens, which are likely associated with alteration of CTD phosphorylation status. This indicates that snRNA/ncRNA extension is a previously undocumented regulatory mechanism for plant gene expression and pol II CTD functions as a hub for this regulation.

Knock-down of CPL4 causes snRNA 3'-extension, which gives rise to translatable snR-DPG fusion

The regulation of snRNA transcription by pol II CTD phosphorylation has been characterized mostly in vertebrates, and very little is known in plants. In animals, Ser2-PO_4/ Ser7-PO_4 double phosphorylation marks facilitate snRNA 3'-end processing

whereas Ser5 phosphorylations were inhibitory by differentially impacting recruitment of the Integrator complex to pol II transcribing snRNA (Egloff et al. 2010). Previous studies identified two conserved CTD phosphatases involved in this process. RNA Pol II-associated protein 2 (RPAP2) is recruited by Ser7-PO_4 and specifically dephosphorylates Ser5, which in turn recruits the Integrator complex (Egloff 2012; Egloff et al. 2012). Another CTD phosphatase, SSU72 dephosphorylates Ser5-PO_4 and Ser7-PO_4, and is required for proper snRNA 3'-end processing in both animals and yeast (Wani et al. 2014). Surprisingly, there have been no implication of functional association of FCP1 (TFIIF-interacting CTD Phosphatase1)-family CTD phosphatases like CPL4, in snRNA transcription cycle. Therefore, my data suggest a new function for FCP1-family CTD phosphatase in snRNA transcription and 3'-end processing, at least in plants. $CPL4_{RNAi}$ calli which accumulate hyperphosphorylated pol II did not show a strong increase of Ser2-PO_4 level on the pol II occupying the U12 3'-extended regions (Figure 3.17d). The effect of *CPL4* knockdown may be more prominent on the non-chromatin-bound fraction of pol II, similarly to the knockdown of *Drosophila melanogaster* FCP1 which is implicated in the pol II recycling process (Fuda et al. 2012). Because both RPAP2 and SSU72 homologs are also present in Arabidopsis genome, it is not clear whether CPL4 replaces all or a part of the RPAP2/SSU72 function in plants, or acts in concert with other phosphatases in plants as well as in other organisms. Considering the observation that pol II occupies more at the downstream of 3'box in $CPL4_{RNAi}$ cells, and distinct substrate specificities of CPL4 and RPAP2/SSU72 (Krishnamurthy et al. 2004; Zhang et al. 2012a) , it is more likely that CPL4 promotes

pol II dissociation from DNA template after 3' box, which is later than the timings for RPAP2/SSU72. This may promote recycling of the pol II in a similar way that FCP1 functions during transcription termination of protein-coding genes (Cho et al. 1999). This model is consistent with our data showing that $CPL4_{RNAi}$ enhances U12-LUC2 reporter gene expression even when the 3'-box was mutated.

PIIsnR is embedded in transposons / snR-DPG in other plants

Though snRNAs and their transcriptional mechanisms are conserved in eukaryotes, the plant genome structures uniquely connect snRNA 3'-extension with the production of snR-DPG transcripts. In Arabidopsis genome, 75 snRNA genes are overall evenly distributed on the chromosomes, with several small gene clusters (Kaul et al. 2000; Wang and Brendel 2004). Thirty-three pol II-dependent snRNA genes are flanked by a protein-coding gene that is located within about 1 kbp downstream of snRNA locus; eighteen and fifteen of them are on the same and the opposite strand, respectively (Table 3.2). In contrast, human snRNA loci are flanked with long intergenic region without protein-coding potentials. All true U1 genes have approximately 2 kb-long conserved flanking sequences in both upstream and downstream regions (Manser and Gesteland 1982; Htun et al. 1984), many of which extend to 20-24 kb extensively conserved intergenic sequences for both directions (Bernstein et al. 1985). Human U2 snRNA genes are clustered as 10-20 tandem repeats of a 6-kb unit (Vanarsdell and Weiner 1984; Westin et al. 1984; Lindgren et al. 1985). Another plant genome feature that facilitates the integration of snRNA transcription mechanism into diverse gene expression systems is the PIIsnR sequences in DNA

transposons, which provides potential mobilization and propagation potentials to PIIsnRs. In the Arabidopsis genome, 75 PIIsnRs locate inside of non-autonomous DNA transposons, particularly in the ATREP5 subclass. These promoters contain a perfectly conserved USE, TATA box and 32-34 bp spacing between the elements, making them indistinguishable from promoters in functional snRNA genes. This may represent an ancient gene capture event(s) by ATREP5, which performs rolling-circle transposition and can capture and spread the flanking sequences (Kapitonov and Jurka 2001; Kapitonov and Jurka 2007). During genome evolution, the snRNA promoters mobilized by transposons could be a source of *de novo* gene expression, mediated by pol II-CTD phosphoregulation. Indeed, the novel ncRNA I identified (*ncRNA*$_{SSP14}$) is expressed from PIIsnR associated with ATREP fragment. *ncRNA*$_{SSP14}$ is an intermediate-ncRNA produced from 5'-region of the *SSP14* locus. Due to its degradation by exosome, *ncRNA*$_{SSP14}$ can accumulate to the detectable level only in the *hen2-4* mutant background, which is compromised in the nuclear exosome pathway (Western et al. 2002). Although *ncRNA*$_{SSP14}$ resembles some previously reported ncRNAs in animals and fungi, its production from a PIIsnR and potential to convert to mRNA make *ncRNA*$_{SSP14}$ distinct. Based on CPL4-regulated conversion to mRNA, transcription termination of *ncRNA*$_{SSP14}$ is likely under control of 3'-box sequence embedded in the *SSP14* coding region. Because the 3'-box are relatively tolerant to deviation from the consensus sequence (Connelly and Filipowicz 1993), 3'-box sequences frequently occur in the Arabidopsis genome sequence, increasing the probability of PIIsnR-3'box pairings during the genome evolution. *ncRNA*$_{SSP14}$ also shares several features with UNT

(upstream noncoding transcripts), which were reported previously as ncRNA that overlap with 5' region of pre-mRNA and are degraded by exosome pathway (Chekanova et al. 2007). Although a 3'-extension of UNT has not been studied, it is possible that conditional conversion of short-lived ncRNA to mRNA by 3'-extension occurs with other types of ncRNAs to regulate gene expression.

Salt stress induces pol II CTD dephosphorylation and snR-DPG accumulation

The salt-inducibility of snRNA 3'-extension and accumulation of snR-DPG in wild-type plants suggest a presence of signaling pathway that fine-tunes snRNA expression mechanism and diverts some pol II complexes to escape snRNA termination/3'-processing. The most likely point where salt signal inputs into snRNA termination/3'-processing is CTD phosphorylation status. Salt treatment caused global dephosphorylation of pol II-CTD (Figure 3.16). The reduced level of pol II with Ser2-PO_4 on the U12 locus in stressed wild-type calli is consistent with the global reduction of phosphorylation level observed in total protein fraction. The reduced Ser2-PO_4 level at the 3'-box region may compromise the recruitment of the snRNA 3'-end processing complex, resulting in 3'-extension of U12 in the salt-stressed wild-type calli. Signal-triggered, gene specific pol II CTD phosphorylation/dephosphorylation are important for several developmental and stress-regulated gene expression systems in animals (Bellier et al. 1997; Dubois and Bensaude 1998; Shim et al. 2002; Walker et al. 2007), and have also been proposed in plant innate immunity responses (Li et al. 2014). In the latter case, signals generated from a perception of microbe-associated molecular patterns (MAMPs) are transduced through MAP kinase cascade and induce rapid and transient

cyclin-dependent kinase C (CDKC)-mediated phosphorylation of pol II CTD in Arabidopsis. By contrast, abiotic stress signal-induced CTD dephosphorylation of plant pol II has not been reported, and stress-specific CTD phosphatase has not been identified. Although my results demonstrated a massive dephosphorylation of pol II CTD after salt stress (Figure 3.16), the identity of CTD phosphatase responsible for the salt-induced CTD dephosphorylation needs to be carefully evaluated. Considering the slower transition of pol II_O to II_A in *CPL4*$_{RNAi}$ calli during salt stress, CPL4 is a strong candidate responsible for salt-induced CTD phosphatase activities. However, I cannot exclude possibilities for other factors contributing salt-induced pol II dephosphorylation, because CTD dephosphorylation still occurred in *CPL4*$_{RNAi}$ after salt treatment. This could be due to partial silencing of CPL4, salt-induced inhibition of CTD-kinase activities, or activation of other CTD-phosphatase activities.

Adapting snRNA 3'-extension to their genome structure, plants produced a unique stress-inducible gene expression system, which is likely activated co-transcriptionally via alteration of pol II CTD phosphorylation level. This system resembles heat-shock activation of paused pol II in animals (Ni et al. 2004) and may allow a quicker response than *de novo* activation of promoters by transcriptional factors. The BLAST survey identified potential 150 snR-DPGs in Arabidopsis and other plant species, including 17 EST-supported snR-DPGs (Table 3.4), suggesting the snRNA-to-mRNA conversion is a ubiquitous process in plants. There are diversities in snR-DPG pairs in different plant genomes, but some conserved combinations, such as *U12-GPI19* and *U1/2-cytochrome c oxidase subunit 5b-1* may have biological significances selected

over the evolution process. However, individual combinations need to be empirically examined for their functions because it is not immediately obvious if snR-DPG-encoded proteins collectively perform defensive functions during the stress. Function through other mechanisms without protein production, including providing physical scaffold (Vilborg et al. 2015) or regulatory RNAs (Uesaka et al. 2014; Vera and Dowell 2016) cannot be excluded. Furthermore, a presence of snRNA structure on mRNA may confer new functionality to the mRNA, like highly mobile dicistronic mRNA-tRNA in Arabidopsis (Zhang et al. 2016). Clearly, $CPL4_{RNAi}$ revealed hidden RNA dynamics orchestrated around plant pol II CTD, which warrant further investigation.

CHAPTER IV

FUNCTIONS OF CPL4 IN *DE NOVO* SHOOT ORGANOGENESIS

4.1 Introduction

Plants have a unique potential to regenerate the whole body from a single somatic cell (Vasil and Hildebrandt 1965; Vasil and Hilderbr.Ac 1965). This trait has been explored in horticultural and agricultural applications and biotechnologies such as asexual propagation, tissue culture and transformation (Newell 2000; Butt et al. 2015). The regeneration process involves *de novo* shoot organogenesis (DNSO), which requires coordinated and balanced actions of phytohormones, auxin and cytokinin (CK) (SKOOG 1950; Skoog 1957). DNSO is typically induced by two-step incubation on media with different auxin:CK ratio (Christianson and Warnick 1983; Che et al. 2007). Firstly, excised plant tissues (explants) are incubated for 3-5 days on a medium containing both auxin and CK with high auxin:CK ratio (Callus Induction Medium; CIM). This process confers explants transient competence to form *de novo* shoots. Subsequent incubation of the CIM pre-incubated explants on higher and lower auxin:CK ratio would lead to *de novo* organogenesis of root and shoot, respectively, whereas prolonged incubation of explants on CIM leads to a formation of callus, a mass of proliferating multipotent cells (Christianson and Warnick 1983).

The callus formation and lateral root primordia (LRP) development share the common auxin-regulated mechanism (Sugimoto et al. 2010), whereas DNSO relies hugely on CK signaling pathways which activate genes essential for shoot apical

meristem (SAM) development and regulation (Che et al. 2007). *De novo* shoots originate from LRP expressing a NAC (No Apical Meristem/Arabidopsis thaliana activating factor/Cup-shaped cotyledon 2) transcription factor *CUP-SHAPED COTYLEDON 2* (CUC2), whose expression is induced during CIM-pre-incubation (Gordon et al. 2007). Subsequent incubation on CK-rich shoot induction media (SIM) activates other transcription factors such as APETALA2/Ethylene Responsive Factor (AP2/ERF) family *ENHANCER OF SHOOT REGENERATION 1* (*ESR1*) and *ESR2*, which also activate expression of *CUC1* (Matsuo et al. 2009). The earlier induction of these transcription factors on SIM result in activation of essential homeobox transcription factors *SHOOT MERISTEMLESS* (*STM*) and *WUSCHEL* (*WUS*), leading to *de novo* shoot meristem initiation from the previous LRP/calli, and eventually DNSO (Cary et al. 2002).

De novo organogenesis from root tissues mainly occurs from differentiated pericycle cells, which were shown to be arrested at G2 phase (Beeckman et al. 2001). Activation of the cell cycle is a key step to initiate LRP development, callus formation and subsequent shoot organogenesis (Himanen et al. 2002; Xu et al. 2012). The cell cycle progression is regulated through coordinated actions of cyclin-dependent kinases (CDKs) and their activator protein family cyclins (Morgan 1995; Francis 2007). The mRNA expression patterns and promoter-reporter gene derivatives of CDKs and cyclins have served as indicators of corresponding cell cycle activities (Colon-Carmona et al. 1999; Kang and Dengler 2002; DiDonato et al. 2004; Adachi et al. 2006; Adachi et al. 2009). For example, the expression of some Arabidopsis Cyclin A (CYCA2;1, CYCA3;1) and Cyclin B (CYCB1;1) are associated with G1-S phase and mitotic

division (G2-M transition), respectively (Hemerly et al. 1992; Niebel et al. 1996; Himanen et al. 2002; Menges et al. 2005). D-type cyclins including CycD3;1 were responsive to CK and shown to be a rate limiting factor of G1-S phase transition and CK response; Constitutive overexpression of cyclin D leads to alleviation of CK requirement for shoot organogenesis (Riou-Khamlichi et al. 1999; Dewitte et al. 2003; Menges et al. 2006; Dewitte et al. 2007). The highest kinase activities of CDKA1 can be observed at G1-S and G2-M phases, and CDKB1;1 activity reaches maxima at G1-S (Joubes et al. 2000). CDKA in Arabidopsis is constitutively expressed, whereas CDKB family expressions get higher during G2-M phases (Francis 2007) .

Previous forward genetics studies have identified additional factors related to general transcription, splicing and RNA metabolism machinery involved in the CK sensitivity and DNSO process. Mutants in two *CYTOKININ-HYPERSENSITIVE* (*CKH1* and *CKH2*) genes were isolated from a screening for rapid callus growth and greening on a low CK media (Kubo and Kakimoto 2000). *CKH1* encodes a TATA-binding protein complex associated factor 12B (TAF12B) homolog, which is a part of general transcription factor TFIID (Kubo et al. 2011). The c*kh1* mutants show CK hypersensitive phenotype in calli without affecting expression pattern of primary CK-responsive genes (Kubo et al. 2011). *CKH2/PICKLE* encodes SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin remodeling factor and functions synergistically with *CKH1* (Furuta et al. 2011). *SHOOT REDIFFERENTIATION DEFECTIVE 2* (*SRD2*) is an essential gene for *in vitro* shoot regeneration from root explants (Yasutani et al. 1994). SRD2 encodes a homolog of SNAP50 subunit in the snRNA activator protein complex

(SNAPc). It has been suggested that the proper production of snRNA is essential for dedifferentiation and regeneration of tissues (Ohtani and Sugiyama 2005; Ohtani et al. 2015). Similarly, a temperature sensitive mutant screening for defects in adventitious root formation from shoot explants identified genes involved in pre-rRNA processing (*ROOT INITIATION DEFECTIVE*, *RID2* and *RID3*) and spliceosome activity (*RID1*, a DEAH-box helicase homologous to yeast Prp22) (Sugiyama 2003; Ohtani et al. 2013).

In the previous chapters, I showed that CPL4 is a genuine pol II-CTD phosphatase (Fukudome et al. 2014) and is involved in co-transcriptional regulation of snRNA in Arabidopsis. In this chapter, I characterize the functions of CPL4 in callus formation and *de novo* organogenesis of root and shoot using the *CPL4* knockdown lines (*CPL4*$_{RNAi}$). *CPL4*$_{RNAi}$ lines are defective in lateral root primordia development without affecting auxin response reporter *DR5::GUS* expression. While callus formation from shoots and root explants are not greatly affected by *CPL4*$_{RNAi}$, DNSO from *CPL4*$_{RNAi}$ root explants was greatly enhanced even with lower CK concentration, with which wild-type explants cannot regenerate shoots. The enhanced DNSO phenotype of *CPL4*$_{RNAi}$ root explants is dependent on *SRD2*. Transcriptome analysis reveals that *CPL4*$_{RNAi}$ root explants overexpress DNA replication and cell cycle regulation related genes after CIM pre-incubation. A co-expression gene network analysis identified several CPL4 and low CK-regulated gene clusters including a thalianol biosynthesis operon-like cluster, suggesting a potential biological function for the terpenoid. Constitutive and/or early activation of the key transcription factors for SAM regulation related genes such as *CUC2*, *WUS* and *STM* were detected in *CPL4*$_{RNAi}$. By contrast, in wild type, expression

levels of type-A *ARRs* and *ESR1* were induced on CIM were similar to *CPL4*$_{RNAi}$, but the SAM regulator activation did not occur. These results indicate that CPL4 negatively regulates DNSO by repressing early activation of SAM regulated genes during CIM pre-incubation.

4.2 Materials and methods

Plant materials and growth condition

Unless otherwise stated, Arabidopsis seeds were germinated and vertically grown on ½ MS media containing 1% sucrose and 1.5% agar (referred to as the basal media throughout this chapter), at 25 ^{o}C under long-day condition. For hypocotyl growth measurements, seedlings were germinated and horizontally grown on the basal media with 0.7% agar. *CPL4*$_{RNAi}$ and CPL4 overexpression transgenic lines are described in Chapter II. *srd2-1* mutant and *CPL4*$_{RNAi}$ *srd2-1* double mutant are described in Chapter III. *DR5::GUS* construct is introduced into *CPL4*$_{RNAi}$ #6 by crossing homozygous parental lines.

Auxin treatment and GUS staining

Seeds were germinated on cellophane membranes placed on the basal media. The seedlings were transferred to a new basal media containing 0 or 100 nM IAA and grown at the same condition for the indicated period. For GUS staining, the tissues are first fixed in ice-cold 90% acetone for 15 min. Then, the fixed tissues were washed once with 100 mM potassium phosphate (pH 7.0). GUS staining solution (100 mM potassium phosphate pH7.0, 5 mM K3 ferrocyanide, 0.5 mM K4 ferrocyanide, 0.1% Triton X-100

and 2 mM X-Gluc) was added to the washed tissues and incubated for 6 hours to overnight. GUS stained tissues were observed under a dissecting microscope (Figure 4.1) or subjected to clearing for Differential Interference Contrast (DIC) imaging.

Semi-synchronous lateral root induction and tissue clearing for DIC imaging

Semi-synchronous lateral root induction was performed essentially as previously described (Ohtani et al. 2010). Shoot portion including cotyledons and root apices were removed from 4-d old vertically grown seedlings. The remaining tissues were cultured on root induction medium (RIM; 1x MS medium containing 1xB5 vitamin, 0.2 g/L KH_2PO_4, 0.1 g/L myo-inositol and 0.5 mg/l indole-3-butyric acid; IBA). Prior to microscopic observation, explants were cleared following a method previously described with a slight modification (Malamy and Benfey 1997). Explants were dipped in solution 1 (0.24N HCl and 20% methanol), vacuum infiltrated for 5 min and incubated at 57 ^{o}C for 15 min. After removal of solution 1, solution 2 (w/v 7% NaOH and 60% ethanol; optional 7% hydroxylamine-HCl is omitted) was added and incubated at room temperature for 15 min. The explants were then rehydrated in 40%, 20%, 10% and 5% ethanol solutions. Each rehydration step was 10 min (15 min for 5% ethanol) at room temperature. Finally, the explants were incubated in 25% glycerol at room temperature for 15 min. The cleared samples were mounted in 50% glycerol on glass plate and observed under DIC objectives using Olympus BX51 microscope.

Callus induction and in vitro shoot regeneration

Explants for callus induction and shoot regeneration were prepared from 10-15 day old seedlings. Root explants of approximately 5 mm in length were excised from a

primary root segment 5 mm away from the root tip. Therefore, one hypocotyl and one root explant each can be prepared from a seedling. Callus formation was induced by incubating explants on callus induction medium (CIM; 1x MS medium containing 1xB5 vitamin, 0.2 g/L KH_2PO_4, 0.1 g/L myo-inositol and 2.0 mg/L 2,4-Dichlorophenoxyacetic acid (2,4-D), 0.05 mg/L kinetin and 0.9% agar) for the indicated amount of time (Doelling and Pikaard 1993). For *in vitro* shoot regeneration, the explants were pre-incubated on the CIM for 4-6 days, then transferred to and incubated on shoot induction medium (SIM) containing 1x MS medium containing 1xB5 vitamin, 0.2 g/L KH_2PO_4, 0.1 g/L myo-inositol and 0.15 mg/L indole-3-acetic acid, 0 – 0.5 mg/L 2-isopentenyladenine (2iP) and 0.9% agar. All incubation was done under long-day condition with 25-35 µmol/m2/s light intensity at 22 ^{o}C or 28 ^{o}C.

Transcriptome analysis of roots incubated on SIM by RNA-seq

Experimental design is depicted in the Figure 4.7a. wild-type (Col-0 *gl1*) and $CPL4_{RNAi}$ line #7 seeds were sown and germinated on the basal media. Aerial portions of ten-day old seedlings were removed, and the root segments were subjected to CIM-pre-incubation for 5-day. Then, the root tissues were transferred to modified SIM containing 1/10 of standard CK (0.05 mg/L 2iP) or none (0 mg/L 2iP and incubated for 72 hours. Subsequently the root tissues were harvested, flash-frozen with liquid nitrogen and stored for RNA extraction. Total RNA including short and small RNAs were extracted from the frozen tissues by miRNeasy Mini Kit (Qiagen, Cat. #217004). After quality check on agarose gel electrophoresis and nanodrop measurements, for each of twelve samples (2 genotypes x 2 conditions x independent biological triplicates), 50 µl (50

ng/µl) of total RNA in RNase-free water was submitted to Genomics and Bioinformatics Service facility in Texas A&M University (TxGen). The TruSeq-Stranded RiboZero library preparation and Illumina-HiSeq 2500v4 High Output run (125-paired end) were performed by the TxGen facility. The raw sequence read files were produced in fastq format.

Bioinformatic analysis

The fastq files were loaded to Galaxy public server (https://usegalaxy.org/) for bioinformatics analysis. For each sample, raw and adapter-free read files in fastq format were produced by the facility; In the differential gene expression analysis (Table 4.1, Figure 4.7, 4.8 and 4.9), the adapter-free reads were used. QC, mapping and differential gene expression analysis were performed as described in Chapter III. Co-expression network analysis and Gene Ontology Overrepresentation analysis were performed as described in Chapter II.

RT-PCR and qPCR

RNA extraction, reverse transcription and qPCR were performed as described in Chapter III. Primers used are listed in Appendix P.

4.3 Results

Knockdown of CPL4 impairs lateral root formation

The root of $CPL4_{RNAi}$ seedlings produced unique growth characteristics. Their primary roots grow slower than wild-type roots and have no lateral roots (Figure 4.1a). The root auxin response of $CPL4_{RNAi}$ was more substantially affected. While the growth

of wild-type and *CPL4*$_{RNAi}$ primary roots were similarly inhibited by Indole-3-acetic acid (IAA), lateral root induction by IAA was observed only in wild-type roots. *CPL4*$_{RNAi}$ did not produce lateral roots even after 100 nM IAA treatment (Figure 4.1a, c). To determine if *CPL4*$_{RNAi}$ is altered for auxin distribution and/or signaling, expression of DR-5 GUS reporter gene was monitored. Root tips of both wild-type and *CPL4*$_{RNAi}$ seedlings show comparative levels of *DR5::GUS* expression with or without IAA-treatment, consistent with the observations that the auxin sensitivity of the *CPL4*$_{RNAi}$ primary root growth is similar to that of the wild type. (Figure 4.1d). Interestingly, *DR5::GUS* staining indicative of lateral root primordia (LRP) was observed not only in wild-type but also in *CPL4*$_{RNAi}$ roots, despite at a lower frequency (Figure 4.1e, f). These observations suggest that *CPL4*$_{RNAi}$ inhibits lateral root development after the initiation of LRP.

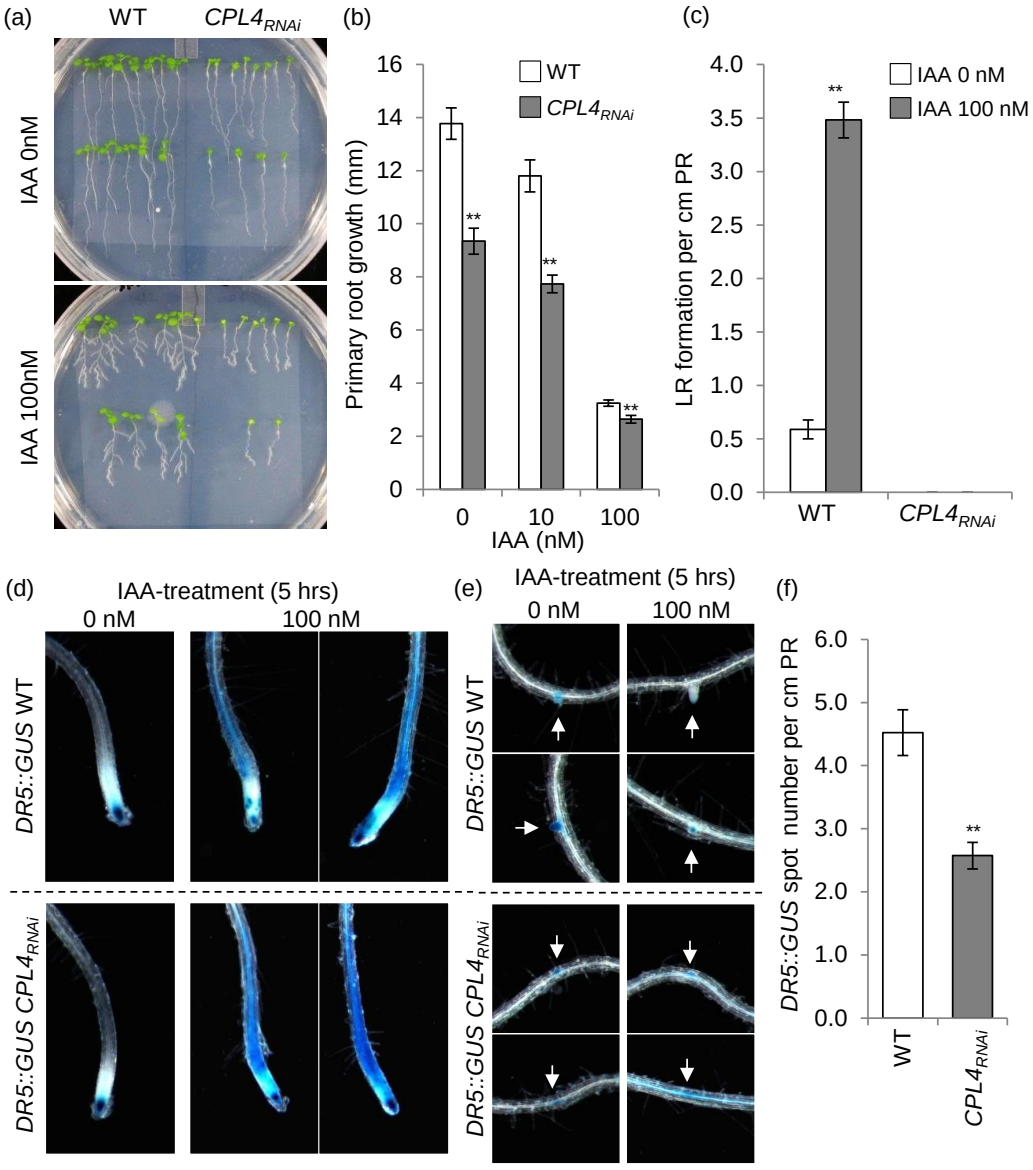


Figure 4.1 *CPL4* knockdown impairs lateral root formation in response to IAA.
(a) 8-day old seedling pictures. 3-day old seedlings germinated and vertically grown on ½ MS media with 1% sucrose, 1.5% agar and cellophane membrane were transferred to ½ MS media containing 0 or 100 nM IAA. Picture was taken at 5-day post transfer. (b) Primary root growth measured at 3-day post transfer (c) Lateral root numbers were counted on 5-day post transfer and normalized by the primary root length. Error bars represent ±SEM. ** $p<0.01$, Student's t-test between mean values of the wild type (WT) and *CPL4*$_{RNAi}$ line. Bars indicate the SEM of 9~13 seedling measurements. (d-e) 7-day old *DR5::GUS* seedlings were transferred to media containing 0 or 100 nM IAA and incubated for 5 hours, followed by GUS-staining. (d) Root tips and (e) lateral root primordia (white arrows) are shown. (f) Frequency of GUS-stained spot on primary roots of 10-day old seedlings. Bars represent the SEM (n=9 and 22 for WT and *CPL4*$_{RNAi}$, respectively).

Knockdown of CPL4 compromises LRP development at later stages

Under standard growth condition, $CPL4_{RNAi}$ seedlings were able to produce LRP but not normal lateral roots. To characterize the alteration of LR development process in $CPL4_{RNAi}$, I conducted semi-synchronous lateral root induction assay. In this assay, root explants without root tip were placed on the root induction medium (RIM) containing a high concentration of auxin (Indole-3-butyric acid; IBA, 0.5 mg/L). Under this condition, the $CPL4_{RNAi}$ root explants can produce LRP with distinctive developmental defects (Figure 4.2). At 48 hours post transfer (hpt), $CPL4_{RNAi}$ LRPs corresponding to stage IV (Malamy and Benfey 1997) are similar to those of WT (Figure 4.2a). However, the many $CPL4_{RNAi}$ LRPs become fasciated and show disorganized cell orientation at 72 hpt while the wild-type LRPs establish the vasculature-like organization in the center (Figure 4.2a). The width of $CPL4_{RNAi}$ LRP becomes significantly wider than wild type, and it is similar to *srd2-1* LRPs at the restrictive temperature (Figure 4.2c)(Ohtani et al. 2010). At 96 hpt, established and elongating lateral roots are observed on wild-type root explants, whereas $CPL4_{RNAi}$ LRPs are not fully established yet (Figure 4.2a). Some fasciated $CPL4_{RNAi}$ LRPs with split meristem are observed at 8-day post transfer (Figure 4.2b). The *DR5::GUS* reporter expression profile shows that $CPL4_{RNAi}$ can establish the auxin localization at the tip; however, the area showing auxin maxima is larger than that of wild type, and weak GUS activity was observed throughout the emerging LR (Figure 4.2d). This suggests that $CPL4_{RNAi}$ is compromised in establishing proper auxin gradient during lateral root development, resulting fasciated LRP.

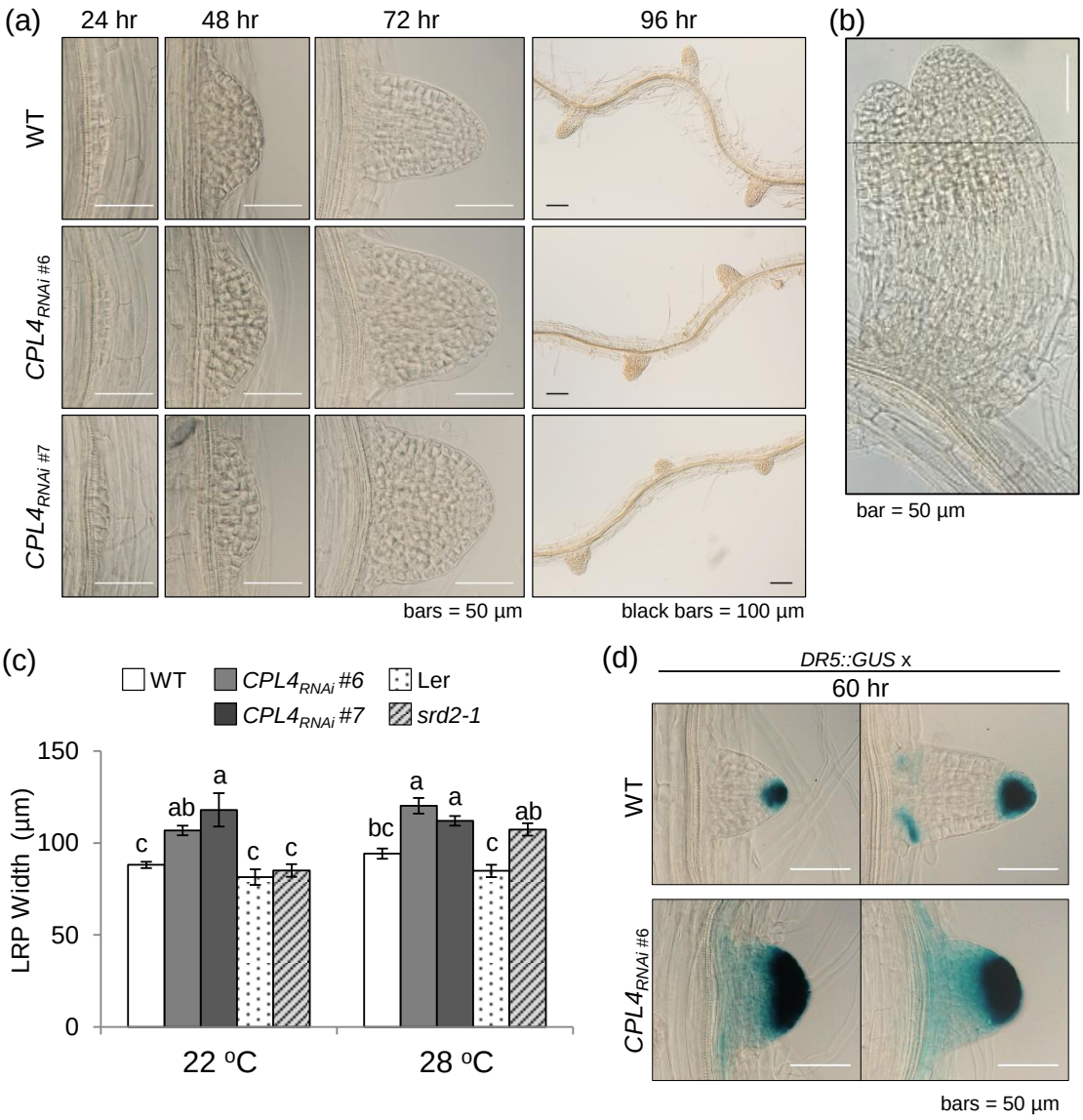


Figure 4.2 Lateral root development of *CPL4$_{RNAi}$* line in semi-synchronous lateral root induction system.

(a) Lateral root induction was performed as previously described (Ohtani et al. 2010). Shoot tips and root apices were removed from 4-d old vertically grown WT or *CPL4$_{RNAi}$* seedlings. The remaining tissues were cultured on MS medium containing 0.5 mg/l indole-3-butyric acid (IBA). DIC images of developing lateral root primordia cleared by the method described in (Malamy and Benfey 1997) were taken at 24-96 hours post transfer. (b) LRP with split/fused meristem from *CPL4$_{RNAi}$* explants incubated on RIM for 8 day. (c) Width of each LRPs was measured. Different letters indicate significant difference by ANOVA Tukey-HSD, $p<0.05$. n=5~18. (d) *DR5::GUS* expression pattern in 60 hours-post-transfer *CPL4$_{RNAi}$* LRPs.

CPL4 enhances hypocotyl elongation in dark

Hypocotyl elongation is another auxin-regulated process (Nagpal et al. 2005; Oh et al. 2014). Light-grown $CPL4_{RNAi}$ seedlings did not show a significant difference in the hypocotyl growth (Figure 4.3); however, that of dark-grown $CPL4_{RNAi}$ seedlings were significantly shorter than that of wild-type seedlings (Figure 4.3a-b). By contrast, when CPL4 was overexpressed using EL2 promoter (CaMV 35S promoter with 2x enhancer), seedlings produced slightly longer hypocotyls in the dark (Figure 4.3a-c). Sucrose content in the media did not affect the phenotype, suggesting that the compromised hypocotyl elongation of $CPL4_{RNAi}$ seedlings is not due to carbohydrate starvation (Figure 4.3c). This is consistent with an alteration in auxin response in $CPL4_{RNAi}$; however, the impact of *CPL4* on this process was less prominent than that on the root phenotype.

Callus formation is not affected by $CPL4_{RNAi}$

The phenotype of $CPL4_{RNAi}$, such as reduced seedling lateral root development, abnormal auxin-induced lateral root formations, and alteration in snRNA metabolism are similar to the symptoms observed with *shoot regeneration defective2* mutant, which shows strong inhibition in *in vitro* cell proliferation and shoot organogenesis at the restrictive temperature (Yasutani et al. 1994; Ohtani and Sugiyama 2005). Therefore, I tested $CPL4_{RNAi}$ for hallmark phenotypes related to *in vitro* organogenesis. First, I examined responses of $CPL4_{RNAi}$ to a callus-inducing condition. $CPL4_{RNAi}$ roots but not hypocotyls were slightly more efficient to form calli in response to exogenous auxin (Figure 4.4). At 20-day post transfer to callus induction medium (CIM), wild-type root explants formed callus tissues only at their termini, while callus formation were

observed on throughout the *CPL4*$_{RNAi}$ root explants (Figure 4.4a-c). On 38th day, calli were observed regardless of genotypes or tissue origins.

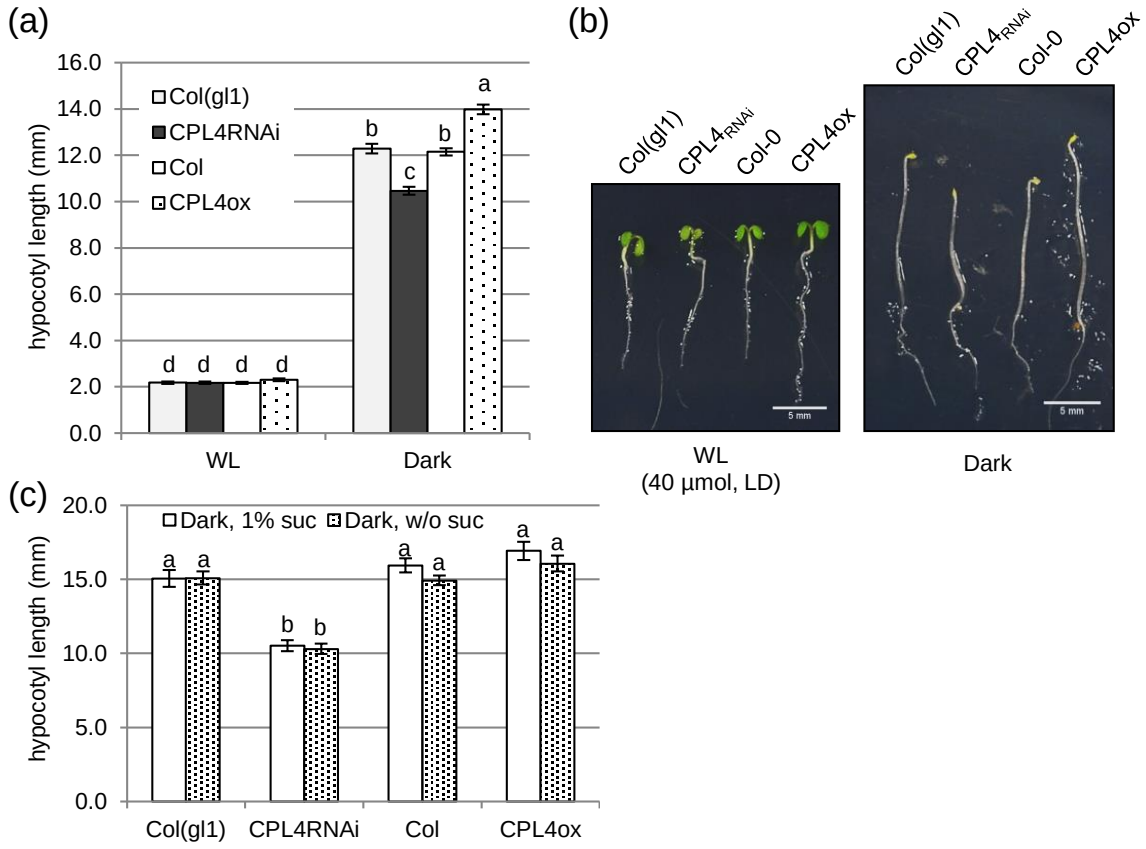


Figure 4.3 Effects of *CPL4*$_{RNAi}$ in hypocotyl elongation.
(a-c) Hypocotyl length (a) and picture (b) of 4-d old dark grown seedlings were measured. Bars represents SEM of 26~33 seedlings. (c) Seedlings were grown on ½ MS medium with or without 1% sucrose under darkness for 5 days. Hypocotyl length was measured. Bars represents SEM of 26~33 seedlings. Different letters indicate significant difference by Tukey-HSD, $p<0.05$.

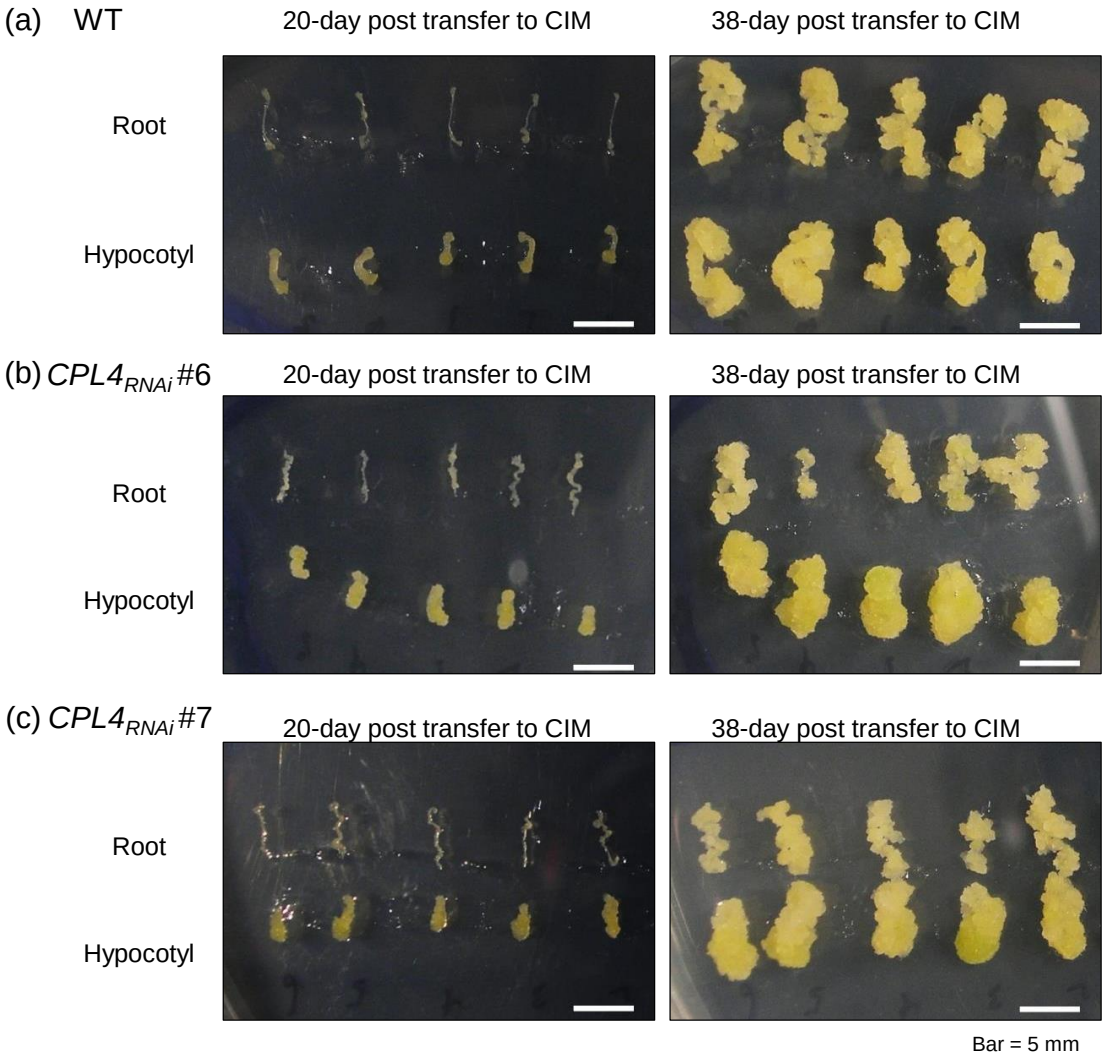


Figure 4.4 Callus induction from hypocotyl and root explants of *CPL4*$_{RNAi}$ lines.

Callus formation was induced by incubating hypocotyl and root explants from 8-day old seedlings on callus induction medium (CIM) containing 2 mg/L 2,4-D and 0.05 mg/L kinetin.(a) wild type, (b) *CPL4*$_{RNAi}$ #6 and (c) *CPL4*$_{RNAi}$ #7. Bars represent 5 mm.

De novo shoot organogenesis (DNSO) from root explants is enhanced in $CPL4_{RNAi}$

Next, I tested the DNSO capability of $CPL4_{RNAi}$ explants using *in vitro* shoot regeneration assay (Figure 4.5). To induce shoot organogenesis, the explants were pre-treated with CIM for 5 days, then shoot organogenesis were monitored on shoot induction medium containing 0.15 mg/L IAA and various concentration of cytokinins, including 0.5 mg/L 2iP (standard), 0.05 mg/L 2iP (1/10 of the standard) and 0 mg/L 2iP. Surprisingly, DNSO from root explants exhibited marked differences between wild type and $CPL4_{RNAi}$. $CPL4_{RNAi}$ root explants formed shoots much more vigorously than those of WT did under the standard condition. At the low 2iP concentration, $CPL4_{RNAi}$ root explants were able to form shoots, whereas WT root explants formed calli and roots but no shoot formation was observed. No callus or shoot formation was observed without 2iP in either genotype (Figure 4.5a-c). In the case of the hypocotyl explants, DNSO from $CPL4_{RNAi}$ were slower than wild type under the standard condition, but similar at low 2iP concentration. Root formations were observed in both genotypes on the media with low or no CK (Figure 4.5b,c). These results indicate that the $CPL4_{RNAi}$ specifically enhances DNSO capability from root explants. $CPL4_{RNAi}$ enhances responsiveness to CK; however, supplying exogenous cytokinin was still necessary.

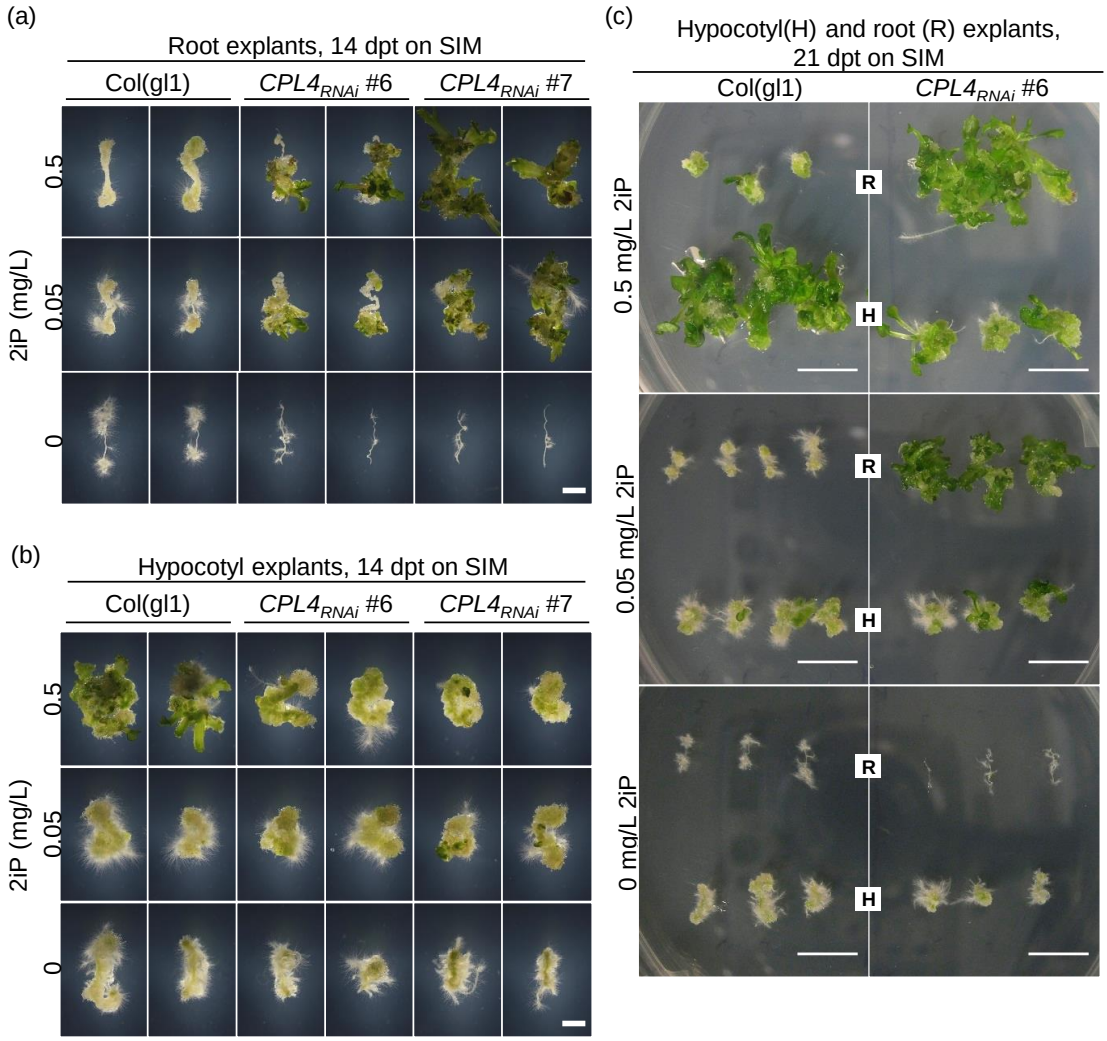


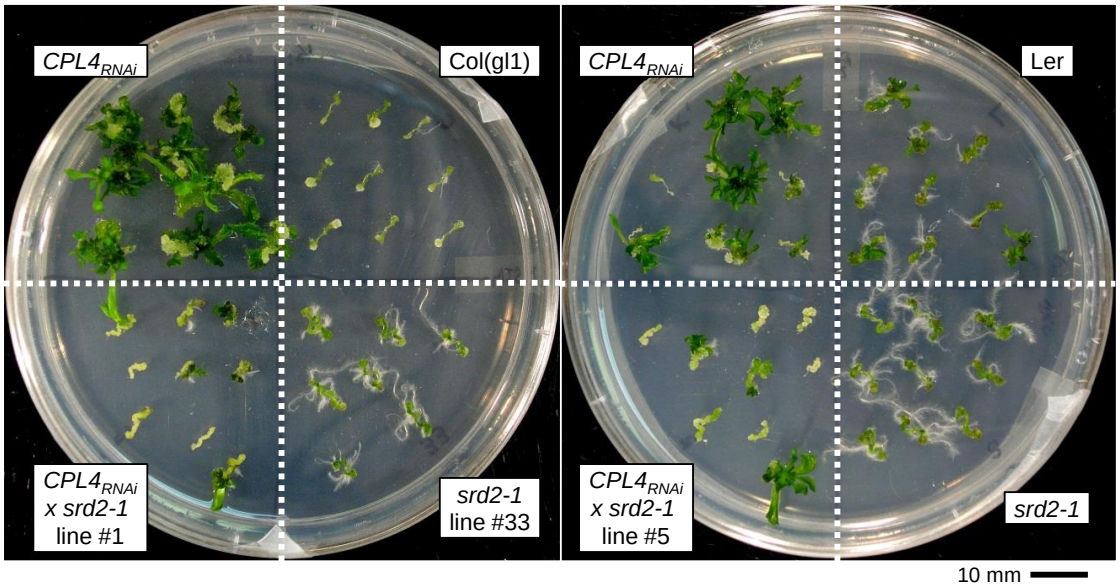
Figure 4.5 *CPL4* knockdown enhances *de novo* shoot organogenesis from root explants.

Hypocotyl and root explants excised from 15-day-old seedlings (see Materials and Methods for detail) were first incubated on callus induction medium containing 1xMS, 2% sucrose, 2.0 mg/L 2,4-D, and 0.05 mg/L Kinetin for 6 days, then transferred to shoot induction medium supplemented with 0.15 mg/L IAA and 0 – 0.5 mg/L 2iP. (a) Root explants incubated on SIM for 14 days. (b) Hypocotyl explants incubated on SIM for 14 days. (c) Explants incubated on SIM for 21 days. White bars indicate 2 mm in (a) and (b), 10 mm in (c).

DNSO enhancement in CPL4$_{RNAi}$ depends on SRD2

Despite that *CPL4$_{RNAi}$* and *srd2-1* share similar LRP development pattern and snRNA-related phenotype, *CPL4$_{RNAi}$* promoted DNSO from root explants rather than inhibiting it. To assess how *CPL4* and SRD2 genetically interact during DNSO, *CPL4$_{RNAi}$ srd2-1* lines (lines #1 and #5 established in Chapter III) were subjected to DNSO assay (Figure 4.6). At the permissive temperature, DNSO was less vigorous from the double mutants than from the parent *CPL4$_{RNAi}$*, but not completely abolished (Figure 4.6a). The DNSO of double mutant was strongly impaired at the restrictive temperature. Not only *CPL4$_{RNAi}$ srd2-1* produced no shoot on the SIM containing 0.5 mg/L 2iP, but also the calli formed from the root explants were less proliferative than parental lines (Figure 4.6b). This indicated that *CPL4* and *SRD2* might synergistically function at an early stage of DNSO, likely at the level of induction of organogenic calli.

(a) 21-dpt on SIM, at permissive temperature **22 ºC**



(b) 21-dpt on SIM, at restrictive temperature **28 ºC**

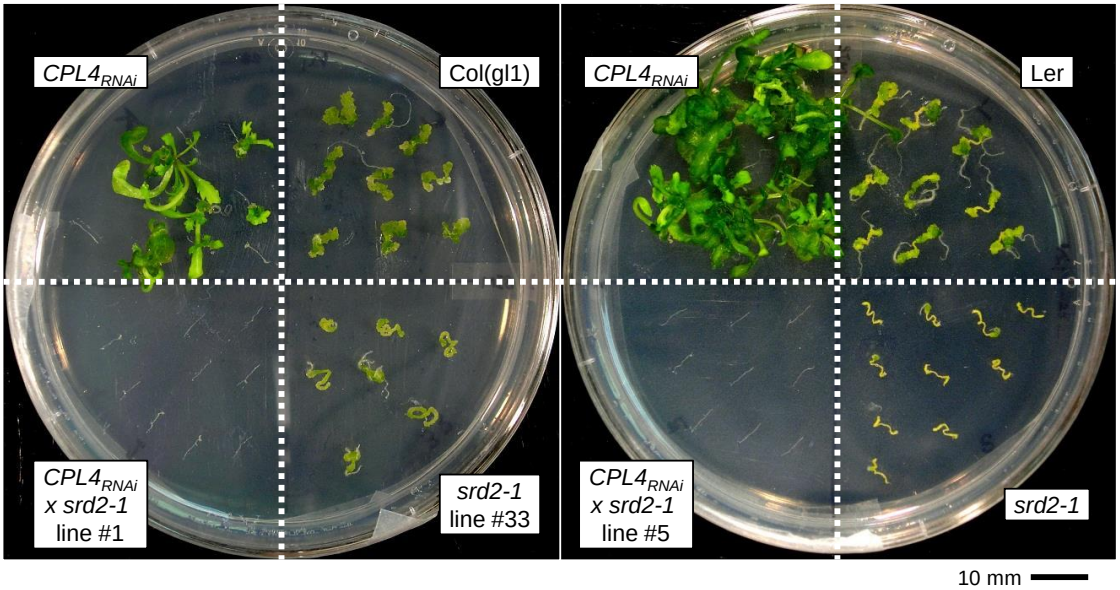


Figure 4.6 ***de novo*** **shoot organogenesis from** ***CPL4$_{RNAi}$ srd2-1*** **double mutant root explants.**

Root explants excised from 10-day-old seedlings were first incubated on CIM for 5 days, then transferred to standard SIM and incubated for 21 days at permissive temperature (a, 22 ^{o}C) or restrictive temperature (b, 28 ^{o}C). Black bars indicate 10 mm.

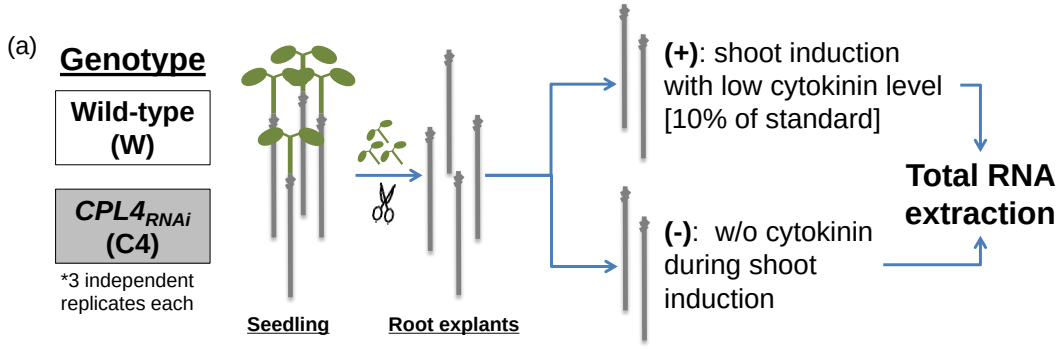
Transcriptome analysis identifies genes synergistically upregulated by $CPL4_{RNAi}$ and CK during early timepoint of DNSO

$CPL4_{RNAi}$ root explants exhibited highly efficient shoot organogenesis with a low concentration of 2iP. Because DNSO is a highly integrated developmental process, it is not immediately obvious in which aspect of DNSO CPL4 influences. To dissect molecular pathways that facilitate low-CK DNSO in $CPL4_{RNAi}$ root explants, transcriptome analysis using RNA-seq was performed on root tissues at the early stage of shoot induction with or without the low concentration (10% of the standard; 0.05 mg/L, Figure 4.7a) of 2iP. Analyses were conducted with samples from four conditions, namely, W(-), wild type on SIM without 2iP; W(+), wild type on SIM with 0.05 mg/L 2iP; C4(-); $CPL4_{RNAi}$ on SIM without 2iP; C4(+), $CPL4_{RNAi}$ on SIM with 0.05 mg/L 2iP, and differentially expressed transcripts relative to W(-) were identified for each condition. Similarly, comparisons were made between $CPL4_{RNAi}$ with or without CK, and $CPL4_{RNAi}$ and WT treated with CK. Because only C4(+) root explants are capable of DNSO (Figure 4.5), the 2,839 genes showing significantly higher expression at least by 1.5 in C4(+) than in W(-) are primarily investigated to identify pathways contributing to the DNSO capability. The 2839 genes are hereafter referred to as "C4(+)UT", C4(+)-Upregulated Transcripts (Figure 4.7b).

C4(+)UTs represent genes up-regulated by $CPL4_{RNAi}$, CK (2iP) or by synergistic effects of $CPL4_{RNAi}$, and CK. Among the 2839 genes, 1595 genes overlap with C4(-)UTs, indicating that this group of genes can be up-regulated solely by $CPL4_{RNAi}$ (Class II, Figure 4.8a). Similarly, 132 C4(+)UTs are found to be up-regulated in W(+), hence CK-

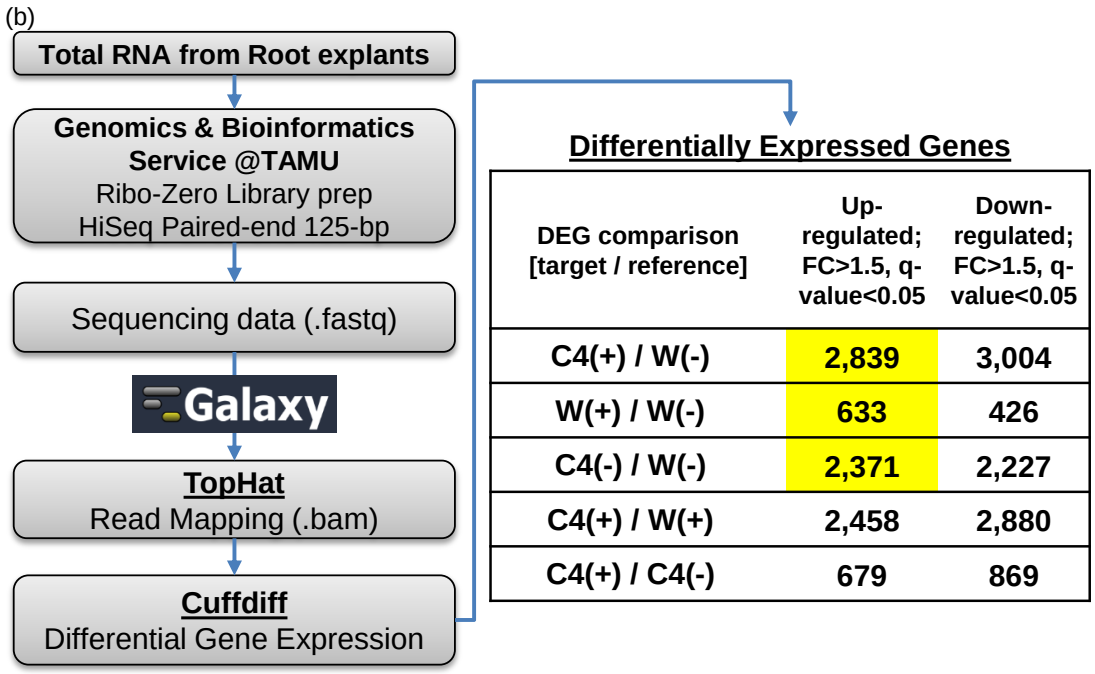
responsive (Class III). 334 C4(+)UTs overlap with both C4(-)UTs and W(+)UTs, indicating that this group of genes can be up-regulated by either $CPL4_{RNAi}$ or CK (Class IV). Class I represents 778 C4(+)UTs not overlapping with C4(-)UTs or W(+)UTs, indicating that their up-regulation requires both $CPL4_{RNAi}$ and CK. As shown in a heat map, however, overall expression pattern of class I genes are more similar to that of class II genes than to class III or class IV, albeit less intense (Figure 4.8b-e). Indeed, 527 out of 778 class I genes are significantly up-regulated in C4(-)/W(-) comparison without reaching the fold-change threshold (1.5 <), therefore can be considered as a subclass of class II. The 527 genes and the rest are termed "class I-a" and "class I-b", respectively.

In addition to the class I-IV criteria based on $CPL4_{RNAi}$ and/or CK – responsiveness, genes or pathways contributing to DNSO in C4(+) may express higher in C4(+) than in C4(-) and W(+) samples, which don't show DNSO. Thus, total 248 C4(+)UTs significantly up-regulated by 1.5 in C4(+)/C4(-) and C4(+)/W(+) comparisons were identified and hereafter referred to as C4(+)HITs, C4(+) Hyper-Induced Transcripts. C4(+)HITs are distributed in all classes I – IV (Figure 4.8a).



	Germination	Callus Induction	Shoot Induction
Cytokinin	none	**Kinetin**; 0.05 mg/L	**2iP**; **0.05 mg/L*** (+) or **0 mg/L** (-)
Auxin	none	**2,4-D**; 2 mg/L	**IAA**; 0.15 mg/L
Treatment	10-day	5-day	72 hours

*10% of standard, 0.5 mg/L



Differentially Expressed Genes

DEG comparison [target / reference]	Up-regulated; FC>1.5, q-value<0.05	Down-regulated; FC>1.5, q-value<0.05
C4(+) / W(-)	2,839	3,004
W(+) / W(-)	633	426
C4(-) / W(-)	2,371	2,227
C4(+) / W(+)	2,458	2,880
C4(+) / C4(-)	679	869

Figure 4.7 Low cytokinin DNSO RNA-seq experiment design and analysis workflow.

(a) Plant materials for total RNA extraction. Root explants were prepared from 10-day old seedlings by removing aerial portion. The root segments were pre-incubated on CIM for 5-day, then incubated on SIM with low 2iP concentration (0.05 mg/L) or without 2iP for 72 hours. (b) Total RNAs extracted from root explants were subjected to RNA-seq. See Materials and Methods for detail. FC, Fold change; W, wild type; C4, *CPL4*$_{RNAi}$; (-) and (+) indicate absence and presence of 0.05 mg/L 2iP, respectively. Yellow highlighted genes are used in the venn diagram analysis in Figure 4.8(a).

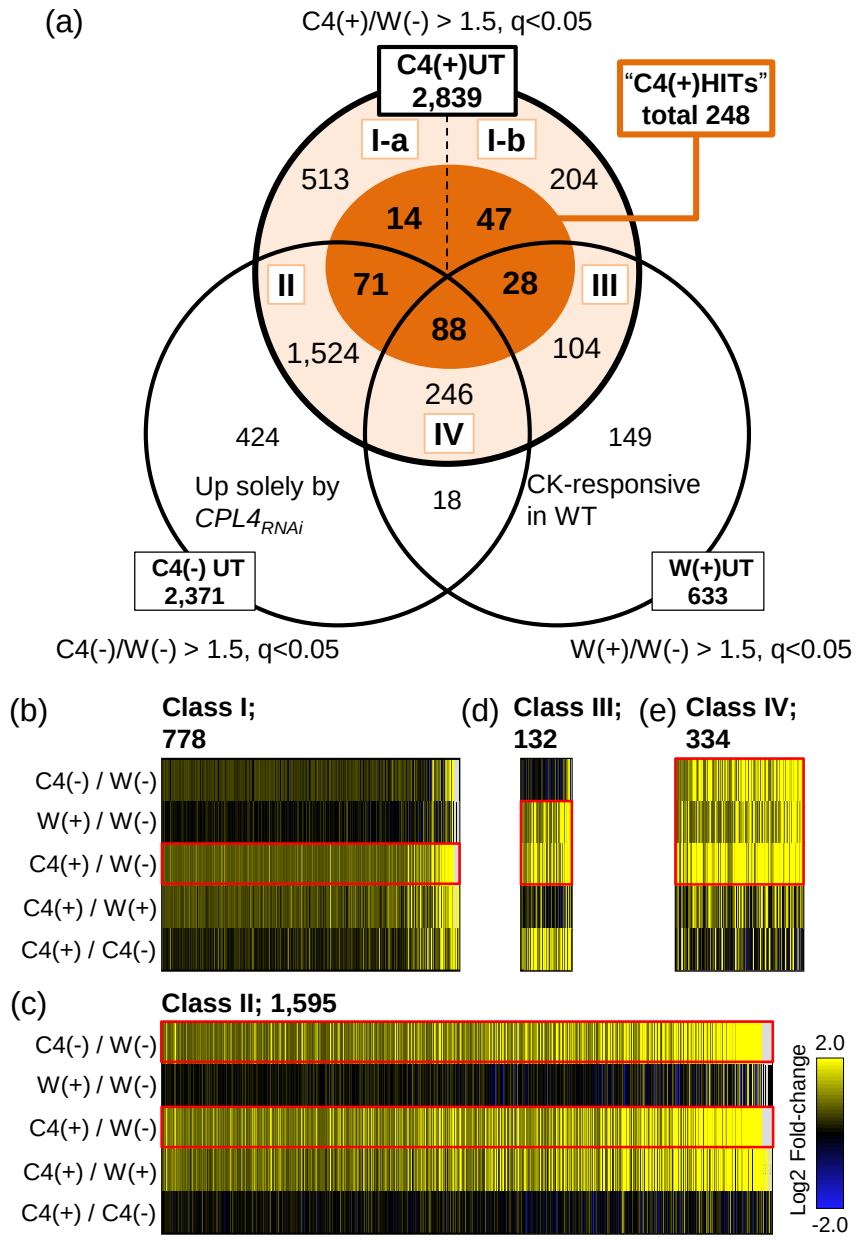
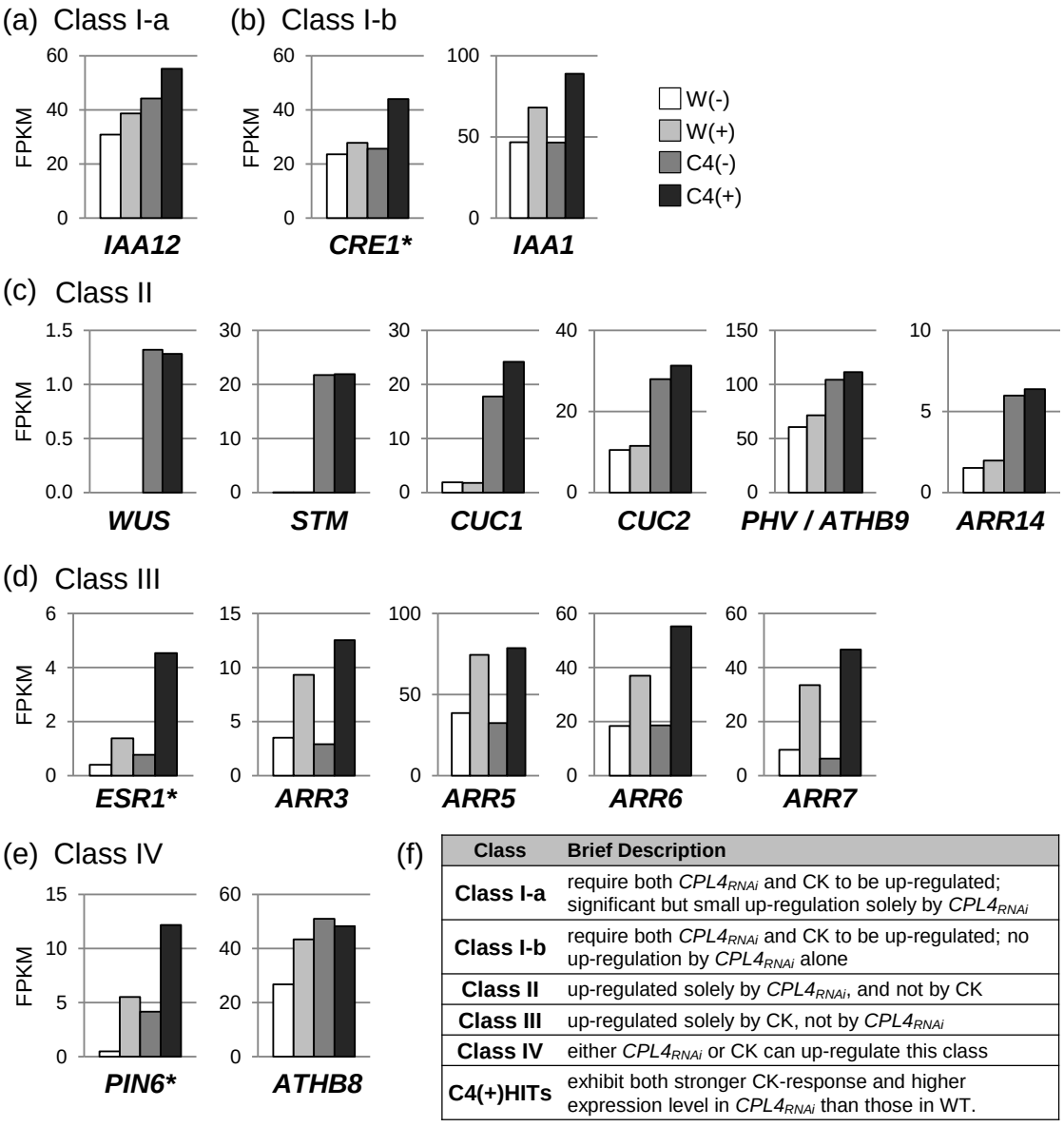


Figure 4.8 Classification of genes up-regulated by *CPL4*$_{RNAi}$ and/or CK.

(a) C4(+)UTs are classified into Class I – IV based on their overlap between C4(-)UTs and/or W(+)UTs. C4(+)UTs fulfilling the following criteria [C4(+)/C4(-) > 1.5, q < 0.05, AND C4(+)/W(+) > 1.5, q <0.05] are identified as C4(+)HITS, C4(+) Hyper Induced Transcripts. Roman numerals indicate the class. (b-e) Heat maps showing expression pattern of class I (b), II (c), III (d) and IV (e) genes in five DEG comparisons. Each column represent one gene, sorted by expression level (C4(-) fpkm large to small, left to right). The comparisons defining each class are highlighted in red.

Well-characterized genes contributing to DNSO include transcription factors and regulators involved in SAM maintenance, auxin-response and CK-response. Each class of C4(+)UTs are found to contain such DNSO-related genes (Figure 4.9). Several DNSO-related genes, such as an AP2/ERF transcription factor *ENHANCER OF SHOOT REGENERATION 1*(*ESR1*), a major CK-receptor, *CYTOKININ RESPONSE 1* (*CRE1*) and an auxin transporter *PIN FORMED* 6 (*PIN6*) are found among C4(+)HITs, implying that these genes contribute to the DNSO in C4(+). Intriguingly, essential SAM regulator genes such as *WUSCHEL* (*WUS*), *SHOOT MERISTEMLESS* (*STM*) and *CUP-SHAPED COTYLEDON* (*CUC*) turn out to be class II but not C4(+)UTs, indicating that their expression in $CPL4_{RNAi}$ roots is not dependent on CK in SIM (Figure 4.9c). This also implies that the constitutive expression of these key SAM regulators is not sufficient for DNSO in C4(-) roots. Class III contains CK-responsive type-A ARABIDOPSIS RESPONSE REGULATORs *ARR3*, *ARR5*, *ARR6* and *ARR7*, in addition to *ESR1* (Figure 4.9d). Auxin responsive genes *IAA1* and *IAA12* can be found in class I, with a slight difference in their expression patterns (Figure 4.9a-b).



(f)

Class	Brief Description
Class I-a	require both *CPL4*$_{RNAi}$ and CK to be up-regulated; significant but small up-regulation solely by *CPL4*$_{RNAi}$
Class I-b	require both *CPL4*$_{RNAi}$ and CK to be up-regulated; no up-regulation by *CPL4*$_{RNAi}$ alone
Class II	up-regulated solely by *CPL4*$_{RNAi}$, and not by CK
Class III	up-regulated solely by CK, not by *CPL4*$_{RNAi}$
Class IV	either *CPL4*$_{RNAi}$ or CK can up-regulate this class
C4(+)HITs	exhibit both stronger CK-response and higher expression level in *CPL4*$_{RNAi}$ than those in WT.

Figure 4.9 DNSO-related genes observed in class I-IV.
Expression levels of DNSO-related genes observed in (a) Class I-a, (b) Class I-b, (c) Class II, (d) Class III and (e) Class IV are shown in FPKM. Asterisks indicate C4(+)HITs. (f) Brief description of each category. See Figure 4.8 for the categorization detail.

Co-expression network of thalianol biosynthesis genes are highly up-regulated in $CPL4_{RNAi}$ root explants during DNSO on low CK

The observation that essential SAM regulators are not C4(+)HITs led us to inquire biological processes represented by this C4(+)-specific hyper-induced class of genes. Among 248 C4(+)HITs, only one co-expression cluster with more than 10 genes can be found by ATTEDII (Figure 4.10a). The C4(+)HITs-cluster contains the oxidosqualene cyclase thalianol synthase 1 (*THAS1*), cytochrome P450 thalianol Hydroxylase (*THAH*) and a BADH acyltransferase (*ACT*), which are the three of four thalianol synthesis genes tandemly arranged in an operon-like organization on chromosome 5 (Field and Osbourn 2008) (Figure 4.10b). The missing gene, thalian-diol desaturase (*THAD*), does not pass the C4(+)HITs criteria but is present in Class IV, with significantly higher expression in C4(+) over W(+) by 1.38-fold (Figure 4.10b). The four tandem genes synthesize and metabolize a tricyclic triterpene thalianol (Fig. 4.10a); THAS1 catalyzes the conversion of oxidosqualene to thalianol; THAH is responsible for converting thalianol into thalian-diol, which can be desaturated by THAD (Field and Osbourn 2008). Overaccumulation of thalianol or thalian-diol results in dwarf phenotype and enhanced root growth (Field and Osbourn 2008). The thalianol operon-like region is under the regulation by repressive histone modification H3K27me3 and chromatin remodeling factors such as *PICKLE* (*PKL*)/*CYTOKININE HYPERSENSITIVE 2* (*CKH2*) (Aichinger et al. 2009; Field et al. 2011; Yu et al. 2016).

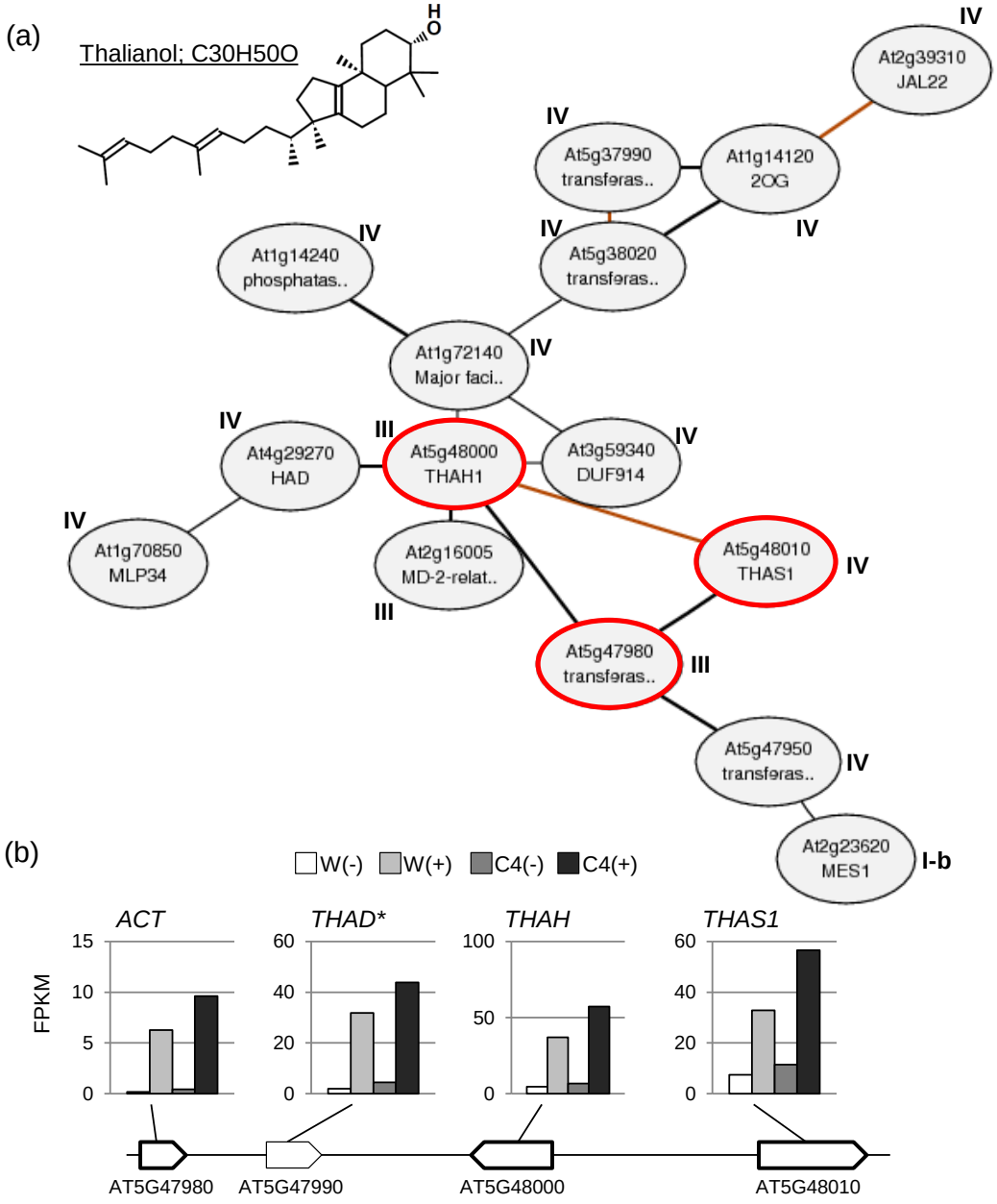


Figure 4.10 Thalianol biosynthesis-related co-expression network observed among C4(+)HITs.

(a) One co-expression cluster comprised of more than 10 genes among 248 C4(+)HITs is detected by NetworkDrawing tool in ATTED-II (http://atted.jp/). An associated class is indicated for each gene. The three genes from a thalianol-biosynthesis operon-like cluster on chromosome 5 are circled in red. The chemical structure of thalianol (C30H50O) is shown, drawn by PubChem Sketcher V2.4, based on the CID25229600. (b) FPKM values of the four thalianol biosynthesis genes arranged in an operon-like manner on chromosome 5. *THAD* (AT5G47990) was not C4(+)HITs, but included in class IV genes.

Because the thalianol network consists of mostly class III and class IV genes, and the class IV *THAD* does not show up as C4(+)HITs, the network analysis was expanded to include class III and IV genes in addition to C4(+)HITs. The expanded network analysis identified a large cluster of 175 genes among the 598 input genes (Figure 4.11). The Thalianol biosynthesis network expands to CK-response network consisting of type A-ARRs and CK dehydrogenases (CKXs), along with multiple branches containing C4(+)HITs. The connection between thalianol biosynthesis and CK-response networks is consistent with the up-regulated thalianol cluster genes in W(+) over W(-) samples (Figure 4.10b), and indicates that the thalianol cluster expression is regulated by cytokinin. The thalianol cluster also expands to several other sub-clusters with scattered C4(+)HITs, including syncytium formation and cell wall biosynthesis clusters.

C4(+)HITs include 26 transcription factors; 14 of them are associated with developmental process-related GOs; they also include 10 homeodomain-like transcription factors (Table 4.2). A few C4(+)HITs transcription factors can be found in the networks detected. WRKY43 is the only C4(+)HITs transcription factor directly connected to the network (Figure 4.11). Two homeobox transcription factors KNAT4 and HAT3 are connected to the network based on protein-protein interactions with ovate family protein 1 (OFP1) and another homeobox transcription factor HAT14, respectively (Figure 4.11). Although the involvement of these transcription factors is not clear due to the distance, these observations indicate that the thalianol biosynthesis pathway is highly up-regulated in $CPL4_{RNAi}$ root on the low 2iP SIM where only $CPL4_{RNAi}$ can exhibit DNSO.

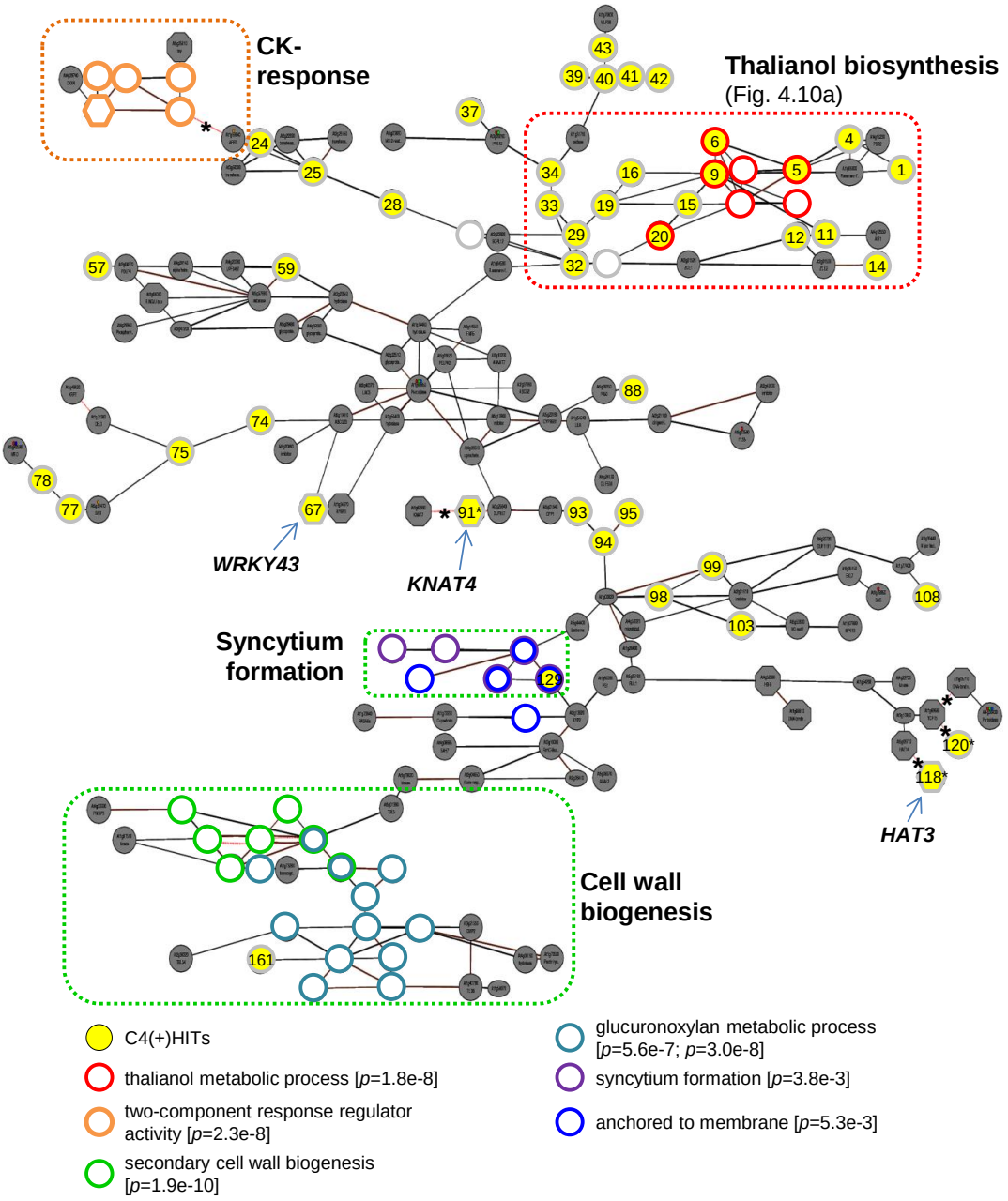


Figure 4.11 Co-expression network of 175 class III, class IV and C4(+)HITs genes.

Co-expression network analysis by ATTED-II identified a co-expression network of 175 genes among 598 genes (Class III, IV and C4(+)HITs combined). Each circle and polygon represents a gene and a transcription factor, respectively. Yellow-filled circle/polygons are C4(+)HITs, and their identities are summarized in Table 4.1. Circles with colored outline constitute sub-categories with significantly enriched GOs. The correspondence of a color to a GO along with *p*-value is described below the chart. Asterisks indicate protein-protein interaction.

Table 4.1 List of C4(+)HITs found in co-expression network analysis in Figure 4.11

# in Fig. 4.11	C4(+)HITs AGI	Class	Gene	Brief description
1	AT2G23620	I-b	MES1	methyl esterase 1
4	AT5G47950	IV	AT5G47950	HXXXD-type acyl-transferase family protein
5	AT5G47980	III	ACT	BADH; HXXXD-type acyl-transferase family protein
6	AT5G48010	IV	THAS1	thalianol synthase 1
9	AT5G48000	III	THAH	Thalianol hydroxylase; cytochrome P450, family 708, subfamily A, polypeptide 2
11	AT4G29270	IV	AT4G29270	HAD superfamily, subfamily IIIB acid phosphatase
12	AT1G70850	IV	MLP34	MLP-like protein 34
14	AT1G70880	I-b	AT1G70880	Polyketide cyclase/dehydrase and lipid transport superfamily protein
15	AT1G14240	IV	AT1G14240	GDA1/CD39 nucleoside phosphatase family protein
16	AT3G59340	IV	AT3G59340	solute carrier family 35 protein (DUF914)
19	AT1G72140	IV	AT1G72140	Major facilitator superfamily protein
20	AT2G16005	III	AT2G16005	MD-2-related lipid recognition domain-containing protein
24	AT1G13420	IV	ST4B	sulfotransferase 4B
25	AT2G25160	IV	CYP82F1	cytochrome P450, family 82, subfamily F, polypeptide 1
28	AT4G11190	IV	AT4G11190	Disease resistance-responsive (dirigent-like protein) family protein
29	AT5G38020	IV	AT5G38020	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
32	AT5G37990	IV	AT5G37990	S-adenosyl-L-methionine-dependent methyltransferase superfamily protein
33	AT1G14120	IV	AT1G14120	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
34	AT2G39310	IV	JAL22	jacalin-related lectin 22
37	AT4G37410	IV	CYP81F4	cytochrome P450, family 81, subfamily F, polypeptide 4
39	AT1G52410	I-b	TSA1	TSK-associating protein 1
40	AT3G16470	I-a	JR1	Mannose-binding lectin superfamily protein
41	AT2G34490	IV	CYP710A2	cytochrome P450, family 710, subfamily A, polypeptide 2
42	AT4G29700	III	AT4G29700	Alkaline-phosphatase-like family protein
43	AT4G23670	III	AT4G23670	Polyketide cyclase/dehydrase and lipid transport superfamily protein
57	AT2G38110	IV	GPAT6	glycerol-3-phosphate acyltransferase 6
59	AT3G44550	IV	FAR5	fatty acid reductase 5
67	AT2G46130	II	WRKY43	WRKY DNA-binding protein 43
74	AT2G35770	II	scpl28	serine carboxypeptidase-like 28
75	AT1G22880	III	CEL5	cellulase 5
77	AT1G23160	IV	AT1G23160	Auxin-responsive GH3 family protein
78	AT3G46170	II	AT3G46170	NAD(P)-binding Rossmann-fold superfamily protein
88	AT5G58910	IV	LAC16	laccase 16
93	AT1G16530	IV	ASL9	ASYMMETRIC LEAVES 2-like 9
94	AT3G02885	IV	GASA5	GAST1 protein homolog 5
95	AT2G19990	II	PR-1-LIKE	pathogenesis-related protein-1-like protein
98	AT4G18630	IV	AT4G18630	hypothetical protein (DUF688)
99	AT3G27400	IV	AT3G27400	Pectin lyase-like superfamily protein
103	AT5G48800	IV	AT5G48800	Phototropic-responsive NPH3 family protein
108	AT3G61920	IV	AT3G61920	UvrABC system protein C
129	AT5G07030	IV	AT5G07030	Eukaryotic aspartyl protease family protein
161	AT1G78120	IV	TPR12	Tetratricopeptide repeat (TPR)-like superfamily protein
91*	AT5G11060	II	KNAT4	homeobox protein knotted-1-like 4
118*	AT3G60390	I-b	HAT3	homeobox-leucine zipper protein 3
120*	AT4G36110	III	SAUR9	SAUR-like auxin-responsive protein family

Each number corresponds to a yellow-filled circle/polygon in Figure 4.11. Asterisks indicate genes connected by protein-protein interaction.

Table 4.2 List of transcription factors in C4(+)HITs

Gene	Class	TF-brief description	Associated biological processes (other than TF)
ESR1	**III**	**Integrase-type DNA-binding superfamily protein**	**animal organ morphogenesis; auxin-activated signaling pathway; cotyledon development; cytokinin-activated signaling pathway; embryonic pattern specification; ethylene-activated signaling pathway; response to cytokinin; specification of plant organ axis polarity**
PTL	IV	Duplicated homeodomain-like superfamily protein	establishment of petal orientation; negative regulation of organ growth; perianth development; petal development; regulation of flower development; response to chitin; sepal development
HB51	II	homeobox 51	bract formation; floral meristem determinacy; regulation of timing of transition from vegetative to reproductive phase
ANT	IV	Integrase-type DNA-binding superfamily protein	animal organ morphogenesis; cell differentiation; flower development; gamete generation; glucosinolate metabolic process; maintenance of shoot apical meristem identity; regulation of cell proliferation
MYB39	I-b	myb domain protein 39	cell differentiation
MEE3	IV	Homeodomain-like superfamily protein	embryo development ending in seed dormancy; gravitropism; response to red light
MYB67	IV	myb domain protein 67	cell differentiation
bHLH093	I-a	beta HLH protein 93	gibberellin catabolic process; multicellular organism development; regulation of gibberellin biosynthetic process
WRKY13	II	WRKY DNA-binding protein 13	positive regulation of sclerenchyma cell differentiation; regulation of lignin biosynthetic process
NAC005	III	NAC domain containing protein 5	multicellular organism development; xylem development
ZFP5	IV	zinc finger protein 5	cytokinin-activated signaling pathway; gibberellic acid mediated signaling pathway; multicellular organism development; xylem development
bZIP44	IV	basic leucine-zipper 44	seed germination
NAC103	IV	NAC domain containing protein 103	multicellular organism development
MYB77	II	myb domain protein 77	cell differentiation; lateral root development; response to chitin; response to ethylene; response to salicylic acid
LHY	II	Homeodomain-like superfamily protein	circadian rhythm; long-day photoperiodism, flowering; negative regulation of circadian rhythm; response to abscisic acid/auxin/cadmium ion/cold/ethylene/GA/JA/SA/salt
ERF71	II	Integrase-type DNA-binding superfamily protein	ethylene-activated signaling pathway; response to anoxia
ESE3	IV	Integrase-type DNA-binding superfamily protein	ethylene-activated signaling pathway
BZIP27	II	basic region/leucine zipper motif 27	response to chitin
HSFA3	II	heat shock transcription factor A3	response to chitin; response to heat;
HAT3	I-b	homeobox-leucine zipper protein 3	-
AT1G14600	II	Homeodomain-like superfamily protein	-
KNAT4	**II**	**homeobox protein knotted-1-like 4**	**response to light stimulus**
HB33	III	homeobox protein 33	abscisic acid-activated signaling pathway; positive regulation of transcription, DNA-templated; response to abscisic acid; seed germination
WRKY66	I-b	WRKY DNA-binding protein 66	-
WRKY43	**II**	**WRKY DNA-binding protein 43**	**-**
AT3G07340	IV	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	-

DNA-replication and cell cycle regulation pathways are up-regulated in CPL4$_{RNAi}$ root explant regardless of CK level in SIM

C4(+)HITs represents 8.4% of all genes up-regulated in C4(+) samples, whereas class II genes, which does not require CK to be up-regulated in C4(+) samples, consists 56.2% of all C4(+)UTs examined (Figure 4.8a). Moreover, this class contains the essential SAM regulator genes such as *WUS*, *STM* and *CUC* transcription factors (Figure 4.9c). To identify other potential DNSO-related genes/pathways which do not require CK in SIM in *CPL4$_{RNAi}$*, I conducted GO enrichment overrepresentation analysis on the 1,595 class II genes (Table 4.3).

The result shows that class II is highly enriched for GOs related to DNA conformation change, DNA replication and cyclin-dependent protein kinase activity, indicating that the *CPL4$_{RNAi}$* roots after CIM pre-activation are actively going through cell divisions which require both DNA replication and cell cycle progression regulated by cyclin-dependent kinases (Table 4.3). In addition to many core histone subunit genes (*HTA*, *HTB*, *HTR* and *HTF*), DNA replication-related class II genes include a suit of basic DNA replication machineries such as a DNA polymerase delta subunit (POL4D), DNA primases (POLA3, EMB2813), a DNA clamp Proliferating Cellular Nuclear Antigen (PCNA1), DNA helicase Minichromosome Maintenance family involved in formation of DNA replication forks (Brewster and Chen 2010). (MCMs), Replication-factor Protein A (RPA), Origin Recognition Complex subunits (ORCs), a GINS complex subunit involved in initiation and progression of DNA replication fork (SLD5), Cell Division Control family (CDC6, CDC45), and a pre-replication complex interacting

Armadillo BTB protein (ABAP1). A ribonucleotide reductase participating in dNDP synthesis during cell division (TSO2) is also observed. The Cyclin-Dependent Kinases (CDK) CDKB1;1 and CDKB2;2, along with co-factor cyclins (CYCA, CYCB, CYCD and CYCPs) are identified in GOs related with CDK activity regulation. It is noteworthy that one of the overexpressed the cyclin D, CYCD3;1, has been shown to be a rate limiting factor in CK response, and its overexpression results in ectopic cell division (Riou-Khamlichi et al. 1999; Dewitte et al. 2007). All these observations clearly indicate that both C4(-) and C4(+) root cells are more proliferative than wild type, regardless of the presence of low cytokinin in the SIM.

$CPL4_{RNAi}$ roots activate DNSO-regulator genes during CIM pre-incubation

The SIM-CK independent expression of SAM regulator in *$CPL4_{RNAi}$* raised a possibility that these genes are activated during CIM-pre-incubation period where a low level of CK (0.05 mg/L Kinetin) is included. To address this, the expression levels of select DNSO-related genes were examined in untreated, CIM-pre-incubated and low-2iP-SIM-incubated roots of wild type and *$CPL4_{RNAi}$* (Figure 4.12a). Interestingly, *WUS* and *STM* expressions are observed in the CIM pre-incubated *$CPL4_{RNAi}$* roots at a level similar to those after the low-2iP-SIM incubation, suggesting that the CIM-pre-incubation alone is enough for *WUS* and *STM* activation in *$CPL4_{RNAi}$*. As for *WUS*, expression was detected even in the untreated *$CPL4_{RNAi}$* root. The CK-responsive *ARR5* expression can be activated by CIM-pre-incubation and *$CPL4_{RNAi}$* roots showed lower response than wild type, confirming that the CK-response itself is not hyper-induced by *$CPL4_{RNAi}$*. CIM-pre-incubation did not strongly induce *ESR1* and *CUC1*, but an early

DNSO marker CUC2 showed highest expression in CIM-pre-incubated $CPL4_{RNAi}$ roots. The mitotic cyclin CYCB1;1 and a histone 3 gene (*HTR14*) show expression patterns similar to *CUC2*, indicating that the activation of cell division and acquisition of competence to DNSO represented by *CUC2* expression during CIM-pre-incubation are enhanced by the knockdown of *CPL4*.

Table 4.3 Gene Ontologies (GO) highly enriched in class II genes

Class II (1,595) Gene Ontologies							
Type	**GO_ID**	**Subcategory**	**expected gene #**	**observed gene #**	**p-value (fdr)**	**Fold-enrich ment**	**Class II Genes**
BP	**GO:0071103**	**DNA conformation change**	**5.07**	**29**	**5.36E-13**	**5.7**	
BP	GO:0006323	DNA packaging	3.57	25	1.97E-13	7.0	
BP	GO:0006333	chromatin assembly or disassembly	4.29	25	1.97E-11	5.8	HTA2, HTA5, HTA10, HTA13, HTB1, HTB2, HTB3, HTB5, HTB6, HTB7, HTB10, HTR14, HTR2, HTR3, HTR6, HTR8, HTR9, HF01, HF05, HF07, GAMMA-H2AX, MCM2, MCM3, MCM4, MCM5
BP	GO:0034728	nucleosome organization	3.29	25	4.54E-14	7.6	
BP	GO:0031497	chromatin assembly	3.40	25	8.88E-14	7.4	
BP	GO:0065004	protein-DNA complex assembly	3.45	25	1.13E-13	7.2	
BP	GO:0006334	nucleosome assembly	3.29	25	4.54E-14	7.6	
BP	**GO:0006260**	**DNA replication**	**6.18**	**31**	**3.29E-12**	**5.0**	POLD4, POLA3, EMB2813, PCNA1, MCM2, MCM3, MCM4, MCM5, MCM10, RPA2, RPA70C, RPA70D, AT4G19130, ORC2, ORC6, CDC6, CDC45, SLD5, ABAP1, ETG1, TSO2, WEE1, CDKB1;1, CKS1, CYCA2;1, SIM, KRP6, ATE2F2, MYB3R-4, FZR2, AT3G02820
BP	GO:0006261	DNA-dependent DNA replication	3.06	19	2.61E-09	6.2	
BP	GO:0006270	DNA-dependent DNA replication initiation	0.67	7	1.95E-05	10.5	
MF	GO:0016538	cyclin-dependent protein kinase regulator activity	1.84	11	1.89E-05	6.0	CYCA1;1, CYCA1;2, CYCA2;1, CYCB1;1, CYCB2;3, CYCD3;1, CYCD4;1, CKS2, ICK6, KRP6, SIM
MF	GO:0004693	cyclin-dependent protein kinase activity	1.78	10	9.32E-05	5.6	CDKB1;1, CDKB2;2, CYCD4;2, CYCD6;1, cycp3;1, CYCP3;2, CYCP4;2, CKS1, CKS2, SDS
BP	**GO:0070192**	**chromosome organization involved in meiosis**	**1.06**	**8**	**6.90E-05**	**7.6**	PHS1, HOP2, SDS, DMC1, ZYP1a, ZYP1b, ASY1, RCK
BP	GO:0007129	synapsis	1.06	8	6.90E-05	7.6	

GOs enriched in the class II (1,595) genes are retrieved by using GeneTrail with following thresholds; false discovery rate (fdr) adjusted p-value < 1E-04; minimum gene # in a category = 5; fold-enrichment (observed # of genes / expected # of genes)> 4. Mutually related GOs are grouped, and genes detected in the largest GO shown in bold in the group are listed. MF, Molecular Function; BP, Biological Process.

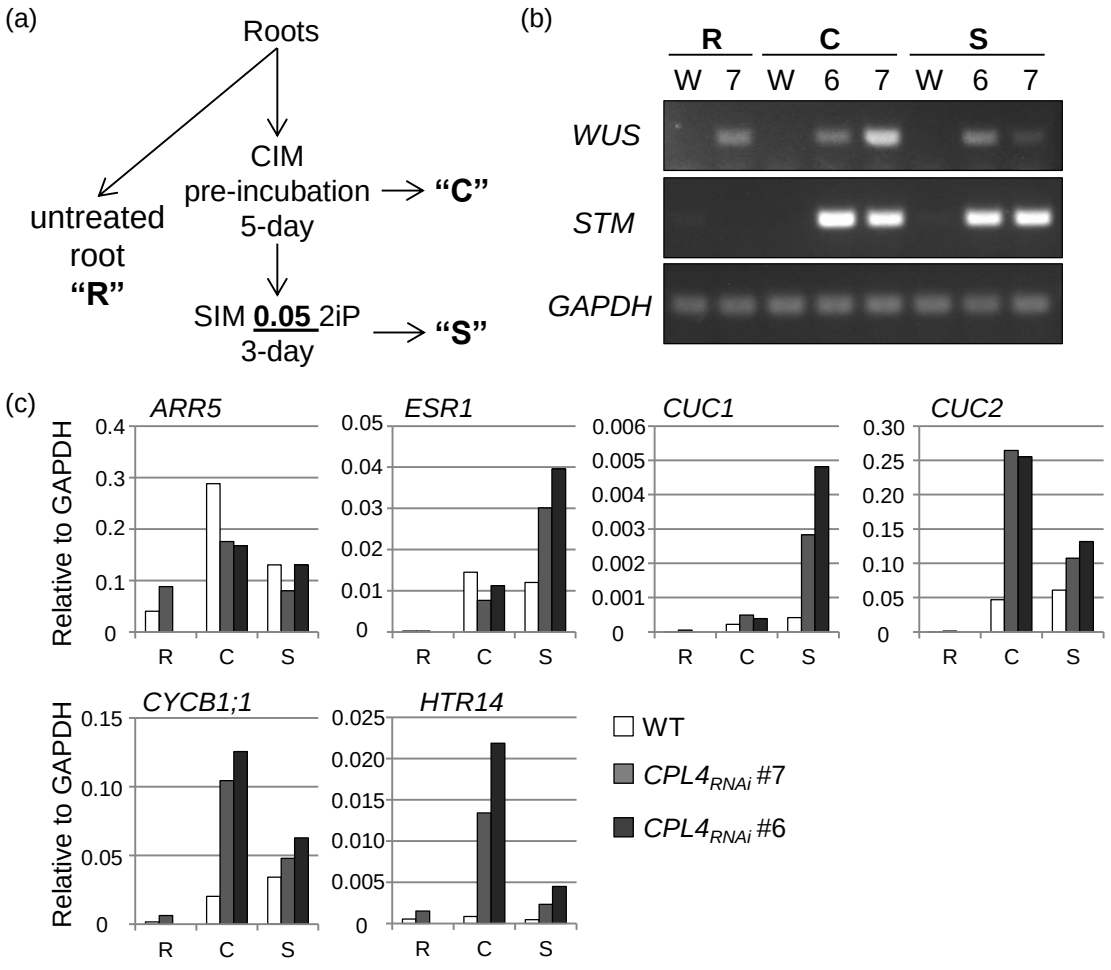


Figure 4.12 DNSO-related gene expression pattern during CIM pre-incubation.
(a) Experimental design. Roots from 14-day old seedling are subjected to pre-CIM incubation for 5 days, then moved to SIM containing 0.05 mg/L 2iP (labeled as "S"). Root samples right after the pre-CIM incubation were also harvested ("C"). Independently harvested roots are used as untreated control ("R"). (b) RT-PCR and (c) RT-qPCR of random-primed cDNA prepared from total RNA of each samples. Expression level of each gene is normalized to that of the reference gene (*GAPDH*).

4.4 Discussion

Recent studies have identified the importance of basic transcriptional machinery and RNA metabolisms in DNSO from *in vitro* cultured explants of *Arabidopsis thaliana* (Furuta et al. 2011; Kubo et al. 2011). In this chapter, I characterized functions of CPL4 in callus formation and *de novo* organogenesis of lateral root and shoot. $CPL4_{RNAi}$ roots exhibit defective lateral root primordia (LRP) development and $CPL4_{RNAi}$ enhances the potential for DNSO from root explants in *SRD2*-dependent manner. The high DNSO potential allows $CPL4_{RNAi}$ root explants to regenerate shoot on SIM with 1/10 dose of CK required for wild-type root explants. Further transcriptome analysis revealed that genes related to SAM regulation, DNA replication and cell cycles are activated in $CPL4_{RNAi}$ roots regardless of the presence of CK in the SIM, while several DNSO-related genes such as *CRE1* and *ESR1*, along with a suite of thalianol biosynthesis genes turn out to be highly expressed in $CPL4_{RNAi}$ roots on the low 2iP concentration SIM. CIM-pre-incubation alone was able to activate DNSO related genes such as *STM* and *CUC2* along with cell division marker genes in $CPL4_{RNAi}$, resulting in higher expression of *ESR1* and *CUC1* in the following low-2iP-SIM incubation. These results indicate that CPL4 negatively regulates cell proliferation and acquisition of competence for DNSO during CIM-pre-incubation stage.

$CPL4_{RNAi}$ seedlings failed to form lateral roots even after 100 nM IAA treatment, while the primary root growth and *DR5::GUS* reporter expression pattern responded to the exogenous IAA as much as wild type does (Figure 4.1b,d). Therefore, $CPL4_{RNAi}$ is unlikely to be defective in auxin perception or response. Local and concentrated

expression of *DR5::GUS* reporter gene is observed in $CPL4_{RNAi}$ primary roots (Figure 4.1e-f), indicating that the initiation of lateral root was not abolished by $CPL4_{RNAi}$. When lateral root formation is semi-synchronously induced from explants by incubating on RIM, lateral root primordia (LRP) from $CPL4_{RNAi}$ became fasciated and showed abnormal cell organization. A previous study found that CK can inhibit lateral root formation, and exogenous CK application can disturb cellular organization in LRPs (Laplaze et al. 2007). As seen in the DNSO experiment, $CPL4_{RNAi}$ lines are more sensitive to endogenous CK (Figure 4.5), and this could be inhibitory to LRP development. Because the lateral root primordia initiation and initial process of callus formation share the same pathway (Sugimoto et al. 2010), it is possible that the $CPL4_{RNAi}$ LRPs are more prone to the callus and DNSO fate rather than establishing lateral roots due to the lower CK requirement. The $CPL4_{RNAi}$ roots could strongly activate *CUC2* expression during CIM-pre-incubation, which is a hallmark of a commitment to DNSO (Gordon et al. 2007).

Transcriptome analysis identified several classes of genes based on their expression pattern (Fig. 4.8). More than half of the genes up-regulated in C4(+) over W(-) samples were CPL4-knockdown dependent (class II), rather than CK-dependent (Class III). The class II genes are highly enriched with basic DNA replication machinery and cell division regulators, in addition to key SAM regulators *WUS*, *STM* and *CUC1/2* (Figure 4.9, Table 4.3). The CIM-pre-incubation experiment showed that their expression can be up-regulated by CIM treatment alone (Figure 4.12). On the other hand, C4(+)HITs showed specific up-regulation in $CPL4_{RNAi}$ root on SIM with low CK. The

major cytokinin receptor *CRE1* is up-regulated by the low CK in $CPL4_{RNAi}$, but not in the wild type. The expression of AP2/ERF transcription factor *ESR1* was up-regulated in wild type in response to the low CK, but the basal expression levels, as well as induction level, are significantly higher in $CPL4_{RNAi}$. The hyperinduction of *ESR1* was not observed in CIM-pre-incubated samples (Figure 4.12c), indicating that CIM-pre-incubation is a prerequisite but not sufficient for *ESR1* hyperactivation in $CPL4_{RNAi}$ on the low-CK-SIM. The hyper-activation of *ESR1* might directly explain the difference in the shoot regeneration capacity at the low CK condition.

One prominent co-expression network found among C4(+)HITs involves a cluster of four tandemly-arranged thalianol biosynthesis and metabolism genes (Figure 4.10). Thalianol is produced from oxidosqualene by THAS, and it is further modified by THAH and THAD resulting in thalian-diol and desaturated thalian-diol, respectively (Field and Osbourn 2008). Thalianol cluster gene expression occurs mainly in roots, and constitutive overexpression of THAS resulting in accumulation of thalianol in shoot portion caused dwarfism and enhanced root growth (Field and Osbourn 2008). Additionally, the operon-like thalianol cluster is a target of repressive histone modification H3K27me3 (Aichinger et al. 2009; Field et al. 2011), and expression of the thalianol cluster is dependent on a chromatin remodeling factor PKL/CKH2 (Yu et al. 2016), whose mutation confers CK-hypersensitivity in terms of proliferation of green calli (Kubo and Kakimoto 2000; Furuta et al. 2011). Therefore, the mutation in *PKL* which is necessary for expression of thalianol cluster confers CK-hypersensitivity in the instance. On the other hand, wild-type root tissues could induce the thalianol cluster

expression in response to the low-CK-SIM, indicating that it can be activated by CK (Figure 4.10b). Although no implication of the tricyclic triterpene compound functions in DNSO has been made, it is tempting to hypothesize that the enhanced thalianol production in C4(+) samples is partially responsible for the enhanced DNSO. Direct application of thalianol to media, or use of genotypes overaccumulating thalianol such as THAS-overexpression or a *thah* mutant in the DNSO experiment will provide us a clue. Further characterization of thalianol cluster expression pattern in the DNSO process and in $CPL4_{RNAi}$ would reveal the function and regulation of the operon-like metabolic gene cluster in development and organogenesis.

Interestingly, $CPL4_{RNAi}$ root explants on SIM without CK could not form shoot despite the highly-activated stem cell regulators, indicating that exogenous CK application is still required for the shoot-regeneration. In addition to mRNA expression level, temporal and spatial regulations of the key SAM regulator genes during *de novo* organogenesis and active meristem formation have been established and well characterized (Gordon et al. 2007). One possibility is that the knockdown of CPL4 allows transcriptional activation of those genes, but cannot bypass the requirement of exogenous CK for their proper spatial regulation. In this regard, $CPL4_{RNAi}$ lines can be a useful genetic resource for studying the function of the low level of CK in DNSO. Combinations of $CPL4_{RNAi}$ and various fluorescence reporter genes would provide a better understanding of how the enhanced DNSO in $CPL4_{RNAi}$ root explants is achieved and how exogenous CK is still required in the $CPL4_{RNAi}$ roots expressing those SAM regulators.

Intriguingly, the double mutant analysis shows that the enhanced DNSO from *CPL4*$_{RNAi}$ root explants is reliant on *SRD2*, an essential snRNA activator complex subunit. The inhibition of DNSO in the double mutant can be observed even at the permissive temperature, indicating that *CPL4*$_{RNAi}$ and *srd2-1* when combined synergize to inhibit the DNSO. As shown in Chapter III, *CPL4*$_{RNAi}$ results in 3'-end processing defects of most pol II-dependent snRNAs. When present, the snRNA extension can run into a downstream protein-coding gene and become a polyadenylated translatable snR-DPG fusion transcript. However, the functions of CPL4 in active snRNA content and formation of snRNPs remain uncharacterized. Although it should be noted that the DNSO defect in the double mutant could be due to pathological growth defects derived from the two mutations both of which affect normal growth and development, the hypostatic nature of *CPL4*$_{RNAi}$ in the double mutant analysis suggests that CPL4 negatively affects the SRD2-mediated regulation of DNSO (Figure 4.6), i.e., controlling dosage of snRNA during DNSO. Precise quantitation and characterization of spliceosomal snRNPs and their activities in *srd2-1*, *CPL4*$_{RNAi}$ and the double mutant lines would reveal the molecular basis of the snRNA requirement for DNSO.

CHAPTER V

CONCLUSIONS

Phosphoregulation of RNA polymerase II C-terminal domain is an integral part of molecular mechanisms that make eukaryotic cell live. Although the molecular detail has been well studied in other systems in vertebrates, little is known about factors involved and consequences of misregulation of pol II-CTD phosphor-regulation in plants. My dissertation characterizes the physiological and molecular functions of an essential pol II-CTD phosphatase in *Arabidopsis thaliana*, C-terminal domain Phosphatase-Like 4 (CPL4).

Chapter II establishes CPL4 as a bona fide pol II-CTD phosphatase by showing a physical interaction between CPL4 and RNA polymerase II complexes, and by detecting catalytic activities of bacterial recombinant proteins and tandem-affinity purified proteins *in vitro*. Overexpression and knock-down (RNAi) of *CPL4* results in reduced and increased level of pol II-CTD phosphorylation status *in vivo*, indicating that the CPL4 is a major pol II-CTD phosphatase in Arabidopsis thaliana. Male gamete lethality of a null mutant allele *cpl4-2* supports the essential function of *CPL4*. Transcriptome analysis by microarray reveals that CPL4 negatively regulates a suite of xenobiotic stress responsive genes. Knock-down of *CPL4* results in basal activation of the xenobiotic stress responsive genes in normal growth condition, which renders resistance to a general herbicide chlorosulfuron during its post-germinative growth.

Chapter III describes the function of CPL4 in transcriptional regulation of a major class of non-coding RNA, snRNA. As revealed by RNA-seq and RT-PCR analysis, knock-down of *CPL4* results in 3'-extension of transcripts from pol II-dependent snRNA loci. The 3'-extended snRNA transcript can fuse into a downstream protein-coding gene if present. Such snRNA-mRNA fusion transcript can be translated into functional protein, and the expression of downstream protein coding gene can be upregulated either by removing the snRNA 3'-end processing signal or by knocking down *CPL4*. Many potential snRNA-mRNA fusion transcripts and several with EST-tag support can be found in many plant species across the kingdom. Such snRNA-mRNA fusion might be a common feature in higher plants, facilitated by transposons embedding pol II-dependent snRNA promoters. I also identified one protein-coding gene locus preceded by the transposon-embedded pol II-dependent promoter, which produces short unstable transcripts in wild-type plants. The short transcripts can be extended into the full-length protein-coding transcript by knocking down *CPL4*. Importantly, the snRNA 3'-extension and snRNA-mRNA fusion production can be induced by salt stress, which also causes global dephosphorylation of pol II-CTD. These results indicate that the pol II-CTD phosphoregulation plays pivotal roles in switching snRNA transcription and snRNA-mRNA fusion transcription at these loci.

Finally, in Chapter IV, I describe the function of CPL4 in *de novo* shoot organogenesis which has been shown to involve not only stem cell and meristem regulator genes but also snRNA transcription activators, general transcription factor components and chromatin remodeling machinery. Knockdown of *CPL4* enhances *de*

novo shoot formation from root explants, requiring less cytokinin. The cytokinin hypersensitiveness allows *CPL4* knockdown explants to activate the expression of shoot apical meristem regulators on callus induction media. The enhanced *de novo* shoot organogenesis capacity might be correlated with a tricyclic triterpene thalianol, whose physiological significance in the pathway is to be determined.

In summary, this dissertation work establishes CPL4 as a genuine and essential pol II-CTD phosphatase in *Arabidopsis thaliana*. CPL4 negatively regulates expression of genes involved in xenobiotic stress responses and *de novo* shoot organogenesis, repressing respective physiological outcome. Also, CPL4 is involved in snRNA 3'-end processing/termination, which can be a novel regulatory switch for a production of non-coding RNA or protein-coding RNA, depending on the pol II-CTD phosphorylation status regulated by CPL4 and environmental cues.

The pol II-CTD consists of more than 30 repeats of a heptapeptide, each with five potential phosphorylation sites and two isomerization sites. The vast possible combinations of post-translational modification status represent a massive regulatory potential centered on the unique C-terminal domain throughout transcription. Although gene expression regulations have been extensively characterized in terms of sequence-specific transcription factors and signaling pathways in plants, the "CTD-code" and its regulatory factors are poorly understood in plants despite their significance. Further investigation into pol II-CTD phosphoregulation and its regulatory factors such as CTD phosphatases and CTD kinases will reveal the uncharted layer of gene expression regulation in higher plants.

REFERENCES

Adachi S, Nobusawa T, Umeda M. 2009. Quantitative And Cell Type-Specific Transcriptional Regulation Of A-Type Cyclin-Dependent Kinase In Arabidopsis Thaliana. *Dev Biol* **329**(2): 306-314.

Adachi S, Uchimiya H, Umeda M. 2006. Expression Of B2-Type Cyclin-Dependent Kinase Is Controlled By Protein Degradation In Arabidopsis Thaliana. *Plant Cell Physiol* **47**(12): 1683-1686.

Aichinger E, Villar CBR, Farrona S, Reyes JC, Hennig L, Kohler C. 2009. CHD3 Proteins And Polycomb Group Proteins Antagonistically Determine Cell Identity In Arabidopsis. *Plos Genet* **5**(8): 12.

Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ. 2009. TFIIH Kinase Places Bivalent Marks On The Carboxy-Terminal Domain Of RNA Polymerase II. *Mol Cell* **34**(3): 387-393.

Aksoy E, Jeong IS, Koiwa H. 2013. Loss Of Function Of Arabidopsis C-Terminal Domain Phosphatase-Like1 Activates Iron Deficiency Responses At The Transcriptional Level. *Plant Physiol* **161**(1): 330-345.

Allison LA, Wong JKC, Fitzpatrick VD, Moyle M, Ingles CJ. 1988. The C-Terminal Domain Of The Largest Subunit Of RNA Polymerase-II Of Saccharomyces-Cerevisiae, Drosophila-Melanogaster, And Mammals - A Conserved Structure With An Essential Function. *Mol Cell Biol* **8**(1): 321-329.

Aloni R, Langhans M, Aloni E, Dreieicher E, Ullrich CI. 2005. Root-Synthesized Cytokinin In Arabidopsis Is Distributed In The Shoot By The Transpiration Stream. *J Exp Bot* **56**(416): 1535-1544.

Aloni R, Langhans M, Aloni E, Ullrich CI. 2004. Role Of Cytokinin In The Regulation Of Root Gravitropism. *Planta* **220**(1): 177-182.

Archambault J, Chambers R, Kobor MS, Ho Y, Bolotin D, Andrews B, Kane CM, Greenblatt J. 1997. An Essential Component Of A C-Terminal Domain Phosphatase That Interacts With Transcription Factor IIF In Saccharomyces Cerevisiae. *Proc Natl Acad Sci USA* **94**: 14300-14305.

Archambault J, Pan G, Dahmus GK, Cartier M, Marshall N, Zhang S, Dahmus ME, Greenblatt J. 1998. FCP1, The RAP74-Interacting Subunit Of A Human Protein Phosphatase That Dephosphorylates The Carboxyl-Terminal Domain Of RNA Polymerase IIO. *J Biol Chem* **273**: 27593-27601.

Azuma Y, Yamagishi M, Ueshima R, Ishihama A. 1991. Cloning And Sequence Determination Of The Schizosaccharomyces Pombe Rpb1 Gene Encoding The Largest Subunit Of RNA Polymerase II. *Nucleic Acids Res* **19**(3): 461-468.

Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA, Muller R, Meese E, Lenhof HP. 2007. Genetrail--Advanced Gene Set Enrichment Analysis. *Nucleic Acids Res* **35**(Web Server Issue): W186-192.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS. 2009. MEME SUITE: Tools For Motif Discovery And Searching. *Nucleic Acids Res* **37**: W202-W208.

Bang W, Kim S, Ueda A, Vikram M, Yun D, Bressan RA, Hasegawa PM, Bahk J, Koiwa H. 2006. Arabidopsis Carboxyl-Terminal Domain Phosphatase-Like Isoforms Share Common Catalytic And Interaction Domains But Have Distinct In Planta Functions. *Plant Physiol* **142**(2): 586-594.

Banno H, Ikeda Y, Niu QW, Chua NH. 2001. Overexpression Of Arabidopsis ESR1 Induces Initiation Of Shoot Regeneration. *Plant Cell* **13**(12): 2609-2618.

Bao WD, Kojima KK, Kohany O. 2015. Repbase Update, A Database Of Repetitive Elements In Eukaryotic Genomes. *Mob DNA* **6**: 6.

Bark C, Weller P, Zabielski J, Janson L, Pettersson U. 1987. A Distant Enhancer Element Is Required For Polymerase-III Transcription Of A U6 RNA Gene. *Nature* **328**(6128): 356-359.

Beeckman T, Burssens S, Inze D. 2001. The Peri-Cell-Cycle In Arabidopsis. *J Exp Bot* **52**: 403-411.

Bellier S, Dubois MF, Nishida E, Almouzni G, Bensaude O. 1997. Phosphorylation Of The RNA Polymerase II Largest Subunit During Xenopus Laevis Oocyte Maturation. *Mol Cell Biol* **17**(3): 1434-1440.

Bernstein LB, Manser T, Weiner AM. 1985. Human-U-1 Small Nuclear-RNA Genes - Extensive Conservation Of Flanking Sequences Suggests Cycles Of Gene Amplification And Transposition. *Mol Cell Biol* **5**(9): 2159-2171.

Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, Tang HX, Pikaard CS. 2015. Identification Of Pol IV And RDR2-Dependent Precursors Of 24 Nt SiRNAs Guiding De Novo DNA Methylation In Arabidopsis. *Elife* **4**: 22.

Blommel PG, Fox BG. 2007. A Combined Approach To Improving Large-Scale Production Of Tobacco Etch Virus Protease. *Protein Expr Purif* **55**(1): 53-68.

Bonnet F, Vigneron M, Bensaude O, Dubois MF. 1999. Transcription-Independent Phosphorylation Of The RNA Polymerase IIC-Terminal Domain (CTD) Involves ERK Kinases (MEK1/2). *Nucleic Acids Res* **27**(22): 4399-4404.

Brewster AS, Chen XJS. 2010. Insights Into The MCM Functional Mechanism: Lessons Learned From The Archaeal MCM Complex. *Crit Rev Biochem Mol Biol* **45**(3): 243-256.

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. 1992. The Human Xist Gene - Analysis Of A 17 Kb Inactive X-Specific RNA That Contains Conserved Repeats And Is Highly Localized Within The Nucleus. *Cell* **71**(3): 527-542.

Buratowski S. 2003. The CTD Code. *Nat Struct Biol* **10**(9): 679-680.

____. 2009. Progression Through The RNA Polymerase II CTD Cycle. *Mol Cell* **36**(4): 541-546.

Buratowski S, Sharp PA. 1990. Transcription Initiation-Complexes And Upstream Activation With RNA Polymerase-II Lacking The C-Terminal Domain Of The Largest Subunit. *Mol Cell Biol* **10**(10): 5562-5564.

Bushnell DA, Kornberg RD. 2003. Complete, 12-Subunit RNA Polymerase II At 4.1-Angstrom Resolution: Implications For The Initiation Of Transcription. *Proc Natl Acad Sci U S A* **100**(12): 6969-6973.

Butaye KM, Goderis IJ, Wouters PF, Pues JM, Delaure SL, Broekaert WF, Depicker A, Cammue BP, De Bolle MF. 2004. Stable High-Level Transgene Expression In Arabidopsis Thaliana Using Gene Silencing Mutants And Matrix Attachment Regions. *Plant J* **39**(3): 440-449.

Butt SJ, Varis S, Nasir IA, Sheraz S, Shahid A, Ali Q. 2015. Micro Propagation In Advanced Vegetable Production: A Review. *Advan Life Sci* **2**(2): 48-57.

Cary AJ, Che P, Howell SH. 2002. Developmental Events And Shoot Apical Meristem Gene Expression Patterns During Shoot Development In Arabidopsis Thaliana. *Plant J* **32**(6): 867-877.

Castillo-Gonzalez C, Liu XY, Huang CJ, Zhao CJ, Ma ZY, Hu T, Sun F, Zhou YJ, Zhou XP, Wang XJ et al. 2015. Geminivirus-Encoded Trap Suppressor Inhibits The Histone Methyltransferase SUVH4/KYP To Counter Host Defense. *Elife* **4**: 31.

Chao SH, Greenleaf AL, Price DH. 2001. Juglone, An Inhibitor Of The Peptidyl-Prolyl Isomerase Pin1, Also Directly Blocks Transcription. *Nucleic Acids Res* **29**(3): 767-773.

Che P, Lall S, Howell SH. 2007. Developmental Steps In Acquiring Competence For Shoot Development In Arabidopsis Tissue Culture. *Planta* **226**(5): 1183-1194.

Chekanova JA, Gregory BD, Reverdatto SV, Chen H, Kumar R, Hooker T, Yazaki J, Li P, Skiba N, Peng Q et al. 2007. Genome-Wide High-Resolution Mapping Of Exosome Substrates Reveals Hidden Features In The Arabidopsis Transcriptome. *Cell* **131**(7): 1340-1353.

Chen T, Cui P, Xiong LM. 2015. The RNA-Binding Protein HOS5 And Serine/Arginine-Rich Proteins RS40 And RS41 Participate In MiRNA Biogenesis In Arabidopsis. *Nucleic Acids Res* **43**(17): 8283-8298.

Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S. 2001. Opposing Effects Of Ctk1 Kinase And Fcp1 Phosphatase At Ser 2 Of The RNA Polymerase II C-Terminal Domain. *Genes Dev* **15**(24): 3319-3329.

Cho EJ, Takagi T, Moore CR, Buratowski S. 1997. MRNA Capping Enzyme Is Recruited To The Transcription Complex By Phosphorylation Of The RNA Polymerase II Carboxy-Terminal Domain. *Genes Dev* **11**(24): 3319-3326.

Cho H, Kim TK, Mancebo H, Lane WS, Flores O, Reinberg D. 1999. A Protein Phosphatase Functions To Recycle RNA Polymerase II. *Genes Dev* **13**(12): 1540-1552.

Christianson ML, Warnick DA. 1983. Competence And Determination In The Process Of In Vitro Shoot Organogenesis. *Dev Biol* **95**(2): 288-293.

Christopher JT, Powles SB, Liljegren DR, Holtum JA. 1991. Cross-Resistance To Herbicides In Annual Ryegrass (Lolium Rigidum) : II. Chlorsulfuron Resistance Involves A Wheat-Like Detoxification System. *Plant Physiol* **95**(4): 1036-1043.

Chung M-H, Chen M-K, Pan S-M. 2000. Floral Spray Transformation Can Efficiently Generate Arabidopsis. *Transgenic Res* **9**(6): 471-486.

Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. 2009. An Architectural Role For A Nuclear Noncoding RNA: NEAT1 RNA Is Essential For The Structure Of Paraspeckles. *Mol Cell* **33**(6): 717-726.

Colon-Carmona A, You R, Haimovitch-Gal T, Doerner P. 1999. Spatio-Temporal Analysis Of Mitotic Activity With A Labile Cyclin-GUS Fusion Protein. *Plant J* **20**(4): 503-508.

Connelly S, Filipowicz W. 1993. Activity Of Chimeric-U Small Nuclear-RNA (SnRNA)/Messenger RNA Genes In Transfected Protoplasts Of Nicotiana-Plumbaginifolia - U-SnRNA 3'-End Formation And Transcription Initiation Can Occur Independently In Plants. *Mol Cell Biol* **13**(10): 6403-6415.

Corden JL. 1990. Tails Of RNA Polymerase-II. *Trends Biochem Sci* **15**(10): 383-387.

Corden JL, Cadena DL, Ahearn JM, Jr., Dahmus ME. 1985. A Unique Structure At The Carboxyl Terminus Of The Largest Subunit Of Eukaryotic RNA Polymerase II. *Proc Natl Acad Sci U S A* **82**(23): 7934-7938.

Cramer P, Bushnell DA, Kornberg RD. 2001. Structural Basis Of Transcription: RNA Polymerase II At 2.8 Angstrom Ngstrom Resolution. *Science* **292**(5523): 1863-1876.

Cui P, Chen T, Qin T, Ding F, Wang ZY, Chen H, Xiong LM. 2016. The RNA Polymerase II C-Terminal Domain Phosphatase-Like Protein FIERY2/CPL1 Interacts With Eif4aiii And Is Essential For Nonsense-Mediated MRNA Decay In Arabidopsis. *Plant Cell* **28**(3): 770-785.

Cui XF, Fan BF, Scholz J, Chen ZX. 2007. Roles Of Arabidopsis Cyclin-Dependent Kinase C Complexes In Cauliflower Mosaic Virus Infection, Plant Growth, And Development. *Plant Cell* **19**(4): 1388-1402.

Curtis MD, Grossniklaus U. 2003. A Gateway Cloning Vector Set For High-Throughput Functional Analysis Of Genes In Planta. *Plant Physiol* **133**(2): 462-469.

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. 2005. Genome-Wide Identification And Testing Of Superior Reference Genes For Transcript Normalization In Arabidopsis. *Plant Physiol* **139**(1): 5-17.

D'Agostino IB, Deruere J, Kieber JJ. 2000. Characterization Of The Response Of The Arabidopsis Response Regulator Gene Family To Cytokinin. *Plant Physiol* **124**(4): 1706-1717.

Dahmus ME. 1996. Reversible Phosphorylation Of The C-Terminal Domain Of RNA Polymerase II. *J Biol Chem* **271**: 19009-19012.

Daimon Y, Takabe K, Tasaka M. 2003. The CUP-SHAPED COTYLEDON Genes Promote Adventitious Shoot Formation On Calli. *Plant Cell Physiol* **44**(2): 113-121.

Davidson L, Muniz L, West S. 2014. 3 ' End Formation Of Pre-MRNA And Phosphorylation Of Ser2 On The RNA Polymerase II CTD Are Reciprocally Coupled In Human Cells. *Genes Dev* **28**(4): 342-356.

Descostes N, Heidemann M, Spinelli L, Schuller R, Maqbool MA, Fenouil R, Koch F, Innocenti C, Gut M, Gut I et al. 2014. Tyrosine Phosphorylation Of RNA Polymerase II CTD Is Associated With Antisense Promoter Transcription And Active Enhancers In Mammalian Cells. *Elife* **3**: 36.

Dewitte W, Riou-Khamlichi C, Scofield S, Healy JMS, Jacqmard A, Kilby NJ, Murray JAH. 2003. Altered Cell Cycle Distribution, Hyperplasia, And Inhibited

Differentiation In Arabidopsis Caused By The D-Type Cyclin CYCD3. *Plant Cell* **15**(1): 79-92.

Dewitte W, Scofield S, Alcasabas AA, Maughan SC, Menges M, Braun N, Collins C, Nieuwland J, Prinsen E, Sundaresan V et al. 2007. Arabidopsis CYCD3 D-Type Cyclins Link Cell Proliferation And Endocycles And Are Rate-Limiting For Cytokinin Responses. *Proc Natl Acad Sci U S A* **104**(36): 14537-14542.

Didonato RJ, Arbuckle E, Buker S, Sheets J, Tobar J, Totong R, Grisafi P, Fink GR, Celenza JL. 2004. Arabidopsis ALF4 Encodes A Nuclear-Localized Protein Required For Lateral Root Formation. *Plant J* **37**(3): 340-353.

Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A. 2007. The Expanding RNA Polymerase III Transcriptome. *Trends Genet* **23**(12): 614-622.

Dietrich MA, Prenger JP, Guilfoyle TJ. 1990. Analysis Of The Genes Encoding The Largest Subunit Of RNA Polymerase II In Arabidopsis And Soybean. *Plant Mol Biol* **15**: 207-223.

Ding Y, Avramova Z, Fromm M. 2011. Two Distinct Roles Of ARABIDOPSIS HOMOLOG OF TRITHORAX1 (ATX1) At Promoters And Within Transcribed Regions Of ATX1-Regulated Genes. *Plant Cell* **23**(1): 350-363.

Doelling JH, Pikaard CS. 1993. Transient Expression In Arabidopsis Thaliana Protoplasts Derived From Rapidly Established Cell Suspension Cultures. *Plant Cell Rep* **12**: 241-244.

Dubois MF, Bellier S, Seo SJ, Bensaude O. 1994. Phosphorylation Of The RNA-Polymerase-II Largest Subunit During Heat-Shock And Inhibition Of Transcription In Hela-Cells. *J Cell Physiol* **158**(3): 417-426.

Dubois MF, Bensaude O. 1998. Phosphorylation Of RNA Polymerase II C-Terminal Domain (CTD): A New Control For Heat Shock Gene Expression? *Cell Stress Chaperones* **3**(3): 147-151.

Egloff S. 2012. Role Of Ser7 Phosphorylation Of The CTD During Transcription Of SnRNA Genes. *Rna Biology* **9**(8): 1033-1038.

Egloff S, Murphy S. 2008. Cracking The RNA Polymerase II CTD Code. *Trends Genet* **24**(6): 280-288.

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S. 2007. Serine-7 Of The RNA Polymerase II CTD Is Specifically Required For SnRNA Gene Expression. *Science* **318**(5857): 1777-1779.

Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S. 2010. The Integrator Complex Recognizes A New Double Mark On The RNA Polymerase II Carboxyl-Terminal Domain. *J Biol Chem* **285**(27): 20564-20569.

Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. 2012. Ser7 Phosphorylation Of The CTD Recruits The RPAP2 Ser5 Phosphatase To SnRNA Genes. *Mol Cell* **45**(1): 111-122.

Eick D, Geyer M. 2013. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chem Rev* **113**(11): 8456-8490.

Espinosa A, Guo M, Tam VC, Fu ZQ, Alfano JR. 2003. The Pseudomonas Syringae Type III-Secreted Protein Hopptod2 Possesses Protein Tyrosine Phosphatase Activity And Suppresses Programmed Cell Death In Plants. *Mol Microbiol* **49**(2): 377-387.

Feng Y, Kang JS, Kim S, Yun DJ, Lee SY, Bahk JD, Koiwa H. 2010. Arabidopsis SCP1-Like Small Phosphatases Differentially Dephosphorylate RNA Polymerase II C-Terminal Domain. *Biochem Biophys Res Commun* **397**(2): 355-360.

Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE. 2011. Formation Of Plant Metabolic Gene Clusters Within Dynamic Chromosomal Regions. *Proc Natl Acad Sci U S A* **108**(38): 16116-16121.

Field B, Osbourn AE. 2008. Metabolic Diversification - Independent Assembly Of Operon-Like Gene Clusters In Different Plants. *Science* **320**(5875): 543-547.

Francis D. 2007. The Plant Cell Cycle - 15 Years On. *New Phytol* **174**(2): 261-278.

Fuda NJ, Buckley MS, Wei WX, Core LJ, Waters CT, Reinberg D, Lisa JT. 2012. Fcp1 Dephosphorylation Of The RNA Polymerase II C-Terminal Domain Is Required For Efficient Transcription Of Heat Shock Genes. *Mol Cell Biol* **32**(17): 3428-3437.

Fukudome A, Aksoy E, Wu XQ, Kumar K, Jeong IS, May K, Russell WK, Koiwa H. 2014. Arabidopsis CPL4 Is An Essential C-Terminal Domain Phosphatase That Suppresses Xenobiotic Stress Responses. *Plant J* **80**(1): 27-39.

Fulop K, Pettko-Szandtner A, Magyar Z, Miskolczi P, Kondorosi E, Dudits D, Bako L. 2005. The Medicago CDKC; 1-CYCLINT; 1 Kinase Complex Phosphorylates

The Carboxy-Terminal Domain Of RNA Polymerase II And Promotes Transcription. *Plant J* **42**(6): 810-820.

Furuta K, Kubo M, Sano K, Demura T, Fukuda H, Liu YG, Shibata D, Kakimoto T. 2011. The CKH2/PKL Chromatin Remodeling Factor Negatively Regulates Cytokinin Responses In Arabidopsis Calli. *Plant Cell Physiol* **52**(4): 618-628.

Gaamouche T, Manes CLD, Kwiatkowska D, Berckmans B, Koumproglou R, Maes S, Beeckman T, Vernoux T, Doonan JH, Traas J et al. 2010. Cyclin-Dependent Kinase Activity Maintains The Shoot Apical Meristem Cells In An Undifferentiated State. *Plant J* **64**(1): 26-37.

Gandia-Herrero F, Lorenz A, Larson T, Graham IA, Bowles DJ, Rylott EL, Bruce NC. 2008. Detoxification Of The Explosive 2,4,6-Trinitrotoluene In Arabidopsis: Discovery Of Bifunctional O- And C-Glucosyltransferases. *Plant J* **56**(6): 963-974.

Ganem C, Devaux F, Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, Facca C, Faye G. 2003. Ssu72 Is A Phosphatase Essential For Transcription Termination Of SnoRNAs And Specific MRNAs In Yeast. *EMBO J* **22**(7): 1588-1598.

Ghosh A, Shuman S, Lima CD. 2008. The Structure Of Fcp1, An Essential RNA Polymerase II CTD Phosphatase. *Mol Cell* **32**(4): 478-490.

____. 2011. Structural Insights To How Mammalian Capping Enzyme Reads The CTD Code. *Mol Cell* **43**(2): 299-310.

Goenka A, Sengupta S, Pandey R, Parihar R, Mohanta GC, Mukerji M, Ganesh S. 2016. Human Satellite-III Non-Coding RNAs Modulate Heat-Shock-Induced Transcriptional Repression. *J Cell Sci* **129**(19): 3541-3552.

Gordon SP, Heisler MG, Reddy GV, Ohno C, Das P, Meyerowitz EM. 2007. Pattern Formation During De Novo Assembly Of The Arabidopsis Shoot Meristem. *Development* **134**(19): 3539-3548.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: Scanning For Occurrences Of A Given Motif. *Bioinformatics* **27**(7): 1017-1018.

Greenleaf AL, Jokerst RS, Zehring WA, Hamilton BJ, Weeks JR, Sluder AE, Price DH. 1987. Drosophila RNA Polymerase II: Genetics And In Vitro Transcription. *RNA Polymerase And The Regulation Of Transcription*, (Ed. WS Reznikoff, RR Burgess, JE Dahlberg, CA Gross, MT Record, MP Wickens), Pp. 459-463. Elsevier Science Publishing Co., New York.

Gu B, Eick D, Bensaude O. 2013. CTD Serine-2 Plays A Critical Role In Splicing And Termination Factor Recruitment To RNA Polymerase II In Vivo. *Nucleic Acids Res* **41**(3): 1591-1603.

Guan QM, Yue XL, Zeng HT, Zhu JH. 2014. The Protein Phosphatase RCF2 And Its Interacting Partner NAC019 Are Critical For Heat Stress-Responsive Gene Regulation And Thermotolerance In Arabidopsis. *Plant Cell* **26**(1): 438-453.

Gullerova M, Barta A, Lorkovic ZJ. 2006. Atcyp59 Is A Multidomain Cyclophilin From Arabidopsis Thaliana That Interacts With SR Proteins And The C-Terminal Domain Of The RNA Polymerase II. *RNA* **12**(4): 631-643.

Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L, Pikaard CS. 2012. In Vitro Transcription Activities Of Pol IV, Pol V, And RDR2 Reveal Coupling Of Pol IV And RDR2 For DsRNA Synthesis In Plant RNA Silencing. *Mol Cell* **48**(5): 811-818.

Hajheidari M, Farrona S, Huettel B, Koncz Z, Koncz C. 2012. CDKF;1 And CDKD Protein Kinases Regulate Phosphorylation Of Serine Residues In The C-Terminal Domain Of Arabidopsis RNA Polymerase II. *Plant Cell* **24**(4): 1626-1642.

Hajheidari M, Koncz C, Eick D. 2013. Emerging Roles For RNA Polymerase II CTD In Arabidopsis. *Trends Plant Sci*.

Hare PD, Moller SG, Huang LF, Chua NH. 2003. LAF3, A Novel Factor Required For Normal Phytochrome A Signaling. *Plant Physiol* **133**(4): 1592-1604.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S et al. 2012. GENCODE: The Reference Human Genome Annotation For The ENCODE Project. *Genome Res* **22**(9): 1760-1774.

Hausmann S, Koiwa H, Krishnamurthy S, Hampsey M, Shuman S. 2005. Different Strategies For Carboxyl-Terminal Domain (CTD) Recognition By Serine 5-Specific CTD Phosphatases. *J Biol Chem* **280**(45): 37681-37688.

Hausmann S, Schwer B, Shuman S. 2004. An Encephalitozoon Cuniculi Ortholog Of The RNA Polymerase II Carboxyl-Terminal Domain (CTD) Serine Phosphatase Fcp1. *Biochemistry* **43**(22): 7111-7120.

Hausmann S, Shuman S. 2002. Characterization Of The CTD Phosphatase Fcp1 From Fission Yeast - Preferential Dephosphorylation Of Serine 2 Versus Serine 5. *J Biol Chem* **277**(24): 21213-21220.

Hemerly A, Bergounioux C, Vanmontagu M, Inze D, Ferreira P. 1992. Genes Regulating The Plant-Cell Cycle - Isolation Of A Mitotic-Like Cyclin From Arabidopsis-thaliana. *Proc Natl Acad Sci U S A* **89**(8): 3295-3299.

Hengartner CJ, Myer VE, Liao SM, Wilson CJ, Koh SS, Young RA. 1998. Temporal Regulation Of RNA Polymerase II By Srb10 And Kin28 Cyclin-Dependent Kinases. *Mol Cell* **2**(1): 43-53.

Hennig L, Christner C, Kipping M, Schelbert B, Rucknagel KP, Grabley S, Kullertz G, Fischer G. 1998. Selective Inactivation Of Parvulin-Like Peptidyl-Prolyl Cis/Trans Isomerases By Juglone. *Biochemistry* **37**(17): 5953-5960.

Heo JB, Sung S. 2011. Vernalization-Mediated Epigenetic Silencing By A Long Intronic Noncoding RNA. *Science* **331**(6013): 76-79.

Hernandez N. 1985. Formation Of The 3' End Of U1 SnRNA Is Directed By A Conserved Sequence Located Downstream Of The Coding Region. *EMBO J* **4**(7): 1827-1837.

____. 2001. Small Nuclear RNA Genes: A Model System To Study Fundamental Mechanisms Of Transcription. *J Biol Chem* **276**(29): 26733-26736.

Hernandez N, Lucito R. 1988. Elements Required For Transcription Initiation Of The Human U2 SnRNA Gene Coincide With Elements Required For SnRNA 3' End Formation. *EMBO J* **7**(10): 3125-3134.

Hernandez N, Weiner AM. 1986. Formation Of The 3' End Of U1 SnRNA Requires Compatible SnRNA Promoter Elements. *Cell* **47**(2): 249-258.

Himanen K, Boucheron E, Vanneste S, Engler JD, Inze D, Beeckman T. 2002. Auxin-Mediated Cell Cycle Activation During Early Lateral Root Initiation. *Plant Cell* **14**(10): 2339-2351.

Hirose Y, Manley JL. 2000. RNA Polymerase II And The Integration Of Nucear Events. *Genes Dev* **14**: 1415-1429.

Ho CK, Shuman S. 1999. Distinct Roles For CTD Ser-2 And Ser-5 Phosphorylation In The Recruitment And Allosteric Activation Of Mammalian MRNA Capping Enzyme. *Mol Cell* **3**(3): 405-411.

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-Directed Mutagenesis By Overlap Extension Using The Polymerase Chain-Reaction. *Gene* **77**(1): 51-59.

Holstege FCP, Vandervliet PC, Timmers HTM. 1996. Opening Of An RNA Polymerase II Promoter Occurs In Two Distinct Steps And Requires The Basal Transcription Factors IIE And IIH. *EMBO J* **15**(7): 1666-1677.

Hou B, Lim EK, Higgins GS, Bowles DJ. 2004. N-Glucosylation Of Cytokinins By Glycosyltransferases Of Arabidopsis Thaliana. *J Biol Chem* **279**(46): 47822-47832.

Howell SH, Lall S, Che P. 2003. Cytokinins And Shoot Development. *Trends Plant Sci* **8**(9): 453-459.

Hsin JP, Manley JL. 2012. The RNA Polymerase II CTD Coordinates Transcription And RNA Processing. *Genes Dev* **26**(19): 2119-2137.

Hsin JP, Sheth A, Manley JL. 2011. RNAP II CTD Phosphorylated On Threonine-4 Is Required For Histone MRNA 3 ' End Processing. *Science* **334**(6056): 683-686.

Hsin JP, Xiang K, Manley JL. 2014. Function And Control Of RNA Polymerase II CTD Phosphorylation In Vertebrate Transcription And RNA Processing. *Mol Cell Biol*.

Htun H, Lund E, Dahlberg JE. 1984. Human U1 RNA Genes Contain An Unusually Sensitive Nuclease S1 Cleavage Site Within The Conserved 3' Flanking Region. *Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences* **81**(23): 7288-7292.

Hwang I, Sheen J. 2001. Two-Component Circuitry In Arabidopsis Cytokinin Signal Transduction. *Nature* **413**(6854): 383-389.

Ikeda Y, Banno H, Niu QW, Howell SH, Chua NH. 2006. The ENHANCER OFSHOOT REGENERATION 2 Gene In Arabidopsis Regulates CUP-SHAPED COTYLEDON 1 At The Transcriptional Level And Controls Cotyledon Development. *Plant Cell Physiol* **47**(11): 1443-1456.

Jensen RC, Wang Y, Hardin SB, Stumph WE. 1998. The Proximal Sequence Element (PSE) Plays A Major Role In Establishing The RNA Polymerase Specificity Of Drosophila U-SnRNA Genes. *Nucleic Acids Res* **26**(2): 616-622.

Jeong IS, Aksoy E, Fukudome A, Akhter S, Hiraguri A, Fukuhara T, Bahk JD, Koiwa H. 2013a. Arabidopsis C-Terminal Domain Phosphatase-Like 1 Functions In MiRNA Accumulation And DNA Methylation. *Plos One* **8**(9): E74739.

Jeong IS, Fukudome A, Aksoy E, Bang WY, Kim S, Guan Q, Bahk JD, May KA, Russel WK, Zhu J et al. 2013b. Regulation Of Abiotic Stress Signalling By Arabidopsis C-Terminal Domain Phosphatase-Like 1 Requires Interaction With A K-Homology Domain-Containing Protein. *Plos One* **8**(11): E80509.

Jiang JF, Wang BS, Shen Y, Wang H, Feng Q, Shi HZ. 2013. The Arabidopsis RNA Binding Protein With K Homology Motifs, SHINY1, Interacts With The C-Terminal Domain Phosphatase-Like 1 (CPL1) To Repress Stress-Inducible Gene Expression. *Plos Genet* **9**(7): 17.

Jin YM, Jung J, Jeon H, Won SY, Feng Y, Kang JS, Lee SY, Cheong JJ, Koiwa H, Kim M. 2011. AtCPL5, A Novel Ser-2-Specific RNA Polymerase II C-Terminal Domain Phosphatase, Positively Regulates ABA And Drought Responses In Arabidopsis. *New Phytol* **190**(1): 57-74.

Jolly C, Metz A, Govin J, Vigneron M, Turner BM, Khochbin S, Vourc'h C. 2004. Stress-Induced Transcription Of Satellite III Repeats. *J Cell Biol* **164**(1): 25-33.

Joubes J, Chevalier C, Dudits D, Heberle-Bors E, Inze D, Umeda M, Renaudin JP. 2000. CDK-Related Protein Kinases In Plants. *Plant Mol Biol* **43**(5-6): 607-620.

Kang CH, Feng Y, Vikram M, Jeong IS, Lee JR, Bahk JD, Yun DJ, Lee SY, Koiwa H. 2009. Arabidopsis Thaliana PRP40s Are RNA Polymerase II C-Terminal Domain-Associating Proteins. *Arch Biochem Biophys* **484**(1): 30-38.

Kang J, Dengler N. 2002. Cell Cycling Frequency And Expression Of The Homeobox Gene ATHB-8 During Leaf Vein Development In Arabidopsis. *Planta* **216**(2): 212-219.

Kapitonov VV, Jurka J. 2001. Rolling-Circle Transposons In Eukaryotes. *Proc Natl Acad Sci U S A* **98**(15): 8714-8719.

____. 2007. Helitrons On A Roll: Eukaryotic Rolling-Circle Transposons. *Trends Genet* **23**(10): 521-529.

Karlsson P, Christie MD, Seymour DK, Wang H, Wang X, Hagmann J, Kulcheski F, Manavella PA. 2015. KH Domain Protein RCF3 Is A Tissue-Biased Regulator Of The Plant MiRNA Biogenesis Cofactor HYL1. *Proc Natl Acad Sci U S A* **112**(45): 14096-14101.

Kaul S Koo HL Jenkins J Rizzo M Rooney T Tallon LJ Feldblyum T Nierman W Benito MI Lin XY et al. 2000. Analysis Of The Genome Sequence Of The Flowering Plant Arabidopsis Thaliana. *Nature* **408**(6814): 796-815.

Kitsios G, Alexiou KG, Bush M, Shaw P, Doonan JH. 2008. A Cyclin-Dependent Protein Kinase, CDKC2, Colocalizes With And Modulates The Distribution Of Spliceosomal Components In Arabidopsis. *Plant J* **54**(2): 220-235.

Knockaert M, Sapkota G, Alarcon C, Massague J, Brivanlou AH. 2006. Unique Players In The BMP Pathway: Small C-Terminal Domain Phosphatases Dephosphorylate Smad1 To Attenuate BMP Signaling. *Proc Natl Acad Sci U S A* **103**(32): 11940-11945.

Kobor MS, Archambault J, Lester W, Holstege FCP, Gileadi O, Jansma DB, Jennings EG, Kouyoumdjian F, Davidson AR, Young RA et al. 1999. An Unusual Eukaryotic Protein Phosphatase Required For Transcription By RNA Polymerase II And CTD Dephosphorylation In S-Cerevisiae. *Mol Cell* **4**(1): 55-62.

Koiwa H. 2006. Phosphorylation Of RNA Polymerase II C-Terminal Domain And Plant Osmotic-Stress Responses. In *Abiotic Stress Tolerance In Plants-Toward The Improvement Of Global Environment And Food*, (Ed. TT Ashwani K. Rai), Pp. 47-57. Springer, Dordrecht, The Netherland.

Koiwa H, Barb AW, Xiong LM, Li F, Mccully MG, Lee BH, Sokolchik I, Zhu JH, Gong ZZ, Reddy M et al. 2002. C-Terminal Domain Phosphatase-Like Family Members (AtCPLs) Differentially Regulate Arabidopsis Thaliana Abiotic Stress Signaling, Growth, And Development. *Proc Natl Acad Sci U S A* **99**(16): 10893-10898.

Koiwa H, Hausmann S, Bang WY, Ueda A, Kondo N, Hiraguri A, Fukuhara T, Bahk JD, Yun DJ, Bressan RA et al. 2004. Arabidopsis C-Terminal Domain Phosphatase-Like 1 And 2 Are Essential Ser-5-Specific C-Terminal Domain Phosphatases. *Proc Natl Acad Sci USA* **101**(40): 14539-14544.

Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M, Rosen BD, Cheng CY, Moreira W, Mock SA et al. 2015. Araport: The Arabidopsis Information Portal. *Nucleic Acids Res* **43**(D1): D1003-D1009.

Krishnamurthy S, He XY, Reyes-Reyes M, Moore C, Hampsey M. 2004. Ssu72 Is An RNA Polymerase II CTD Phosphatase. *Mol Cell* **14**(3): 387-394.

Kubo M, Furuta K, Demura T, Fukuda H, Liu YG, Shibata D, Kakimoto T. 2011. The CKH1/EER4 Gene Encoding A TAF12-Like Protein Negatively Regulates Cytokinin Sensitivity In Arabidopsis Thaliana. *Plant Cell Physiol* **52**(4): 629-637.

Kubo M, Kakimoto T. 2000. The CYTOKININ-HYPERSENSITIVE Genes Of Arabidopsis Negatively Regulate The Cytokinin-Signaling Pathway For Cell Division And Chloroplast Development. *Plant J* **23**(3): 385-394.

Kuhlman TC, Cho H, Reinberg D, Hernandez N. 1999. The General Transcription Factors IIA, IIB, IIF, And IIE Are Required For RNA Polymerase II Transcription From The Human U1 Small Nuclear RNA Promoter. *Mol Cell Biol* **19**(3): 2130-2141.

Lange H, Zuber H, Sement FM, Chicher J, Kuhn L, Hammann P, Brunaud V, Berard C, Bouteiller N, Balzergue S et al. 2014. The RNA Helicases Atmtr4 And HEN2 Target Specific Subsets Of Nuclear Transcripts For Degradation By The Nuclear Exosome In Arabidopsis Thaliana. *Plos Genet* **10**(8): 21.

Laplaze L, Benkova E, Casimiro I, Maes L, Vanneste S, Swarup R, Weijers D, Calvo V, Parizot B, Herrera-Rodriguez MB et al. 2007. Cytokinins Act Directly On Lateral Root Founder Cells To Inhibit Root Initiation. *Plant Cell* **19**(12): 3889-3900.

Lauressergues D, Couzigou JM, Clemente HS, Martinez Y, Dunand C, Becard G, Combier JP. 2015. Primary Transcripts Of MicroRNAs Encode Regulatory

Peptidesprimary Transcripts Of MicroRNAs Encode Regulatory Peptides. *Nature* **520**(7545): 90-U205.

Lavoie SB, Albert AL, Thibodeau A, Vincent M. 1999. Heat Shock-Induced Alterations In Phosphorylation Of The Largest Subunit Of RNA Polymerase II As Revealed By Monoclonal Antibodies CC-3 And MPM-2. *Biochemistry And Cell Biology-Biochimie Et Biologie Cellulaire* **77**(4): 367-374.

Lee Y, Kim M, Han JJ, Yeom KH, Lee S, Baek SH, Kim VN. 2004. MicroRNA Genes Are Transcribed By RNA Polymerase II. *EMBO J* **23**(20): 4051-4060.

Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. 1980. Are Snrnps Involved In Splicing. *Nature* **283**(5743): 220-224.

Lerner MR, Steitz JA. 1979. Antibodies To Small Nuclear RNAs Complexed With Proteins Are Produced By Patients With Systemic Lupus-Erythematosus. *Proc Natl Acad Sci U S A* **76**(11): 5495-5499.

Li FJ, Cheng C, Cui FH, De Oliveira MVV, Yu X, Meng XZ, Intorne AC, Babilonia K, Li MY, Li B et al. 2014. Modulation Of RNA Polymerase II Phosphorylation Downstream Of Pathogen Perception Orchestrates Plant Immunity. *Cell Host & Microbe* **16**(6): 748-758.

Li LG, He ZY, Pandey GK, Tsuchiya T, Luan S. 2002. Functional Cloning And Characterization Of A Plant Efflux Carrier For Multidrug And Heavy Metal Detoxification. *J Biol Chem* **277**(7): 5360-5368.

Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, Mcneil JB, Bentley DL. 2002. Functional Interaction Of Yeast Pre-MRNA 3' End Processing Factors With RNA Polymerase II. *Mol Cell* **9**(5): 1101-1111.

Lin PS, Dubois MF, Dahmus ME. 2002. TFIIF-Associating Carboxyl-Terminal Domain Phosphatase Dephosphorylates Phosphoserines 2 And 5 Of RNA Polymerase II. *J Biol Chem* **277**(48): 45949-45956.

Lindell TJ, Weinberg F, Morris PW, Roeder RG, Rutter WJ. 1970. Specific Inhibition Of Nuclear RNA Polymerase II By Alpha-Amanitin. *Science* **170**(3956): 447-&.

Lindgren V, Ares M, Weiner AM, Francke U. 1985. Human Genes For U2 Small Nuclear-RNA Map To A Major Adenovirus-12 Modification Site On Chromosome-17. *Nature* **314**(6006): 115-116.

Liu JF, Gough J, Rost B. 2006. Distinguishing Protein-Coding From Non-Coding RNAs Through Support Vector Machines. *Plos Genet* **2**(4): 529-536.

Liu YF, Li SJ, Chen Y, Kimberlin AN, Cahoon EB, Yu B. 2016. SnRNA 3 ' End Processing By A CPSF73-Containing Complex Essential For Development In Arabidopsis. *Plos Biol* **14**(10): 22.

Lobo SM, Hernandez N. 1989. A 7-Bp Mutation Converts A Human RNA Polymerase-II SnRNA Promoter Into An RNA Polymerase-III Promoter. *Cell* **58**(1): 55-67.

Loraine AE, Mccormick S, Estrada A, Patel K, Qin P. 2013. RNA-Seq Of Arabidopsis Pollen Uncovers Novel Transcription And Alternative Splicing. *Plant Physiol* **162**(2): 1092-1109.

Lu H, Flores O, Weinmann R, Reinberg D. 1991. The Nonphosphorylated Form Of RNA Polymerase-II Preferentially Associates With The Preinitiation Complex. *Proc Natl Acad Sci U S A* **88**(22): 10004-10008.

Malamy JE, Benfey PN. 1997. Organization And Cell Differentiation In Lateral Roots Of Arabidopsis Thaliana. *Development* **124**(1): 33-44.

Manabe Y, Tinker N, Colville A, Miki B. 2007. CSR1, The Sole Target Of Imidazolinone Herbicide In Arabidopsis Thaliana. *Plant Cell Physiol* **48**(9): 1340-1358.

Manavella PA, Hagmann J, Ott F, Laubinger S, Franz M, Macek B, Weigel D. 2012. Fast-Forward Genetics Identifies Plant CPL Phosphatases As Regulators Of MiRNA Processing Factor HYL1. *Cell* **151**(4): 859-870.

Manser T, Gesteland RF. 1982. Human U1 Loci - Genes For Human U1 RNA Have Dramatically Similar Genomic Environments. *Cell* **29**(1): 257-264.

Marker C, Zemann A, Terhorst T, Kiefmann M, Kastenmayer JP, Green P, Bachellerie JP, Brosius J, Huttenhofer A. 2002. Experimental Rnomics: Identification Of 140 Candidates For Small Non-Messenger RNAs In The Plant Arabidopsis Thaliana. *Curr Biol* **12**(23): 2002-2013.

Martin M, Medina FJ. 1991. A Drosophila Anti-RNA Polymerase-II Antibody Recognizes A Plant Nucleolar Antigen, RNA Polymerase-I, Which Is Mostly Localized In Fibrillar Centers. *J Cell Sci* **100**: 99-107.

Matsuda O, Sakamoto H, Nakao Y, Oda K, Iba K. 2009. CTD Phosphatases In The Attenuation Of Wound-Induced Transcription Of Jasmonic Acid Biosynthetic Genes In Arabidopsis. *Plant J* **57**(1): 96-108.

Matsuo N, Mase H, Makino M, Takahashi H, Banno H. 2009. Identification Of ENHANCER OF SHOOT REGENERATION 1-Upregulated Genes During In Vitro Shoot Regeneration. *Plant Biotechnol* **26**(4): 385-393.

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. 2012. CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment To RNA Polymerase II. *Science* **336**(6089): 1723-1725.

Mcstay B, Paule M, Schultz MC, Willis I, Pikaard CS. 2002. Oddpols United: New Insights Into Transcription By RNA Polymerases I And III. *Gene Expr* **10**(5-6): 263-269.

Medlin JE, Uguen P, Taylor A, Bentley DL, Murphy S. 2003. The C-Terminal Domain Of Pol II And A DRB-Sensitive Kinase Are Required For 3' Processing Of U2 SnRNA. *EMBO J* **22**(4): 925-934.

Menges M, De Jager SM, Gruissem W, Murray JAH. 2005. Global Analysis Of The Core Cell Cycle Regulators Of Arabidopsis Identifies Novel Genes, Reveals Multiple And Highly Specific Profiles Of Expression And Provides A Coherent Model For Plant Cell Cycle Control. *Plant J* **41**(4): 546-566.

Menges M, Samland AK, Planchais S, Murray JAH. 2006. The D-Type Cyclin CYCD3;1 Is Limiting For The G1-To-S-Phase Transition In Arabidopsis. *Plant Cell* **18**(4): 893-906.

Morgan DO. 1995. Principles Of CDK Regulation. *Nature* **374**(6518): 131-134.

Motte H, Verstraeten I, Werbrouck S, Geelen D. 2011. CUC2 As An Early Marker For Regeneration Competence In Arabidopsis Root Explants. *J Plant Physiol* **168**(13): 1598-1601.

Mudge JM, Harrow J. 2015. Creating Reference Gene Annotation For The Mouse C57BL6/J Genome Assembly. *Mamm Genome* **26**(9-10): 366-378.

Myer VE, Young RA. 1998. RNA Polymerase II Holoenzymes And Subcomplexes. *J Biol Chem* **273**(43): 27757-27760.

Nagaya S, Kawamura K, Shinmyo A, Kato K. 2010. The HSP Terminator Of Arabidopsis Thaliana Increases Gene Expression In Plant Cells. *Plant Cell Physiol* **51**(2): 328-332.

Nagpal P, Ellis CM, Weber H, Ploense SE, Barkawi LS, Guilfoyle TJ, Hagen G, Alonso JM, Cohen JD, Farmer EE et al. 2005. Auxin Response Factors ARF6 And ARF8 Promote Jasmonic Acid Production And Flower Maturation. *Development* **132**(18): 4107-4118.

Navarro P, Pichard S, Ciaudo C, Avner P, Rougeulle C. 2005. Tsix Transcription Across The Xist Gene Alters Chromatin Conformation Without Affecting Xist Transcription: Implications For X-Chromosome Inactivation. *Genes Dev* **19**(12): 1474-1484.

Nawrath C, Schell J, Koncz C. 1990. Homologous Domains Of The Largest Subunit Of Eucaryotic RNA Polymerase II Are Conserved In Plants. *Mol Gen Genet* **223**(1): 65-75.

Nelson JD, Denisenko O, Bomsztyk K. 2006. Protocol For The Fast Chromatin Immunoprecipitation (Chip) Method. *Nat Protoc* **1**(1): 179-185.

Newell CA. 2000. Plant Transformation Technology - Developments And Applications. *Mol Biotechnol* **16**(1): 53-65.

Ni ZY, Schwartz BE, Werner J, Suarez JR, Lis JT. 2004. Coordination Of Transcription, RNA Processing, And Surveillance By P-TEFb Kinase On Heat Shock Genes. *Mol Cell* **13**(1): 55-65.

Niebel A, Engler JD, Hemerly A, Ferreira P, Inze D, Vanmontagu M, Gheysen G. 1996. Induction Of Cdc2a And Cyc1at Expression In Arabidopsis Thaliana During Early Phases Of Nematode-Induced Feeding Cell Formation. *Plant J* **10**(6): 1037-1043.

Nishimura C, Ohashi Y, Sato S, Kato T, Tabata S, Ueguchi C. 2004. Histidine Kinase Homologs That Act As Cytokinin Receptors Possess Overlapping Functions In The Regulation Of Shoot And Root Growth In Arabidopsis. *Plant Cell* **16**(6): 1365-1377.

Nowack MK, Harashima H, Dissmeyer N, Zhao XA, Bouyer D, Weimer AK, De Winter F, Yang F, Schnittger A. 2012. Genetic Framework Of Cyclin-Dependent Kinase Function In Arabidopsis. *Dev Cell* **22**(5): 1030-1040.

Oakenfull EA, Riou-Khamlichi C, Murray JAH. 2002. Plant D-Type Cyclins And The Control Of G1 Progression. *Philos Trans R Soc Lond Ser B-Biol Sci* **357**(1422): 749-760.

Obayashi T, Nishida K, Kasahara K, Kinoshita K. 2011. ATTED-II Updates: Condition-Specific Gene Coexpression To Extend Coexpression Analyses And Applications To A Broad Range Of Flowering Plants. *Plant Cell Physiol* **52**(2): 213-219.

Oh E, Zhu JY, Bai MY, Arenhart RA, Sun Y, Wang ZY. 2014. Cell Elongation Is Regulated Through A Central Circuit Of Interacting Transcription Factors In The Arabidopsis Hypocotyl. *Elife* **3**: 47.

Ohshima Y, Itoh M, Okada N, Miyata T. 1981. Novel Models For RNA Splicing That Involve A Small Nuclear-RNA. *Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences* **78**(7): 4471-4474.

Ohtani M, Demura T, Sugiyama M. 2010. Particular Significance Of SRD2-Dependent SnRNA Accumulation In Polarized Pattern Generation During Lateral Root Development Of Arabidopsis. *Plant Cell Physiol* **51**(12): 2002-2012.

____. 2013. Arabidopsis ROOT INITIATION DEFECTIVE1, A DEAH-Box RNA Helicase Involved In Pre-MRNA Splicing, Is Essential For Plant Development. *Plant Cell* **25**(6): 2056-2069.

Ohtani M, Sugiyama M. 2005. Involvement Of SRD2-Mediated Activation Of SnRNA Transcription In The Control Of Cell Proliferation Competence In Arabidopsis. *Plant J* **43**(4): 479-490.

Ohtani M, Takebayashi A, Hiroyama R, Xu B, Kudo T, Sakakibara H, Sugiyama M, Demura T. 2015. Cell Dedifferentiation And Organogenesis In Vitro Require

More SnRNA Than Does Seedling Development In Arabidopsis Thaliana. *J Plant Res* **128**(3): 371-380.

Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, Pikaard CS. 2005. Plant Nuclear RNA Polymerase IV Mediates SiRNA And DNA Methylation-Dependent Heterochromatin Formation. *Cell* **120**(5): 613-622.

Page RDM. 1996. Treeview: An Application To Display Phylogenetic Trees On Personal Computers. *Comput Appl Biosci* **12**(4): 357-358.

Palancade B, Bellier S, Almouzni G, Bensaude O. 2001. Incomplete RNA Polymerase II Phosphorylation In Xenopus Laevis Early Embryos. *J Cell Sci* **114**(Pt 13): 2483-2489.

Paroo Z, Ye XC, Chen S, Liu QH. 2009. Phosphorylation Of The Human MicroRNA-Generating Complex Mediates MAPK/Erk Signaling. *Cell* **139**(1): 112-122.

Patturajan M, Schulte RJ, Sefton BM, Berezney R, Vincent M, Bensaude O, Warren SL, Corden JL. 1998. Growth-Related Changes In Phosphorylation Of Yeast RNA Polymerase II. *J Biol Chem* **273**(8): 4689-4694.

Paule MR, White RJ. 2000. Transcription By RNA Polymerases I And III. *Nucleic Acids Res* **28**(6): 1283-1298.

Peterlin BM, Price DH. 2006. Controlling The Elongation Phase Of Transcription With P-TEFb. *Mol Cell* **23**(3): 297-305.

Phatnani HP, Greenleaf AL. 2006. Phosphorylation And Functions Of The RNA Polymerase II CTD. *Genes Dev* **20**(21): 2922-2936.

Pikaard CS, Haag JR, Ream T, Wierzbicki AT. 2008. Roles Of RNA Polymerase IV In Gene Silencing. *Trends Plant Sci* **13**(7): 390-397.

Poppenberger B, Berthiller F, Lucyshyn D, Sieberer T, Schuhmacher R, Krska R, Kuchler K, Glossl J, Luschnig C, Adam G. 2003. Detoxification Of The Fusarium Mycotoxin Deoxynivalenol By A UDP-Glucosyltransferase From Arabidopsis Thaliana. *J Biol Chem* **278**(48): 47905-47914.

Poppenberger B, Fujioka S, Soeno K, George GL, Vaistij FE, Hiranuma S, Seto H, Takatsuto S, Adam G, Yoshida S et al. 2005. The UGT73C5 Of Arabidopsis Thaliana Glucosylates Brassinosteroids. *Proc Natl Acad Sci U S A* **102**(42): 15253-15258.

Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolic L, Pikaard CS. 2009. Subunit Compositions Of The RNA-Silencing Enzymes Pol IV And Pol V Reveal Their Origins As Specialized Forms Of RNA Polymerase II. *Mol Cell* **33**(2): 192-203.

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. 2002. MicroRNAs In Plants. *Genes Dev* **16**(13): 1616-1626.

Reyes-Reyes M, Hampsey M. 2007. Role For The Ssu72 C-Terminal Domain Phosphatase In RNA Polymerase II Transcription Elongation. *Mol Cell Biol* **27**(3): 926-936.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E et al. 2007. Functional Demarcation Of Active

And Silent Chromatin Domains In Human HOX Loci By Noncoding RNAs. *Cell* **129**(7): 1311-1323.

Riou-Khamlichi C, Huntley R, Jacqmard A, Murray JAH. 1999. Cytokinin Activation Of Arabidopsis Cell Division Through A D-Type Cyclin. *Science* **283**(5407): 1541-1544.

Roeder RG, Rutter WJ. 1969. Multiple Forms Of DNA-Dependent RNA Polymerase In Eukaryotic Organisms. *Nature* **224**(5216): 234-&.

Ruiz MT, Voinnet O, Baulcombe DC. 1998. Initiation And Maintenance Of Virus-Induced Gene Silencing. *Plant Cell* **10**(6): 937-946.

Sadowski CL, Henry RW, Lobo SM, Hernandez N. 1993. Targeting Tbp To A Non-Tata Box Cis-Regulatory Element - A Tbp-Containing Complex Activates Transcription From SnRNA Promoters Through The Pse. *Genes Dev* **7**(8): 1535-1548.

Saleh A, Alvarez-Venegas R, Avramova Z. 2008. An Efficient Chromatin Immunoprecipitation (Chip) Protocol For Studying Histone Modifications In Arabidopsis Plants. *Nat Protoc* **3**(6): 1018-1025.

Shim EY, Walker AK, Shi Y, Blackwell TK. 2002. CDK-9/Cyclin T (P-TEFb) Is Required In Two Postinitiation Pathways For Transcription In The C-Elegans Embryo. *Genes Dev* **16**(16): 2135-2146.

Shimotohno A, Matsubayashi S, Yamaguchi M, Uchimiya H, Umeda M. 2003. Differential Phosphorylation Activities Of CDK-Activating Kinases In Arabidopsis Thaliana. *FEBS Lett* **534**(1-3): 69-74.

Shimotohno A, Umeda-Hara C, Bisova K, Uchimiya H, Umeda M. 2004. The Plant-Specific Kinase CDKF;1 Is Involved In Activating Phosphorylation Of Cyclin-Dependent Kinase-Activating Kinases In Arabidopsis. *Plant Cell* **16**(11): 2954-2966.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, Mcwilliam H, Remmert M, Soding J et al. 2011. Fast, Scalable Generation Of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol Syst Biol* **7**: 6.

SKOOG F. 1950. Chemical Control Of Growth And Organ Formation In Plant Tissues. *L'Annee Biologique* **54**(10): 545-562.

____. 1957. Chemical Regulation Of Growth And Organ Formation In Plant Tissues Cultured In Vitro. In *Symp Soc Exp Biol*, Vol 11, Pp. 118-131.

Solymosy F, Pollak T. 1993. Uridylate-Rich Small Nuclear RNAs (UsnRNAs), Their Genes And Pseudogenes, And Usnrnps In Plants - Structure And Function - A Comparative Approach. *Crit Rev Plant Sci* **12**(4): 275-369.

Sontheimer EJ, Steitz JA. 1993. The U5 And U6 Small Nuclear RNAs As Active-Site Components Of The Spliceosome. *Science* **262**(5142): 1989-1996.

Spahr H, Calero G, Bushnell DA, Kornberg RD. 2009. Schizosacharomyces Pombe RNA Polymerase II At 3.6-Angstrom Resolution. *Proc Natl Acad Sci U S A* **106**(23): 9185-9190.

Sugimoto K, Jiao YL, Meyerowitz EM. 2010. Arabidopsis Regeneration From Multiple Tissues Occurs Via A Root Development Pathway. *Dev Cell* **18**(3): 463-471.

Sugiyama M. 2003. Isolation And Initial Characterization Of Temperature-Sensitive Mutants Of Arabidopsis Thaliana That Are Impaired In Root Redifferentiation. *Plant Cell Physiol* **44**(6): 588-596.

Suh H, Hazelbaker DZ, Soares LM, Buratowski S. 2013. The C-Terminal Domain Of Rpb1 Functions On Other RNA Polymerase II Subunits. *Mol Cell* **51**(6): 850-858.

Sukegawa Y, Yamashita A, Yamamoto M. 2011. The Fission Yeast Stress-Responsive MAPK Pathway Promotes Meiosis Via The Phosphorylation Of Pol II CTD In Response To Environmental And Feedback Cues. *Plos Genet* **7**(12): 12.

Szczesniak MW, Rosikiewicz W, Makalowska I. 2016. Cantatadb: A Collection Of Plant Long Non-Coding RNAs. *Plant Cell Physiol* **57**(1): 7.

Thaller MC, Schippa S, Rossolini GM. 1998. Conserved Sequence Motifs Among Bacterial, Eukaryotic, And Archaeal Phosphatases That Define A New Phosphohydrolase Superfamily. *Protein Sci* **7**: 1651-1656.

Tietjen JR, Zhang DW, Rodriguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X, Chapman RD, Shokat K, Keles S et al. 2010. Chemical-Genomic Dissection Of The CTD Code. *Nat Struct Mol Biol* **17**(9): 1154-U1116.

To JPC, Haberer G, Ferreira FJ, Deruere J, Mason MG, Schaller GE, Alonso JM, Ecker JR, Kieber JJ. 2004. Type-A Arabidopsis Response Regulators Are Partially Redundant Negative Regulators Of Cytokinin Signaling. *Plant Cell* **16**(3): 658-671.

Tognetti VB, Van Aken O, Morreel K, Vandenbroucke K, Van De Cotte B, De Clercq I, Chiwocha S, Fenske R, Prinsen E, Boerjan W et al. 2010. Perturbation Of Indole-3-Butyric Acid Homeostasis By The UDP-Glucosyltransferase UGT74E2 Modulates Arabidopsis Architecture And Water Stress Tolerance. *Plant Cell* **22**(8): 2660-2679.

Tommasini R, Vogt E, Fromenteau M, Hortensteiner S, Matile P, Amrhein N, Martinoia E. 1998. An ABC-Transporter Of Arabidopsis Thaliana Has Both Glutathione-Conjugate And Chlorophyll Catabolite Transport Activity. *Plant J* **13**(6): 773-780.

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. 2010. Long Noncoding RNA As Modular Scaffold Of Histone Modification Complexes. *Science* **329**(5992): 689-693.

Tsugama D, Liu S, Takano T. 2011. A Rapid Chemical Method For Lysing Arabidopsis Cells For Protein Analysis. *Plant Methods* **7**: 22.

Ueda A, Li P, Feng Y, Vikram M, Kim S, Kang CH, Kang JS, Bahk JD, Lee SY, Fukuhara T et al. 2008. The Arabidopsis Thaliana Carboxyl-Terminal Domain Phosphatase-Like 2 Regulates Plant Growth, Stress And Auxin Responses. *Plant Mol Biol* **67**(6): 683-697.

Uesaka M, Nishimura O, Go Y, Nakashima K, Agata K, Imamura T. 2014. Bidirectional Promoters Are The Major Source Of Gene Activation-Associated Non-Coding RNAs In Mammals. *BMC Genomics* **15**: 14.

Uguen P, Murphy S. 2003. The 3 ' Ends Of Human Pre-SnRNAs Are Produced By RNA Polymerase IICTD-Dependent RNA Processing. *EMBO J* **22**(17): 4544-4554.

Umeda M, Bhalerao RP, Schell J, Uchimiya H, Koncz C. 1998. A Distinct Cyclin-Dependent Kinase-Activating Kinase Of Arabidopsis Thaliana. *Proc Natl Acad Sci U S A* **95**(9): 5021-5026.

Valgardsdottir R, Chiodi F, Giordano M, Cobianchi F, Riva S, Biamonti G. 2005. Structural And Functional Characterization Of Noncoding Repetitive RNAs Transcribed In Stressed Human Cells. *Mol Biol Cell* **16**(6): 2597-2604.

Valgardsdottir R, Chiodi I, Giordano M, Rossi A, Bazzini S, Ghigna C, Riva S, Biamonti G. 2008. Transcription Of Satellite III Non-Coding RNAs Is A General Stress Response In Human Cells. *Nucleic Acids Res* **36**(2): 423-434.

Van Leene J, Witters E, Inze D, De Jaeger G. 2008. Boosting Tandem Affinity Purification Of Plant Protein Complexes. *Trends Plant Sci* **13**(10): 517-520.

Vanarsdell SW, Weiner AM. 1984. Human Genes For U2 Small Nuclear-RNA Are Tandemly Repeated. *Mol Cell Biol* **4**(3): 492-499.

Varon R, Gooding R, Steglich C, Marns L, Tang H, Angelicheva D, Yong KK, Ambrugger P, Reinhold A, Morar B et al. 2003. Partial Deficiency Of The C-Terminal-Domain Phosphatase Of RNA Polymerase II Is Associated With Congenital Cataracts Facial Dysmorphism Neuropathy Syndrome. *Nat Genet* **35**(2): 185-189.

Vasil V, Hildebrandt AC. 1965. Differentiation Of Tobacco Plants From Single Isolated Cells In Microcultures. *Science* **150**(3698): 889-+.

Vasil V, Hilderbr.Ac. 1965. Growth And Tissue Formation From Single Isolated Tobacco Cells In Microculture. *Science* **147**(3664): 1454-&.

Venetianer A, Dubois MF, Nguyen V, Bellier S, Seo SJ, Bensaude O. 1995. Phosphorylation State Of The RNA-Polymerase-II C-Terminal Domain (Ctd) In Heat-Shocked Cells - Possible Involvement Of The Stress-Activated Mitogen-Activated Protein (Map) Kinases. *Eur J Biochem* **233**(1): 83-92.

Vera JM, Dowell RD. 2016. Survey Of Cryptic Unstable Transcripts In Yeast. *BMC Genomics* **17**: 14.

Vilborg A, Passarelli MC, Yario TA, Tycowski KT, Steitz JA. 2015. Widespread Inducible Transcription Downstream Of Human Genes. *Mol Cell* **59**(3): 449-461.

Visconti R, Palazzo L, Della Monica R, Grieco D. 2012. Fcp1-Dependent Dephosphorylation Is Required For M-Phase-Promoting Factor Inactivation At Mitosis Exit. *Nat Commun* **3**: 10.

Visvanathan J, Lee S, Lee B, Lee JW, Lee SK. 2007. The MicroRNA Mir-124 Antagonizes The Anti-Neural REST/SCP1 Pathway During Embryonic CNS Development. *Genes Dev* **21**(7): 744-749.

Waibel F, Filipowicz W. 1990. RNA-Polymerase Specificity Of Transcription Of Arabidopsis U-SnRNA Genes Determined By Promoter Element Spacing. *Nature* **346**(6280): 199-202.

Walker AK, Boag PR, Blackwell TK. 2007. Transcription Reactivation Steps Stimulated By Oocyte Maturation In C-Elegans. *Dev Biol* **304**(1): 382-393.

Wang BB, Brendel V. 2004. The ASRG Database: Identification And Survey Of Arabidopsis Thaliana Genes Involved In Pre-MRNA Splicing. *Genome Biology* **5**(12).

Wang W, Cho HS, Kim R, Jancarik J, Yokota H, Nguyen HH, Grigoriev IV, Wemmer DE, Kim SH. 2002. Structural Characterization Of The Reaction Pathway In Phosphoserine Phosphatase: Crystallographic "Snapshots" Of Intermediate States. *J Mol Biol* **319**(2): 421-431.

Wang WM, Chen XM. 2004. HUA ENHANCER3 Reveals A Role For A Cyclin-Dependent Protein Kinase In The Specification Of Floral Organ Identity In Arabidopsis. *Development* **131**(13): 3147-3156.

Wang YQ, Wang XC, Deng W, Fan XD, Liu TT, He GM, Chen RS, Terzaghi W, Zhu DM, Deng XW. 2014. Genomic Features And Regulatory Roles Of Intermediate-Sized Non-Coding RNAs In Arabidopsis. *Mol Plant* **7**(3): 514-527.

Wani S, Yuda M, Fujiwara Y, Yamamoto M, Harada F, Ohkuma Y, Hirose Y. 2014. Vertebrate Ssu72 Regulates And Coordinates 3'-End Formation Of RNAs Transcribed By RNA Polymerase II. *Plos One* **9**(8): 13.

Weiss SB, Gladstone L. 1959. A Mammalian System For The Incorporation Of Cytidine Triphosphate Into Ribonucleic Acid. *J Am Chem Soc* **81**(15): 4118-4119.

Wen YX, Shatkin AJ. 1999. Transcription Elongation Factor Hspt5 Stimulates MRNA Capping. *Genes Dev* **13**(14): 1774-1779.

Western TL, Cheng YL, Liu J, Chen XM. 2002. HUA ENHANCER2, A Putative Dexm-Box RNA Helicase, Maintains Homeotic B And C Gene Expression In Arabidopsis. *Development* **129**(7): 1569-1581.

Westin G, Zabielski J, Hammarstrom K, Monstein HJ, Bark C, Pettersson U. 1984. Clustered Genes For Human-U2 RNA. *Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences* **81**(12): 3811-3815.

Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding Transcription By RNA Polymerase Pol Ivb/Pol V Mediates Transcriptional Silencing Of Overlapping And Adjacent Genes. *Cell* **135**(4): 635-648.

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA Polymerase V Transcription Guides ARGONAUTE4 To Chromatin. *Nat Genet* **41**(5): 630-634.

Wrighton KH, Willis D, Long J, Liu F, Lin X, Feng XH. 2006. Small C-Terminal Domain Phosphatases Dephosphorylate The Regulatory Linker Regions Of Smad2 And Smad3 To Enhance Transforming Growth Factor-Beta Signaling. *J Biol Chem* **281**(50): 38365-38375.

Xiang CB, Han P, Lutziger I, Wang K, Oliver DJ. 1999. A Mini Binary Vector Series For Plant Transformation. *Plant Mol Biol* **40**(4): 711-717.

Xie ZX, Allen E, Wilken A, Carrington JC. 2005. DICER-LIKE 4 Functions In Trans-Acting Small Interfering RNA Biogenesis And Vegetative Phase Change In Arabidopsis Thaliana. *Proc Natl Acad Sci U S A* **102**(36): 12984-12989.

Xiong LM, Lee H, Ishitani M, Tanaka Y, Stevenson B, Koiwa H, Bressan RA, Hasegawa PM, Zhu JK. 2002. Repression Of Stress-Responsive Genes By FIERY2, A Novel Transcriptional Regulator In Arabidopsis. *Proc Natl Acad Sci U S A* **99**(16): 10899-10904.

Xu K, Liu J, Fan MZ, Xin W, Hu YX, Xu CY. 2012. A Genome-Wide Transcriptome Profiling Reveals The Early Molecular Events During Callus Initiation In Arabidopsis Multiple Organs. *Genomics* **100**(2): 116-124.

Xu YX, Hirose Y, Zhou XZ, Lu KP, Manley JL. 2003. Pin1 Modulates The Structure And Function Of Human RNA Polymerase II. *Genes Dev* **17**(22): 2765-2776.

Yasutani I, Ozawa S, Nishida T, Sugiyama M, Komamine A. 1994. Isolation Of Temperature-Sensitive Mutants Of Arabidopsis Thaliana That Are Defective In The Redifferentiation Of Shoots. *Plant Physiol* **105**(3): 815-822.

Yeo M, Lee SK, Lee B, Ruiz EC, Pfaff SL, Gill GN. 2005. Small CTD Phosphatases Function In Silencing Neuronal Gene Expression. *Science* **307**(5709): 596-600.

Yeo M, Lin PS, Dahmus ME, Gill GN. 2003. A Novel RNA Polymeras II C-Terminal Domain Phosphatase That Preferentially Dephosphorylates Serine 5. *J Biol Chem* **278**: 26078-26085.

Yu N, Nutzmann HW, Macdonald JT, Moore B, Field B, Berriri S, Trick M, Rosser SJ, Kumar SV, Freemont PS et al. 2016. Delineation Of Metabolic Gene Clusters In Plant Genomes By Chromatin Signatures. *Nucleic Acids Res* **44**(5): 2255-2265.

Yu X, Chini CC, He M, Mer G, Chen J. 2003. The BRCT Domain Is A Phospho-Protein Binding Domain. *Science* **302**(5645): 639-642.

Yukawa Y, Sugita M, Sugiura M. 1997. Efficient In Vitro Transcription Of Plant Nuclear TRNA(Ser) Genes In A Nuclear Extract From Tobacco Cultured Cells. *Plant J* **12**(4): 965-970.

Yuo CY, Ares M, Weiner AM. 1985. Sequences Required For 3' End Formation Of Human U2 Small Nuclear-RNA. *Cell* **42**(1): 193-202.

Zehring WA, Lee JM, Weeks JR, Jokerst RS, Greenleaf AL. 1988. The C-Terminal Repeat Domain Of RNA Polymerase-II Largest Subunit Is Essential Invivo But Is Not Required For Accurate Transcription Initiation Invitro. *Proc Natl Acad Sci U S A* **85**(11): 3698-3702.

Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ. 2012a. Ssu72 Phosphatase-Dependent Erasure Of Phospho-Ser7 Marks On The RNA Polymerase II C-Terminal Domain Is Essential For Viability And Transcription Termination. *J Biol Chem* **287**(11): 8541-8551.

Zhang MM, Wang XJ, Chen X, Bowman ME, Luo YH, Noel JP, Ellington AD, Etzkorn FA, Zhang Y. 2012b. Structural And Kinetic Analysis Of Prolyl-Isomerization/Phosphorylation Cross-Talk In The CTD Code. *ACS Chem Biol* **7**(8): 1462-1470.

Zhang WN, Thieme CJ, Kollwig G, Apelt F, Yang L, Winter N, Andresen N, Walther D, Kragler F. 2016. TRNA-Related Sequences Trigger Systemic MRNA Transport In Plants. *Plant Cell* **28**(6): 1237-1249.

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb Proteins Targeted By A Short Repeat RNA To The Mouse X Chromosome. *Science* **322**(5902): 750-756.

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. 2004. GENEVESTIGATOR. Arabidopsis Microarray Database And Analysis Toolbox. *Plant Physiol* **136**(1): 2621-2632.

APPENDIX

Appendix A2.1 Functional GO Annotation of genes affected by *CPL4*$_{RNAi}$ (Chapter II; Microaray)

GO id	GO: Molecular Function	UP	DOWN
0016787	hydrolase activity	26	31
0016301	kinase activity	12	5
0016740	transferase activity	33	23
0003824	other enzyme activity	55	55
0003700	transcription factor activity	25	10
0003677///0003723	DNA or RNA binding	28	11
0003676	nucleic acid binding	9	1
0000166	nucleotide binding	30	15
0005515	protein binding	23	27
0005102///0004872	receptor binding or activity	2	2
0005488	other binding	64	60
0005198	structural molecule activity	0	3
0005215	transporter activity	17	10
0005554	unknown molecular functions	35	44
0003674	other molecular functions	22	30

GO id	GO: Cellular Component	UP	DOWN
0005739	mitochondria	21	10
0009507	chloroplast	22	29
0009536	plastid	3	8
0005840	ribosome	0	0
0005829	cytosol	9	8
0005783	ER	7	4
0005794	Golgi apparatus	4	5
0005737	other cytoplasmic components	53	57
0005634	nucleus	73	44
0005622	other intracellular components	33	32
0005886	plasma membrane	31	26
0016020	other membranes	28	34
0008372	unknown cellular components	5	6
0005576	extracellular	34	64
0005618	cell wall	10	23
0005575	other cellular components	14	17

(A2.1 continued to next page)

Appendix A2.1 Functional GO Annotation of genes affected by *CPL4*$_{RNAi}$ (Chapter II; Microaray) continued

GO id	GO: Biological Process	UP	DOWN
0000004	unknown biological processes	36	29
0007275	developmental processes	36	49
0006810	transport	36	33
0007165	signal transduction	24	29
0016043	cell organization and biogenesis	14	19
0009987	other cellular processes	130	134
0006259///0006403	DNA or RNA metabolism	4	1
0019538	protein metabolism	30	23
0006118///0006091	electron transport or energy pathways	4	1
0006351	transcription,DNA-dependent	28	15
0008152	other metabolic processes	132	118
0009628///0009607	response to abiotic or biotic stimulus	72	62
0006950	response to stress	72	77
0008150	other biological processes	63	67

"GOid" columns indicate the identifier of the GO term. "GO:" columns indicate GO slim term, a group of GO terms defined by TAIR, under three aspects (Molecular Function, Cellular Component and Biological Process). "UP/DOWN" columns represent number of up- or down-regulated genes annotated to the particular GO slim term. The total number in one GO aspect may exceed total number of genes used because a gene can be annotated more than once.

Appendix A2.2 Over-represented GO associated with clusters C1-C3 (Chapter II; Microaray)

Cluster 1

Subcategory name	p-value	Expected # genes	Observed # genes	GeneIDs of test set in subcategory
transcription factor import into nucleus	0.00132	0.00478	2	CRF6 CRF5
drug transmembrane transporter activity	0.00141	0.05263	3	AT2G04050 AT2G04070 TX1
antiporter activity	0.00882	0.11722	3	AT2G04050 AT2G04070 TX1
cotyledon development	0.00882	0.02312	2	CRF6 CRF5
protein import into nucleus	0.00929	0.0311	2	CRF6 CRF5

Cluster 2

Subcategory name	p-value	Expected # genes	Observed # genes	GeneIDs of test set in subcategory
response to abscisic acid stimulus	2.5E-07	0.23603	7	ATHB-7 RD20 ATHB-12 GBF3 AT5G59220AFP1 ABF3
response to water deprivation	3.7E-07	0.15469	6	AtGolS2 ATHB-7 RD20 ATHB-12 AT5G59220ABF3
response to water	3.7E-07	0.16187	6	AtGolS2 ATHB-7 RD20 ATHB-12 AT5G59220ABF3
response to hormone stimulus	6E-05	0.64748	7	ATHB-7 RD20 ATHB-12 GBF3 AT5G59220AFP1 ABF3
response to endogenous stimulus	8.3E-05	0.7025	7	ATHB-7 RD20 ATHB-12 GBF3 AT5G59220AFP1 ABF3

Cluster 3

Subcategory name	p-value	Expected # genes	Observed # genes	GeneIDs of test set in subcategory
secondary metabolic process	9.6E-10	0.14172	7	DFR PAP1 ATGSTF12 F3H TT4ERD9 FMO_GS-OX2
flavonoid biosynthetic process	2.1E-07	0.0232	4	DFR PAP1 F3H TT4
flavonoid metabolic process	2.5E-07	0.02682	4	DFR PAP1 F3H TT4
anthocyanin biosynthetic process	7.4E-07	0.00725	3	DFR PAP1 TT4
phenylpropanoid biosynthetic process	1.6E-06	0.04893	4	DFR PAP1 F3H TT4

Appendix P List of primers used in each chapter

Chapter	Usage	Primer ID	Primer Name	F/R	Seq
II	qPCR	F77	UGT74E2-AT1G05680.1_F	F	TTTCCCTTCGTTCCCGATGCTG
II	qPCR	F78	UGT74E2-AT1G05680.1_R	R	TTCGGGTATGAGGACGATTCGC
II, III, IV	qPCR	E506	GAPDH_QPCR_F	F	TTGGTGACAACAGGTCAAGCA
II, III, IV	qPCR	E507	GAPDH_QPCR_R	R	AAACTTGTCGCTCAATGCAATC
II	qPCR	F145	ANAC13-AT1G32870.1_F	F	CCTGAGGAATTACCTGAGAAAGCG
II	qPCR	F146	ANAC13-AT1G32870.1_R	R	TCCCTGTTGCTTTCCAATAGCC
II	qPCR	F79	PAP1-AT1G56650.1_F	F	TCCTGTAAGAGCTGGGCTAAACC
II	qPCR	F80	PAP1-AT1G56650.1_R	R	CCCTAGAAGCCTATGAAGGCGAAG
II	qPCR	F137	GPI19-AT1G61280.1_F(2)	F	TCGATGGTGACTCTGTTAGTAGC
II	qPCR	F138	GPI19-AT1G61280.1_R(2)	R	TTGGCCTGTCTTCCCCATTC
II	qPCR	F81	TX1-AT2G04040.1_F	F	TCCCATCAGCTGCAATGATTTGTC
II	qPCR	F82	TX1-AT2G04040.1_R	R	AGGTCTCGAGTTTCGGGTTAGG
II, III	qPCR	F83	MATE1-AT2G04050.1_F	F	ATGTGAGGTACTCCAGCTCCTG
II, III	qPCR	F84	MATE1-AT2G04050.1_R	R	ATAGCCACCATTCTAGGCAAACC
II	qPCR	F85	MATE2-AT2G04070.1_F	F	GTCTTCTGTCAAGCAGTTCTTCCG
II	qPCR	F86	MATE2-AT2G04070.1_R	R	TAGCCACCATTCTAGGCAAAGC
II	qPCR	F105	C4U11-AT2G18190.1_F	F	TGTAGTAGTGCCGAGGTGGTAGAC
II	qPCR	F106	C4U11-AT2G18190.1_R	R	AAGCCCTGATAGAGTCACCCTTCC
II	qPCR	F151	OxiM-AT2G21640.1_F	F	TGCCTCAATGGATGAACACAAGAC
II	qPCR	F152	OxiM-AT2G21640.1_R	R	GAGAAGCTCCCGAATATCTTGTCC
II	qPCR	F173	UGT73C1-AT2G36750.1_F	F	TCCTTGACGGAATGACAGAAGGG
II	qPCR	F174	UGT73C1-AT2G36750.1_R	R	AGCTGGCTCGAGCTCTTCAAAC
II	qPCR	F153	DOGT1-AT2G36800.1_F	F	CATCAGTTGGAGGGTTCCTAACAC
II	qPCR	F154	DOGT1-AT2G36800.1_R	R	TGTAAGTAGCGGTAGACCAGCAG
II	qPCR	F155	C4U17-AT2G41730.1_F	F	TCAAATACCACCGGAGAAAGCG
II	qPCR	F156	C4U17-AT2G41730.1_R	R	TTGTGGCCCGACTTGATAGCTG
II	qPCR	F157	AtMRP3-	F	ATGGTTCTGCTTCTAAGCAATGGG

			AT3G13080.1_F		
II	qPCR	F158	AtMRP3-AT3G13080.1_R	R	TTGAGGTGTACTCAGCCACAAGC
II	qPCR	F87	AOX1a-AT3G22370.1_F	F	AGCATCATGTTCCAACGACGTTTC
II	qPCR	F88	AOX1a-AT3G22370.1_R	R	GCTCGACATCCATATCTCCTCTGG
II	qPCR	F91	CRF6-AT3G61630.1_F	F	TCAAAGGACCTAAAGCGCTCACG
II	qPCR	F92	CRF6-AT3G61630.1_R	R	TGGAGATCGATAACCGGCGTTG
II	qPCR	E502	UBQ10_QPCR_F	F	GGCCTTGTATAATCCCTGATGAATAAG
II	qPCR	E503	UBQ10_QPCR_R	R	AAAGAGATAACAGGAACGGAAACATAGT
II	qPCR	F175	CYP81D8-AT4G37370.1_F	F	CCAACCGGATTACTTCACGGATCG
II	qPCR	F176	CYP81D8-AT4G37370.1_R	R	TCGGTCCCTGCTAGTATCAAAGCG
II	qPCR	F67	SAG29-AT5G13170.1_F	F	TAAGCGCCGTTATGTGGTTCGC
II	qPCR	F68	SAG29-AT5G13170.1_R	R	ATCCCACCACGTTTGGAATCGC
II	qPCR	F97	DFR-AT5G42800.1_F	F	AGGAAGGAAGCTACGATGATGCC
II	qPCR	F98	DFR-AT5G42800.1_R	R	TGTCGGCTTTATCACTTCGTTCTC
II	qPCR	F159	2OGdiox-AT5G43450.1_F	F	TTGCTCCAGATCCTCCGAATCC
II	qPCR	F160	2OGdiox-AT5G43450.1_R	R	TGTATTCGATCACCGCACTCCTG
II	qPCR	F129	HSP20L-AT5G51440.1_F	F	ACTCGTGGAATGGGAGCTTCTG
II	qPCR	F130	HSP20L-AT5G51440.1_R	R	GATGCAACGCGTCGTCTTTCTC
II	qPCR	E360	cpl4-RNAi_F	F	CCAAGGCAAAGCCAGAAGAT
II	qPCR	E361	cpl4-RNAi_R	R	GCTGCATCTATCCATCCTCT
II	qPCR	F131	C3HC4LR-AT5G60250.1_F	F	TCTTCGACTTGCCAGATGGTGTG
II	qPCR	F132	C3HC4LR-AT5G60250.1_R	R	CCGGCAAGTACAATTGCTGACG
II	QuikChange	F30	CPL4D128A(+)	F	gttgtatttagttctcgCtctagaccacacactgc
II	QuikChange	F31	CPL4D128A(-)	R	gcagtgtgtggtctagaGcgagaactaaatacaac
II	Cloning	-	SalI-CPL4FCPH	-	TCTAAGTCGACATGAGCGTAGC
II	Cloning	-	CPL4FCPH-EcoRI (for CPL41-338)	-	ACTAAGAATTCACGCACTTGTTTCAGCATTAACC
II	Cloning	-	CPL4BRCT-EcoRI (for full length)	-	GAGCGGAATTCCGAGTTACGTAT
II	Cloning	-	12attB1CPL4full-F	F	AAAAAGCAGGCTCTATGAGCGTAGCAAGTGATTCTCC
II	Cloning	-	12attB2CPL4full-R_+stop	R	AGAAAGCTGGGTCTCACTCTTCTTCGGTCAATTGCT
II	Cloning	-	12attB1CPL4-BRCT-F_-atg (for CPL4297-440)	F	AAAAAGCAGGCTCTAGTGAACCAGATGGGGCACTT
II	Cloning	-	12attB1CPL4-FCPH-R_+stop (for CPL41-296)	R	AGAAAGCTGGGTCCTACTCATCACTCTTCAACTCCGAT

II	Cloning	-	attB1 adapter	-	CGGGGACAAGTTTGTACAAAAAAGCAGGCT
II	Cloning	-	attB2 adapter	-	GGGGACCACTTTGTACAAGAAAGCTGGGT
II	Genotyping	F212	Locus-specific_84mm_CPL4_R	R	TTACGCACTTGTTTCAGCATTAAC
II	Genotyping	F213	T-DNA primer_8474_R	F	ATAATAACGCTGCGGACATCTACATTTT
II	Genotyping	F214	T-DNA primer_8409_R	R	ATATTGACCATCATACTCATTGC
III	ChIP	F500	U12upst1_F	F	gttcatcgtgtgttatcccatgtcg
III	ChIP	F501	U12upst1_R	R	caaacatgagtctatttaacataatgcccaaag
III	ChIP	F502	U12prom_F	F	cggaactataactcgttggagacc
III	ChIP	F503	U12prom_R	R	GGCAagtattccagatgctgtctatcac
III	qPCR	F189	U12GPI19_1_F	F	GGAAAACAAAGCGTCCGGTG
III	qPCR	F190	U12GPI19_2_R	R	AACACATCGGCCAGCCTATAAC
III	ChIP/qPCR	F332	U12-GPI19_region2_F	F	GCAGAACTCACCTCGTGCGG
III	ChIP/qPCR	F333	U12-GPI19_region2_R	R	cccactaatggcctcaaaatcttcg
III	qPCR	F334	U12-GPI19_region3_F	F	ggtttcataatcgcattttgggattaagc
III	qPCR	F335	U12-GPI19_region3_R	R	ccaatttgaaggacgagacttgctaac
III	ChIP/qPCR	F336	U12-GPI19_region4_F	F	cttaagaaaaaacacatcgcttatcttcgtc
III	ChIP/qPCR	F337	U12-GPI19_region4_R	R	GATCCTACAAAACCATAAACCTCGGAAG
III	RT-PCR	F297	atU1s_F	F	cgcgatcaagaagacgagtgg
III	RT-PCR	F298	atU1s_R	R	GGCTGCGCGAACGCAAACC
III	RT-PCR	F295	atU1-6ext_R	R	ctgaagctcgttagtcgttaagctgg
III	RT-PCR	F296	atU1-7ext_R	R	caggtccaatagatcttcaatgaataacg
III	RT-PCR	F343	atU1-5ext-q_R	R	gccacaatcggagaactaatcaaagactgtcac
III	RT-PCR	F349	atU1-4ext-q_R	R	TGGGACTCTATATACCCAAAGACAAATATCAAAAGGTTGTC
III	RT-PCR	F197	U2.5MtHSC70_1_F	F	CGGCCTTTTGGCTAAGATCAAG
III	RT-PCR	F303	atU2s_R	R	CGACTCGTGAAAGCCCAGA
III	RT-PCR	F200	U2.5MtHSC70_4_F	R	CCCACTGCTTGAAAgGCATCTG
III	RT-PCR	F300	atU2.7ext_R	R	CTTACTACTCATACCCATTAGCTTACATG
III	RT-PCR	F301	atU2-10ext_R	R	ccacttcaaatcaaagtgtttctgaactac
III	RT-PCR	F347	atU2-11ext-q _R	R	ggtttagctttcatttggtatgctctccaatc
III	RT-PCR	F305	atU4-9/s_F	F	ATGAGGTACTAACCGAGGCG
III	RT-PCR	F308	atU4s_R	R	TCTCGAAGGCCCAAGACAAG
III	RT-PCR	F304	atU4.1ext_R	R	ccaaaagatcggataaacataatctctcc
III	RT-PCR	F307	atU4-7ext_R	R	cgaaaccaatctcagttacataagaaagtgg
III	RT-PCR	F306	atU4-9ext_R	R	caactttagagaattgaaaagacaaataccttcg
III	RT-PCR	F359	atU4atac-q_R	R	ccactcgccttgaaaattatgttcgtc
III	RT-PCR	F355	atU4atac-ext-q_R	R	tggttcagtttaattggtttgatcctgttc
III	qPCR	F448	AT3G55850.2_LAF3ISF1_F	F	ACGGGCTTTAGGCTAATTTAACGG
III	qPCR	F449	AT3G55850.1_LAF3ISF2_F	F	TTAAGAATTTCTCCGGCGACTGAC

III	qPCR	F450	AT3G55850_LAF3ISF 1&2_R	R	GAGACTTGAGAGTTGACCAATATAAGTCG
III	RT-PCR	F486	KAKU4_AT4G31430. 2(RT)_F	F	CACACTAGATCTCGTGGAGTTGGTTTG
III	RT-PCR	F487	KAKU4_AT4G31430. 2(RT)_R	R	CCGCTGGTGAACTAGGTGATGAAG
III	RT-PCR	F472	PIIsnR011 SSP14-F1	F	GTCTTTGATTGAAAACTCATGTCTAGAACC AG
III	RT-PCR	F452	AT1G20320_SSP14 5'R	R	CCGTAGTCTCTGTCTACCGTCG
III	RT-PCR	F475	PIIsnR011 SSP14-R3	R	CCAAGCAATGTTCCTCGGATGATG
III	qPCR	F400	SSP14_AT1G20320.1 _F	F	TGTGTTGGCTGATGAGTGTGGAG
III	qPCR	F401	SSP14_AT1G20320.1 _R	R	TCAGGCCAAACATGACGAGTGTC
III	qPCR	E157	AtRD29A_F	F	CAGCACCCAGAAGAAGTTGAACA
III	qPCR	E158	AtRD29A_R	R	TCTTGCTCATGCTCATTGCTT
IV	RT-PCR	F366	STM_AT1G62360.1_ F	F	ACCTTCCTCTTTCTCCGGTTATGG
IV	RT-PCR	F367	STM_AT1G62360.1_ R	R	GCGCAAGAGCTGTCCTTTAAGC
IV	RT-PCR	F368	WUS_AT2G17950.1_ F	F	TCATCACGGTGTTCCCATGCAG
IV	RT-PCR	F369	WUS_AT2G17950.1_ R	R	CCCGTTATTGAAGCTGGGATATGG
IV	qPCR	F389	ARR5_AT3G48100.1 _F	F	AGTTCGGTTGGATTTGAGGATCTG
IV	qPCR	F390	ARR5_AT3G48100.1 _R	R	TCCAGTCATCCCAGGCATAGAG
IV	qPCR	F360	ESR1_AT1G12980.1 _F	F	CATTTGACACGGCGGAACAAGC
IV	qPCR	F361	ESR1_AT1G12980.1 _R	R	TTTGCTCCACGAAAGGCACGAG
IV	qPCR	F372	CUC1_AT3G15170.1 _F	F	ATCGCCTTGACGGCAAATTCTC
IV	qPCR	F373	CUC1_AT3G15170.1 _R	R	CAGAGAACCCATTCATCCTTAGCG
IV	qPCR	F378	CUC2_AT5G53950.1 _F	F	AACTTCCCGGGAGAGCTAAGATGG
IV	qPCR	F379	CUC2_AT5G53950.1 _R	R	TACTTCCGGTCACGGAGGCTAAAG
IV	qPCR	F440	CYCB1;1_AT4G3749 0.1_F	F	ACCTCGCAGCTGTGGAATATGTG
IV	qPCR	F441	CYCB1;1_AT4G3749 0.1_R	R	ATCTCGTGGCCTCCATTCACTCTC
IV	qPCR	F444	H3.3_AT1G75600.1_ F	F	ATCTGTGTGCCATTCATGCCAAG
IV	qPCR	F445	H3.3_AT1G75600.1_ R	R	ATTCGAATCTAAGCACGCTCTCC