UNIVERSITY OF CALIFORNIA
RIVERSIDE

Uncovering Factors and Molecular Mechanisms in Transcriptional Regulation in *Arabidopsis thaliana*

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

Thanh Thu Dinh

September 2012

Dissertation Committee:

Dr. Xuemei Chen, Chairperson
Dr. Patricia S. Springer
Dr. Sean Cutler

The Dissertation of Thanh Thu Dinh is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

This work would not have been possible without the help and support of everyone around me.

First of all, I am deeply grateful to my adviser, Dr. Xuemei Chen for her unconditional support throughout my graduate career. Mere words could not simply describe how indebted I am to her. Her unrelenting guidance has shaped me into the scientist that I am today. From her, I learned how to think and function as an independent scientist. Under her tutelage, I learned how to be a leader, shape opinions, and be a critical thinker. In addition to the science aspect, Dr. Chen has also been a pillar for my personal life as well. She has always been supportive of my non-scientific endeavors and was always willing to lend an ear when life issues occurred. All of these positive aspects that I have learned from her, including the ones I have failed to mention, has provided me with a strong foundation for success in my future scientific endeavors and goals.

I would also like to thank the other members of my dissertation committee—Drs. Patricia Springer and Sean Cutler--for their support, advice and encouragement. Both professors have offered invaluable advice throughout my graduate career. Specifically, Dr. Springer has always been supportive of my work and was pivotal in making my transition to graduate school go more smoothly. Work done for Chapter One would not have been possible without the support from her and her lab members. As we did not have any biochemical experience, she allowed me to come to her lab and learn techniques

firsthand from her graduate student, Aman Husbands. Dr. Sean Cutler has always been honest. He has always pushed me think outside my comfort zone, whether it was with my project or in class. He continually challenged me and offered constructive criticism thereby pushing me to become a better scientist.

In addition, I would like to thank Dr. Julia Bailey-Serres for giving me the opportunity to participate in the NSF ChemGen IGERT program. Through this program, not only did my science improve, I have forged life-long relationships with my colleagues. I would also like to thank Drs. Harley Smith, Thomas Eulgem, Shou-Wei Ding, who were also on my various other committees and provided me with advice and support during my graduate career.

Furthermore, I would like to thank every member in Dr. Chen's lab—especially Drs. Binglian Zheng, Xigang Liu and YunJu Kim. Binglian mentored me when I first came to the lab and taught me a lot about protein expression and purification. Xigang is a very supportive post-doc and was always willing to lend a helping hand. YunJu was not only my lab-mate, but she was also my closest friend in the lab. She was the first one to congratulate me when I found something exciting and was there to hold me when I failed. YunJu—I don't know how I could have made it through graduate school without you—thank you very much from the bottom of my heart. To my fellow labmates, both past and present, thank you for making everyday so eventful; I could not have ask for better lab-mates.

Last but not least, I would like to thank my family for their unconditional support. They have been so patient with me and motivated me throughout my graduate career. I

want to especially like to thank my parents, brothers and sisters for all the sacrifices that they have had to make in order for me to attain my academic goals and dreams. In addition, I would like to thank mon puce, Mostapha Elhams, for motivating me, "picking me up" when things got tough or when I am being difficult.

Thanh T Dinh

Riverside, California

May 2012

AUTHOR CONTRIBUTIONS

The following is a document stating the author contributions for each chapter of my thesis. For Chapter One, I performed all the experiments except for the following. Dr. Xigang Liu made the AP2-pENTRY vector that I used as a donor vector for Gateway-based recombination into the GR-inducible system. Dr. Thomas Girke performed all of the bioinformatics analyses and I was present for the majority of the analyses. Dr. Levi Yant offered unpublished data from his ChIP-seq experiment regarding the *AG* 2nd intron and gave me his ChIP protocol.

Chapter Two was the start of the new field of study for our lab. Thus, I would like to give credit to all my lab-mates, as the results from this paper, as well as others, were due a wonderful collaboration amongst all of us. Specifically, for Chapter Two, Dr. Binglian Zheng generated the reporter line (*LUCL*), transformed it into plants and performed the segregation analysis for single loci transgenic plants. I bulked up the seeds, and worked out the conditions for the chemical screen. Michael O'Leary (an summer REU student who eventually stayed on as an undergraduate researcher) and I performed the chemical screen. I designed the experiments and helped him perform the secondary screen and McrBC experiments. I went on an internship to Dr. Detlef Weigel's lab and with the assistance and guidance from Dr. Pablo Manavella, I used their TopCount machine to look at the effect of several chemicals on four different luciferase-based reporter lines. Dr. Shengben Li performed the preliminary analysis (McrBC experiments) on *LUCL*. So Youn Won, a fellow graduate student, conducted the 5'aza 2-deoxycytidine experiments.

Dr. Xigang Liu isolated the *topIα* mutant and made all the crosses to RdDM mutants and tagged lines. I got the F1 seeds from him and performed all the subsequent genotyping experiments to obtain the double mutants. Undergraduate (Janet Mojica and Lorena Arroyo) and high school (Ashley Bianco) students assisted in DNA isolation. Yuanyuan Zhao, a fellow graduate student, designed the primers and I performed the bisulfite analysis at the promoter and part of the coding region of *LUC* coding region of *LUCH* and *LUCL*. For the whole genome analysis, since we did not have capacity of do these experiments in our lab, we formed a collaboration with Dr. Joseph Ecker. His post-doc, Dr. Robert Schmitz, made the library, submitted the sequencing runs and gave us the raw data. Dr. Lei Gao, a post-doc bioinformatics expert in the lab, performed all subsequent analyses with the whole genome data under the direction of me and, mainly, Dr. Xuemei Chen. I performed all the single loci analyses (small RNA Northern Blots, Real-Time RT-PCR and Southern Blots). For the whole-genome small RNA seq analysis, Dr. Dongming Li, now a post-doc in the lab, made the library. Lastly, Dr. Xigang Liu performed the ChIP with regards to the AGO4 occupancy at various loci in *topIα*.

ABSTRACT OF THE DISSERTATION

Uncovering Factors and Molecular Mechanisms in Transcriptional Regulation in *Arabidopsis thaliana*

by

Thanh Thu Dinh

Doctor of Philosophy, Graduate Program in Plant Biology

University of California, Riverside, September 2012

Dr. Xuemei Chen, Chairperson

Although every cell contains a full complement of genes, not every gene is expressed equally. It is through different types of regulation that each cell is able to differentiate into various cell types to generate a multi-cellular organism. One such type of regulation is at the transcriptional level. In my thesis work, I have gained a better understanding of how transcriptional regulation occurs in *Arabidopsis thaliana* through two divergent projects. First, I have characterized the molecular mechanism by which *APETALA2* (*AP2*), a floral homeotic gene, acts in floral development, by identifying its elusive target binding sequence, TTTGTT and/or AACAAA. In addition, I have found that AP2 modulates the expression of *AGAMOUS* (*AG*), another floral homeotic gene involved in

floral organ specification as well as stem cell maintenance, through this site *in vivo*. This finding provided a molecular link between *AP2* and *AG,* thereby answering a long-standing question in the field of floral development. Second, using a forward chemical genetics screen for players involved in transcriptional gene silencing, I have identified a chemical, Camptothecin (CPT), which released a transcriptionally silenced luciferase-based reporter. CPT is a well-known anti-cancer compound known to target *DNA Topoisomerase Iα*. Topoisomerases are used to relieve torsional stress on the DNA double helix during replication. Here in my thesis work, I have found that in *Arabidopsis, TOP1α* is also involved in transcriptional gene silencing through several different modes: body methylation and RNA-directed DNA methylation. Addition of the chemical resulted in the release of not only DNA methylation at the reporter transgene, but also endogenous loci as well. These findings highlight the diverse functions of *TOP1α* in development and provide implications in the study of cancer biology.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Appendix A

Appendix B

Appendix C

List of Tables

Chapter 1

Chapter 2

Appendix A

Appendix C

CHAPTER 1

The floral homeotic protein APETALA2 recognizes and acts through an AT-rich sequence element in *Arabidopsis*

ABSTRACT

Cell fate specification in development requires transcription factors for proper regulation of gene expression. In *Arabidopsis,* transcription factors encoded by four classes of homeotic genes, A, B, C, and E, act in a combinatorial manner to control proper floral organ identity. The A-class gene *APETALA2 (AP2)* promotes sepal and petal identities in whorls 1 and 2 and restricts the expression of the C-class gene *AGAMOUS* (*AG*) from whorls 1 and 2. However, it is unknown how *AP2* performs these functions. Unlike the other highly characterized floral homeotic proteins containing MADS domains, AP2 has two DNA binding domains referred to as the AP2 domains and its DNA recognition sequence is still unknown. Here, we show that AP2 binds a non-canonical AT-rich target sequence, and utilizing a GUS reporter system, we demonstrate that the presence of this sequence in the *AG* 2nd intron is important for the restriction of *AG* expression *in vivo.* Furthermore, we show that AP2 binds the *AG* 2nd intron and directly regulates *AG* expression through this sequence element. Computational analysis reveals that the binding site is highly conserved in the second intron of *AG* orthologs throughout *Brassicaceae*. By uncovering a biologically relevant AT-rich target sequence, this work shows that AP2 domains have wide-ranging target specificities and provides a

missing link in the mechanisms underlying flower development. It also sets the foundation for understanding the basis of the broad biological functions of *AP2* in *Arabidopsis* as well as the divergent biological functions of *AP2* orthologs in dicotyledonous plants.

INTRODUCTION

Transcription factors are proteins that directly bind DNA and modulate gene expression either in a positive or negative manner. The flower is an evolutionary innovation that contributes to the success of angiosperms. Floral organogenesis occurs post-embryonically and is initiated at specific regions called the inflorescence meristem. Dicotyledonous flowers are composed of four major types of organs: sepal, petal, stamen and carpel; each developing in a distinct, concentric whorl. A myriad of transcription factors are responsible for floral organ identity in *Arabidopsis.* Here in Chapter One, I will briefly discuss the genes that contribute to floral organogenesis, with a major emphasis on one of these genes, *APETALA2* (*AP2*) and its biochemical function in *Arabidopsis*.

ABCE Model

Transcription factors can govern many different processes in an organism. One type of transcription factor is a homeodomain protein. A homeotic gene refers to a gene that is involved in developmental patterning such that its loss results in the replacement of one organ type by another in that region. Homeotic mutations were first noted in 1894 when William Bateson observed that floral stamens appeared in the wrong place; he found examples where stamens grew where the petals were supposed to be (Bateson, 1894). Homeotic mutations were also studied in the late 1940's by Edward Lewis, who observed of bizarre rearrangements of body parts in *Drosophila melanogaster*. This led to the finding of the HOX genes and the notable mutation in *ANTENNAPEDIA*. Loss of

ANTENNAPEDIA results in legs developing in lieu of antennae in *Drosophila* (Lewis, 1998). Studies in *Drosophila* for genes involved in developmental patterning paved the way for the generation of the field of developmental biology; in plants, based on research in *Antirrhinum* and *Arabidopsis thaliana* in the early 1990's, researchers identified homeotic genes involved in floral development and formulated a model through which floral organogenesis occurs, called the "ABC Model" (Schwartz and Sommer, 1990; Bowman et al., 1991; Coen and Meyerowitz, 1991).

For a decade, it was thought that three major classes of floral homeotic genes, A, B, and C, specify the four floral organ types in a combinatorial and cadastral manner. However, this model was revised to be the ABCE model with the discovery of the E-class genes, which interact in a cooperative nature with the other homeotic genes to specify proper floral organ identity (Pelaz et al., 2000; Ditta et al., 2004).

A-class function is encoded by two genes: *APETALA1* (*AP1*) (Irish and Sussex, 1990; Mandel et al. 1992) and *APETALA2 (AP2*) (Bowman et al., 1989; Kunst et al., 1989; Jofuku et al. 1994). *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) are the B-class genes (Bowman et al., 1989), *AGAMOUS* (*AG*) is the C-class gene (Bowman et al., 1989) and E-class genes are comprised of *SEPALLATA* (*SEP*) *1-4* (Pelaz et al., 2000; Homna and Goto, 2001; Ditta et al., 2004). A-class genes confer sepal identity in the first whorl with E-class gene. Petal identity is conferred by the activities of A-, and B-, and E-class genes in the second whorl. The C-class gene, *AGAMOUS* (*AG*), together with B- and E-class genes, specifies stamen identity in the third whorl. Carpel identity is conferred by C- and E-class activities in the fourth whorl.

These interactions were first observed through genetic analysis of the corresponding mutants and later supported through molecular studies. For example, in severe mutants of *ap2,* flowers have carpelloid and stamen-like organs in the outer two whorls (Kunst et al., 1989). In *ap1* mutants, sepal identity is lost and depending on the severity of the allele, a range of phenotypic aberrations can be observed in the second whorl (Irish and Sussex, 1990; Bowman et al., 1993). Interestingly, *AP1* was not initially included in the ABC model because it was not shown to regulate C-function. However, in 1994, it was included in the ABC model as *AG* was found to negatively regulate *AP1* function from the inner two whorls (Gustafon-Brown et al., 1994). The two B-class genes *AP3* and *PI* contribute to petal and stamen identity in *Arabidopsis. In situ* hybridization experiments indicate that the expression patterns for these two genes are highly spatially and temporally specific as the highest levels of expression are restricted to the second and third whorl, corresponding to petal and sepal identity, respectively (Jack et al., 1992; Goto and Meyerowitz, 1994). Ectopic expression of *AP3/PI* leads to the homeotic transformation of sepals and carpels to petalloid and staminoid tissue (Jack et al., 1994) and its loss results in flowers with sepals in the first two whorls and carpels in the third and fourth whorls (Goto and Meyerowitz 1994). The C-class gene *AG* is restrictively expressed in the inner two whorls and mutations result in a flower with perianth-like organs in the inner two whorls (Bowman et al., 1989; Yanofsky et al., 1990; Drews et al. 1991). Constitutive ectopic expression of *AG* leads to flowers with carpelloid sepals and staminoid petals (Mizukami and Ma, 1992). A family of functionally redundant genes (*SEPALLATA1-4*) encodes E-class function. Single *sep*

mutants have very subtle phenotypes, whereas as *sep1 sep2 sep3* triple mutants result in flowers with primarily sepal-like organs and *sep1 sep2 sep3 sep4* quadruple mutants result in a flower with leaf-like organs (Pelaz et al. 2000; Ditta et al. 2004). Interestingly, over-expression of *SEP3* does not result in alterations of floral organ identity, though those plants flower early and have curled rosette leaves like *35S::AP1* (Pelaz et al. 2001). Ectopic expression of the A-, B-, C-class genes alone is not sufficient to convert leaves to floral organs (Mizukami and Ma, 1992; Krizek and Meyerowitz, 1996). However, when *35S::SEP3* is combined with constitutively expressed *AP1, AP3* and *PI*, rosette leaves are completely converted to petals (Pelaz et al., 2001). This finding shows that it is the combination of the conventional ABC genes in conjunction with E genes that confers proper floral organ identity.

In addition to promoting specific organ types, a hallmark feature of the ABCE model is its cadastral nature. In particular, A and C genes restrict one another to their domains of action within the flower. For example, loss of *AG* results in the expansion of *AP1* expression and *AP2* activity into the inner two whorls, consequently producing flowers with sepals and petals in succession in all four whorls (Bowman et al., 1991; Gustafson-Brown et al., 1994). Conversely, in an *ap2* mutant, *AG* expression expands into the outer two whorls, resulting in the conversion of perianth organs to reproductive organs (Bowman et al., 1991; Drews et al., 1991) and a suppression of this phenotype is seen in an *ap2 ag* double mutant (Bowman et al. 1991). Previous studies using a GUS reporter system have shown that the 3-kb *AG* second intron contains sequence elements required for its proper expression, including responsiveness to repression by *AP2*

(Bomblies et al., 1999; Deyholos and Sieburth, 2000). Recent studies have shown that AP2 directly modulates *AG* expression (Yant et al. 2010 and this study), however two different genome-wide experiments failed to identity the AP2 binding motif. (Yant et al. 2010).

MADS-domain and the floral quartet model

With the exception of *AP2*, all floral homeotic genes encode MADS domain-containing proteins for which DNA binding and dimerization/multimerization specificities have been extensively characterized (reviewed in (Theissen and Saedler, 2001), Ditta et al., 2004; Immink et al. 2009). MADS refers to the four "founding genes": *MCM1* (in yeast), *AG* (in *Arabidopsis*), *DEF* (in *Antirrhinum*) and *SRF* (in humans). There are 107 MADS box genes in plants and they fall into two different families: 1) Type-I class, which groups with human *SRF* and 2) Type-II class, which groups with yeast *MEF2*. ABCE MADS-box proteins belong to this group. Members in this class are characterized by having four distinct domains: MIKC. The M-domain is the minimal DNA binding domain. The I-domain is the intervening domain, the K-domain is the Keratin-like domain and the C-domain is C-terminal to the K domain. I- and K-domains have been shown to mediate interactions between other MADS proteins; however, forming higher order complexes requires the C-domain (Melzer and Theissen, 2011; Dornelas et al., 2011). MADS-box proteins have been shown to bind CArG boxes ($CC(A/T)_6GG$) as homodimers or heterodimers (Dolan, 1991; Treisman, 1992).

Although the ABCE model provides a nice framework for our understanding of floral development, it does not explain the molecular mechanism by which the floral homeotic genes work in conjunction with one another, especially in species that that may not follow the four-whorl scheme; though MADS-box homologs have been identified in those species (Saedler et al. 2001). Thus, to get at the molecular mechanism by which the MADS-box homeobox proteins function in floral development, Theissen in 2001, proposed the floral quartet model. The basic premise of this model is that organ specification is conferred by the formation and subsequent activity of complexes among the four classes of floral homeotic genes. For instance, petal identity is conferred by a higher order complex of AP1, SEP, AP3 and PI proteins (Theissen 2001). *In vitro* yeast three- and four-hybrid assays, DNA-binding assays and protoplast FRET-FLIM experiments (Honma and Goto, 2001; Immink et al. 2009; Melzer and Theissen, 2009; Melzer et al., 2009) provided support for this model. However, it wasn't until a recent study in which Smaczniak and colleagues were able to identify, *in vivo,* MADS domain complexes through immunoprecipitation followed by mass spectrometry, which *truly* demonstrated the validity of the "floral quartet model" (Smaczniak et al. 2012).

APETALA2 and its diverse roles

While the floral quartet model is able to explain the molecular function of the MADS-box proteins, it neglects the A-class gene, *AP2*. To date, there are no studies that indicate that the presence of AP2 in the floral quartet. *AP2* has always been the outlier of the ABCE model. In addition to its role in sepal and petal identity, *AP2* is also involved

in the initiation and establishment of the floral meristem with at least three other meristem genes, *AP1, LEAFY,* and *CAULIFLOWER* (Irish and Sussex, 1990; Huala and Sussex, 1992; Bowman et al., 1993; Schultz and Haughn, 1993; Shannon and Meeks-Wagner, 1993). Thus, it does not strictly function in floral organ identity.

In addition, unlike that the other floral homeotic genes whose activity is tightly coupled with its expression domain, *AP2* RNA is found throughout all four whorls as well as leaves and stems of *Arabidopsis.* AP2 protein, however, is not present in the inner two whorls of the flower, probably due to targeted translational repression by miR172 (Aukerman and Sakai, 2003 and Chen, 2004).

In addition to floral development, *AP2* also impacts stem cell maintenance (Wurshum et al., 2006), seed coat development and embryo size (Jofuku et al., 1994; Ohto et al., 2005; Ohto et al., 2009), floral transition (Yant et al., 2010), and fruit development (Ripoll et al., 2012). The diversity of its functional roles highlights its importance as a broad regulator and may explain why its DNA binding sequence may still be elusive.

ap2 mutants show a wide range of phenotypic defects depending on the allele and the temperature at which it grows. For instance, in partial loss of function *ap2* alleles, such as *ap2-1, ap2-5, ap2-8, ap2-9* and *ap2-13*, sepals are transformed into leaves and petals are transformed into staminoid organs (Bowman et al., 1989 and 1991). However grown at higher temperatures, the phenotypes of weak alleles mirror that of stronger alleles (such as *ap2-2, ap2-6, ap2-7, ap2-10, ap2-11, ap2-12* and the dominant negative allele, *I28*)—conversion of sepal tissue to carpelloid organs and the suppression of petal

development is observed (Komaki et al., 1988; Kunst et al., 1989; Bowman et al., 1991; Jofuku et al. 1994).

AP2 DNA Binding Domain

The *AP2* gene is 2.5kb in length and contained within a 3.7 kb *Bgl* fragment within the upper region of the fourth chromosome. The full-length protein is 432 amino acids in length with a molecular weight of 48 kD (Jofuku et al. 1994). Amino acids 14-50 constitute a highly acidic and serine-rich region thought to be the transcriptional activation domain and amino acids 119 to 128 make up a highly basic region containing the KKSR nuclear localization signal sequence (Jofuku et al. 1994). The central core of AP2 contains two copies of a 68 amino acid repeat (R1 and R2), referred to as the AP2 domain. Previously thought to be plant specific, it is now known that this domain is also present in diverse species such as *Tetrahymena* (Wuitschick et al., 2004) and *Plasmodium* (Yuda et al., 2009).

In *Arabidopsis,* about 1,600 genes encode transcription factors (~6% of the genome) and AP2-domain containing proteins comprise the largest group except for the MYB- and bHLH- family (Melzer and Theissen, 2011). This multi-gene family is collectively called the AP2/ERF (Ethylene Responsive Element Binding Factor) multi-gene family and AP2 is the founding member of this family of 145 genes that encode at least one AP2 DNA binding domain in *Arabidopsis* (Shigyo et al., 2006); the biological

functions of this family range from development to stress and defense responses (Jofuku et al., 1994; Weigel, 1995; Okamuro et al., 1997; Riechmann and Meyerowitz, 1998).

The AP2 domain is characterized by two hallmark features--the YRG and RAYD elements. The YRG element is comprised of 22-23 highly basic and hydrophilic amino acids at the N-terminal region (of the AP2 domain) (Okamuro et al. 1997); whereas the RAYD element is about 40 amino acids long in the C-terminal region of the AP2 domain. This element contains an 18 amino acid stretch thought to form an amphipathic α-helix and is important for proper structure and function. In addition, this element is thought to mediate protein-protein interactions or bind the major groove of DNA through the hydrophobic groove (Jofuku et al., 1994; Okamuro et al., 1997).

This large family is divided into two classes, denoted by the number of AP2 DNA binding domains it contains: *AP2*-like (generally two domains) and *ERF*-like (one domain). The *ERF*-like class is further divided into four subclasses: *ERF*-like, *DREB*-like, *RAV*-like and others; members of this class are involved in abiotic and biotic stress responses (Sakuma et al., 2002). The *ERF*-, *DREB*-, and others have one AP2 DNA binding domain, whereas members of the *RAV*-like class have one AP2 DNA binding domain and one B3 binding domain (as reviewed in (Saleh and Pagés, 2003)). The *ERF*-like class is marked by a conserved WAAIERD motif in the YRG element whereas in the *AP2*-like class is marked by a WEAR/WESH motif. Sequence alignment of the AP2 domain among extended to close family members range from less than 30% to >76% identical, respectively (Fujimoto et al., 2000). Previous studies have shown that the AP2

domain generates three β-sheets and a α-helix *in silico*. Crystal experiments (Allen et al. 1998; Krizek, 2003) and computational analysis also showed that the most conserved residues are found in a part of the α-helix region and the first β-sheet (Kim et al., 2006). It has been shown that *ERF* proteins bind the GCC-box via the β-sheets (Allen et al., 1998).

Members of the *AP2*-like class modulate developmental processes in plants. This *AP2*-like class is sub-categorized into two sub-classes or lineages: eu*AP2* and *ANT* (named after its founding member, *AINTEGUMENTA*). The eu*AP2 lineage* is comprised of major developmental regulators of reproductive and vegetative organs whereas the *ANT* lineage controls lateral organ development by controlling cell size, number and regulates ovule development, fusion of the gynoecium margin, mega-gametophyte formation, and floral growth (Elliot et al., 1996; Klucher et al., 1996; Sanders et al., 1999, Krizek, 1999). Furthermore, some petal identity loss is observed in *ant.* In an *ap2 ant* double mutant, an enhanced *ap2* second whorl phenotype is observed and there is an increase in ectopic *AG* expression, but this is not observed in *ant* single mutants (Krizek, 2000). Thus, loss of petal identity is not a result of ectopic *AG* expression so *ANT* promotes petal identity in a different pathway and not through repression of *AG*. Also, phenotypic defects are only seen in the second whorl, therefore, *ANT* does not play a role in sepal formation.

In *Arabidopsis,* there are six members in the eu*AP2* lineage, *AP2, TOE1-3, SMZ* and *SNZ.* Interestingly, *SNZ* and *SMZ* only have one AP2 domain but are also involved

in floral transition (Mathieu et al., 2009). *AP2, TOE1-3* as well as all members of the ANT lineage all have two AP2 domains. However, there are several differences between the two lineages: first, genes in the eu*AP2* lineage are targets of a microRNA, miR172 and second, there is a single amino acid insertion in the R2 domain and a ten amino acid insertion in the R1 domain of proteins in the *ANT* lineage (Kim et al. 2006). Between the two AP2 domains there is a 25 amino acid linker region. This linker region is 40% identical between members of the eu*AP2* and *ANT* lineages and the linker region has been shown to be critical for its function as mutations in the linker region abolishes the ability for the ANT to bind DNA (Klucher et al., 1996).

The *ANT* lineage could be further divided into two groups: the eu*ANT* and basal*ANT* lineages. In *Arabidopsis*, the eu*ANT* lineage has eight members (including the founding the member *ANT*) and the basal*ANT* lineage has four members. The difference between these two lineages is that the *euANT* lineage contains four highly conserved ten amino acid stretches, three in the pre-domain region and one in R1 (Kim et al. 2006). Interestingly, the ten amino acid insertion is located between the second and third β-sheets in the AP2 domain, thereby forming a longer linker region than that of genes from the eu*AP2* lineage (Kim et al., 2006). Thus, this insertion could attribute to differences in binding specificities and functions between the two lineages.

Computational studies have shed additional light on the AP2 domains of the two lineages (eu*AP2* and *ANT*). Interestingly, the ANT-R1 vs. AP2-R1 and ANTR2 vs. AP2R2 have a higher amino acid identity than AP2-R1 and AP2-R2, even with the extra

amino acid insertions present in the *ANT* lineage. In the eu*AP2* lineage, AP2-R1 and R2 share 53% amino acid identity and 69% amino acid similarity (Shigyo et al. 2006); whereas, in the *ANT* lineages, ANTR1 and ANTR2 have 39% amino acid identity and 48% similarity (Okamuro et al., 1997).

The DNA binding properties of AP2 domain proteins from various subfamilies have been studied. Members of the *ERF*-like and *DREB*-like subfamily bind well-documented GC-rich motifs (Ohme-Takagi and Shinshi, 1995; Stockinger et al., 1997; Hao et al., 1998; Liu et al., 1998). Notably, residues that bind the GCC-box are not conserved in AP2. The AP2 domain in a *RAV*-like family member binds a CAACA motif (Kagaya et al., 1999). ANT is the only protein in the *AP2*-like subfamily for which DNA binding properties have been studied (Nole-Wilson and Krizek, 2000; Krizek, 2003). Unlike members of the *ERF*-like family, ANT binds a loose and long consensus sequence. This could be explained by the ten amino acid insertion in R1. Despite the crucial role of AP2 in flower development and the diversity of its targets (Yant et al., 2010), the DNA binding property of AP2 has never been characterized.

To better understand the molecular functions of AP2, we sought to determine its binding consensus sequence *in vitro* and characterize the relevance of the sequence *in vivo*. Here, we report a novel, non-canonical AP2 binding sequence, TTTGTT or AACAAA, which binds specifically to AP2R2 *in vitro*. We show that these motifs within the 2nd intron of *AG* are important for restricting *AG* expression to the inner two whorls *in vivo*. *In silico* analysis of 2nd intron sequences from *AG* orthologs uncovers strong conservation of this element in the *Brassicaceae* family. Further, we found that

AP2 directly regulates *AG* in young flowers through these elements. These findings establish a missing link in the mechanisms underlying flower development, shed light on the molecular function of *AP2*, and set the foundation for further appreciation of the molecular basis for the broad biological functions of *AP2*.

RESULTS

AP2R2 binds a novel consensus sequence

To begin uncovering the molecular mechanisms underlying the role of *AP2* in development, we sought to identify its binding sequence by performing a SAAB assay with AP2R1, AP2R2 or AP2R1R2 doubly purified based on their N- and C-terminal tags.

When the SAAB assay was performed for the purified AP2R2 (Fig. 1A), amplified DNA from the bound fraction could be detected starting from cycle 4 (Fig. 1B). After cloning and sequencing of the AP2R2-bound DNA after cycle 6, 20 unique sequences were obtained, 17 of which contained two perfect AT-rich motifs - AACAAA or the complementary TTTGTT (Fig. 1C), whereas the other three clones had either one site (clone 11), a single nucleotide change at each site (clone 12), or a single nucleotide change at one of the sites (clone 7) (Fig. 1C). Therefore, all recovered clones contained at least one copy of the consensus sequence with no more than one nucleotide change.

When the same procedure was applied to purified AP2R1 (Fig. 2A), no bound DNA was detectable by PCR, indicating that AP2R1 did not bind DNA *in vitro* (Fig. 2B). The lack of DNA recovered from the AP2R1 SAAB assay was not due to loss of the

protein during the procedure because the AP2R1 protein was present on the beads throughout the experiment (Fig. 2C). The SAAB assay was also performed for purified AP2R1R2 (Fig. 3A). DNA bound to AP2R1R2 was detectable from cycles 3 to 6 (Fig. 3B), however, sequencing of cloned DNA bound to AP2R1R2 from cycle 6 did not reveal any obvious consensus sequence (Fig. 3C). 10 out of 25 unique sequences contained one of the sites bound by AP2R2 (Fig. 3C and data not shown), one had a site with one nucleotide change (clone 3), and some of the other sequences were GC-rich.

Next, to confirm the SAAB assay results, we performed EMSAs with the AP2R2 protein. Sense and antisense strands corresponding to one of the sequences obtained from the SAAB assay containing both AACAAA (termed "α" site) and TTTGTT (termed "β" site) were used as a probe (Fig. 4A). DNA binding as revealed by a shift in mobility was observed with as little as 50ng of AP2R2 (data not shown), and with 100 or 200ng of AP2R2, which yielded more strong and consistent binding (Fig. 4B, lanes 1 and 2, arrow). To confirm the observed binding, 20-fold cold competitor was added to the binding reaction. Indeed, the intensity of the shift was greatly diminished (Fig. 4B, lane 3). Since the experiments were conducted with AP2R2 purified from *E. coli*, there was a possibility that the observed shift may be due to a contaminating protein instead of AP2R2 (although AP2R2 was the only protein detected by Coomassie staining in the protein fraction; Fig. 1A). If the observed binding was specifically caused by AP2R2, the addition of His and T7 antibodies would generate a supershift as AP2R2 had both a T7 and a His tag. We observed that the inclusion of the His and T7 antibodies resulted in super-shifted bands as compared to AP2R2 alone (Fig. 4B, compare lanes 4 and 5 to lane

2, stars), confirming that it was indeed AP2R2 itself that bound the probe. As a control, we also added 400ng His and T7 antibodies to the reaction in the absence of AP2R2 and we did not observe any of the same band shifts as seen in the presence of AP2R2 (Fig. 4B, lane 6).

Next, we tested whether both sites were necessary for AP2R2 binding. We mutated the AACAAA site to AGGTGA and the TTTGTT site to TCCACT (Fig. 4A). The EMSA showed that AP2R2 was still able to bind probes that had one intact site (Fig. 4C, lanes 2 and 5, arrow). The shift was lost upon addition of the corresponding cold competitor (Fig. 4C, lanes 3 and 6). Loss of both sites, however, completely abolished binding (Fig. 4C, lane 8). These results demonstrate that AP2R2 binds AACAAA and/or TTTGTT *in vitro*.

ANT, the only protein characterized in terms of DNA binding properties in the *AP2*-like subfamily of AP2 domain-containing proteins, has been shown to bind a loose and long consensus sequence, gCAC(A/G)N(A/T)TcCC(a/g)ANG(c/t) (Nole-Wilson and Krizek, 2000). We wondered how tight the AP2R2 consensus binding sequence might be. Since AP2R2 was able to bind probes containing one site, probes containing just one site were used for testing the nucleotide specificity at each position. The nucleotide at each position was converted into one of the other three nucleotides while the other five positions remained unchanged. All possible perturbations were made and EMSAs were performed with AP2R2. Our results indicated that the binding site was extremely tight as any mutation at any position greatly compromised binding by AP2R2 (Fig. 5, 6). Notably, any change at the fourth position resulted in loss of binding (Fig. 5, 6).

Proteins in the *ERF*-like and *DREB*-like subfamilies containing a single AP2 DNA-binding domain have been shown to bind specific GC-rich consensus sequences (Ohme-Takagi and Shinshi, 1995; Stockinger et al., 1997; Hao et al., 1998; Liu et al., 1998). Thus, we wanted to test whether AP2R2 could bind the conventional GCC-box (Hao et al., 1998), GCCGCC, or other random sequences. We performed EMSAs with a GCC-box, two random probes (Fig. 7A), and the Δα probe as a positive control. We observed binding of AP2R2 to the Δα probe (Fig. 7B, lane 2) but not to the canonical GCC-box or two other random sequences (Fig. 7B, lanes 5, 7 and 9).

AP2 full-length protein lacks obvious DNA binding specificity *in vitro*

Next, to determine whether the full-length AP2 protein could bind the consensus sequence, TTTGTT, we cloned the AP2 full-length protein into a vector containing an N-terminal His and MBP tag (MBP-AP2), and purified and desalted it (Fig. 8A). Due to the large size of MBP-AP2, 400ng was used to perform the EMSA. Using the same probes as in the AP2R2 EMSAs (Fig. 4A), we found that MBP-AP2 was able to bind probes containing one or both consensus sequences (Fig. 9A, lanes 2 and 6, arrow). Interestingly, MBP-AP2 was also able to bind the probe with both sites mutated (Fig. 9A, lane 9, arrow). To further assess the binding specificity of the AP2 full-length protein, we performed EMSAs with a probe containing the canonical GCC-box and two random probes in Figure 7A. Interestingly, MBP-AP2 could bind all of the probes *in vitro* (Fig. 9B, lanes 2, 5, and 8, arrow). To address whether the MBP tag may be binding the probes, an equal amount of MBP was added to the reactions for the EMSA (Fig. 8B).

MBP alone did not bind any DNA sequences (Fig. 9A, lanes 3, 7 and 10; 9B, lanes 3, 6 and 9).

AP2R2 binds *AG* 2nd intron *in vitro*

Considering that AP2R2, but not AP2 full-length protein, specifically bound the consensus sequence, we sought to test whether the AP2R2 binding site had biological relevance. Previous studies have shown that *AP2* represses *AG* expression (Drews et al., 1991), but it is still unknown whether *AG* is a direct target of AP2, although AP2 has been shown to bind the AG 2nd intron *in vivo* in our previous ChIP-seq analysis (Yant et al., 2010). Characterization of the *AG* 2nd intron has identified a 750-bp region that, when fused to the GUS reporter in a construct termed KB31, confers *AP2* responsiveness to GUS (Bomblies et al., 1999; Deyholos and Sieburth, 2000). *In silico* analysis indicated that the KB31 region contained two AP2R2 binding sites at the 3' end, which we termed A and B (Fig. 10A). Thus, we proceeded to test whether AP2R2 was able to bind this region of the *AG* second intron *in vitro*. Using primers encompassing this region (Table 1), we generated a 167-bp probe (Fig. 10A) to perform an EMSA. Indeed, a shift was found with AP2R2 and the binding was stronger as we added increasing amounts of the protein (Fig. 10B, lanes 2-7, arrow). Furthermore the binding was lost upon the addition of 20x or 40x cold competitor (Fig. 10B, lanes 8 and 9, respectively) to the binding reaction. To test whether the observed binding required the two sites within the 167-bp sequence, we performed site-directed mutagenesis to mutate both sites (Fig. 10A). With the ΔAB probe, binding was diminished greatly (Fig. 10B), showing that

AP2R2 binds the 167-bp region via the two elements *in vitro.* In addition, and consistent with prior results, gel shifts showed binding of the MBP-AP2 full-length protein to both the wild-type and the ΔAB probe (Fig. 10C, lanes 3-8, 12 and 13, arrow). This binding was abolished upon addition of 40x cold competitor (Fig. 10C, lanes 9 and 14) and MBP alone did not bind either probe (Fig. 10C, lanes 2 and 11).

The AP2R2 binding sites are important for the restriction of *AG* expression *in vivo*

To evaluate the importance of the AP2R2 binding sites in the *AG* 2nd intron *in vivo,* we utilized the KB31 GUS reporter, which had been shown to faithfully report the endogenous domains of *AG* expression and to respond to the regulation by *AP2* (Bomblies et al., 1999; Deyholos and Sieburth, 2000). We cloned the 750-bp KB31 region from the *AG* 2nd intron containing either the wild-type or mutant (ΔAB) sites into a GUS expression vector with a minimal (-60) 35S promoter (Tilly et al., 1998). The constructs were introduced into *rdr6-11* to prevent post-transcriptional gene silencing of the transgenes (Dalmay et al., 2000; Mourrain et al., 2000). For the wild-type construct, GUS staining of inflorescences from 99 independent T1 transgenic plants showed that 74 recapitulated the proper *AG* expression patterns (Fig. 11A, C and E). 22 inflorescences did not show any GUS staining and three did not recapitulate the proper *AG* expression patterns. For the ΔAB construct, however, 71% of the 27 independent transformants showed expansion of the GUS expression domain to the outer two whorls (Fig. 11B, D and F) in all stages of flower development, with the remainder showing the correct

expression patterns. Therefore, the A and B sites are important for the restriction of *AG* expression to the inner two floral whorls.

AP2 directly regulates *AG* in young flowers through the binding sites

The fact that the AP2R2 binding sites in the 2nd intron of *AG* are important for the restriction of *AG* expression to the inner two whorls implies that *AP2* is a direct regulator of *AG*. To address whether AP2 acts on *AG* directly, we used a rat glucocorticoid receptor (GR)-induction system that has been widely utilized in *Arabidopsis* as a method to establish direct relationships between a transcription factor and its targets (Sablowski and Meyerowitz, 1998; Wagner et al., 1999; Ito et al., 2004; William et al., 2004). Because *AP2* is targeted by miR172 and transgenes containing miRNA target sites are readily silenced *in vivo*, we fused a miR172-resistant *AP2* cDNA (*AP2m3*) (Chen, 2004) to GR. The *35S::AP2m3-GR* construct was transformed into the progeny of *ap2-2*/+ plants. After obtaining single-locus insertion transformants of the *ap2-2*/+ genotype, single and continuous treatments of 10μM DEX were performed to determine the functionality of the transgene. A single DEX treatment of inflorescences was not sufficient to induce the *AP2m3* phenotype (data not shown; (Chen, 2004)). Continuous treatments (once a day for one week), however, led to the induction of the *AP2m3* phenotype thus showing that the transgene was functional (Fig. 12A and B).

To determine whether AP2 directly represses *AG* expression, we subjected *35S::AP2m3-GR ap2-2* inflorescences to a single treatment of cyclohexamide (CHX) with or without DEX. After six hours, inflorescence tissue was micro-dissected to remove

stage 8 and older flowers. RT-PCR was performed to measure *AG* mRNA levels. We found that upon induction of AP2m3-GR, *AG* mRNA levels decreased in young flowers (Fig. 12C). Real time RT-PCR of three biological replicates revealed a 50% decrease in *AG* transcript levels upon AP2m3-GR induction (Fig. 12D). Therefore, *AG* is likely a direct target of *AP2*.

Next, we sought to determine whether *AP2* regulates *AG* through the two AP2R2 binding sites. If AP2 acts through the two sites, we would expect *KB31*, but not *KB31ΔAB*, to be repressed by AP2. *KB31* and *KB31ΔAB* transgenic lines harboring a single transgene locus were identified and crossed into the *35S::AP2m3-GR ap2-2* background. Homozygous *KB31 35S::AP2m3-GR ap2-2* or *KB31ΔAB 35S::AP2m3-GR ap2-2* inflorescences were treated with DMSO or DEX for six hours and GUS expression was determined by real time RT-PCR. DEX induction caused a decrease in GUS mRNA levels in *KB31* but not in *KB31ΔAB* (Fig. 12E and F). Therefore, *AP2* represses *AG* through the two AP2R2 binding sites *in vivo*.

AP2 binds *AG* 2nd intron *in vivo*

To test whether AP2 is associated with the *AG* 2nd intron *in vivo*, we performed ChIP assays using anti-AP2 antibodies (Mlotshwa et al., 2006). The antibodies were directed against a C-terminal portion of AP2 that is predicted to be absent in the *ap2-2* mutant. From a ChIP-seq experiment conducted with these antibodies on whole inflorescences, genome-wide AP2 binding sites were uncovered (Yant et al., 2010). The ChIP-seq effort identified a region in the 5' end of the KB31 fragment that was bound by

AP2, which we named region II (Fig. 13A). This region did not overlap with the region containing our binding sites, AB (Fig. 13A). To specifically test whether AP2 binds the AB region, especially in young flowers, we performed ChIP experiments with dissected inflorescences containing stages 7 and younger flowers and used region II as the positive control. We were able to find enrichment of AP2 within the AB region as well as region II in two biological replicates (Fig. 13B and data not shown). Since the our ChIP assays showed that AP2 bound both sites but the genome-wide study did not show enrichment in our AB site (Yant et al. 2010), we further dissected the II-region to test which part of that region may be bound by AP2. We divided the region into four parts (1→4, Fig. 14A and Table 1), including a region directly upstream and downstream. We found that AP2R2 bound part 3 of region II (Fig. 14B) and upon further inspection, the 5' end of part 3 contained an AP2R2 consensus sequence with one nucleotide change (TTTGTT→TTTGTG). Consistent with our site mutagenesis analysis (Fig. 5, 6), AP2R2 is still able to bind the TTTGTG sequence in a medium manner. Binding was not visibly observed with this probe (p3) until 400ng of protein was added (Fig. 14, lane 5) as compared to 200ng of the wild-type probe (Fig. 14, lane 2). Consistent with all of our previous results, MBP-AP2 was able to bind all four probes (Fig. 14C).

The consensus sequence TTTGTT is highly conserved in *Brassicaceae*

The repression of *AG* expression by *AP2* is specific to *Arabidopsis* as the *AP2* orthologs in *Antirhinnum, LIPLESS 1* and *2* (*LIP1* and *2*), do not negatively regulate C-class function (Keck et al., 2003). However, it is possible that AP2-mediated restriction

of *AG* expression is conserved in species closer to *Arabidopsis*. Thus, we sought to determine whether the AP2R2 binding consensus sequence was conserved in *Brassicaceae.* Using *AG* 2nd intron sequences from 29 *Brassicaceae* species (Hong et al., 2003), we performed multiple sequence alignment as well as a sliding window conservation analysis to identify regions in the 2nd intron that are conserved both in sequence and in position. Although the TTTGTT (or AACAA) motif was present multiple times in each of the introns (Fig. 15), the A-site (Fig. 10A) was embedded within a short region that was conserved throughout *Brassicaceae* both in sequence and in position within the introns as revealed by the sliding window analysis (Fig. 15, 16). The B-site was not found at invariant positions among the introns (Fig. 15). At the A-site, 28 of the 29 *Brassicaceae* species showed a perfect match to the TTTGTT pattern. The apparent exception appeared to be *Thysanocarpus* (AY253255) with a single nucleotide change in the motif. When the analyses included the 2nd intron of *AG* homologs from *Antirrhinum majus* (AY935269), *Lycopersicon esculentum* (AY254705), or *Cucumis sativus* (AY254702 and AY254704) belonging to *Veronicaceae*, *Solanaceae*, or *Cucurbitaceae*, respectively, the divergence in these sequences was too high to compute reliable multiple sequence alignments of the introns, thus precluding any conclusions on the conservation of this motif outside of *Brassicaceae*.

DISCUSSION

DNA binding specificities of AP2 domain proteins

Genes encoding one or more AP2 DNA binding domains are categorized under five subfamilies: *DREB*-like, *ERF*-like, *RAV*-like, *AP2*-like, and others (Sakuma et al., 2002). The *AP2*-like subfamily, which can be further divided into two lineages, *ANT* and *euAP2*, is the only subfamily that contains two AP2 domains. Single AP2 domain containing proteins of the other subfamilies bind to highly specific, mostly GC-rich sequence motifs (Ohme-Takagi and Shinshi, 1995; Stockinger et al., 1997; Hao et al., 1998; Liu et al., 1998). Only the target sequence of a single member of the *AP2*-like subfamily, ANT, has been reported. ANT binds a long and loose consensus sequence that is also GC-rich (Nole-Wilson and Krizek, 2000). In contrast to conventional GC-rich target sequences of these characterized AP2-domain proteins, AP2R2 is highly specific for the AT-rich consensus sequences TTTGTT or AACAAA. The AP2 domain in RAV1 also binds a non-GC rich sequence CAACA (Kagaya et al., 1999). Two AP2 domain proteins from *Plasmodium* were found to bind the consensus sequences TGCATGCA and GTGCAC, which are different from the target sequences of all plant AP2 domain proteins characterized to date (De Silva et al., 2008). Collectively, these studies show that AP2 domains have wide ranging binding specificities. Consistent with this, the AP2 domain of AtERF1 and the AP2 domains of ANT appear to use largely non-conserved amino acids for DNA binding (Allen et al., 1998; Krizek, 2003) (Fig. 17). The

fact that AP2R1 does not appreciably bind any DNA sequences *in vitro* raises the possibility that some AP2 domains function in processes other than DNA binding.

It is useful to compare and contrast the DNA binding specificities of ANT and AP2 as representatives of the two lineages within the *AP2*-like subfamily. *In vitro* selection of DNA sequences bound by ANT-AP2R1R2 led to the identification of a long consensus sequence (Nole-Wilson and Krizek, 2000; Krizek, 2003). In our study, we found that AP2R1R2 bound DNA *in vitro,* but no consensus sequence could be identified. We note that both ANT-AP2R1R2 and AP2R1R2 have poor DNA binding specificities (as exemplified by the loose ANT consensus sequence and the lack of obvious consensus motifs from AP2R1R2-bound sequences). We also note that ANT-AP2R1R2 and AP2R1R2 have differences in their binding preferences. While ANT-AP2R1R2 binds GC-rich sequences, AP2R1R2 probably prefers the TTTGTT or AACAAA motif since this motif was in 10 out of 25 clones from the SAAB assay. Moreover, not all clones from the SAAB assay were GC-rich. Consistently as well, this could be explained by the addition of 10 amino acids in ANTR1 vs. AP2R1 and a single amino acid addition in ANTR2 vs. AP2R2 (Fig. 17).

AP2 DNA binding specificities *in vivo*

AP2 full-length protein was able to bind all probes that were tested *in vitro* (Fig. 9, 10, 14). This led us to question the specificity of AP2 DNA binding *in vivo*, especially in relation to its biological functions. *AP2* has diverse biological functions such as seed development (Jofuku et al., 2005; Ohto et al., 2009), shoot apical meristem maintenance

(Würschum et al., 2006), control of floral timing (Yant et al., 2010), preventing replum overgrowth during fruit development (Ripoll et al., 2011), establishment of floral meristem identity (Schultz and Haughn, 1993; Shannon and Meeks-Wagner, 1993), floral organ specification (Bowman et al., 1989; Kunst et al., 1989), and the regulation of homeotic gene expression (Drews et al., 1991). Perhaps the lack of strong, inherent DNA binding specificities underlies *AP2*'s diverse biological roles. Whole genome ChIP-seq experiments identified more than 2000 sites that are bound by AP2 *in vivo* (Yant et al., 2010), highlighting the potential of *AP2* in influencing the expression of a large number of genes. However, computational analyses failed to uncover a consensus sequence that is enriched in regions bound by AP2 *in vivo* (data not shown). Also, due to its AT-rich nature, we were not able to state that the AP2R2 binding site was statistically significant among the targets. However, it is interesting to note that in the sum sequence space of the 2275 bound sites there are 445 instances of AACAAA and 473 of TTTGTT (Yant et al., 2010). The lack of ability to find an AP2 consensus sequence could be reflective of its diverse roles in development. In addition, the discrepancy in the two sites (II and AB) that we found in the two ChIP experiments ((Yant et al., 2010) and this study) could be due to tissue differences.

Despite the large number of *in vivo* binding sites, AP2 is still selective in its DNA binding *in vivo*, in contrast to its largely non-specific DNA binding *in vitro*. One potential mechanism underlying the *in vivo* specificity is that it might be conferred by other DNA binding proteins that interact with AP2. In this scenario, the largely non-specific DNA binding by AP2 enhances the binding of other transcription factors at

specific sites. The promiscuous binding of MBP-AP2 to all DNA sequences *in vitro* lends itself to this hypothesis as it could be feasible that full-length AP2 itself, as a regulator of diverse functions, would have specific binding abilities depending on its protein binding partners that may modulate its activity *in vivo.* Another potential mechanism is that other factors interact with AP2R1 to allow AP2R2 to specifically interact with DNA. We prefer this mechanism since the AP2R2 binding sites in the *AG* 2nd intron are indeed important for the function of *AP2 in vivo*. Moreover, the AP2R2 consensus sequence was recovered in 10 out of 25 clones in the AP2-R1R2 SAAB assay, suggesting that there is some inherent affinity of the AP2 domains for the TTTGTT or AACAAA consensus sequence. Both scenarios may occur *in vivo*, in which case the AP2R2 consensus sequence would only be present at some of the *in vivo* AP2 binding sites.

AP2 directly regulates *AG*

AP2 has long been known to be essential in establishing the inner two whorl-specific pattern of *AG* expression (Drews et al., 1991). In *ap2* loss of function mutants, *AG* expression expands into the outer two whorls. Using the GUS reporter system, elements responsive to *AP2* regulation have been mapped to at least two regions in the *AG* 2nd intron (Bomblies et al., 1999; Deyholos and Sieburth, 2000). However, it was not known whether *AP2* regulates *AG* expression directly. We found that a 750-bp *AP2*-responsive region contains two AP2R2 consensus sequences. Site-directed mutagenesis experiments indicated that the two sites were important for the negative regulation of *AG*

by AP2. In addition, we found that this negative regulation was direct through an inducible system (AP2m3-GR) as well as ChIP experiments. Therefore, *AP2* is a direct, negative regulator of *AG*.

Our data also suggest that *AP2* represses *AG* most effectively during early stages of flower development. Initially, when *AP2m3-GR* whole inflorescences (composed of both young and old flowers) were used in the induction experiments, no obvious changes in *AG* mRNA levels were seen. However, upon micro-dissection of the inflorescences after induction to retain only flowers of stages 7 and younger, we observed a 50% decrease in *AG* mRNA levels upon AP2 induction. It is feasible that AP2 only negatively regulates *AG* during early stages of flower development as it has been shown that a myriad of other genes, such as *CURLY LEAF* (*CLF*), *LEUNIG*, *SEUSS*, and *RABBIT EARS* also negatively regulate *AG* (Goodrich et al., 1997; Liu et al., 1998; Franks et al., 2002; Krizek et al., 2006). It is possible that in the outer two whorls, *AP2* establishes the initial repression of *AG*, and other mechanisms such as *CLF*-mediated histone modifications are responsible for the maintenance of the repressed state throughout flower development.

In addition, it has been shown that AP2 acts through different regions of the *AG* 2nd intron at different developmental time points (Bomblies et al., 1999; Deyholos and Sieburth, 2000). Thus, it is feasible that region II and the AB site may both be important at different time points. In the Bomblies study, KB31 is sufficient to confer *AG* expression and contain elements by which AP2 negatively regulates *AG,* and this construct has both region II and the AB site. Constructs that do not have region II or the

AB site did not provide much information as although they did not respond to the loss of *AP2,* there was no basal *AG* expression anyhow (KB28—encompasses p3 and the A site only). In this study, we show that the AB site is functional as loss of AB resulted in expansion of GUS expression (Fig. 11). Interestingly, AP2R2 was able to bind to both region II and the AB site albeit at a higher affinity with the AB site.

The "A"-Site is highly conserved in *Brassicaceae*

The eu*AP2* lineage predates the divergence of gymnosperms and angiosperms, but the biological functions of *AP2* and its orthologs differ amongst flowering plants (reviewed in (Litt, 2007)). In *Arabidopsis*, *AP2* specifies perianth identities and restricts C-function to the inner two whorls. However, characterized *AP2* orthologs from *Antirrhinum* and petunia do not appear to share *AP2*'s role in flower development (Maes et al., 2001; Keck et al., 2003). For example, *LIP1* and *LIP2, AP2* orthologs in *Antirrhinum,* promote sepal identities but do not control petal identity or restrict the expression of *PLENA* (C-class gene) (Keck et al., 2003). In fact, mutations with ectopic C function in the outer whorls in *Antirrhinum* and petunia map to a microRNA, miR169 (Cartolano et al., 2007). Interestingly, the petunia ortholog of *LIP*/*AP2*, *PhAP2A*, was able to rescue the *ap2-1* mutant when expressed in *Arabidopsis* (Maes et al., 2001). The ability of the petunia AP2 protein to regulate *AG* in the *Arabidopsis* context suggests that the DNA binding properties of the petunia AP2 are similar to those of *Arabidopsis* AP2 and implies that divergence in C-class regulatory sequences or in AP2-interacting proteins may be responsible for the divergence in AP2's ability to regulate C-class genes.

In this study, we show that AP2R2 recognizes an AT-rich motif *in vitro* and that two such motifs within the *AG* 2nd intron mediate the regulation of *AG* by *AP2 in vivo*. Given the AT-richness of introns, this sequence motif is present multiple times in the introns of *AG* and *AG* orthologs from other species. The positions of the motifs relative to other transcription factor binding sites may influence the ability of AP2 to act upon them. We show that the A site recognized by AP2R2 in the *AG* 2nd intron is conserved both in sequence and in position in *Brassicaceae*. This implies that AP2-mediated regulation of C-class gene expression is conserved in this family. Although the motif is present in the 2nd introns of *AG* orthologs from non-*Brassicaceae* species, the overall large divergence in 2nd intron sequence precluded confident alignments to determine whether the positions of the motifs are conserved.

MATERIALS AND METHODS

Plasmid Construction

To express the AP2R1, AP2R2 and AP2R1R2 domains of AP2 in *E. coli*, the corresponding coding regions from the *AP2* cDNA were amplified by PCR (Table 1) and cloned into the pET21-A vector using *BamH*I and *EcoR*I sites (Novagen). To express the full-length AP2 protein, the entire coding region of *AP2* was cloned in-frame to an N-terminal MBP and His tag using *BamH*I and *EcoR*I sites in the pmCSG7 XF0510 MBP-LIC vector (a gift from Dr. Xiaofeng Cao, Institute of Genetics and Developmental Biology, Beijing, China).

For *in vivo* analysis of the *AG* 2nd intron, the portion of the *AG* 2nd intron in the KB31 construct (Bomblies et al., 1999) was amplified and cloned into PCR2.1 (Invitrogen). Site-directed mutagenesis was performed (Table S1) to introduce mutations into each of the two AP2 binding sites. The wild-type and mutant KB31 fragments were then cloned into pD991 (Tilly et al., 1998) using *BamH*I and *Hind*III sites. The *35S::AP2m3-GR* construct was generated as described (Yant et al., 2010).

Protein Expression and Purification

The pET21A-AP2R1, AP2R2, and AP2R1R2 and the MBP-AP2 full-length protein plasmids were transformed into *E. coli* BL21. Protein expression and purification were done as previously described (Smith et al., 2002; Husbands et al., 2007) and purified proteins were quantified against BSA.

Selection Affinity and Amplification Binding (SAAB) Assay

200ng or 500ng of doubly affinity-purified (with Ni^{2+} beads and T7 antibody) and desalted AP2R2, AP2R1, or AP2R1R2 was subjected to a SAAB assay as previously described (Smith et al., 2002). Briefly, the protein-bead mixture was divided into six tubes. In the first tube, a pool of random, double-stranded oligonucleotides (Table 1) was added and incubated for four hours with the protein-bead mixture. The DNA bound by the protein-bead complex was eluted and PCR was performed to amplify the bound sequences. An aliquot of the PCR reaction was added to the second tube of the protein-bead mixture to allow protein-DNA binding to occur. This process of affinity binding

and amplification was reiterated a total of six times. The PCR product from either cycle 5 and/or 6 was cloned via TA cloning and sequenced. The sequences were analyzed with the motif finding program MEME to identify consensus motifs.

Electrophoretic Mobility Shift Assays (EMSAs)

EMSAs were performed as described (Husbands et al., 2007) with some modifications. For EMSAs shown in Figures 4, 5, 6, 7 and 9, probes were generated by annealing 100pmol of sense and anti-sense oligonucleotides (Table S1) and 1-2pmol of probe was used in each reaction. For gel-shifts shown in Figures 10B, C, and 14, the DNA fragment was amplified by PCR from wild-type or mutant versions of KB31, or from Col genomic DNA, and 0.1-0.2pmol of labeled probe was used in each reaction. Probes were prepared as previously described (Broitman-Maduro et al., 2005).

Gel shift reactions were conducted at 4°C in 20% Glycerol, 20mM Tris (pH 8.0), 10mM KCl, 1mM DTT, 12.5ng poly dI/C, 6.25pmol of random, single-stranded oligonucleotides, Herring sperm DNA, BSA and the probe in the amount specified above.

All samples involving AP2R2 were loaded on an 8% gel, whereas as those involving the AP2 full-length protein were loaded on a 6% gel to resolve protein-DNA complexes. Gels were then dried and either exposed to X-ray films or imaged and quantified using the Typhoon PhosphorImager. For quantification analyses in Figures 5 and 6, the percentage bound was calculated by dividing the shifted amount after incubation with the mutated probes by the shifted amount after incubation with the Δα-probe.

In reactions with cold competitors, 5-40x unlabeled probes were included in the reactions. Further, anti-His and anti-T7 antibodies were added to some reactions at 1-2x the amount of the AP2R2 protein to obtain super-shifts.

GUS Staining and Microscopy

Inflorescences were stained for GUS activity and processed for sectioning as previously described (Sieburth and Meyerowitz, 1997). Slides were viewed under a Leica DMR compound microscope and images were taken with a Spot digital camera (Diagnostic Instruments).

Induction and Expression Analysis of *35S::AP2m3-GR*

The inflorescences of *35S::AP2m3-GR ap2-2* plants were treated once with a solution of 10μM Dexamethasone (DEX)/0.015% Silwet with or without 10μM cyclohexamide (CHX) (Fisher). Six hours later, the treated inflorescences were dissected to remove stage 8 and older flowers. Total RNA was isolated from the dissected inflorescences and subjected to DNaseI treatment and reverse transcription. RT-PCR was performed on the cDNAs using primers specific for *AG* and *UBQ5* (Table 1). Real time RT-PCR was performed on the same cDNAs using a Biorad real time PCR SYBR Green system (Biorad). Three technical replicates were performed for each real time RT-PCR. Three biological replicates of DEX induction and real time RT-PCR were performed. Error bars represent the standard deviation from three technical replicates.

Chromatin Immunoprecipitation (ChIP)

ChIP experiments were performed on two biological replicates following previously described protocols (Gomez-Mena et al., 2005; Mathieu et al., 2009; Yant et al., 2010). The input and ChIP samples were subjected to real time PCR (Biorad). Three technical replicates were performed. The data were analyzed as previously described (Wierzbicki et al., 2008a).

Sequence Analysis

All 2nd intron sequences from *AG* orthologs were downloaded from GenBank (Table 2). The start and end positions, provided by (Hong et al., 2003) and those specified in the GenBank annotations files, were used to parse the introns from their source sequences and to bring them into their proper sense orientation (Table 2). Sequence manipulations and analyses were performed with custom scripts that are based on the Biostrings package of the statistical programming environment R (Morgan et al., 2009; Team, 2010). Multiple sequence alignments (MSAs) were computed with the diaglign2-2 software from Morgenstern (Morgenstern, 2004) using the default parameters in the DNA mode. A sliding window analysis was performed to visualize the degree of conservation in the final MSA. For this, the relative conservation of each base was calculated at each position where a value of 1.0 indicates perfect conservation of one base (disregarding gaps) and a value of 0 indicates equal representation of all four bases. For plotting purposes, these conservation values were smoothed by calculating their mean for a sliding window size of 20 nucleotides along all MSA positions. Pattern searches were

performed with the matchPattern function of the Biostrings package (Morgan et al., 2009).

REFERENCES

Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M. and Suzuki, M. (1998). A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *Embo J* 17, 5484-5496.

Aukerman, M.J. and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *The Plant Cell* 15, 2730-2741.

Bateson, W. (1894). Materials for the study of variation, treated with especial regards to discontinuity in the origin of species. *London: Macmillan xv*, 598.

Bomblies, K., Dagenais, N. and Weigel, D. (1999). Redundant enhancers mediate transcriptional repression of *AGAMOUS* by *APETALA2*. *Dev Biol* 216, 260-264.

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

Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* 112, 1-20.

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M. and Smyth, D.R. (1993). Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* 119:721-743.

Broitman-Maduro, G., Maduro, M. F. and Rothman, J. H. (2005). The noncanonical binding site of the MED-1 GATA factor defines differentially regulated target genes in the C. elegans mesendoderm. *Dev Cell* 8, 427-433.

Cartolano, M., Castillo, R., Efremova, N., Kuckenberg, M., Zethof, J., Gerats, T., Schwarz-Sommer, Z. and Vandenbussche, M. (2007). A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity. *Nat Genet* 39, 901-905.

Coen, E. and Meyerowitz, E. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31-37.

Chen, X. (2004). A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* 303, 2022-2025.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S. and Baulcombe, D. C. (2000). An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* 101, 543-553.

De Silva, E. K., Gehrke, A. R., Olszewski, K., Leon, I., Chahal, J. S., Bulyk, M. L. and Llinas, M. (2008). Specific DNA-binding by apicomplexan AP2 transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* 105, 8393-8398.

Deyholos, M. K. and Sieburth, L. E. (2000). Separable whorl-specific expression and negative regulation by enhancer elements within the *AGAMOUS* second intron. *Plant Cell* 12, 1799-1810.

Ditta, G., Pinyopich, A., Pelaz, S. and Yanofsky, M. (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr Biol* 14, 1935-1940.

Dolan, J.W. and Fields, S. (1991). Cell-type-specific transcription in yeast. *Biochim Biophys Acta* 1088, 155-169.
Dornelas, M.C., Patreze, C.M., Angenent, G.C. and Immink, R.G. (2011). MADS: the missing link between identity and growth? *Trends Plant Sci* 16, 89-97.
Drews, G. N., Bowman, J. L. and Meyerowitz, E. M. (1991). Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* 65, 991-1002.
Elliot, R.C., Betzner, A.S., Huttner, E., Oakes, M., Tucker, W.Q. Gerentes, D., Perez, P. and Smyth, D.R. (1996). *AINTEGUMENTA,* an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *The Plant Cell* 8, 155-168.
Franks, R. G., Wang, C., Levin, J. Z. and Liu, Z. (2002). SEUSS, a member of a novel family of plant regulatory proteins, represses floral homeotic gene expression with LEUNIG. *Development* 129, 253-263.
Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H. and Ohme-Takagi, M. (2000). Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC-mediated gene expression. *The Plant Cell* 12, 393-405.
Gomez-Mena, C., de Folter, S., Costa, M. M., Angenent, G. C. and Sablowski, R. (2005). Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. *Development* 132, 429-438.
Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E. M. and Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* 386, 44-51.
Goto, K., and Meyerowitz, E.M. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes Dev* 8, 1548-1560.
Gustafson-Brown, C., Savidge, B. and Yanofsky, M. (1994). Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Cell* 76, 131-143.
Hao, D., Ohme-Takagi, M. and Sarai, A. (1998). Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. *J Biol Chem* 273, 26857-26861.
Hong, R. L., Hamaguchi, L., Busch, M. A. and Weigel, D. (2003). Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell* 15, 1296-1309.
Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525-529.
Huala, E. and Sussex, I.M. (1992). LEAFY interacts with floral homeotic genes to regulate *Arabidopsis* floral development. *The Plant Cell* 4, 901-913.
Husbands, A., Bell, E. M., Shuai, B., Smith, H. M. and Springer, P. S. (2007). LATERAL ORGAN BOUNDARIES defines a new family of DNA-binding transcription factors and can interact with specific bHLH proteins. *Nucleic Acids Res* 35, 6663-6671.
Immink, R. G., Kaufmann, K. and Angenent, G. C. (2010). The 'ABC' of MADS domain protein behaviour and interactions. *Semin Cell Dev Biol* 21, 87-93.
Irish, V.F. and Sussex, I.M. (1990). Function of the *apetala-1* gene during *Arabidopsis* floral development. *The Plant Cell* 2, 741-753.

Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J. L. and Meyerowitz, E. M. (2004). The homeotic protein *AGAMOUS* controls microsporogenesis by regulation of *SPOROCYTELESS*. *Nature* 430, 356-360.
Jack, T., Brockman, L.L., and Meyerowitz, E.M. (1992). The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* 68, 683-697.
Jack, T., Fox, G.L. and Meyerowitz, E.M. (1994). *Arabidopsis* homeotic gene *APETALA3* ectopic expression: transcriptional and post-transcriptional regulation determine floral organ identity. *Cell* 76, 703-716.
Jofuku, K. D., den Boer, B. G., Van Montagu, M. and Okamuro, J. K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell* 6, 1211-1225.
Jofuku, K. D., Omidyar, P. K., Gee, Z. and Okamuro, J. K. (2005). Control of seed mass and seed yield by the floral homeotic gene *APETALA2*. *Proc Natl Acad Sci U S A* 102, 3117-3122.
Kagaya, Y., Ohmiya, K. and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res* 27, 470-478.
Keck, E., McSteen, P., Carpenter, R. and Coen, E. (2003). Separation of genetic functions controlling organ identity in flowers. *Embo J* 22, 1058-1066.
Kim, S., Soltis, P. S., Wall, K. and Soltis, D. E. (2006). Phylogeny and domain evolution in the *APETALA2*-like gene family. *Mol Biol Evol* 23, 107-120.
Klucher, K.M., Chow, H., Reiser, L. and Fischer, R.L. (1996). The *AINTEGUMENTA* gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene *APETALA2*. *The Plant Cell* 8, 137-153.
Komaki, M.K., Okada, K., Nishino, E., and Shimura, Y. (1988). Isolation and characterization of nove1 mutants of Arabidopsis thaliana defective in flower development. *Development* 104, 195–203.
Krizek, B.A. and Meyerowitz, E.M. (1996). The *Arabidopsis* homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. *Development* 122, 11-22.
Krizek, B.A., Prost, V. and Macias, A. (2000). *AINTEGUMENTA* promotes petal identity and acts as a negative regulator of *AGAMOUS. The Plant Cell* 12, 1357-1366.
Krizek, B.A. (1999) Ectopic expression of *AINTEGUMENTA* in Arabidopsis plants results in increased growth of floral organs. *Dev. Gen.* 25, 224-236.
Krizek, B. A. (2003). AINTEGUMENTA utilizes a mode of DNA recognition distinct from that used by proteins containing a single AP2 domain. *Nucleic Acids Res* 31, 1859-1868.
Krizek, B. A. and Fletcher, J. C. (2005). Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet* 6, 688-698.
Krizek, B. A., Lewis, M. W. and Fletcher, J. C. (2006). *RABBIT EARS* is a second-whorl repressor of *AGAMOUS* that maintains spatial boundaries in *Arabidopsis* flowers. *Plant J* 45, 369-383.

Kunst, L., Klenz, J. E., Martinez-Zapater, J. and Haughn, G. W. (1989). *AP2* gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana*. *Plant Cell* 1, 1195-1208.
Lewis, E. (1998). The bithorax complex: the first fifty years. *Int. J. Dev. Biol.* 42, 403-415.
Litt, A. (2007). An evaluation of A-funciton: evidence from the *APETALA1* and *APETALA2* gene lineages. *Int. J. Plant Sci.* 168, 73-91.
Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell* 10, 1391-1406.
Mandel, M.A., Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F. (1992). Molecular characterizatio of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* 69, 273-277.
Maes, T., Van de Steene, N., Zethof, J., Karimi, M., D'Hauw, M., Mares, G., Van Montagu, M. and Gerats, T. (2001). Petunia *Ap2*-like genes and their role in flower and seed development. *Plant Cell* 13, 229-244.
Mathieu, J., Yant, L. J., Murdter, F., Kuttner, F. and Schmid, M. (2009). Repression of flowering by the miR172 target SMZ. *PLoS Biol* 7, e1000148.
Melzer, R., Verelst, W. and Theissen, G. (2009). The class E floral homeotic protein SEPALLAT3 is sufficient to loop DNA in "floral quartet"-like complexes *in vitro*. *Nucleic Acids Res* 37, 144-157.
Melzer, R. and Theissen, G. (2009). Reconstitution of "floral quartets" *in vitro* involving class B and class E floral homeotic proteins. *Nucleic Acids Res* 37, 2723-2736.
Melzer, R. and Theissen, G. (2011). MADS and more: transcription factors that shape the plant. *Methods Mol Biol* 754, 3-18.
Mizukami, Y. and Ma, H. (1992). Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* 71, 119-131.
Mlotshwa, S., Yang, Z., Kim, Y. and Chen, X. (2006). Floral patterning defects induced by *Arabidopsis APETALA2* and microRNA172 expression in *Nicotiana benthamiana*. *Plant Mol Biol* 61, 781-793.
Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pages, H. and Gentleman, R. (2009). ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* 25, 2607-2608.
Morgenstern, B. (2004). DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Res* 32, W33-36.
Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Jouette, D., Lacombe, A. M., Nikic, S., Picault, N. et al. (2000). *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* 101, 533-542.
Nole-Wilson, S. and Krizek, B. A. (2000). DNA binding properties of the *Arabidopsis* floral development protein AINTEGUMENTA. *Nucleic Acids Res* 28, 4076-4082.
Ohme-Takagi, M. and Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell* 7, 173-182.

Ohto, M.A., Fischer, R.L., Goldberg, R.B., Nakamura, K. and Harada, J.J. (2005). Control of seed mass by APETALA2. *Proceedings of the National Academy of Sciences of the United States of America* 102, 3123-3128.
Ohto, M. A., Floyd, S. K., Fischer, R. L., Goldberg, R. B. and Harada, J. J. (2009). Effects of *APETALA2* on embryo, endosperm, and seed coat development determine seed size in *Arabidopsis*. *Sex Plant Reprod* 22, 277-289.
Okamuro, J. K., Caster, B., Villarroel, R., Van Montagu, M. and Jofuku, K. D. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc Natl Acad Sci U S A* 94, 7076-7081.
Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405, 200-203.
Pelaz, S., Tapia-Lopez, R., Alvarez-Buylla, E.R. and Yanofsky, M. (2001). APETALA1 and SEPALLATA3 interact to promote flower development. *Curr Biol* 11, 182-184.
Riechmann, J. L. and Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. *Biol Chem* 379, 633-646.
Ripoll, J. J., Roeder, A. H., Ditta, G. S. and Yanofsky, M. F. (2011). A novel role for the floral homeotic gene APETALA2 during Arabidopsis fruit development. *Development* 138, 5167-5176.
Sablowski, R. W. and Meyerowitz, E. M. (1998). A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3*/*PISTILLATA*. *Cell* 92, 93-103.
Saedler, H., Becker, A., Winter, K.U., Kirchner, C. and Theissen, G. (2001). MADS-box genes are involved in floral development and evolution. *Acta Biochm Pol* 48, 351-358.
Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem Biophys Res Commun* 290, 998-1009.
Saleh, A. and Pagés, M. (2003). Plant AP2/ERF transcription factors. *Genetika* 35, 37-50.
Sanders, P.M., Bui, A.Q., Weterings, K., McIntire, K.N., Hsu, Y.-C., Lee, P.Y., Truong, M.T., Beals, T.P., and Goldberg, R.B. (1999) Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. *Sex. Plant Reprod.* 11, 297-322
Schultz, E. A. and Haughn, G. W. (1993). Genetic analysis of the floral initiation process (FLIP) in *Arabidopsis*. *Development* 119, 745-765.
Schwartz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250, 931-936.
Shannon, S. and Meeks-Wagner, D. R. (1993). Genetic interactions that regulate inflorescence development in *Arabidopsis*. *Plant Cell* 5, 639-655.
Shigyo, M., Hasebe, M. and Ito, M. (2006). Molecular evolution of the *AP2* subfamily. *Gene* 366, 256-265.
Sieburth, L. E. and Meyerowitz, E. M. (1997). Molecular dissection of the *AGAMOUS* control region shows that cis elements for spatial regulation are located intragenically. *Plant Cell* 9, 355-365.

Smaczniak, C., Immink, R.G., Muiño, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J.,, Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles C.C., Angenent, G.C. and Kaufmann, K. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proceedings of the National Academy of Sciences of the United States of America* 109, 1560-1565.
Smith, H. M., Boschke, I. and Hake, S. (2002). Selective interaction of plant homeodomain proteins mediates high DNA-binding affinity. *Proceedings of the National Academy of Sciences of the United States of America* 99, 9579-9584.
Stockinger, E. J., Gilmour, S. J. and Thomashow, M. F. (1997). Arabidopsis thaliana *CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proc Natl Acad Sci U S A* 94, 1035-1040.
Team, R. D. C. (2010). R: a language and environment for statistical computing.
Theissen, G. (2001). Genetics of identity. *Nature* 414, 491.
Theissen, G. and Saedler, H. (2001). Plant biology. Floral quartets. *Nature* 409, 469-471.
Tilly, J. J., Allen, D. W. and Jack, T. (1998). The CArG boxes in the promoter of the *Arabidopsis* floral organ identity gene *APETALA3* mediate diverse regulatory effects. *Development* 125, 1647-1657.
Treisman, R. (1992). The serum response element. *Trends Biochem Sci* 17, 423-426.
Wagner, D., Sablowski, R. W. and Meyerowitz, E. M. (1999). Transcriptional activation of *APETALA1* by *LEAFY*. *Science* 285, 582-584.
Weigel, D. (1995). The APETALA2 domain is related to a novel type of DNA binding domain. *Plant Cell* 7, 388-389.
Wierzbicki, A. T., Haag, J. R. and Pikaard, C. S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135, 635-648.
William, D. A., Su, Y., Smith, M. R., Lu, M., Baldwin, D. A. and Wagner, D. (2004). Genomic identification of direct target genes of *LEAFY*. *Proc Natl Acad Sci U S A* 101, 1775-1780.
Wuitschick, J. D., Lindstrom, P. R., Meyer, A. E. and Karrer, K. M. (2004). Homing endonucleases encoded by germ line-limited genes in *Tetrahymena thermophila* have APETALA2 DNA binding domains. *Eukaryot Cell* 3, 685-694.
Würschum, T., Gross-Hardt, R. and Laux, T. (2006). *APETALA2* regulates the stem cell niche in the *Arabidopsis* shoot meristem. *Plant Cell* 18, 295-307.
Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A. and Meyerowitz, E.M. (1990). The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. *Nature* 346, 35-39.
Yant, L., Mathieu, J., Dinh, T. T., Ott, F., Lanz, C., Wollmann, H., Chen, X. and Schmid, M. (2010). Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. *Plant Cell* 22, 2156-2170.
Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G. R., Janse, C. J., Waters, A. P., Kato, T. and Kaneko, I. (2009). Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol* 71, 1402-1414.

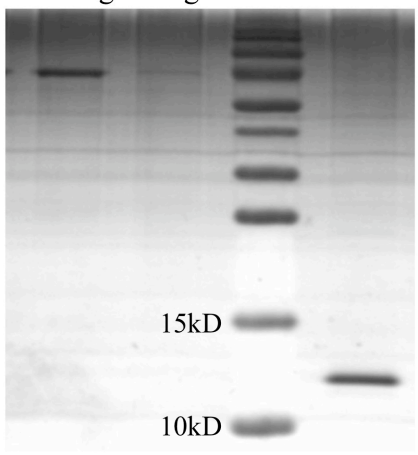
Figure 1.1 Identification of an AP2R2 binding sequence

(A) Purified His-AP2R2-T7 (~13kD) resolved by SDS-PAGE. BSA was used as a standard for estimating the amount of the AP2R2 protein. M, molecular weight standards.

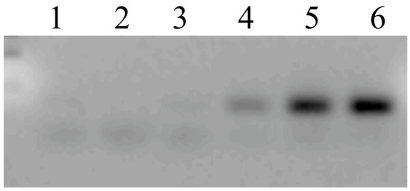
(B) PCR amplification of the AP2R2-bound DNA during the SAAB assay. Numbers on top of the gel represent cycle number.

(C) PCR product from cycle 6 were cloned, sequenced and found to contain the consensus sequence TTTGTT and/or AACAAA. The orange sequences differ from these motifs by a single nucleotide.

A



B



C

```
CLONE1     AACAAAAAGTTTGTTCACCG
CLONE2     AAACAAAAAGTTTGTTTGGC
CLONE3     GAACAAAAAGTTTGTTCCGG
CLONE4     GAACAAAATGTTTGTTTCGG
CLONE5     GAACAAACTGTTTGTTTGTG
CLONE6     AAAACAAACTTTTTGTTCGC
CLONE7     AAAACAAACTTTTTGTGCCG
CLONE8     AACAAATATTTTGTTCCGTC
CLONE9     GGTGAACAAATTATTTGTTC
CLONE10    TAACAAAATTTTGTTACGCG
CLONE11    CTTGGATCGAAAAGTTTGTT
CLONE12    AACAATATTCATTATTGTTC
CLONE13    GGGTTTTGTTCTTTTTTGTT
CLONE14    ATTTGTTCCAGATTTGTTTG
CLONE15    AATTTGTTTTTGTTTGTTTG
CLONE16    GTTTGTTTCGATTTGTTTCG
CLONE17    GTTTGTTTCGATTTGTTTCG
CLONE18    GTTTGTTCGATTTTGTTTCG
CLONE19    GTTTGTTTTAGTTTGTTTTG
CLONE20    AAACAAACGGGAACAAAGTC
```

Figure 1.2 AP2R1 does not bind DNA *in vitro*

(A) Electrophoresis of purified His-AP2R1-T7. The gel was stained with Coomassie Blue. BSA was used as a standard for a rough estimation of protein quantities.

(B) PCR amplification of bound DNA after each cycle during the SAAB assay. The numbers above the gel image indicate the SAAB assay cycle numbers. Even with 2μg of AP2R1 protein used (as opposed to 200ng for AP2R2), no DNA was recovered in the SAAB assay.

(C) Western blotting using a monoclonal anti-T7 antibody shows the presence of the AP2R1protein (indicated by a star) after each cycle of the SAAB assay. Numbers=cycle; AP2R1=13kD. M, molecular weight standards.

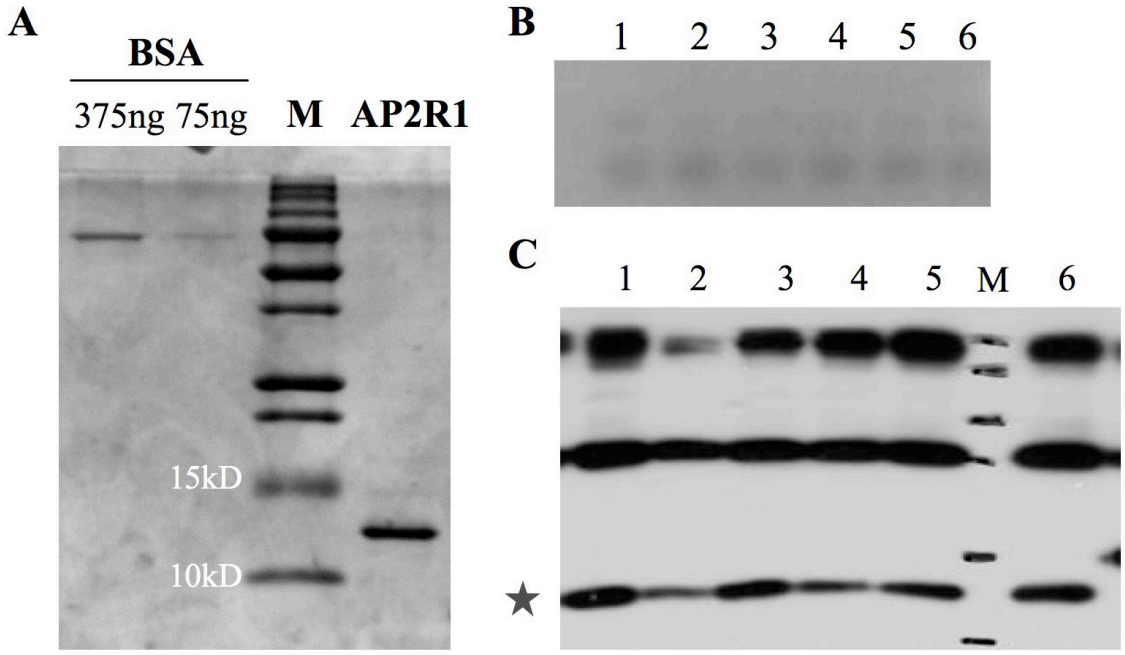
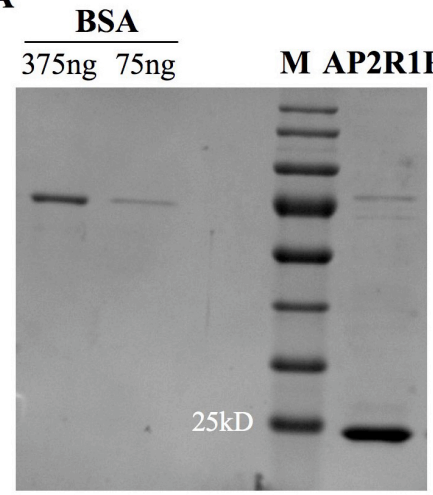

A
BSA
375ng 75ng
M
AP2R1
15kD
10kD
B
1 2 3 4 5 6
C
1 2 3 4 5 M 6
★

Figure 1.3 AP2R1R2 binds DNA in a SAAB assay but no consensus sequence can be identified

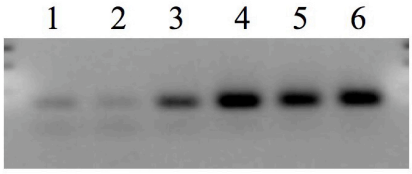
(A) Electrophoresis of purified His-AP2R1R2-T7 (~23kD). The gel was stained with Coomassie Blue. BSA was used as a standard for a rough estimation of protein quantities.

(B) PCR amplification of the AP2R1R2-bound DNA after each cycle of the SAAB assay. Numbers on top of the gel image represent SAAB assay cycle numbers. (**C**) Sequences of 17 out of 25 clones from the PCR products after cycle 6 of the SAAB assay. No consensus sequence was identified. However, 10 out of 25 clones contained one of the AP2R2 consensus sequences (some not shown) and 1 out of 25 (clone 3) contained a site with one nucleotide change. M, molecular weight standards.

A



B



C

```
CLONE1    ATAGCTTTAGTACATGCCAC
CLONE2    GTACGAAACAAACTGTTGTG
CLONE3    GACCGCTTTAGTTTTTGGGC
CLONE4    TTAGAACCTCGTACTTTGTT
CLONE5    CGACCCGTCGAAGCAGCGTT
CLONE6    TGGGGGTGGGGGCGGAGGGC
CLONE7    CAAACCTAGGATTGTCGAAC
CLONE8    AAATAACAGGGTGGGGGATT
CLONE9    CGACAAACAAAACTCTGTTC
CLONE10   GGACGAACAAAAACCTCGGC
CLONE11   CCAGGCGATGGAAAACTGTG
CLONE12   GGTCGCAGGTGTACTTATGC
CLONE13   TGTGGCATCGGGGGGGTGTC
CLONE14   GTACGAAACAAACTGTTGTG
CLONE15   TACGAGATGGTACGAGGGTC
CLONE16   AGTGCCTCGTACAGAAGGAC
CLONE17   AAGGTGGATATTGTGCGATG
```

Figure 1.4 AP2R2 binds the TTTGTT and/or AACAAA motif *in vitro*

(A) The sequences of the DNA probes (only one strand is shown). The consensus sites are underlined. αβ: a probe containing two sites; Δα or Δβ: probes containing only one site.

(B) AP2R2 binds the αβ probe containing both sites. The shifted band that represents binding (lanes 1 and 2) is indicated by an arrow. The binding was lost upon the addition of 20x cold competitor (lane 3). The + sign indicates that 200ng of AP2R2 protein was included; for lane 1, 100ng AP2R2 was used. In lanes 4 and 5, 200ng and 400ng, respectively, of T7 and His antibodies were added to the reactions. The bands indicated by the stars and double stars likely represented the super-shifts. As a negative control, 400ng His and T7 antibodies were added without AP2R2 (lane 6). Lane 7 represents the free probe lane.

(C) One site is sufficient for binding by AP2R2. A shift was observed, as indicated by the arrow, with probes containing one site (lanes 2 and 5). Binding was lost upon addition of the cold competitors (lanes 3 and 6). AP2R2 binding was abolished when both sites were mutated (lane 8). Lanes 1, 4 and 7 are the free probe lanes.

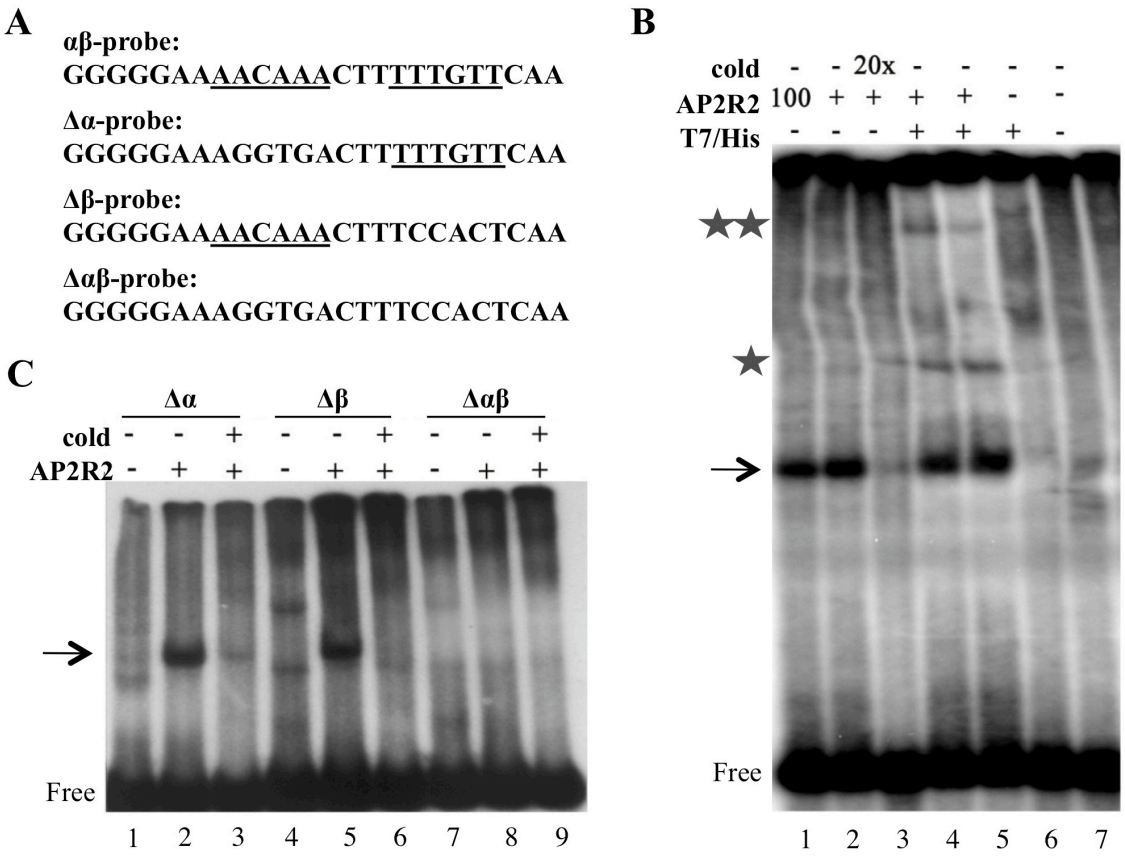

A
αβ-probe:
GGGGGAAAACAAACTTTTTGTTCAA
Δα-probe:
GGGGGAAAGGTGACTTTTTGTTCAA
Δβ-probe:
GGGGGAAAACAAACTTTCCACTCAA
Δαβ-probe:
GGGGGAAAGGTGACTTTCCACTCAA
B
cold - - 20x - - - -
AP2R2 100 + + + + - -
T7/His - - - + + + -
Free
1 2 3 4 5 6 7
C
Δα Δβ Δαβ
cold - - + - - + - - +
AP2R2 - + + - + + - + +
Free
1 2 3 4 5 6 7 8 9

Figure 1.5 Dissection of the AP2R2 consensus sequence

EMSAs were performed upon mutational perturbations at each position of the six-mer consensus sequence to the other three nucleotides while keeping the other five positions unchanged. Green boxes represent the original site, yellow and orange boxes represent medium and minimal binding, respectively, whereas red boxes indicate complete loss of binding. Medium and minimal binding are defined as 20-30% and 1-5% of binding in comparison to the wild-type sequence, respectively.

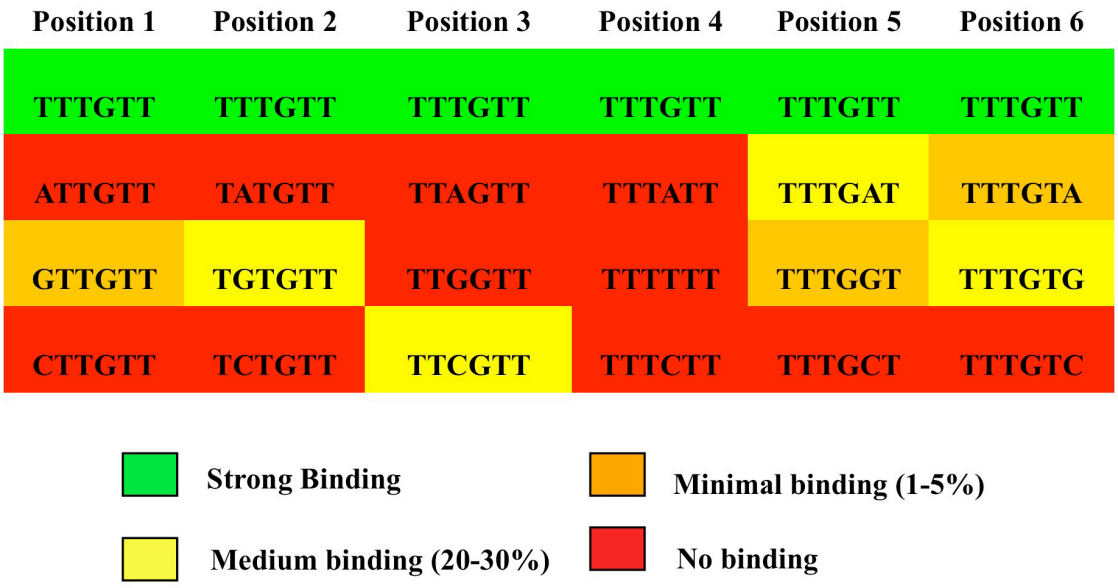

Position 1
Position 2
Position 3
Position 4
Position 5
Position 6
TTTGTT
TTTGTT
TTTGTT
TTTGTT
TTTGTT
TTTGTT
ATTGTT
TATGTT
TTAGTT
TTTATT
TTTGAT
TTTGTA
GTTGTT
TGTGTT
TTGGTT
TTTTTT
TTTGGT
TTTGTG
CTTGTT
TCTGTT
TTCGTT
TTTCTT
TTTGCT
TTTGTC
Strong Binding
Minimal binding (1-5%)
Medium binding (20-30%)
No binding

Figure 1.6 Analysis of the AP2R2 consensus sequence

Raw gel images of table depicted in Figure 2. In each gel, the first two lanes were shifts done with the Δα probe and were used as a positive control in the quantification analyses. For the other lanes, gel shifts with mutations of each nucleotide position were performed as indicated. The number 1 represents free probe. The number 2 represents the addition of AP2R2 protein and the number 3 represents a failed AP2 full-protein preparation (low and dirty protein yield) from an early construct. Yellow and orange arrowheads represent medium and minimal binding, respectively, and are defined as 20-30% and 1-5% binding in comparison to the wild-type sequence, respectively.

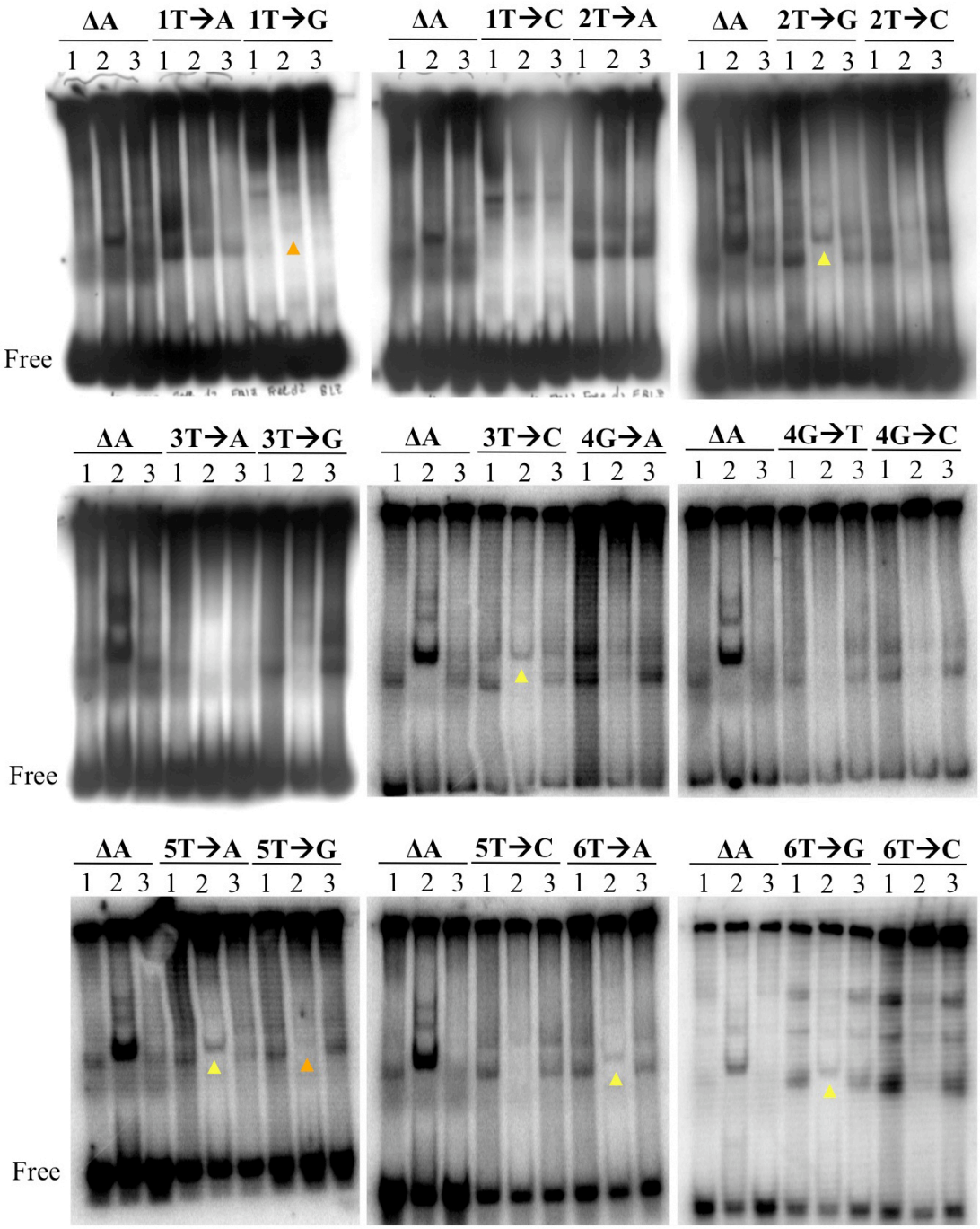

ΔA
1T→A
1T→G
1T→C
2T→A
2T→G
2T→C
3T→A
3T→G
3T→C
4G→A
4G→T
4G→C
5T→A
5T→G
5T→C
6T→A
6T→G
6T→C
1 2 3
Free

Figure 1.7 AP2R2 does not bind the canonical GCC-box or two random probes

(A) Sequences of the probes (only one strand is shown). A five-G overhang was included for radiolabeling via Klenow reaction. The canonical GCC-box is indicated by the underline.

(B) Gel shifts with the Δα, the GCC-box, and two random probes. AP2R2 was able to bind the Δα-probe (the shift in lane 2 is indicated by an arrow) but not the GCC-box and two random probes (lanes 5, 7, and 9). The + sign indicates the presence of 200ng of AP2R2. The free probe lanes are 1, 4, 6 and 8. 20x cold competitor was added to the reaction mixture in lanes 3 and 10.

A

GCC box:
GGGGGTAAGAGCCGCCATGATCGATT

Random Probe 1 (RP-1):
GGGGGGCACGGCAGGTCATCGTACCA

Random probe 2 (RP-2):
GGGGGAAAGGTGACTTTGGGCCCAA

B

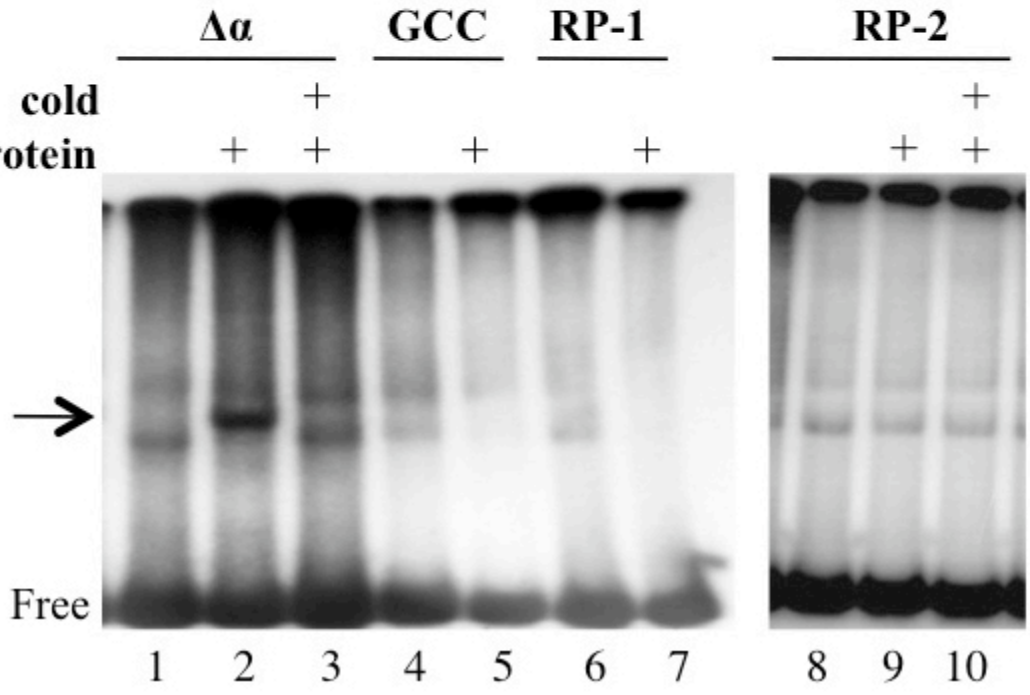


Figure 1.8 Purified MBP-AP2 full-length protein (A) and MBP (B)

The purified MBP-AP2 and MBP were resolved by SDS-PAGE and stained with Coomassie Blue. BSA was used as a standard for a rough estimation of protein quantities. M, molecular weight standards.

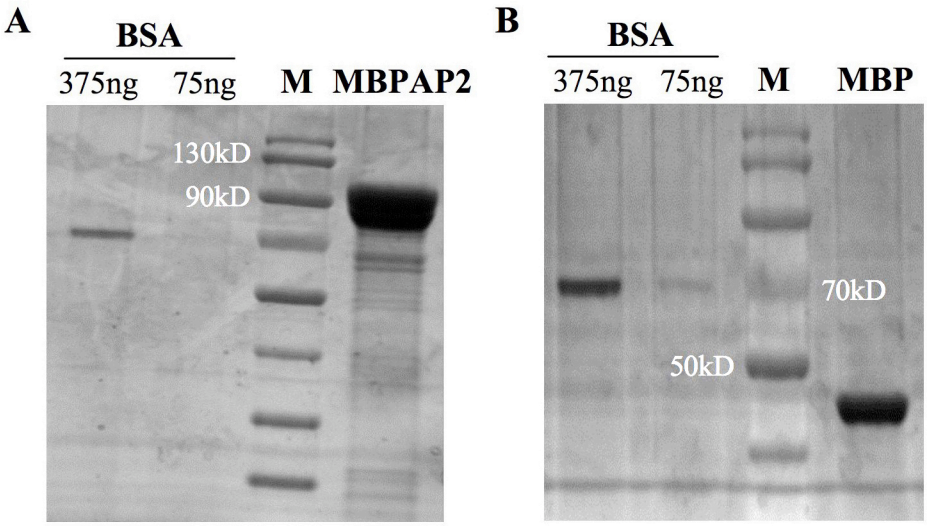
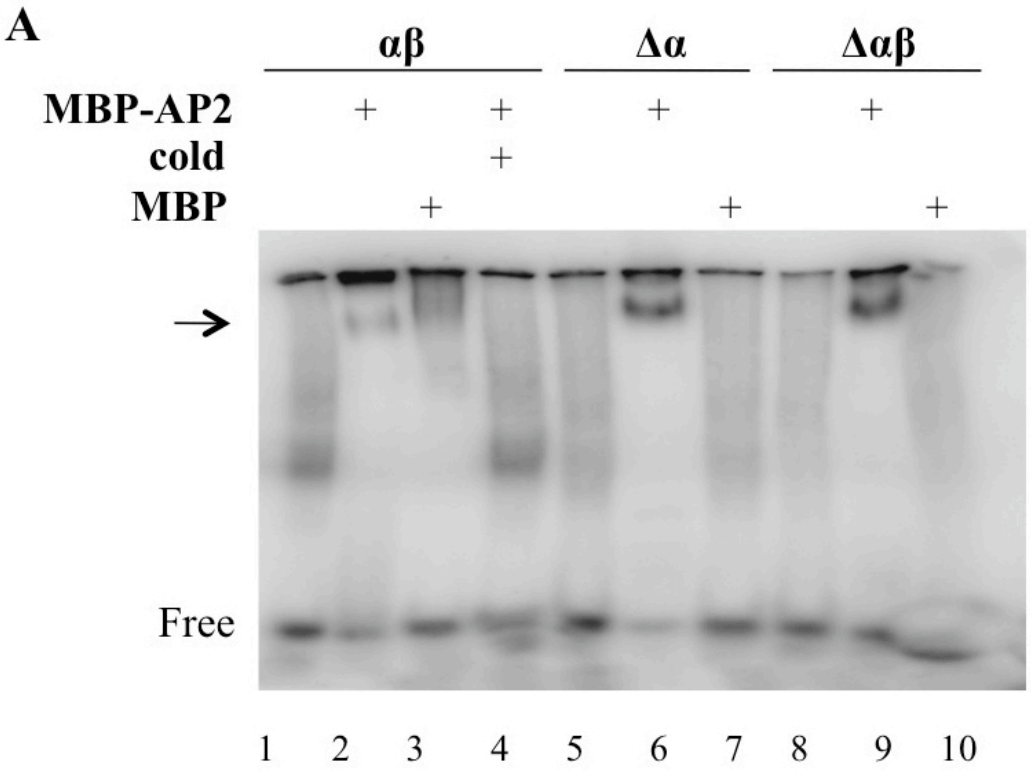

A
BSA
375ng
75ng
M
MBPAP2
130kD
90kD
B
BSA
375ng
75ng
M
MBP
70kD
50kD

Figure 1.9 AP2 full-length protein binds DNA *in vitro*

(A) EMSAs on αβ, Δα, and Δαβ probes. MBP-AP2 full-length protein (labeled as MBP-AP2) was able to bind all three probes (lanes 2, 6 and 9). This binding was abolished upon addition of 10x cold competitor (lane 4). MBP alone did not bind DNA (lanes 3, 7 and 10). There seems to be some MBP-AP2 bound αβ probe still stuck at the well (lane 2).

(B) MBP-AP2 full-length protein was able to bind the canonical GCC-box and two other random probes (lanes 2, 5 and 8). MBP alone was not able to bind these probes (lanes 3, 6 and 9). 400ng of MBP-AP2 full-length protein or MBP alone was used in all gel shifts. The shifted bands are indicated by arrows.


A
αβ
Δα
Δαβ
MBP-AP2
cold
MBP
Free
1 2 3 4 5 6 7 8 9 10

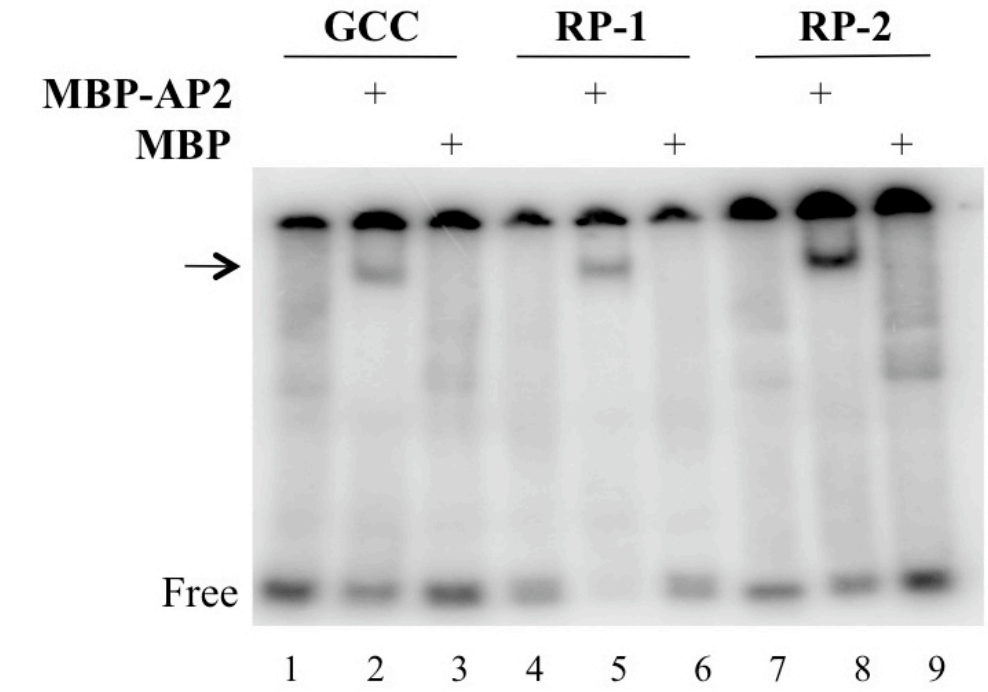

B
GCC
RP-1
RP-2
MBP-AP2
MBP
Free
1 2 3 4 5 6 7 8 9

Figure 1.10 AP2R2 and AP2 full-length protein bind *AG* 2nd intron *in vitro*

(A) A diagram of a partial *AG* genomic region. Solid dark gray boxes indicate the 2nd and 3rd exons, lines represent introns, and the solid light gray box indicates the 750-bp region (KB31) that still responds to negative regulation by *AP2*. Within KB31, a 167-bp region, for which the sequence is shown, contains two AP2R2 consensus sites (rectangles). This 167-bp region was amplified with primers (arrows) and used as a probe for EMSA. Several nucleotides in the two consensus sequences were mutated (in gray) to generate the ΔAB probe.

(B) EMSA with AP2R2 on the 167-bp wild-type (AB) or ΔAB probe. AP2R2 bound the wild-type probe (lanes 2-7; arrow) and the binding was lost upon addition of the cold competitor (lanes 8 and 9). c=20x cold competitor; C=40x cold competitor. The protein amounts are as follows: 35ng (lane 2), 75ng (lanes 3 and 11), 107ng (lane 4), 160ng (lanes 5 and 12), and 240ng (lanes 6 and 13). Binding was minimal for the ΔAB probe containing mutated versions of both sites (lanes 11-13). Free= 0.2 pmol/reaction (lanes 1 and 10).

(C) EMSA with MBP-AP2 full-length protein (labeled as MBP-AP2) on the 167-bp fragment. MBP-AP2 full-length protein bound both the wild-type (lanes 3-8) and mutated (lanes 12 and 13) probes, and the binding was lost upon addition of the cold competitor (lanes 9 and 14). Binding was observed starting at 243ng of the protein (lane 6), and was most obvious with 355ng (lane 7) and 429ng (lane 8) of the protein. MBP

(M) alone did not bind any probe (lanes 2 and 11). Free=0.1 pmol/reaction (lanes 1 and 10). Triangles depict increasing amounts of protein added. The arrow marks the shift.

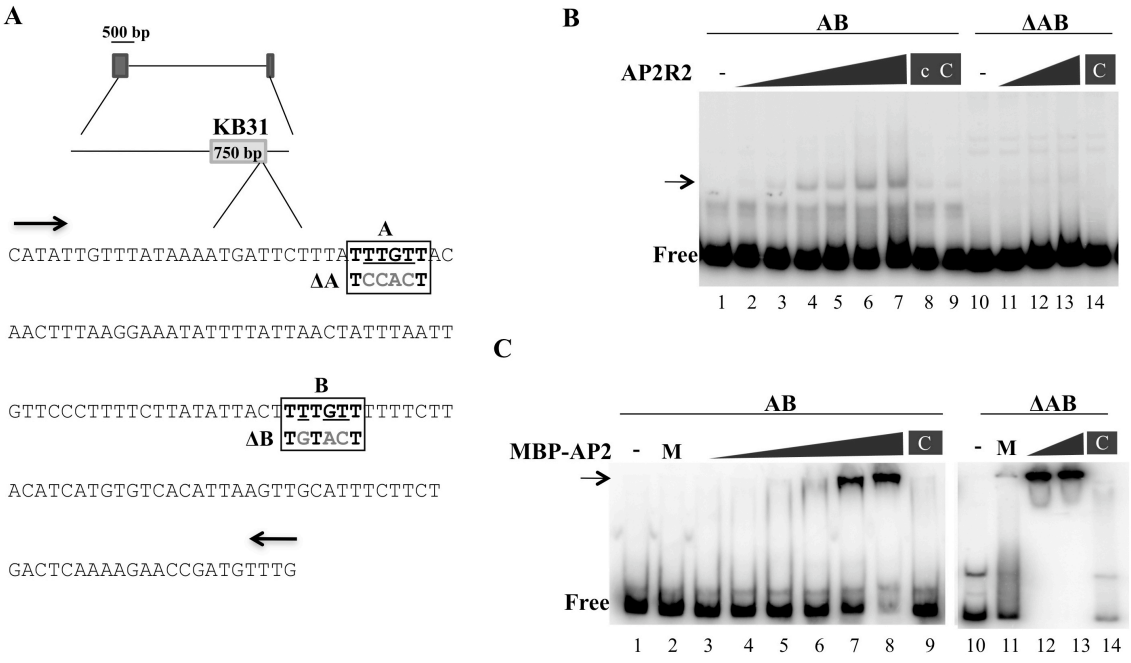

A
500 bp
KB31
750 bp
A
CATATTGTTTATAAAATGATTCTTTATTTGTTAC
ΔA TCCACT
AACTTTAAGGAAATATTTTATTAACTATTTAATT
B
GTTCCCTTTTCTTATATTACTTTTGTTTTTTCTT
ΔB TGTACT
ACATCATGTGTCACATTAAGTTGCATTTCTTCT
GACTCAAAAGAACCGATGTTTG
B
AB
ΔAB
AP2R2
-
c C
-
C
Free
1 2 3 4 5 6 7 8 9 10 11 12 13 14
C
AB
ΔAB
MBP-AP2
-
M
C
-
M
C
Free
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 1.11 The AP2R2 binding sites are essential in restricting *AG* expression in floral meristems

Either the wild-type *KB31* fragment or a mutant version lacking the two AP2R2 binding sites (*KB31ΔAB*) was fused to GUS and GUS expression was evaluated *in vivo* in T1 transgenic lines. Representative images of typical T1 inflorescences are shown.

(A-B) Stages 3-4 floral meristems.

(C-D) Stages 6-7 flowers.

(E-F) Stages 8-12 flowers.

(A, C, and E) Wild-type *KB31*.

(B, D and F) Mutant *KB31* lacking both A and B sites (*KB31ΔAB*).

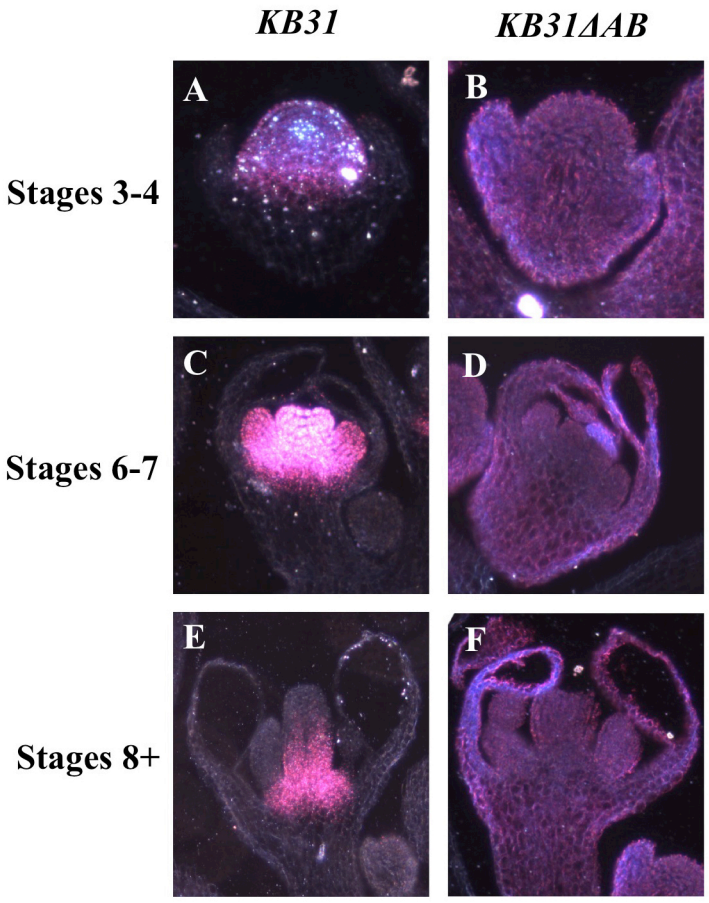

KB31
KB31ΔAB
A
B
Stages 3-4
C
D
Stages 6-7
E
F
Stages 8+

Figure 1.12 Induction of AP2 leads to decreased *AG* mRNA levels in young floral meristems

(A) An uninduced *35S::AP2m3-GR* flower.

(B) A *35S::AP2m3-GR* flower after continuous DEX induction.

(C) RT-PCR analysis of *AG* mRNA levels from *35S::AP2m3-GR ap2-2* inflorescences after six hours of treatments with CHX alone (6C) or CHX + DEX (6DC). 33 and 25 cycles of PCR were performed for *AG* and *UBQ5* (an internal control), respectively.

(D) Real time RT-PCR analysis of three biological replicates of *35S::AP2m3-GR ap2-2* C- or DC-treated inflorescence tissue.

For both (**C**) and (**D**), dissected inflorescences containing stages 7 and younger flowers were used.

(E) Real time RT-PCR analysis of *KB31 35S::AP2m3-GR ap2-2* C- or DC-treated inflorescence tissue.

(F) Real time RT-PCR analysis of *KB31ΔAB 35S::AP2m3-GR ap2-2* C- or DC-treated inflorescence tissue.

Primers corresponding to two regions in the *GUS* coding region (*GUSp1* and *GUSp3*) were used for the real time RT-PCR. For (**E**) and (**F**), three biological replicates were performed and yielded similar results. One representative image is shown. Error bars represent the standard deviation from three technical replicates.

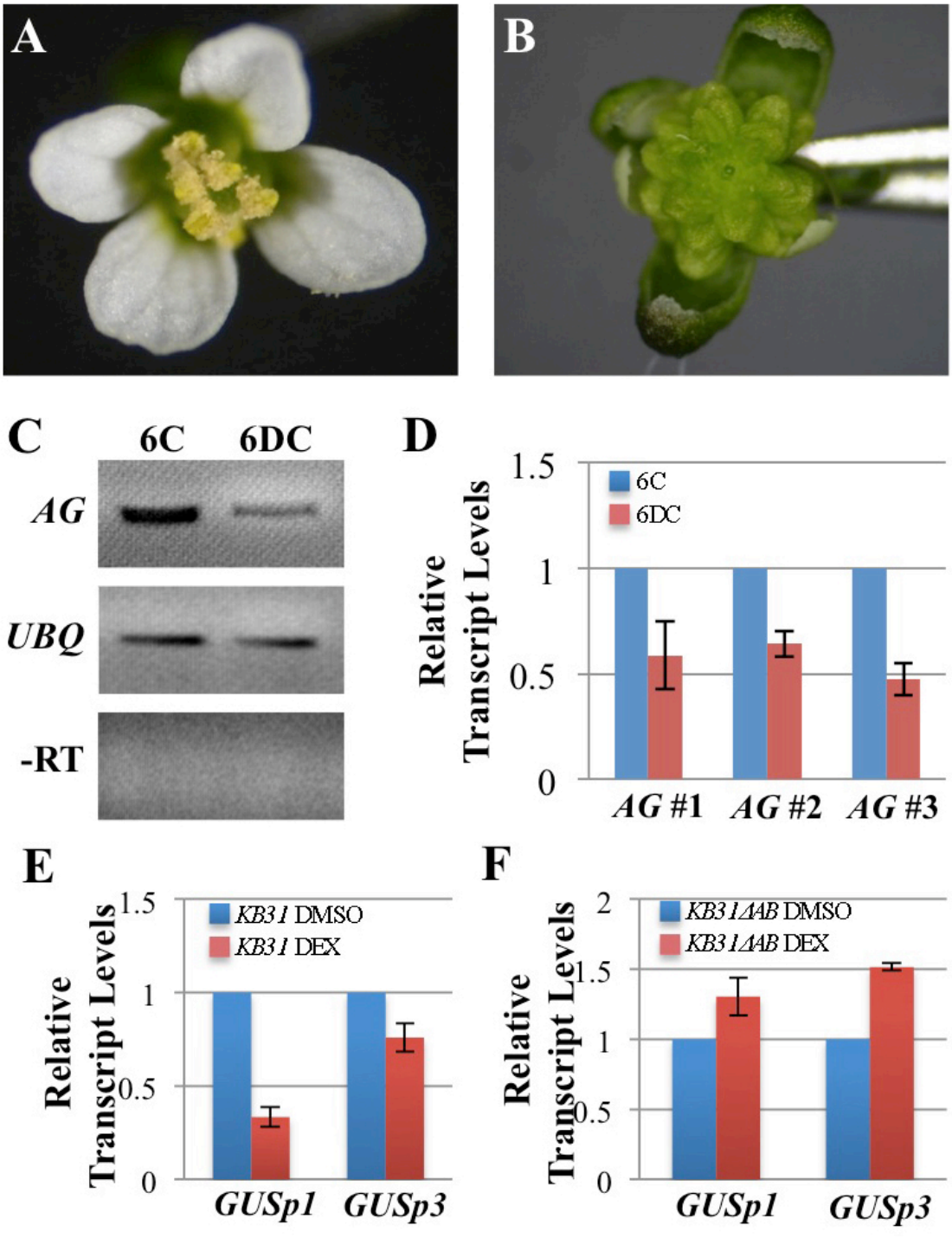

A
B
C
6C
6DC
AG
UBQ
-RT
D
Relative Transcript Levels
1.5
1
0.5
0
6C
6DC
AG #1
AG #2
AG #3
E
Relative Transcript Levels
1.5
1
0.5
0
KB31 DMSO
KB31 DEX
GUSp1
GUSp3
F
Relative Transcript Levels
2
1.5
1
0.5
0
KB31ΔAB DMSO
KB31ΔAB DEX
GUSp1
GUSp3

Figure 1.13 AP2 binds *AG* 2nd intron *in vivo*

(A) A diagram of a partial *AG* genomic region highlighting the second intron. Dark gray boxes indicate the 2nd and 3rd exons, whereas the line represents the 2nd intron.

(B) AP2 is associated with the *AG* 2nd intron *in vivo*. ChIP was performed on dissected inflorescence tissue from wild type and *ap2-2* using anti-AP2 antibodies. Real time PCR was performed on input and bound fractions and the percentage of bound DNA relative to input was calculated. The AB region, as well as region II (a positive control), was found to be enriched in the L*er* versus *ap2-2* sample. eIF4A1 was used as a negative control. Error bars represent the standard deviation from three technical replicates.

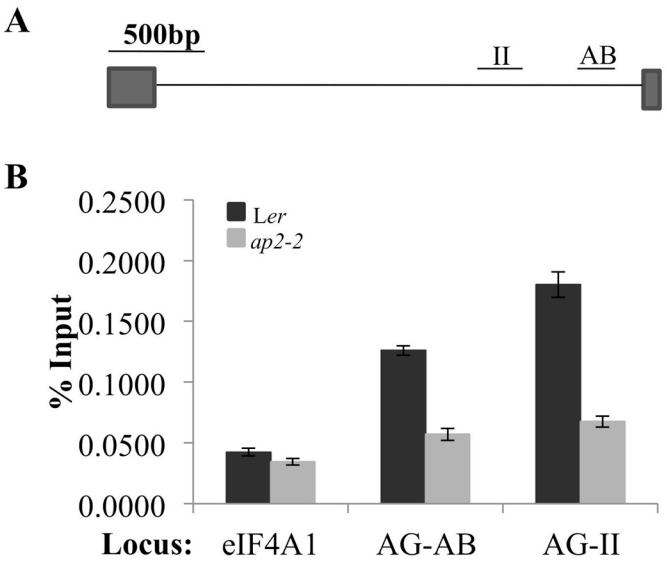

A
500bp
II
AB
B
0.2500
0.2000
0.1500
0.1000
0.0500
0.0000
% Input
Ler
ap2-2
Locus:
eIF4A1
AG-AB
AG-II

Figure 1.14 Dissection of region II within the *AG* 2nd intron

Region II found by Yant et al. through a ChIP assay to bind to AP2 was further dissected to test whether AP2R2 and/or MBP-AP2 bound the region *in vitro.*

(A) A diagram of part of the AG genomic region with the line representing the second intron. The positions of region II and the AB site are indicated. Region II was divided into four parts, 1 to 4. Part 1 encompassed a 134-bp region upstream of the II site (2040-2174 from the beginning of the 2nd intron (from here on, all nucleotide numbers in this figure refers to the nucleotide position from the start of the *AG* 2nd intron)). Part 2 encompassed nucleotides 2175-2292, part 3 encompassed nucleotides 2293-2418, and part 4 encompassed nucleotides 2419-2531.

(B) AP2R2 binds the ΔA probe (indicated by arrows) but does not bind the part 1, 2, or 4 of the *AG* 2nd intron. AP2R2 binds part 3 (indicated by the star).

(C) MBP-AP2 binds probes 1 to 4 (as indicated by a star).

The + symbol indicates 200ng of protein or 5x cold competitor added, ++ indicates 400ng of protein added, and M= 400ng MBP alone added.

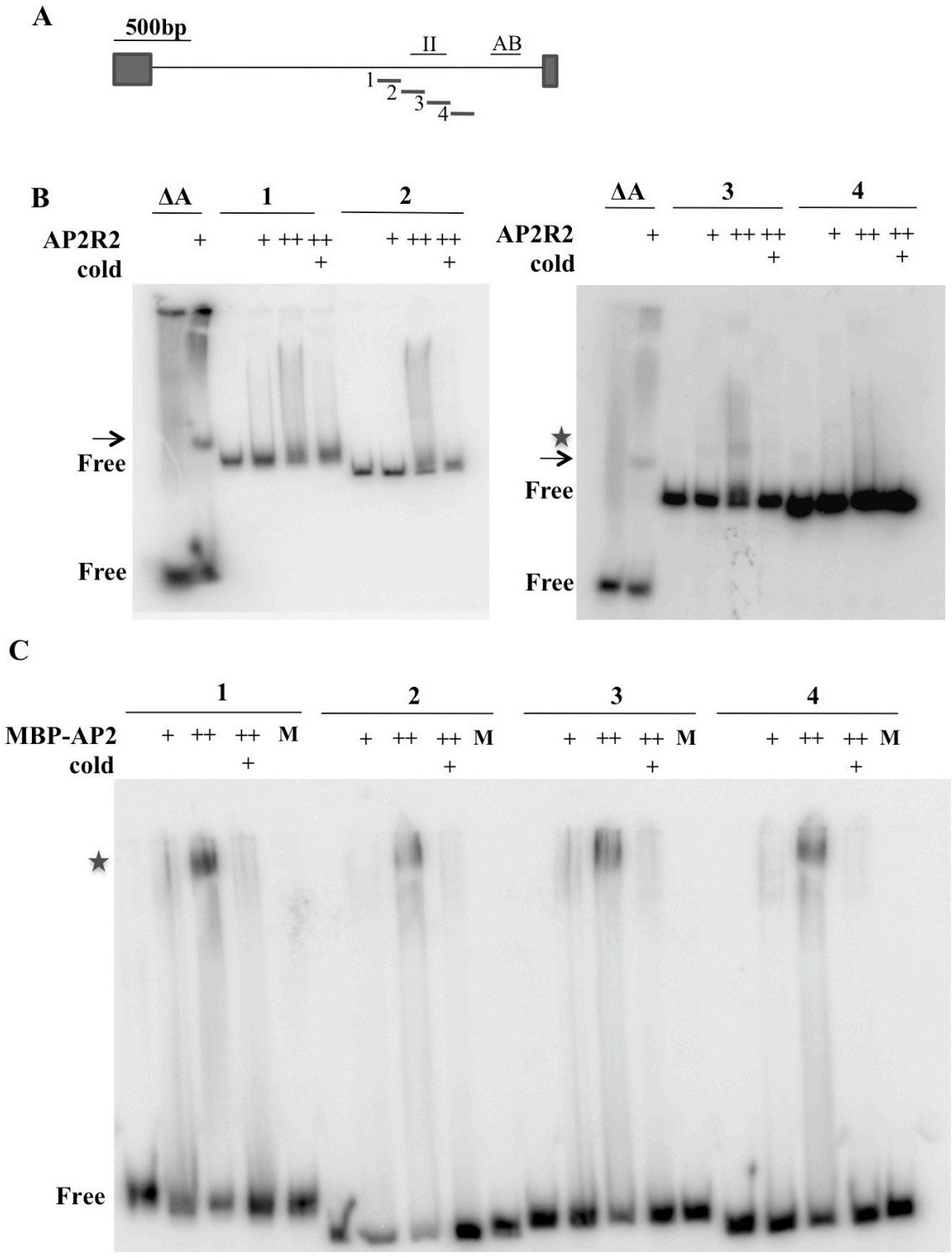

A
500bp
II
AB
1
2
3
4
B
ΔA
1
2
AP2R2
cold
Free
Free
ΔA
3
4
AP2R2
cold
Free
Free
C
1
2
3
4
MBP-AP2
cold
M
Free

Figure 1.15 Distribution of the TTTGTT consensus sequence in *Brassicaceae*

The consensus sequence TTTGTT was mapped to the multiple alignment of the 2nd introns of *AG* and *AG* orthologs from 29 *Brassicaceae* species. The aligned sequences are represented as a gray box and the positions of the pattern matches (TTTGTT) within the alignment are given in black. The sequence identifiers and their numbers of residues are given on the left and right, respectively. The scale on the bottom refers to the position in the multiple alignment. "A" and "B" sites are denoted with arrows and the corresponding letters.

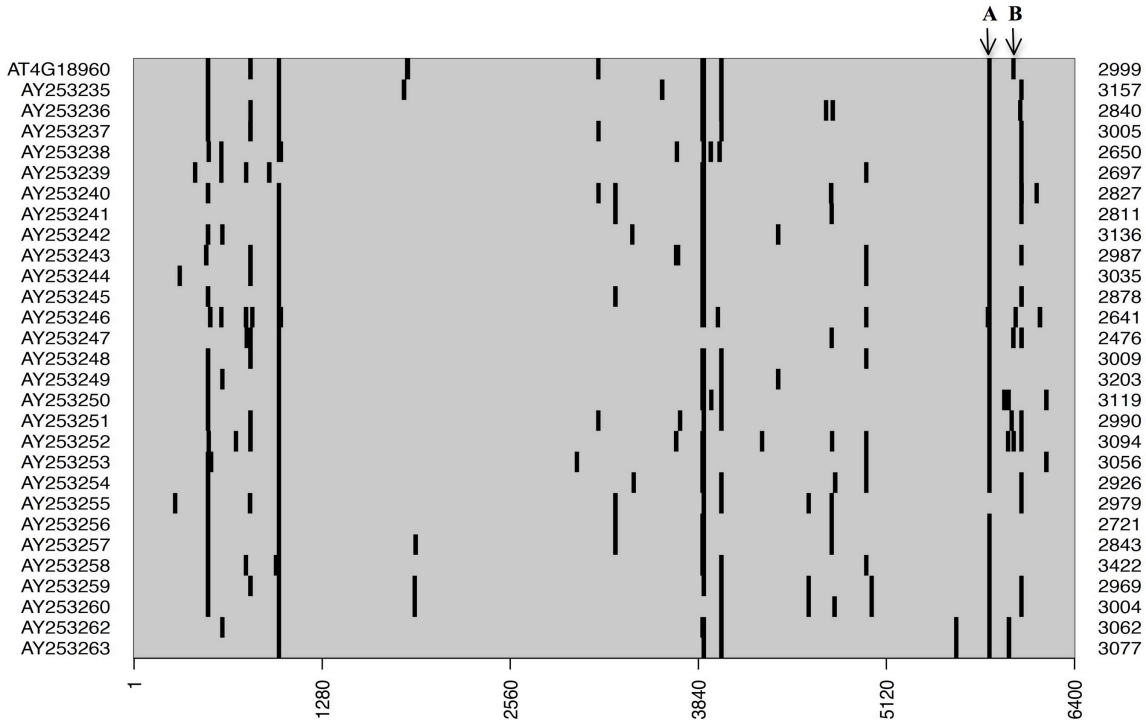

A
B
AT4G18960
AY253235
AY253236
AY253237
AY253238
AY253239
AY253240
AY253241
AY253242
AY253243
AY253244
AY253245
AY253246
AY253247
AY253248
AY253249
AY253250
AY253251
AY253252
AY253253
AY253254
AY253255
AY253256
AY253257
AY253258
AY253259
AY253260
AY253262
AY253263
2999
3157
2840
3005
2650
2697
2827
2811
3136
2987
3035
2878
2641
2476
3009
3203
3119
2990
3094
3056
2926
2979
2721
2843
3422
2969
3004
3062
3077
1
1280
2560
3840
5120
6400

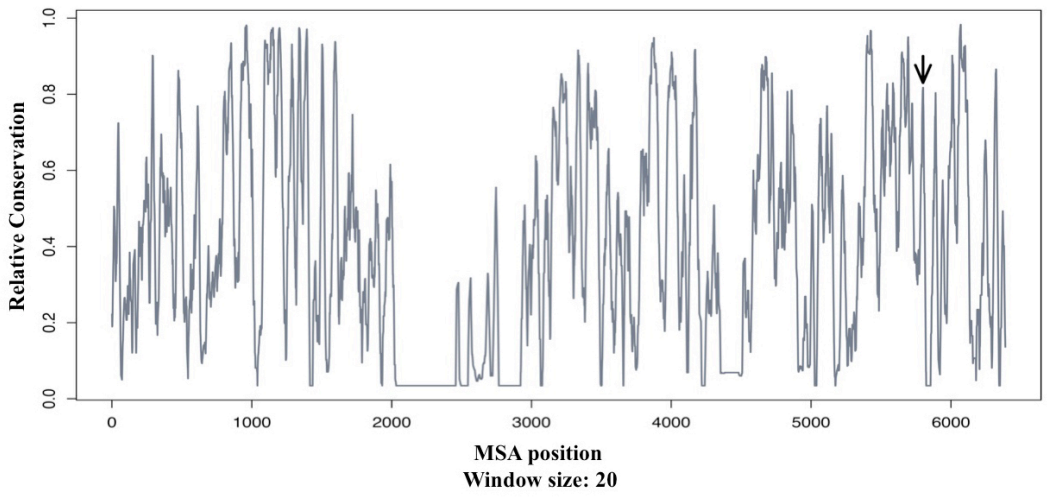
Figure 1.16 The "A" site within the *AG* 2nd intron is evolutionarily conserved

A multiple sequence alignment (MSA) of 2nd intron sequences from *AG* and *AG* orthologs from 29 *Brassicaceae* species was computed with dialign2-2.

(A) The profile of conserved alignment regions is shown in the form of a sliding window plot using a window size of 20 nucleotides. Values close to 1 indicate a high degree of conservation whereas values close to zero indicate low conservation. The position of the "A" site is indicated by an arrow. The scale on the bottom indicates the relative position in the MSA.

(B) A section of the multiple alignment that contains the "A" site, which is highlighted in red. Only the sequence from AY253255 contains a single nucleotide substitution (T to C) in this site (indicated in orange).

A



B

```
                                  A-site
AT4G18960   ATGATTCT...TTATTTGTTA.C.
AY253235    ATGACTCT...TTATTTGTTA.C.
AY253236    ATGATTCT...TTATTTGTTA.C.
AY253237    ATGATTCT...TTATTTGTTA.C.
AY253238    ATGATTTCTA..AATTTGTTA.T.
AY253239    ATGATTCT...TTATTTGTTA.G.
AY253240    ATGATTCT...TTATTTGTTA.C.
AY253241    ATGATTCT...TTATTTGTTA.C.
AY253242    ATGATTCT...TTATTTGTTGTT.
AY253243    ATGATTCT...TTATTTGTTA.C.
AY253244    ATGATTCT...TTATTTGTTA.C.
AY253245    ATGATTCTTTATTATTTGTTA.C.
AY253246    GTTTTTGT...TAATTTGTTA.T.
AY253247    ATGATTCT...TTATTTGTTA.T.
AY253248    ATGATTCT...TTATTTGTTA.C.
AY253249    ATGATTCT...TTATTTGTTA.TT
AY253250    ATGATTCT...TTATTTGTTA.C.
AY253251    ATGATTCT...TTATTTGTTA.C.
AY253252    ATGATTCT...TTATTTGTTA.C.
AY253253    ATGTTTCT...TTATTTGTTT.T.
AY253254    ATGATTCT...TTATTTGTTA.C.
AY253255    ATCATTCT...TTGCTTGTTA.C.
AY253256    ACGATTCT...TTATTTGTTA.C.
AY253257    ATGATTCT...TTATTTGTTA.C.
AY253258    TTGATTCT...TTATTTGTTA.C.
AY253259    ATGGTTCT...TTATTTGTTA.C.
AY253260    ACGATTCT...TTATTTGTTA.C.
AY253262    AGGTTTCT...TTTTTTGTTA.C.
AY253263    AGGTTTCT...TTATTTGTTA.C.
```

Figure 1.17 Alignment of ERF, ANT-AP2R1 & -AP2R2 and AP2R1 & R2

Alignments between AtERF1, ANT-AP2R1 and ANT-AP2R2; AP2R1 vs. ANT-AP2R1; AP2R2 vs. ANT-AP2R2; AP2R1 vs. AP2R2 and AP2- vs. ANT-linker regions. Known mutations of *AP2* in R1 and R2 are indicated. White arrowheads indicate single nucleotide mutations whereas black arrowheads indicate deletions or T-DNA insertions. The mutation for *ap2-6* (gray arrowhead) is due to a single G·C to A·T mutation at position 1342 that causes a mis-splicing event that causes a frameshift mutation of several amino acids before resulting in a premature stop codon (Wakem and Kohalmi, 2003). This mutation altered the RAYD element and truncated the AP2 protein after that element.

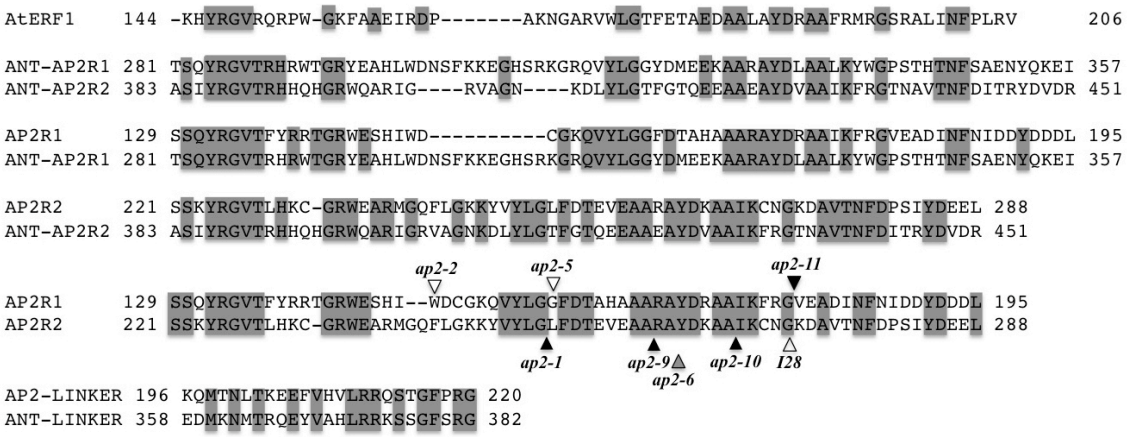

AtERF1 144 -KHYRGVRQRPW-GKFAAEIRDP-------AKNGARVWLGTFETAEDAALAYDRAAFRMRGSRALINFPLRV 206
ANT-AP2R1 281 TSQYRGVTRHRWTGRYEAHLWDNSFKKEGHSRKGRQVYLGGYDMEEKAARAYDLAALKYWGPSTHTNFSAENYQKEI 357
ANT-AP2R2 383 ASIYRGVTRHHQHGRWQARIG----RVAGN----KDLYLGTFGTQEEAAEAYDVAAIKFRGTNAVTNFDITRYDVDR 451
AP2R1 129 SSQYRGVTFYRRTGRWESHIWD----------CGKQVYLGGFDTAHAAARAYDRAAIKFRGVEADINFNIDDYDDDL 195
ANT-AP2R1 281 TSQYRGVTRHRWTGRYEAHLWDNSFKKEGHSRKGRQVYLGGYDMEEKAARAYDLAALKYWGPSTHTNFSAENYQKEI 357
AP2R2 221 SSKYRGVTLHKC-GRWEARMGQFLGKKYVYLGLFDTEVEAARAYDKAAIKCNGKDAVTNFDPSIYDEEL 288
ANT-AP2R2 383 ASIYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFGTQEEAAEAYDVAAIKFRGTNAVTNFDITRYDVDR 451
ap2-2
ap2-5
ap2-11
AP2R1 129 SSQYRGVTFYRRTGRWESHI--WDCGKQVYLGGFDTAHAAARAYDRAAIKFRGVEADINFNIDDYDDDL 195
AP2R2 221 SSKYRGVTLHKC-GRWEARMGQFLGKKYVYLGLFDTEVEAARAYDKAAIKCNGKDAVTNFDPSIYDEEL 288
ap2-1
ap2-9
ap2-6
ap2-10
I28
AP2-LINKER 196 KQMTNLTKEEFVHVLRRQSTGFPRG 220
ANT-LINKER 358 EDMKNMTRQEYVAHLRRKSSGFSRG 382

Table 1.1 Oligonucleotide Sequences

Plasmid Construction

Name	Oligonucleotide Sequence	Application
AP2R1-F	5'-CGGGATCCCCAAGATCAAGAAGTTCTC-3'	Cloning of AP2R1
AP2R1-R	5'TCGGAATTCCGAGTCATCTGTTTCAAGTCATCATCAT3'	Cloning of AP2R1
AP2R2-F	5'-CGGGATCCCGAGGAAGTTCGAAGTATAGAGGTGT-3'	Cloning of AP2R2
AP2R2-R	5'-TCGGAATTCCGGGCATTGAGTTCCTCATCGTAAATAC-3'	Cloning of AP2R2
AP2-F	5'-CACCTTAGGCCCGACCTATCGTC-3'	Cloning of *AP2* into pENTR
AP2-R	5'-AGAAGGTCTCATGAGAGGAGGTTGG-3'	Cloning of *AP2* into pENTR
AP2 full length-F	5'-TACTTCCAATCCAATGCGATGTGGGATCTAA-3'	Cloning of AP2 cDNA into MBP vector
AP2 full length-R	5'-TTATCCACTTCCAATGCGCTAAGAAGGTCTCATG-3'	Cloning of AP2 cDNA into MBP vector
AGKB31-F	5'-TACGGATCCAATAGTTTAAGAGTTTGGT-3'	Cloning of KB31
AGKB31-R	5'-TCTAAGCTTAATGTACGCTTAAATCTGC-3'	Cloning of KB31
AGKB31ΔA-F	5'-CATAAAATGATTCTTTATCCACTACAACTTTAAGGAAATA-3'	Mutagenesis of the A site in KB31
AGKB31ΔA-R	5'-ATATTTCCTTAAAGTTGTAGTGGATAAAGAATCATTTATG-3'	Mutagenesis of the A site in KB31

AGKB31ΔB-F	5'-GTTTTCTTATATTACTTGTACTTTTTCTTCACATCATGTG-3'	Mutagenesis of the B site in KB31
AGKB31ΔB-R	5'-CACATGATGTGAAGAAAAAGTACAAGTAATATA AGAAAAC-3'	Mutagenesis of the B site in KB31

** AP2R1-F and AP2R2-R were used for the cloning of AP2R1R2

Genotyping

Name	**Oligonucleotide Sequence**	Application
AP2p29-F	5'-CCAAGGAAGAGTTCGTACACGTAC-3'	Genotyping of AP2m3-GR transgene
GR junction-R	5'-TTCTCCATGCTGAATCTGG-3'	Genotyping of AP2m3-GR transgene
35Sseq-R	5'-CTCTCCAATGAAATGAACTTGAAG-3'	Genotyping transgene
AP2intron-F	5'-CGTTAATCGATCGTACTTTAGA-3'	Genotyping *ap2-2*
AP2-2-R	5'-GATATCCGCTTCTACTCCACGG-3'	Genotyping *ap2-2*

RT-PCR and Real-time PCR

Name	**Oligonucleotide Sequence**	**Application**
UBQ5-N	5'-GGTGCTAAGAAGAGGAAGAAT-3'	Real-time RT-PCR
UBQ5-C	5'-CTCCTTCTTTCTGGTAAAGCT-3'	Real-time RT-PCR
GUSp1-F	5'-GTCACTCATTACGGCAAAGT-3'	Real-time RT-PCR
GUSp1-R	5'-CCAGTTCAGTTCTTGTTCA-3'	Real-time RT-PCR
GUSp3-F	5'-ATCTCTTTGATGTGCTGTGC-3'	Real-time RT-PCR
GUSp3-R	5'-ACACTGATACTCTTCACTCCA-3'	Real-time RT-PCR

eIF4A1-F	5'-TCTTGGTGAAGCGTGATGAG-3'	Real-time PCR in ChIP experiment
eIF4A1-R	5'-GCTGAGTTGGGAGATCGAAG-3'	Real-time PCR in ChIP experiment
AG2617-F	5'-TCGATAAATTTAAGCTTTCAGAGG-3'	Real-time PCR in ChIP experiment
AGp7-R	5'-CAAACATCGGTTCTTTTGAGTC-3'	Real-time PCR in ChIP experiment
AG2617-F	5'-TCGATAAATTTAAGCTTTCAGAGG-3'	Real-time PCR in ChIP experiment
AG2190Levi-F	5'-AGAGTTTGGTCTGCCTTCTACGATC-3'	Real-time PCR in ChIP experiment
AG2418Levi-R	5'-GTCCGAGTAACATCACAACGTTC-3'	Real-time PCR in ChIP experiment
AG-F	5'-TTCTTTGTGATGCTGAAGTC-3'	Real-time RT-PCR; RT-PCR
AG-R	5'-ATGCTGATTATTTGTTGACG-3'	Real-time RT-PCR; RT-PCR

SAAB and EMSA

Name	**Oligonucleotide Sequence**	**Application**
SAAB library	5'-GAGAGGATCCAGTCAGCATG$(N)_{20}$CTCAGCCTCGAGAATT CCAA-3'	SAAB assay; N= A,T,C or G
AGp1-F	5'-GGTGTTGATAGATTTATGCAATTTCTC-3'	EMSA
AGp1-R	5'-GTACTAAAAATCTCACTTTCCTCTCAGT-3'	EMSA
AGp2-F	5'-GGATCCAATAGTTTTAAGAGTTTGGTCTG-3'	EMSA
AGp2-R	5'-TTACAATGACATATTATAGACTTTGATGTCTG-3'	EMSA
AGp3-F	5'-ACGTATTTGTGTATATATATCTATGTACAAGTAC-3'	EMSA
AGp3-R	5'-GTCCGAGTAACATCACAACGTTCCATACTTT-3'	EMSA

AGp4-F	5'- AAGTCATTTAGTTACATCCATCACGTT-3'	EMSA
AGp4-R	5'-CATGTGTCAACAACCCATTAACACATTGGG-3'	EMSA
AGp6-F	5'-CATATTGTTTCATAAAATGATTCTT-3'	EMSA
AGp7-R	5'-CAAACATCGGTTCTTTTGAGTC-3'	EMSA
αβ-F	5'-GGGGGAAAACAAACTTTTTGTTCAA-3'	EMSA
αβ-R	5'-GGGGGTTGAACAAAAAGTTTGTTTT-3'	EMSA
Δα-F	5'-GGGGGAAAGGTGACTTTTTGTTCAA-3'	EMSA
Δα-R	5'-GGGGGTTGAACAAAAAGTCACCTTT-3'	EMSA
Δβ-F	5'-GGGGGAAAACAAACTTTCCACTCAA-3'	EMSA
Δβ-R	5'-GGGGGTTGAGTGGAAAGTTTGTTTT-3'	EMSA
Δαβ-F	5'-GGGGGAAAGGTGACTTTCCACTCAA-3'	EMSA
Δαβ-R	5'-GGGGGTTGAGTGGAAAGTCACCTTT-3'	EMSA
GCCbox-F	5'-GGGGGTAAGAGCCGCCATGATCGATT-3'	EMSA
GCCbox-R	5'-GGGGGAATCGATCATGGCGGCTCTTA-3'	EMSA
GC-1-F	5'-GGGGGGCACGGCAGGTCATCGTACCA-3'	EMSA
GC-1-R	5'-GGGGGGTGGTACGATGACCTGCCGAG-3'	EMSA
GC-2-F	5'-GGGGGAAAGGTGACTTTGGGCCCAA-3'	EMSA
GC-2-R	5'-GGGGGTTGGGCCCAAAGTCACCTTT-3'	EMSA
1A-F	5'-GGGGGAAAGGTGACTTATTGTTCAA-3'	EMSA-site analysis
1A–R	5'-GGGGGTTGAACAATAAGTCACCTTT-3'	EMSA-site analysis
1C-F	5'-GGGGGAAAGGTGACTTCTTGTTCAA-3'	EMSA-site analysis
1C-R	5'-GGGGGTTGAACAAGAAGTCACCTTT-3'	EMSA-site analysis
1G-F	5'-GGGGGAAAGGTGACTTGTTGTTCAA-3'	EMSA-site analysis
1G-R	5'-GGGGGTTGAACAACAAGTCACCTTT-3'	EMSA-site analysis
2A-F	5'-GGGGGAAAGGTGACTTTATGTTCAA-3'	EMSA-site analysis
2A-R	5'-GGGGGTTGAACATAAAGTCACCTTT-3'	EMSA-site analysis
2C-F	5'-GGGGGAAAGGTGACTTTCTGTTCAA-3'	EMSA-site analysis
2C-R	5'-GGGGGTTGAACAGAAAGTCACCTTT-3'	EMSA-site analysis
2G-F	5'-GGGGGAAAGGTGACTTTGTGTTCAA-3'	EMSA-site

		analysis
2G-R	5'-GGGGGTTGAACACAAAGTCACCTTT-3'	EMSA-site analysis
3A-F	5'-GGGGGAAAGGTGACTTTTAGTTCAA-3'	EMSA-site analysis
3A-R	5'-GGGGGTTGAACTAAAAGTCACCTTT-3'	EMSA-site analysis
3C-F	5'-GGGGGAAAGGTGACTTTTCGTTCAA-3'	EMSA-site analysis
3C-R	5'-GGGGGTTGAACGAAAAGTCACCTTT-3'	EMSA-site analysis
3G-F	5'-GGGGGAAAGGTGACTTTTGGTTCAA-3'	EMSA-site analysis
3G-R	5'-GGGGGTTGAACCAAAAGTCACCTTT-3'	EMSA-site analysis
4A-F	5'-GGGGGAAAGGTGACTTTTTATTCAA-3'	EMSA-site analysis
4A-R	5'-GGGGGTTGAATAAAAAGTCACCTTT-3'	EMSA-site analysis
4T-F	5'-GGGGGAAAGGTGACTTTTTTTTCAA-3'	EMSA-site analysis
4T-R	5'-GGGGGTTGAAAAAAAAGTCACCTTT-3'	EMSA-site analysis
4C-F	5'-GGGGGAAAGGTGACTTTTTCTTCAA-3'	EMSA-site analysis
4C-R	5'-GGGGGTTGAAGAAAAAGTCACCTTT-3'	EMSA-site analysis
5A-F	5'-GGGGGAAAGGTGACTTTTTGATCAA-3'	EMSA-site analysis
5A-R	5'-GGGGGTTGATGAAAAAGTCACCTTT-3'	EMSA-site analysis
5C-F	5'-GGGGGAAAGGTGACTTTTTGCTCAA-3'	EMSA-site analysis
5C-R	5'-GGGGGTTGAGGAAAAAGTCACCTTT-3'	EMSA-site analysis
5G-F	5'-GGGGGAAAGGTGACTTTTTGGTCAA-3'	EMSA-site analysis
5G-R	5'-GGGGGTTGACGAAAAAGTCACCTTT-3'	EMSA-site analysis
6A-F	5'-GGGGGAAAGGTGACTTTTTGTACAA-3'	EMSA-site analysis

6A-R	5'-GGGGGTTGTAGAAAAAGTCACCTTT-3'	EMSA-site analysis
6C-F	5'-GGGGGAAAGGTGACTTTTTGTCCAA-3'	EMSA-site analysis
6C-R	5'-GGGGGTTGGAGAAAAAGTCACCTTT-3'	EMSA-site analysis
6G-F	5'-GGGGGAAAGGTGACTTTTTGTGCAA-3'	EMSA-site analysis
6G-R	5'-GGGGGTTGCAGAAAAAGTCACCTTT-3'	EMSA-site analysis

Table 1.2 Accession numbers of sequences of *AG* orthologs

Species	Family	GenBank Accession Number
Antirrhinum majus L.	*Veronicaceae*	AY935269
Lycopersicon esculentum L. cv. Microtom	*Solanaceae*	AY254705 (TAG1)
Arabidopsis thaliana Col	*Brassicaceae*	AL161549 (AT4G18960)
Arabidopsis arenosa	*Brassicaceae*	AY253237
Arabidopsis lyrata	*Brassicaceae*	AY253251
Arabis gunnisoniana Rollins	*Brassicaceae*	AY253244
Arabis pumila Jacq.	*Brassicaceae*	AY253243
Erysimum capitatum (Douglas ex. Hook.) E. L. Greene	*Brassicaceae*	AY253248
Cheiranthus cheiri L.	*Brassicaceae*	AY253258
Barbarea vulgaris R. Br.	*Brassicaceae*	AY253235
Nasturtium officinale R. Br.	*Brassicaceae*	AY253250
Thlaspi arvense L.	*Brassicaceae*	AY253254
Camelina sativa (L.) Crantz	*Brassicaceae*	AY253253
Capsella bursa-pastoris (L.) Medikus	*Brassicaceae*	AY253262
Capsella rubella Reut.	*Brassicaceae*	AY253263
Conringia orientalis (L.) Dumort	*Brassicaceae*	AY253252
Thysanocarpus sp. Hook	*Brassicaceae*	AY253255
Streptanthus insignis Jepson	*Brassicaceae*	AY253259
Guillenia flavescens (Hook.) E. Greene	*Brassicaceae*	AY253260
Lobularia maritima (L.) Desv.	*Brassicaceae*	AY253242
Alyssum saxatile L.	*Brassicaceae*	AY253249
Eruca sativa Mill.	*Brassicaceae*	AY253240
Brassica oleracea L. var. oleracea cv. A12	*Brassicaceae*	AY253241

Raphanus sativus L. cv. Cherry Bell	*Brassicaceae*	AY253245
Diplotaxis catholica (L.) DC.	*Brassicaceae*	AY253257
Cakile maritima Scop.	*Brassicaceae*	AY253256
Berteroa incana (L.) DC	*Brassicaceae*	AY253236
Draba corrugata S. Wats. car. Corrugata	*Brassicaceae*	AY253247
Lepidium africanum (burm. f.) DC	*Brassicaceae*	AY253238
Coronopus squamatus (Forsk.) Ashers.	*Brassicaceae*	AY253246
Lepidium phlebopetalum (F. Muell.) F. Muell	*Brassicaceae*	AY253239
Cucumis sativusL.	*Cucurbitaceae*	AY254704
Cucumis sativus L.	*Cucurbitaceae*	AY254702

CHAPTER 2

DNA TOPOISOMERASE Iα promotes DNA methylation in *Arabidopsis*

ABSTRACT

Cytosine methylation is an epigenetic modification by which a methyl group is added to the 5 position of the cytosine pyrimidine ring, oftentimes preventing transcription. Cytosine methylation is important to the success of an organism, as it prevents the expression of transposable elements and is involved in key developmental processes, such as genomic imprinting and X-chromosome inactivation. In the past decade, the role of small RNAs in Transcriptional Gene Silencing (TGS) has sparked intense interest and research, and through a collaborative effort in the field, scientists are making headway into understanding this phenomenon. Recent studies show that small RNAs guide cytosine methylation in order to repress the expression of transposable elements and heterochromatic regions through a process termed RNA-directed DNA Methylation (RdDM). However, the molecular mechanisms by which the small RNA machinery is recruited to various loci as well as the composition of the machinery are not fully understood. Here, using a forward chemical genomics approach, we have identified a novel component of RdDM, *DNA TOPOISOMERASE Iα* (*TOP1α*). We show that *TOPIα* affects genome-wide DNA methylation in two different contexts: CHH and CG. In terms of CHH methylation, as most are under RdDM regulation in *Arabidopsis*, we found that although *TOPIα* does not play a role in siRNA biogenesis, it does partake in

the repression of several siRNA target loci and preliminary results suggest that it may play a role in the production of Pol V-dependent scaffold transcripts. Further, genome-wide studies indicate that CHH methylation specific DMRs (Differentially Methylated Regions) between wild type and *topIα* come from transposable elements. In terms of CG methlyation, we found that the majority of affected loci are gene bodies, and gene body methylation is poorly understood both mechanistically and functionally. We found that longer genes with higher expression levels are preferentially affected. In addition, gene body methylation was more affected at the 3' than the 5' end. Taken together, these findings implicate a well-known anti-cancer target in DNA methylation; thereby potentially finding the missing link between cancer and DNA methylation, as during carcinogenesis, tumor suppressor genes are often methylated. Our findings transcend the plant model species and lay groundwork for future methylation studies in cancer biology. Furthermore, other recent studies have identified novel roles for *TOPIα* therefore, our findings in conjunction with others, unveil the intricate roles played by this gene previously thought to only be involved in DNA topology.

INTRODUCTION

Cytosine methylation is an epigenetic modification pivotal to the development and success of many different organisms. Here in Chapter Two, I discuss the different modes of DNA methylation, the role of small RNAs in DNA methylation, the identification of two compounds from a forward chemical genetics screen, Methotrexate (MTX) and Camptothecin (CPT), and the characterization of the target of CPT, *DNA TOPOISOMERASEIα* and its role in DNA methylation in *Arabidopsis*.

DNA methylation

Cytosine DNA methylation is the process by which a methyl group is placed on the five position of the pyrimidine cytosine. This mark in intergenic regions or at transposable elements (TEs) is often associated with transcriptional gene silencing (TGS). Cytosine DNA methylation (from now on, referred to only as DNA methylation) occurs in three different contexts: symmetric (CG or CpG in animals (the "p" stands for the phosphate that links the C and G nucleotides together) and CHG (H being any other nucleotide besides G) and asymmetric (CHH). In animals, DNA methylation occurs almost exclusively in the CpG context in somatic cells whereas CHH methylation is observed in embryonic stem cells. The animal genome contains 70-80% CpG methylation whereas the rest of the unmethylated CpG dinucleotides are located in clusters called CpG islands (Suzuki and Bird, 2008a). Interestingly, these islands are located near the promoters of genes, and their methylation status affects the expression of the genes. Thereby DNA methylation contributes to the success of the organism, as it

provides an avenue for facile gene regulation. In contrast to animals, DNA methylation occurs in plants in three sequence contexts: CG, CHG and CHH. The *Arabidopsis* genome contains 24% CG, 6.7% CHG, and 1.7% CHH methylation (Henderson and Jacobsen, 2007; Cokus, 2008); methylation usually occurs in transposons and other repetitive elements (Zhang, 2006). In addition, approximately 4,361 genes (20%-30% of all genes) exhibit gene body methylation, which occurs mainly in the CG context (Zhang, 2006; Vaughn et al., 2007; Zilberman et al., 2007; Cokus, 2008; Lister et al., 2008; Takuno and Gaut, 2012). In this section, I will discuss *de novo* methylation, methylation maintenance, and demethylation in animals and plants.

De novo methylation

In mammals, patterns of DNA methylation patterning are established during early embryogenesis around the time of implantation (Monk et al., 1987; Reik, 2007). *De novo* methylation or methylation initiation is governed by the activities of *DNA METHYLTRANSFERASE3A* and *B* (*DNMT3A* and *DNMT3B*, respectively) (Goll and Bestor, 2005). After implantation, further methylation that occurs at transposable elements and imprinted genes is controlled by Piwi-interacting RNAs (piRNAs) and *DNMT3A* and its paralog *DNA METHYLTRANSFERASE3-like* (DNMT3L) (Aravin et al., 2007a; Aravin et al., 2007b). First identified in *Drosophila* (Aravin, 2001), piRNAs are 25-30 nucleotide small RNAs that are bound by the Piwi-clade of ARGONAUTE (AGO) proteins and guide the cleavage of transposon transcripts (Aravin et al., 2007a; Aravin et al., 2007b). *AGOs* encode a family of highly conserved proteins that are pivotal in small

RNA-mediated TGS and post-transcriptional gene silencing (PTGS) (Hutvagner and Simard, 2008). In mammals, piRNAs are enriched at very early stages of mouse development and through PTGS, guide the cleavage of targets mRNAs at this stage. In addition, piRNAs also function in TGS in that loss of two *Piwi* clade genes, *Mili* and *Miwi2*, result in a decrease in DNA methylation and an increase in transposon expression at several loci (Aravin et al., 2007b). Moreover, it has been shown that piRNAs participate in *de novo* methylation as piRNA populations are enriched in transposon sequences at the stage of development in which *de novo* methylation occurs, and piRNAs are present in *dnmt3l* mutants, suggesting that piRNAs act upstream of *de novo* methylation (Aravin and Bourc'his, 2008; Kuramochi-Miyagawa, 2008).

De novo methylation of imprinted genes in mammals is controlled by the concerted effort of DNMT3A and DNMTL. DNMTL is a non-catalytic paralogue of *DNMT3A*. Biochemical analyses show that DNMTL interacts with unmethylated H3K4 tails through its cysteine-rich ATRX-DNMT3-DNMT3L (ADD) domain (Ooi, 2007; Otani, 2009). It has been proposed that this interaction leads to the recruitment of DNMT3A (with which DNMT3L also interacts), which in turn promotes DNA methylation. Several lines of evidence support this model, such as 1) oocyte-specific H3K4 demethylase is required for the establishment of DNA methylation, and 2) H3K4 dimethylation and trimethylation are anti-correlated with DNA methylation (Fournier, 2002; Delaval, 2007; Guenther et al., 2007; Okitsu and Hsieh, 2007; Ciccone and Chen, 2009).

In plants, small RNAs (specifically 24-nt small interfering RNAs (siRNAs)), also play a role in *de novo* methylation through a process called RNA-directed DNA Methylation (RdDM) (Wassenegger et al., 1994). RdDM requires the DICER and AGO proteins, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (the plant homolog of DNMT3), two plant specific RNA polymerases, Pol IV and Pol V (Pol II is also required at certain loci), chromatin remodeling factors, as well as other proteins (Law and Jacobsen, 2010). First, through an unknown mechanism, Pol IV is recruited to a certain locus and generates single-stranded RNA (ssRNA) from that locus. Then, the ssRNA is made into double-stranded RNA (dsRNA) by RNA-DEPENDENT RNA POLYMERASE2 (RDR2). The dsRNA is then further processed into 24-nt siRNAs by DCL3, and the siRNAs are methylated on their 3' ends by HUA ENHANCER 1 (HEN1). The siRNAs are loaded in AGO4, which associates with Pol V, setting off a cascade of events that is not yet fully understood. Pol V generates intergenic non-coding (IGN) or scaffold transcripts, and the transcription activity of Pol V is imperative to RdDM. The biogenesis of the scaffold transcripts does not depend on Pol IV, DCL3 or RDR2, thereby suggesting that the generation of these transcripts and their function are independent of siRNA biogenesis. These transcripts have been hypothesized to act as a scaffold for the recruitment of the silencing machinery through base pairing with the siRNA in the siRNA-AGO4 complex. Another factor, SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L), a transcription elongation factor, has been shown to interact with AGO4 via the GW/WG motif and proposed to be the bridge between the IGN transcripts and the AGO4-siRNA complex. Interestingly, NRPE1 (a subunit of Pol V) also binds AGO4 via

this motif. INVOLVED IN DE NOVO 2 (IDN2) is a downstream effector of RdDM. IDN2 contains an XS domain that can recognize dsRNAs with 5' overhangs. Researchers hypothesize that IDN2 may act upon the AGO-bound siRNA pairing with Pol V dependent transcripts, thereby acting as a signal to recruit DRM2 to establish DNA methylation, however, this theory has not been verified (Law and Jacobsen, 2010). How Pol IV and Pol V are recruited to certain loci is still unclear but several mutants have been identified that may provide a clue on the recruitment of these two polymerases. These mutants are: DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1) (Kanno, 2004) and CLASSY 1 (CLSY1) (Smith, 2007), both of which are putative chromatin-remodeling factors, and DMS3 (Kanno, 2008) an RdDM component with similarity to structural maintenance of chromosome (SMC) proteins. In addition to these proteins, which affect chromatin structure, another component that has been found to play a role in RdDM and co-purifies with Pol IV is the SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) (Law et al., 2011) SHH1 contains a cryptic homeodomain as well as a SAWADEE domain of unknown function (Law et al., 2011).

In addition to the activities of Pol IV and Pol V, Pol II has been to shown to also be involved in TGS at certain loci (Zheng et al., 2009). A weak allele of Pol II was identified and studies show that Pol II-dependent transcripts were required for TGS at certain loci, and it has been proposed that these non-coding transcripts also act as scaffolds for the recruitment of AGO4, Pol IV and Pol V. Further support for the role of Pol II in RdDM is the identification of DEFECTIVE IN MERISTEM SILENCING 4 (DMS4), which has been shown to interact with Pol II in yeast (He, 2009; Kanno, 2009).

Methylation Maintenance

After initial methylation, CG methylation in mammals is maintained by DNMT1. DNMT1 is able to restore hemi-methylated DNA to a fully methylated state during replication (Kim et al., 2009). DNMT interacts with UHRF1 (ubiquitin-like plant homeodomain and RING finger domain 1) and this interaction is required for the DNMT to associate with chromatin (Bostick, 2007; Sharif, 2007). Loss of *UHRF1* results in a drastic decrease in DNA methylation (Bostick, 2007). *LSH1* (Chromatin-remodeling factor lymphoid-specific helicase 1) is also another factor involved in DNA methylation, however, its molecular function is still unknown (Dennis et al., 2001; Huang, 2004).

The three types of cytosine methylation in plants (CG, CHG and CHH) are each maintained by a different set of genes. *MET1* (*DNA METHYLTRANSFERASE 1*), the Dnmt1 homolog, plays a central role in the maintenance of CG methylation (Vongs et al., 1993). Other players include the VARIANT IN METHYLATION (VIM) family of SRA proteins (Woo et al., 2007; Woo et al., 2008)and DDM1 (Vongs et al., 1993; Hirochika et al., 2000). *VIM 1-5* are the UHRF1 orthologs in plants and *VIM 1-3* have been shown to function redundantly in maintaining CG methylation in *Arabidopsis* (Woo et al., 2008). It is still nebulous as to how these genes maintain methylation on a molecular level. Interestingly, about one-third of genes have CG methylation in their coding region in *Arabidopsis* (this number is higher in mammals), which is maintained by MET1. CG methylation in gene bodies does not cause silencing, unlike methylation at transposons (Zilberman et al., 2007). In fact, genes harboring body methylation are moderately

actively expressed. The purpose of CG body methylation is still unclear, however, there are three hypotheses as to why CG body methylation occurs. **One**, body methylation suppresses expression from cryptic promoters within coding regions thus preventing leaky expression of genes, which is costly to the organism (Zilberman et al., 2007; Maunakea et al., 2010). **Two**, body methlyation enhances accurate splicing of primary transcripts (Lorincz et al., 2004; Luco et al., 2010). **Three**, body methylation has no functional significance and perhaps is just a by-product of transcription (Roudier et al., 2009; Teixeira et al., 2009). If the first two hypotheses are true, DNA methylation should be associated with essential genes because aberrant promoter expression or mis-splicing events would lead to detrimental effects. Support for the first two hypotheses include: body methylation and H3K36 methylation are pre-dominantly distributed at exons and alternatively spliced exons tend to have lower levels of methylation (Choi, 2002; Ball et al., 2009; Hodges et al., 2009; Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Feng et al., 2010). In addition, recent comparative bioinformatics analyses show that not only are body methylated genes longer in mean length (3.349 kb versus 1.595 kb), 55.7% of assessed knock-out mutants with body methylation have phenotypic effects as compared with 26.2% in non-body methylated genes (Takuno and Gaut, 2012). Body methylated genes were found to also evolve at a slower rate (Takuno and Gaut, 2012), thereby alluding to their importance as more important genes are highly regulated thus less prone to "evolution." However, support for the third hypothesis includes: only minor, but positive effects on levels of gene expression were observed in genes containing body methylation when the methylation is lost, and that body methylation can

be highly polymorphic among individuals (Zhang et al., 2006; Vaughn et al., 2007; Zilberman et al., 2007; Zhang et al., 2008; Zemach et al., 2010). There is more support for the first two hypotheses, however, without definitive mutants, we cannot exclude the importance of gene body methylation.

In *Arabidopsis,* the major player involved in maintaining CHG methylation is CMT3 (CHROMOMETHYLASE 3) (Chan et al., 2005). CHG methylation maintenance is reinforced by a loop involving DNA methylation and histone modifications (Johnson, 2007). In support of this, loss of *SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4* (*SUVH4*, *5* and *6* comprise this family), the histone methyltransferase responsible for H3K9 dimethylation, results in a decrease in DNA methylation (Jackson et al., 2002).

CHH methylation is normally maintained by constant *de novo* methylation by DRM2 and the RdDM pathway. However, at some loci, its maintenance requires both CMT3 and DRM2 (Cao, 2003). In addition, CHH methylation also requires SRA-domain containing proteins, such as the SUVH family for maintenance, however the mechanism by which these proteins affect CHH methylation is still unclear (Johnson et al., 2008).

Active DNA Demethylation

In mammals, genome-wide decreases in DNA methylation had only been observed in primordial germ cells and on the paternal zygotic genome soon after fertilization (Reik, 2007; Sasaki and Matsui, 2008). Recently however, putative active DNA demethylation pathways have been discovered. Early work in mammals have

identified factors, such as activation-induced cytosine deaminase (AID) and apolipoprotein B mRNA-editing enzyme 1 (APOBEC1), that *may* be involved in DNA demethylation (Morgan et al., 2004). In support of the hypothesis that these genes may be involved in demethylation, orthologs of these genes in zebrafish have been found to partake in DNA demethylation (Rai, 2008). In addition, recent studies have implicated hydroxylation, formylation, and carboxylation as intermediate steps in DNA demethylation (Wu and Zhang, 2011). Ten-eleven translocation 1-3 (Tet1-3) proteins are DNA hydroxylases that have been shown, using fully methylated or hemi-methylated DNA as substrates, to be able to hydroxylate 5mC to 5hmC (5-hydroxylmethylcytosine) and further oxidize it to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), sequentially (Tahiliani et al., 2009; Ficz et al., 2011; He et al., 2011; Ito et al., 2011; Pastor et al., 2011).

Tet proteins were identified in acute myeloid lymphoma (AML) as a fusion partner of the histone H3K4 methyltransferase MLL (mixed lineage leukemia) (Ono et al., 2002; Lorsbach et al., 2003). Tet proteins (1-3) all (except Tet 2) contain a CXXC domain, which has high affinity for clustered, unmethylated CpG islands; and a catalytic domain, which contains a cysteine rich region and a DSBH (double-stranded β helix) region, reminiscent of Fe(II)- and 2-oxoglutarate (2OG)-dependent dioxygenases (Wu and Zhang, 2011).

Three enzymatic pathways that these proteins act in have been proposed in DNA demethylation: **one**, the iterative 5mC oxidation reactions are followed by DNA

glycosylase activity/base excision repair (BER). Base excision repair refers to the mechanism by which a cell removes damaged DNA. This process initiated by DNA glycosylases and continued through two pathways, short- or long- patch, depending on the length of the lesion or damage (Fromme and Verdine, 2004). In this pathway, the conversion of 5mC to 5caC occurs through the iterative oxidations of 5mC. In support of this mode, thymine DNA glycosylases (TDG) have been shown to be able to excise 5fC and 5caC at CpG sites (He et al., 2011; Maiti and Drohat, 2011a; Maiti and Drohat, 2011b). In mice, depletion of TDG leads to embryonic lethality and increased levels of DNA methylation at promoters (Maiti and Drohat, 2011b). However, TDG has no activity for 5hmC excision, so other glycosylases may be required. The **second** possible pathway is that 5hmC deamination occurs first, followed by BER. In this mode, the 5hmC is first deaminated by AID/APOBEC to 5hmU, followed by a 5hmU:G mismatch that is repaired through the action of DNA glycosylases and BER. In support of this mode, TDG and SMUG1 (another DNA glycosylase) have been shown to exhibit excision activity against 5hmU:G in dsDNA and have no activity on 5hmC (Cortellino et al., 2011; He et al., 2011). Further, it has been shown that TDG can interact directly with AID (Cortellino et al., 2011). However, there has been some skepticism with regards to this mode of action as there is no direct biochemical evidence that deaminases exhibit robust activity for 5hmC and there is no accumulation of 5hmUs. In addition, although AID may contribute to the demethlyation of primordial germ cells (PGCs) *in vivo*, a large portion of demethylation still occurs in the absence of AID (Popp et al., 2010). Although several lines of evidence support the first two pathways, it is not feasible for whole

genome demethylation to occur in this manner as mismatch repair processes take time to complete, which puts the genome at risk for instability. The **third** possible pathway is that iterative 5mC oxidation reactions are followed by decarboxylation to generate 5caC molecules. Decarboxylation reactions that convert 5mC to C would only require two enzymes and negate the need for DNA strand breakage, however, no decarboxylases have been identified that could convert 5caC to C.

Tet1 exhibits strong preference for CpG islands, which do not exhibit DNA methylation and overlap with transcription start sites (TSS) (Wu et al., 2011a; Wu et al., 2011b; Xu et al., 2011; Williams et al., 2012). Tet1 acts to keep the area hypomethylated and a deficiency of Tet1 leads to increased levels of 5mC at normally Tet1-enriched regions (Wu et al., 2011b; Xu et al., 2011). Further, Tet proteins exhibit dual functionality, as it has been found that Tet1 and 5hmC are enriched at gene promoters that are associated with bivalent domains; these domains are generally associated with poised developmentally regulated genes, particularly lineage-specific transcription factor genes that are specifically controlled by PRC2 repressive complexes (Pastor et al., 2011). Moreover, it was found that Tet1 and 5hmC contribute to the maintenance of an undifferentiated state in mouse ES (embryonic stem) cells by facilitating PRC2-mediated repression of lineage specific genes (Wu et al., 2011b). Tet1 acts as a major modulator in orchestrating the balance between a pluripotent state and cellular differentiation initiation. In addition, Tet1 proteins and 5hmC have also been shown to control gene expression through the regulation of enhancer functions (Ficz et al., 2011; Pastor et al., 2011; Wu et al., 2011b; Williams et al., 2012). Furthermore, global conversion of 5mC

to 5hmC in the paternal pronucleus, followed by a replication-dependent passive loss of 5hmC during pre-implantation development (this latter mechanism is not quite clear) have also been discovered (Inoue and Zhang, 2011) . Thus, the discovery of these Tet proteins has further enhanced our understanding of DNA demethylation, as it has added another layer to the complexity of epigenetic regulation.

In *Arabidopsis*, a family of DNA glycosylases, whose founding members include *DEMETER* (*DME*) and *REPRESSOR OF SILENCING 1* (*ROS1*), governs active DNA demethylation (Choi, 2002; Gong, 2002). Glycosylases usually recognize and remove mutagenic sequences through BER, thereby ensuring genome integrity (Baute and Depicker, 2008). DME and ROS1 recognize and remove methylated cytosines from dsDNA oligonucleotides, independent of sequence context (Penterman, 2007). Although both are gylcosylases, they have different biological roles in DNA demethylation. DME plays a pivotal role in imprinting in endosperms; it has been shown to activate the maternal alleles of *MEDEA* (*MEA*), *FLOWERING WAGENINGEN* (*FWA*) and *FERTILIZATION INDEPENDENT SEED 2* (*FIS2*) (Gehring et al., 2009). However, recent studies have shown that *DME* also plays a role in genome-wide demethylation through global studies comparing endosperm and embryo tissues (Gehring et al., 2009). On the other hand, *ROS1*, *DEMETER-LIKE 2* and *3* (*DML2* and *DML3*) are expressed in the vegetative tissue and act at heterochromatic regions as well boundaries between euchromatin and heterochromatin, thereby providing the genome with more flexibility to adapt (Penterman et al., 2007). Although several major players have been identified, the molecular mechanism by which these glycosylases are targeted to specific loci is not

fully understood. A recent study has identified ZDP, a 3' phosphatase, whose functions, in mammals, include binding and repairing single stranded and double stranded breaks (SSBs and DSBs) (Whitehouse et al., 2001; Chappell et al., 2002; Petrucco et al., 2002; Martinez-Macias et al., 2012). In plants, only binding of ZDP to SSBs and DSB has been observed, but the repair mechanism has not been reported (Petrucco et al., 2002). *ZDP* gets rid of the 3' phosphate at the breaks, such as those made by ROS1, so that an unmethylated C could be added prior to re-ligation (Martinez-Macias et al., 2012). Loss of *ZDP* results in genome-wide hypermethylation similar to *rdd* (the triple mutant *ros1 dml2 dml3)* (Martinez-Macias et al., 2012).

Thus, DNA methylation and demethylation are active processes that contribute to the success of an organism by 1) suppressing the expression of possibly deleterious genes/transposons and 2) allowing the organism a tool to be more malleable to the ever-changing environment and selective pressures. This is quite apparent in the complexity of DNA methylation and de-methylation pathways in plants, which, being sessile organisms, have to be more adaptable in their means of survival. While key players in this pathway have been found, the molecular mechanism by which these players recognize the location or timing to initiate their function is still nebulous. Therefore, to better understand these processes, we set out to identify other genes that may function in transcriptional gene silencing, specifically in DNA methylation.

Screening for Players involved in TGS

Forward genetics has been used as a powerful tool to identify genes involved in a particular biological process. However, some of the caveats of conventional forward genetic screens have been the inability to identify mutants due to genetic redundancy, lethality and the length of time it takes to identify the gene responsible for the observed aberrant phenotype (although this aspect has recently been facilitated by deep sequencing availabilities).

The use of chemicals as a tool to identify new genes involved is not a new concept--the pharmacy industry has been screening small molecules to identify new drugs for decades. The use of small molecules has advantages over conventional genetic screens in that it can be used to 1) identify family members functioning in the same genetic pathway thereby circumventing genetic redundancy, 2) perturb active sites of a protein, and 3) further probe the function of a gene by reversibly altering the chemical dosage and time.

Although there are many advantages to using small molecules, a bottleneck, especially to those nascent to the field, has been target identification. There has been intense research on the modes of target identification, which I will briefly describe here. As for direct approaches, one could utilize Time of Flight (TOF) followed by mass spectrometry (MS) analysis. MS analysis has been the major mode of target identification following chemical screens. Basically, the "hit" chemical is immobilized with a biotin affinity tag and whole cell lysate is passed over the beads. Bound proteins are eluted and subjected to MS sequencing (Galat et al., 1992). However, the

disadvantage of this technique is that the immobilization may cause the compound to become inactive due to loss of chirality or unique scaffolding. Thus, to overcome these problems, several modifications have been made. The Schreiber group developed an Affi-gel tag to lower non-specific binding (Yang et al., 2007). In addition, CLICK chemistry was developed in which once an alkyne-derivatized molecule is covalently bound to its target protein, a tag (fluorescein or rhodamine azide) gets attached. This allows for minimal structural modifications so the small molecule does not lose activity. This method has been used to identify the target of origamicin, an inhibitor of Hepatitis C Virus replication (Rakic et al., 2006). Another method that one could use is photo cross-linking, which was also developed to overcome loss of activity. Aryldiazirine groups are covalently attached to a solid support and upon UV irradiation, are transformed into highly reactive carbenes that are able to bind proximal compounds irreversibly. Bound proteins were then purified and analyzed by MALDI-TOF (Kanoh et al., 2003). FG beads have also been used to overcome non-specific binding and stability of matrices (Sakamoto et al., 2009). Another direct approach that does not use proteomics is to use conventional mutagenesis followed by map-based positional cloning approaches (identifying the target by looking for plants that are resistant to the activity of the chemical). This approach, although laborious, has been used to identify a family of ABA receptors in *Arabidopsis* (Park et al., 2009).

These methods are considered as direct approaches, however, indirect approaches could also be used to identify targets. For instance, drug affinity responsive target stability (DARTS) works on the premise that binding of the target to the drug will

stabilize the target, therefore an SDS-GEL could be run to identify bands protected from thermolysin digestion and analyzed by MS. The target of resveratrol was found using this method (Wood et al., 2004). Other indirect approaches include connectivity map databases (Lamb et al., 2006), proteomic profiling (Muroi et al., 2010), high content screening/ morphological profiling (Li et al., 1994), metabolomics profiling (Kitagawa et al., 2010), and chemical genetic interactions, which also include synthetic lethal viability tests in yeast (Parsons et al., 2004; Parsons et al., 2006). Indirect approaches use the working premise that one could infer target by comparing many datasets of related biological processes.

Several research groups have successfully used chemical genetics to identify new players and/or gene families in various signaling pathways in *Arabidopsis* (Robert et al., 2008; Zabotina et al., 2008; De Rybel et al., 2009; Park et al., 2009; Zhang et al., 2012). We sought to gain a better understanding of factors involved in TGS, specifically DNA methylation, by performing a chemical genetics screen using a transcriptionally silent, methylated, luciferase-based reporter line (*LUCL*). We performed a chemical screen with 24,970 compounds with which we obtained two confirmed hits, Methotrexate (MTX) and Camptothecin (CPT). They were both from a screen of the Library of Active Compounds (LATCA) in *Arabidopsis* (http://cutlerlab.blogspot.com/2008/05/latca.htm).

MTX is an anti-metabolite and anti-folate drug. It is commonly used as an anti-cancer compound at high doses, but at low doses, is used against autoimmune diseases including rheumatoid arthritis and lupus, (Cronstein, 2005b; Cronstein, 2005a). At high doses, MTX allosterically inhibits dihydrofolate reductase (DHFR), an enzyme that

participates in tetrahydrofolate synthesis (Rajagopalan et al., 2002). Tetrahydrofolate is essential for the biosynthesis of purines, thymidylate and several amino acids. In addition, MTX inhibits cytosine methylation through an indirect mechanism. DNA methyltransferases use S-adenosylmethionine (SAM) as a donor of methyl (to be placed at the five position). The generation of SAM is mainly modulated by folic acid. Folic acid is metabolized to dihydrofolate (DHF) and then tetrahydrofolate (THF); the latter step is inhibited by MTX. THF is then converted to methyl-THF by vitamin B12, which then catalyzes the conversion of homocysteine to methionine (Fig. 4AA). Methionine is further metabolized to S-adenosylmethionine, the substrate for Dnmt (Huber et al., 2007). Deficiencies in the enzymes involved in these processes result in hypomethylation of DNA and genomic instability (Huber et al., 2007). MTX led to the reduction in DNA methylation at *LUCL* probably by reducing the levels of SAM. Since the role of MTX in DNA methylation has been studied, we focused our on attention on CPT.

CPT is a natural quinoline alkaloid first discovered in 1966 in a screen for natural products that had anti-cancer properties (Wall, 1966). CPT is derived from the bark and stem of *Camptotheca acuminata*, a native tree of China. Unfortunately, CPT has very low solubility and high drug adverse reactions, but anti-cancer properties were effective, so structure and activity (SAR) studies were performed in order to identify a better compound for drug therapy (Zunino et al., 2002). Today, two CPT analogues are used in anti-cancer therapy: Topotecan and Irinotecan (Creemers et al., 1994), both of which are over a hundred times more potent than CPT.

CPT was found to target the action of DNA Topoisomerase I (Jaxel et al., 1991; Levin et al., 1993). DNA topoisomerase I functions to maintain DNA topology by generating nicks in the DNA strand during replication to alleviate torsional stress caused by its double helical nature (Wang and Droge, 1996). Usually, TopI forms a ternary complex with the DNA duplex, makes a single nick on the 5' strand and momentarily attaches to the cleaved phosphate via a covalent bond while the free strand is able to swivel around the intact strand. After, a nucleophillic attack of the 5'-OH of the free strand displaces the phosphotyrosine bond and the enzyme is displaced so that religation of the two DNA ends may occur. However, CPT stabilizes the ternary complex by positioning itself between the 3'phospho-tyrosine covalent intermediate and the free 5'-OH end, thus increasing the distance of the 5' and 3' termini thereby preventing re-ligation. Cell death probably occurs after the collision of the TopI cleavage complex with moving replication forks (Jaxel et al., 1991; Staker et al., 2002; Staker et al., 2005). This finding leads us to question why TopI may be found in our genetic screen, as it would be counterintuitive that a chemical that induces cell death may cause the activation of transcription of our normally methylated reporter line. In this next section, I discuss the structure and functions of the different topoisomerases and their diverse roles.

DNA Topoisomerases

As organisms began to evolve from single-stranded RNA forms to large double-helical DNA strands as carriers of genetic information, DNA topology during replication presents a huge problem. Due to its helical nature, when the replication forks move along

the DNA strand, positive supercoils are generated in front of the fork while negative supercoils are generated behind it. This excessive supercoiling could have deleterious effects on the organism. DNA topoisomerases evolved to solve this problem by generating nicks in the DNA strand thereby allowing the DNA to unwind to relieve the torsional stress. All topoisomerases contain a nucleophillic tyrosine, which is used to promote strand scission. There are two types of DNA topoisomerases that span across three kingdoms, Archaea, Bacteria, and Eukarya, Type I and Type II. Odd number topoisomerases belong to the Type I family, whereas even number topoisomerases belong to the Type II family. Topoisomerases are classified by the number of cuts that they generate during the cleavage process. Type I enzymes produce one nick and do not utilize ATP during the cleavage process, whereas Type II enzymes generate two nicks and the cleavage/re-ligation process utilizes ATP hydrolysis (Forterre et al., 2007).

Type I DNA Topoisomerases

Type I topoisomerases are also further divided into three classes: A, B and C. They are differentiated by their mode of cleavage and/or function. Type IA topoisomerases are widespread throughout all three kingdoms: Eukarya, Archaea and Bacteria. Type IA topoisomerases have three distinct domains: the cleavage/strand passage domain, the Zn(II) binding domain and the C-terminal domain. The Zn(II) binding domain is required for the strand passage activity (Tse-Dinh, 1991). The C-domain is dispensable for activity *in vitro*, however, it is highly basic and contributes to substrate binding (Beran-Steed and Tse-Dinh, 1989; Zhu et al., 1995). Like the B and C

classes of Type I topoisomerases, they generate one cut but cleavage of the DNA strand is accompanied by the covalent attachment of the enzyme's tyrosine active site to one DNA strand through a 5' phosphodiester bond. This action subsequently releases a free hydroxylated strand. Type IA topoisomerases are considered single-stranded-DNA "strand passage" enzymes because after cleavage occurs, the DNA strand is physically opened and the cut strand is navigated through the opening. In essence, the enzyme has a toroidal structure after binding, which acts like a padlock so that the cut strand could be guided through (Tse et al., 1980). After, the hydroxyl end of the free strand (3'-OH) attacks the phosphotyrosine bond, thereby restoring the phosphodiester bond so the enzyme is released. In addition, all of these topoisomerases require magnesium (Mg^{2+}) for the DNA relaxation reaction. Their substrates include plasmids that contain negative but not positive supercoils. Further, the relaxation of negative supercoils does not go to completion. In addition to having the ability to relax negative supercoils, Type IA topoisomerases all function as monomers and can catalyze the knotting, unknotting and interlinking of single stranded DNA as well as DNA duplexes (Champoux, 2001; Vos et al., 2011). Interestingly, the functions of Type IA topoisomerases vary depending on the species. For example, Top1 in bacteria can relax native plasmids directly from the bacteria, whereas Top1s in eukaryotes require a hypernegatively supercoiled plasmid DNA substrate (Hanai et al., 1996; Hiasa and Marians, 1996; Wilson et al., 2000). The diversity of topoisomerases in different species highlights the evolutionary pressures and events placed upon each individual species.

Type IB topoisomerases are structurally and mechanistically different from Type IA topoisomerases. These topoisomerases interestingly share more structural and functional properties with tyrosine recombinases than other types of topoisomerases. Actually, they share no sequence or structural homology with the topoisomerases, even those from Type IA (Stewart et al., 1998). The most well characterized Type IB topoisomerase is human TopI; however, Type IB topoisomerases are ubiquitous throughout eukaryotes and could also be found in some viruses and bacteria (Krogh and Shuman, 2002).

Type IB topoisomerase proteins have four distinct domains: N-terminal, core, linker, and C-terminal domains. The N-terminal domain is dispensable for relaxation activity *in vitro*, poorly conserved, highly charged and unstructured. However, this domain does contain four nuclear localization signals and sites for protein-protein interactions (Bharti et al., 1996; Stewart et al., 1996a; Stewart et al., 1996b). The core domain is highly conserved and contains all the catalytic residues except for the active tyrosine (Redinbo et al., 1998). The linker domain is poorly conserved and prone to protease activity and has been found to be dispensable *in vitro* (Redinbo et al., 1998). The C-terminal domain is conserved and contains the active tyrosine (Redinbo et al., 1998).

Top1 in humans has been shown to prefer supercoiled over relaxed plasmid DNA substrates (Camilloni et al., 1989). In addition, it has also been shown to associate with DNA nodes, the region of the DNA duplex where the two duplexes cross (Zechiedrich

and Osheroff, 1990). Type IB topoisomerases have a preference for a combination of nucleotides on the scissile strand—5'-(A/T)(G/C)(A/T)T-3', with the enzyme forming a covalent attachment to the 3' end. Sometimes a C could be found at that end but it prefers a T (Been et al., 1984; Bonven et al., 1985; Tanizawa et al., 1993). In fact, the -1 position (whether it is a T or C) dictates which amino acids of the enzyme will bind the DNA strand in order to stabilize it (after cleavage).

Type IB topoisomerases cleave DNA strands by a nucleophillic attack of the O4 oxygen of the active site tyrosine on the scissile phosphate. This action breaks the DNA strand to generate a phosphodiester link between tyrosine and the 3' phosphate, releasing a 5'-hydroxyl. The alignment and subsequent binding of the enzyme with the DNA duplex depend on the -1 position of the cleavage site. If the -1 position is a T, then the three amino acids, R, R and H, bind the DNA duplex in order to stabilize the TopI cleavage complex. However, if the -1 position is a C, then amino acids R, H, and K are used to stabilize the ternary complex until the passage strand is able to rotate, with respect to the other strand, around the intact phosphodiester bond (Redinbo et al., 1998; Stewart et al., 1998; Champoux, 2001). The mechanism by the free strand rotates or swivels around the intact strand gives Type IB enzymes their name as DNA "swivelases" (Koster et al., 2005). The swivel rotation passage is controlled by friction between the DNA strand and the enzyme, which also helps the re-ligation of the 5'-hydroxyl DNA strand to it 3' end (Champoux and Dulbecco, 1972; Koster et al., 2005). Like Type IA topoisomerases, Type IB topoisomerases do not utilize ATP hydrolysis for the re-ligation reaction and function as monomers.

Type IB topoisomerases have the ability to relax both positive and negative supercoiled DNA. This allows them to play a major role in both transcription and DNA replication by relaxing the positive supercoils that accumulate in front of the moving polymerases (the exception is that this role is fulfilled by Top IIA in bacteria) (Kim and Wang, 1989). There are two different Type IB topoisomerases present in vertebrates, one localized in the nucleus and the other in the mitochondria (Zhang et al., 2004). Interestingly, the nuclear Top I cannot replace the Top I in the mitochondria.

There is only one member in the Type IC sub-class: Top1 from the archael genus *Methanopyrus*, which was initially named Topo V (Forterre, 2006). It acts similarly to Type IB topoisomerases in that it can relax both positive and negative supercoils through a swiveling mechanism (Slesarev et al., 1993; Taneja et al., 2006). However, it was moved into its own class as it is structurally different from Type IB topoisomerases and has a different evolutionary lineage (Slesarev et al., 1993; Taneja et al., 2006). The N-terminal region of this protein is responsible for the Topo I activity, but at the C-terminal domain, Topo V also has an apurinic/apyrimidic (AP) site-processing domain (Belova et al., 2002), which is an essential component of BER. In addition, unlike the other Type I topoisomerases, Topo V does not function as a monomer (Krah et al., 1996).

Type II topoisomerases differ from Type I topoisomerases in that they break both strands of the DNA duplex and use ATP hydrolysis to power passage of another intact duplex (Gellert et al., 1976; Brown and Cozzarelli, 1979; Goto and Wang, 1982). They can relax both positive and negative supercoils, as well as disentangle long and

intertwined chromosomes and DNA catenenes (Hsieh and Brutlag, 1980; Mizuuchi et al., 1980). Type IIA topoisomerases are found in all cellular organisms and some viruses (Forterre et al., 2007), whereas Type IIB topoisomerases are found in archea, plants, some bacteria, protists and algae(Bergerat et al., 1997; Malik et al., 2007). Type IIA and IIB catalysis mechanisms are similar but the division occurred because topo VI (the only member who is categorized as a Type IIB enzyme) is structurally different from other Type IIA topoisomerases.

Diverse Roles of Topoisomerases

In addition to its role in DNA topology, DNA topoisomerases have other roles as well. For example, they function in chromosome compaction by working with structural maintenance of chromosome (SMC) proteins (Maeshima and Laemmli, 2003; Tadesse et al., 2005). They also are required for chromatin assembly during mitosis in yeast (Salceda et al., 2006). Further, topoisomerases is linked to RNA polymerase II activity (Capranico et al., 2007; Durand-Dubief et al., 2010; Durand-Dubief et al., 2011). In terms of Top I's relationship with Pol II, it has been proposed that Top1 1) regulates transcription-coupled processes such as recombination, and 2) mediates the chromatin binding sites of Pol II by regulating transcription pausing at promoter-proximal sites (Khobta et al., 2006), and 3) plays a role in mRNA spicing as it has been shown that Top1 depletion results in impaired exonic enhancer-dependent splicing (Soret et al., 2003). Further, in plants, Top VI, a Type IIB topoisomerase, is required for

endoreduplication, chromatin condensation and transcriptional silencing (Sugimoto-Shirasu et al., 2002; Breuer et al., 2007; Kirik et al., 2007).

In addition, inactivation of *topo IB* in *S. cerevisiae* leads to histone acetylation and methylation events that increase the transcription of a telomere proximal gene (Lotito et al., 2008). Also, Top IB in *S. pombe* has been found to influence nucleosome assembly and disassembly at certain promoter regions (Durand-Dubief et al., 2010). In yeast, Top I was found to be responsible for a 2-bp deletion in transcription-associated mutagenesis (TAM). During high levels of transcription, the Top 1 cleavage complex gets "stuck" and its removal causes a 2-nt gap, then slippage of the opposite strand (this generates a loop in the opposite strand which is later removed by the mismatch repair machinery) results in the conversion of a gap to a nick, ligase reseals the nick and the top strand continues on to replicate, thus a 2-bp deletion is generated and perpetuated in future generations (Lippert et al., 2011; Takahashi et al., 2011).

A chemical genetics screen implicated the role of Top I in epigenetic silencing of genes in mammals. In neurons, the *UBE3A* (Ubiquitin Protein Ligase E3A) gene is expressed from the maternal allele; the paternal allele is intact but epigenetically silenced. Deletion of, or mutations in, the maternal allele of this gene lead to the Angelman syndrome. A chemical genetics screen was conducted to identify compounds that revert the silencing of the paternal allele. From 2,306 compounds, 12 Top I and 9 Top II inhibitors were identified. The Top I inhibitors include CPT analogs, irinotecan and topotecan (Huang et al., 2012). This result provides support that Top I enzymes play a

role in epigenetic silencing of genes. In addition, another chemical genetics screen in rice identified etoposide, a Top II inhibitor, which causes changes in DNA methylation in certain rice ecotypes and the mobilization of mPing (a transposable element in rice) after the 2nd or 3rd generation of chemical application (Yang et al., 2012).

In *Arabidopsis*, several DNA topoisomerases from all types have been identified (Kieber et al., 1992; Corbett and Berger, 2004; Makarevitch and Somers, 2006; Hartung et al., 2007). Type IA topoisomerases are important for suppressing somatic crossovers and essential for the resolution of meiotic recombination intermediates (Hartung et al., 2007). Mutants in type IA enzymes exhibit severely impaired growth including deformed cotyledons and lack of roots (Hartung et al., 2007). Mutants in Type IIA and Type IIB topoisomerases have severe growth defects. Using the yeast Top I cDNA, researchers first found the Top I homolog in *Arabidopsis* through hybridization experiments (Kieber et al., 1992). In *Arabidopsis*, there are two Type IB topoisomerases that are tandemly arrayed with each other on the fifth chromosome, *TOP Iα* and *TOP Iβ*. *topIα* plants have gross morphological defects such as fasciated stems, and altered phyllotaxy and plant architecture (Takahashi et al., 2002). *top Iβ* plants are phenotypically normal, however RNAi lines that down-regulate both *TOP Iα* and *TOP Iβ* are lethal, thus these two genes are functionally redundant (Takahashi et al., 2002). Recent studies have shown that *TOP Iα* regulates the stem cell niche in *Arabidopsis* the shoot apical meristem by acting synergistically with chromatin remodeling factors (Graf et al., 2010). In addition, it was shown to be required to maintain the expression state of two epigenetically–regulated genes: *AGAMOUS* and *BREVIPEDICELLUS*, both of

which are PRC2 targets; however it did not affect *TSI*, *MULE*, and *CACTA-like* repeats (Graf et al., 2010).

DNA topoisomerases comprise a large family of diverse and structurally distinct enzymes. Initially thought only to be involved in maintaining DNA topology, recent studies have shown that these enzymes partake in many different biological processes. Small molecule screens have been fruitful in identifying novel targets in canonical signaling and developmental pathways. Here, we show that the *TOP Iα* inhibitor CPT is able to activate a transcriptionally silenced reporter, *LUCL*, by releasing DNA methylation of the reporter transgene. In addition, we show that *TOP1α* promotes DNA methylation at both CHH and CG contexts not only at the reporter transgene but also on a genome-wide scale. Interestingly, CHH differentially methylated regions (DMRs) are enriched in TEs while CG DMRs are mostly found in gene bodies and correlate with gene length. Further, siRNA accumulation is not affected, however, siRNA target loci are de-repressed in the *top Iα* mutant, thereby suggesting a role for *TOP Iα* in RdDM downstream of siRNA biogenesis. Moreover, preliminary studies suggest that *TOP Iα* may contribute to the production of Pol V-dependent transcripts. Taken together, these results provide a new function for a previous well-characterized enzyme. As this enzyme is the target for popular anti-cancer compounds, this study transcends the plant model organism and provides a novel molecular basis for carcinogenesis.

RESULTS

LUCL is silenced through DNA methylation

A forward genetics screen is a powerful tool used to identify genes in a particular biological process. Using small molecules as the chemical probe, we sought to identify new genes involved in cytosine methylation through a *LUCIFERASE* (*LUC*)- based reporter transgene. We have established two independent *Arabidopsis* lines containing a luciferase transgene driven by dual *35S* reporters (Fig. 1A), *LUCH* (Won, 2012) and *LUCL* (*LUC* repressed by CG methylation, Low). Sequence-wise, the two transgenes are identical. However, *LUCH* has a higher basal level of LUC activity (Fig. 1B), probably due to higher methylation levels in *LUCL* (see below). After digestion of genomic DNA with McrBC, which cuts methylated DNA, no PCR product was obtained for *LUCL* at the *35S* and *LUC* regions, whereas with *LUCH*, PCR product was observed in the *LUC* region (Fig. 1C). Therefore, *LUCH* and *LUCL* both harbor 35S promoter methylation and *LUCL* also contains coding region methylation. In addition, treatment of *LUCH* and *LUCL* seedlings with 5-aza-2′-deoxycytidine, a chemical inhibitor of DNA methlytransferase activity, resulted in higher and nearly equal levels of luciferase luminescence from the two transgenes (Fig. 1B, D). The expression of the luciferase transgene as well as the nearby *NPTII* transgene was de-repressed as determined by RT-PCR (Fig. 1E). These data support the conclusion that DNA methylation is responsible for the repression of *LUCH* and *LUCL* transgenes. Moreover, bisulfite sequencing of the *35S* promoter of *LUCH* and *LUCL* revealed that *LUCL* contained extensive CG

methylation (81%) at this region (Fig. 1F). Furthermore, CG and CHG methylation is higher in *LUCL* than *LUCH*; *LUCH* harbors higher levels of CHH methylation than *LUCL* (Fig. 1F). The bisulfite conversion rates of these samples in the three contexts are all over 96% (Fig. 2). Unlike *LUCH*, in which TAIL-PCR was able to identify the genome insertion site of the transgene, multiple attempts doing so for *LUCL* failed (data not shown). However, southern blot analysis with BamHI, which has a single site on the transgene (Won et al., 2012), as well as the segregation pattern of Kan resistance, show that there is a single insertion in the genome, however, there are multiple copies of the transgene tandemly arrayed (Fig. 3).

Using *LUCL*, we set up a forward chemical screen to find small molecules that de-repressed LUC activity. We screened 24, 970 compounds (1200 from Life Sciences; 2000 from Spectrum; 400 from Myria/Sigma; 4204 from a triazine-tagged library; 2768 from CLICK; 3580 from LATCA) against seedlings at the two-leaf stage and obtained two confirmed hits, Amethopterin (renamed as Methotrexate (MTX)) (Fig. 4A-H) and Camptothecin (CPT) (Fig. 5A).

MTX is an allosteric inhibitor of DHFR (dihydrofolate reductase) and blocks the pathway involved in methyl biogenesis (Huber et al., 2007) (Fig. 4AA). Upon performing secondary screens of the chemical, we found that the MTX in the library was the less active form (Fig. 4A-H, Y)**,** which releases LUC activity at a higher concentration than the active form (Fig. 4A-H versus Fig. 4Q-X). The difference between the two forms is their chirality; the less active form exhibits D-chirality whereas the active form exhibits L-chirality (Fig, 4Y). As chemical acquisition may present a

problem, (the company currently no longer sells the less active compound), we also tested the racemic form and found that it could also release LUC activity, however, the optimal chemical concentration by which it does so varies among different experiments (data not shown) and different chemical concentrations (Fig. 4I-P). This is probably due to the fact that active and less active isoforms are found in the powder and with each experiment the powder has to be reconstituted so the concentration between active and less active forms may vary between batches. The less active form of MTX was able to release DNA methylation at *35S* (Fig. 4Z). Since MTX probably inhibits DNA methylation by reducing the amount of the methyl donor (Huber et al., 2007), we decided not to further our studies on this chemical and concentrated on CPT.

Camptothecin releases the silencing of *LUCL*

CPT is a natural quinoline alkaloid derived from the bark and stems of *Camptotheca acuminata*, a tree native to China (Fig. 5B) (Wall, 1966). This chemical was identified in a screen looking for natural compounds that possessed anti-cancer properties (Wall, 1966). We found that CPT released LUC activity from *LUCL* in a concentration- (Fig. 5A) and time-dependent (Fig. 5E, F) manner. RT-PCR and Real-Time RT-PCR experiments showed that the observed release of LUC activity was due to elevated transcript levels and not through arbitrary reactions of the luciferin substrate and CPT (Fig. 5C and D). Furthermore, we found that cell division was probably required for the release of LUC activity, as LUC activity was not observed until a day after incubation with the chemical compound (Fig. 5F).

As *LUCL* harbors a miR172 binding site, we wanted to know whether the increase of LUC activity may be attributed to loss of miRNA activity. Using a microRNA-based reporter line (*PM*), we found that the addition of CPT had no effect on LUC activity levels in that line (Fig 5E) (the transgene is described in Manavella et al., submitted).

A *DNA Topoisomerase Iα* (*topIα*) mutant is defective in DNA methylation

As indicated by the bisulfite sequencing data, *LUCL* is highly methylated, thus as CPT was able to release LUC activity, we sought to define its relationship with DNA methylation. CPT is a well known anti-cancer compound that prevents the activity of *TOPI* by binding the cleavage site so that Top1 cannot be displaced and thus, the DNA strand cannot re-ligate, thereby leading to cell death (Jaxel et al., 1991; Levin et al., 1993). Therefore, our data imply a role for *TOPIα* in DNA methylation. Coincidentally, through a screen for genes involved in floral determinacy in *Arabidopsis*, a *topIα* mutant, *topIα-2*, was isolated in my lab (Liu et al., in prep), therefore, I used the *topIα-2* mutant to determine whether *TOPIα* plays a role in DNA methylation. The *topIα-2* mutation led to prolonged floral stem cell activity to result in bulged gynoecia with ectopic floral organs inside. Treating *Arabidopsis* inflorescences with CPT phenocopied *topIα-2* (data not shown; Liu et al., in prep), confirming that CPT inhibits TOP1α activity *in vivo*.

First, we tested whether *topIα* mutants can release the silencing of *LUCL*, so we crossed *topIα-2*^{Col} to *LUCL*. *topIα-2* is in the L*er* background, thus was backcrossed five times to Col-0 to obtain the introgressed *top1α-2*^{Col} background suitable for crossing into *LUCL*. *topIα-2*^{Col} did not lead to the de-repression in LUC activity in *LUCL* (Fig. 6).

This is probably due to genetic redundancy as there are two *TOPIα* genes tandemly arrayed on chromosome five in *Arabidopsis* (Takahashi et al., 2002).

TOP Iα and CHH methylation

As *LUCL* harbors CHH methylation (Fig. 1F), which is established and maintained by 24-nt siRNAs, we wanted to know whether *TOPIα* plays a role in siRNA biogenesis. Northern blotting showed that siRNAs from three endogenous loci accumulated to wild-type levels in *topIα-2* (Fig. 7A). This was further validated through whole-genome small RNA analyses. The length distribution of small RNA was virtually identical in wild type and *topIα-2* (Fig. 8A). In total, there were only 37 500 bp-windows in the genome in which small RNAs showed 2-fold changes between the two genotypes --25 of which were 21-nt small RNAs and 10 24-nt small RNAs (Fig. 8B; Table 1). The differentially expressed small RNAs (DSRs) are derived from mostly transposons and some genes (Fig. 8B-E). However, the small number of DSRs as compared to the vast number of differentially methylated regions (DMRs; see below) between wild type and *topIα-2* suggests that *TOPIα* impacts DNA methylation downstream of siRNA biogenesis. In addition, loss of *topIα* does not affect microRNA levels at a single locus (miR173, Fig. 7A) or genome-wide (Fig. 8). Taken together, *TOPIα* does not play a role in miRNA biogenesis.

Although small RNA levels were not altered, *top Iα-2* or addition of CPT to wild-type seedlings led to siRNA target loci de-repression at *Cluster4, AtCopia, IG/LINE* and *AtMuI* (Fig. 7B). This shows that *TOPIα* does play a role in siRNA target loci repression,

even though it may not play a role in siRNA biogenesis or accumulation. Pol V generates transcripts called scaffold/Pol V/intergenic (IGN) transcripts that function downstream of siRNA biogenesis to promote/recruit DRM2 to methylate the region of interest (Wierzbicki et al., 2008b). Thus, we wanted to see whether *TOPIα* plays a role in the production of these transcripts. Real-Time RT-PCR experiments show that in the *topIα-2* mutant, scaffold transcript levels are slightly affected (consistently at 0.7 in *topIα-2*) at *MEA-ISR* (Fig. 7C) and *IGN5* (data not shown). This slight reduction could be due to genetic redundancy.

Therefore, in order to gain a better understanding of the role of *TOPIα* in CHH methylation, we performed whole genome analyses of L*er* and *topIα*. First, in terms of CHH methylation, we found 3,086 differentially methylated regions (DMRs)--732 of which are "increased" and 2,354 are "reduced" in the mutant (Fig. 9A, 10). In addition, most of the DMRs are located in transposable elements (78% of increased class, 79% of reduced clas), whereas only 15%-12% (increased and reduced, respectively) and 7%-9% are found in the intergenic and gene body regions, respectively (Fig. 9B-D).

As most CHH methylated loci are controlled by Pol IV and/or Pol V (Wierzbicki et al., 2008b), we sought to determine whether *TOPIα* acts through Pol IV and/or Pol V. Whole genome analyses show that *TOPIα* functions mainly independently of Pol IV and/or Pol V, as 77% (of increased DMRs) and 76% (of reduced DMRs) did not overlap with the DMRs between wild type (Col) and *sde4-3* (a Pol IV mutant) or between Col and *nrpe-1* (a Pol V mutant) (these are reduced in DNA methylation in the Pol IV or Pol V mutant; Fig. 11A). However, a subset of DMRs (20% of increased; 21% of reduced)

overlapped with *sde4-3* and *nrpe-1* DMRs (Fig. 11A). Figure 11B lists the raw numbers of the percentages represented in Figure 11A. In addition, we examined the chromosomal distribution of DMRs for *nrpe-1* (reduced in DNA methylation in this mutant relative to Col-0), *rdd* (increased in DNA methylation relative to Col-0), and *topIα* (both increased and reduced DMRs were included in the analysis). There was a mild and strong enrichment of *topIα* increased DMRs and reduced DMRs, respectively, at the pericentromeric regions (Fig 12, 13). *nrpe-1* and *rdd* DMRs are also enriched in the pericentromeric regions (Fig 12, 13). This enrichment could simply reflect the concentration of transposable elements, which tend to harbor CHH methylation, in the pericentromeric regions. The small overlap between *topIα* reduced DMRs with regions acted upon by Pol IV/Pol V suggests that *TOPIα* functions in CHH methylation mainly in a Pol IV- or Pol V-independent manner.

<u>*TOP Iα* acts independently of the *ROS1* pathway</u>

In *Arabidopsis,* DNA demethlyation is governed by a class of DNA glycosylases, *ROS1*, *DML2*, and *DML3* (Choi, 2002; Gong, 2002). In the *rdd* triple mutant, whole genome analyses showed that there was widespread hypermethylation at many loci (Choi, 2002; Gong, 2002). As a subset of DMRs was increased in *topIα*, this suggests that *TOPIα* also promotes DNA demethylation at certain loci. Thus, we wonder whether *TOPIα* acts through the *ROS1* pathway. Whole genome analysis of the increased DMRs in *topIα* (as compared to Ler) and *rdd* (as compared to Col) revealed little overlap (10%) (Fig. 14). Therefore, *TOPIα* probably does not function through the *ROS1* pathway.

TOP Iα and CHG methylation

In *Arabidopsis*, cytosine methylation occurs in all three contexts, however, *TOPIα* plays a smaller role in CHG methylation, as only 645 DMRs (277 of which are increased and 368 are reduced) were identified in the *topIα* mutant (Fig. 9A, 10). In the increased DMRs, 55% were from transposable elements, 26% from intergenic and 19% from genic regions (Fig. 9B, D, top panel). In the reduced DMRs, 52% were derived from transposable elements, 26% from intergenic regions and 22% from genic regions (Fig. 9C, D, bottom panel). While DMRs unique to *topIα* still made up the large portion of *topIα* DMRs (46% in the increased subset, 55% in the reduced subset), a surprisingly high percentage was seen for the *topIα & sde4-3 & nrpe1-11* overlapping DMRs (44% for increased and 34% for reduced DMRs) (Fig. 11A, B). Taken together, *TOPIα* does not have a large impact on CHG methylation (Fig. 9A), but for regions that it is required, due to the large overlap with Pol IV and Pol V at those regions, it is possible that *TOPIα* may act through these genes at certain loci (Fig. 11).

TOP Iα and CG methylation

Since *LUCL* harbors extensive CG methylation (Fig. 1F), we sought to identify *TOPIα*'s role in the methylation of this sequence context. We initially tested the highly repeated 5S loci using a CG methylation sensitive enzyme, *Hpa*II (which cuts unmethylated DNA). We found that *topIα-2* reduced methylation at these repeats, albeit not as strongly as *nrpe-1* (a Pol V mutant). The weaker allele of *TOP1α*, *topIα-3* and a

TOP1β allele, *topIβ-1*, both SALK lines (described in Liu et al., in prep), did not show a release of methylation (Fig. 15). We also tested the same loci (5S) with *Hae*III, an enzyme that detects CHH methylation, however, no changes were observed between L*er* and *top1α* (Fig. 16C). In addition, no changes were observed at *MEA-ISR* and *180S* when digested with MspI, an enzyme that detects CNG methylation (Fig. 16A, B).

From the whole genome analyses of DNA methylation in *topIα*, 3,060 CG DMRs were found, 1,167 were increased and 1,893 were reduced in *topIα* (Fig. 9A, 10). Interestingly, in the increased subset of DMRs, 50% were found in genic regions, while 34% were derived from transposable elements and 16% from intergenic regions (Fig. 9B, D, top panel). In the reduced subset of DMRs, 61% of DMRs were from genic regions, 28% from transposable elements and 11% from intergenic regions (Fig. 9C, D, bottom panel). The majority of *topIα* DMRs (76% and 78% in the increased and reduced populations, respectively) did not overlap with regions showing Pol IV- and/or Pol V-dependent methylation (Fig. 11A, B).

Gene body methylation occurs mainly in the CG context and is an ancient phenomenon found in plants, invertebrates and vertebrates, yet is poorly understood in terms of its mechanisms of deposition and function (Lister et al., 2008; Suzuki and Bird, 2008b; Zemach et al., 2010). As our whole genome analysis revealed changes in CG methylation at thousands of genes (Fig. 9A), we further examined the CG DMRs to better understand what types of genes tend to be affected by *top1α*. First, we examined the distribution of the CG DMRs (both increased and reduced in *top1α*) within the gene body. 90% of the DMRs are in the coding region as opposed to the UTRs (Fig. 17A).

Interestingly, for those DMRs found in the UTR, significantly more was found in the 3' UTR than the 5'UTR (Fig. 17A). Furthermore, DMR density increased as the distance from the transcription start site (TSS) increased, especially in the reduced DMR population (Fig. 17B). This shows that *TOPIα* is important for methylation towards the 3' end of the transcription unit (Fig. 17B).

There has been conflicting evidence in terms of the relationship between body methylation and genic expression levels. One study found that body methylation was associated with high transcription (Ball et al., 2009), whereas another linked it with moderate levels of transcription (Zemach et al., 2010). We took advantage of the gene body DMRs between wild type and *top1α* to examine the correlation between gene body methylation and expression. We plotted the proportion of genes (as a percentage) according to their transcript levels as determined by microarray-based expression profiling (Schmid et al., 2003) (Fig. 18A). We found that genes with moderate levels (8 or 9) levels of expression are preferentially affected in gene body methylation in *top1α*. This finding is consistent with data looking at transcription and body methylation in rice (Zemach et al., 2010).

In addition to gene expression levels, gene length has also been correlated with body methylation (Takuno and Gaut, 2012). A previous bioinformatics analysis has shown that the mean length of body-methylated genes is 3.3495 kb, whereas unmethylated genes are, on average, 1.5953 kb (Takuno and Gaut, 2012). We plotted the proportion of genes (as a percentage) in comparison to its length (x-axis). We found that the longer genes (from 3 kb and up) tend to show DMRs in *top1α* (Fig. 18B).

In wild-type plants, exons tend to be more methylated than introns (Lister et al., 2008). Thus, we mapped the *topIα* DMRs in relation with their exon or intron distribution. We found that the *top1α* DMRs are enriched in exons (Fig. 19), especially the "decreased" DMRs (Fig. 19B).

In addition, it has been hypothesized that gene body methylation may be important for alternative splicing (Jeltsch, 2010), thus we examined the *topIα* DMRs in relation with the number of spliced isoforms. We plotted the proportion of genes (y-axis) with *i* (*i*=1 to xx) spliced isoforms. In the population of increased DMRs in the CG context, 5.8% compared to 3.6% of the total genes had three spliced isoforms (Fig. 20A, C, E). In the population with reduced DMRs in the CG context, a higher percentage (than the total percentages) was observed for genes with two or more spliced isoforms (Fig. 20B, D, F). Thus, we found that genes with DMRs had a tendency to have more spliced forms (Fig. 20).

DISCUSSION

LUCL can be used as a tool to identify genes involved in cytosine methylation, especially gene body methylation

The molecular mechanism by which RdDM functions, such as the recruitment of Pol IV or DRM2 to the target loci, is still nebulous. Hence, to gain a better understanding of this pathway, we established two luciferase-based reporter systems: *LUCH* (Won, 2012) and *LUCL* (this study). The sequence of the reporter transgene in the two lines is

identical. *LUCH* has been shown to be a reporter of RdDM and the relatively high basal level of expression renders it suitable for bi-directional screens to isolate mutants with defective as well as enhanced DNA methylation (Won, 2012). Known as well as novel players in RdDM have been identified after subjecting this line to T-DNA and EMS mutagenesis ((Won, 2012); data not shown).

LUCL has been more difficult to characterize in that although this reporter line has a single locus insertion (from Kan resistance segregation ratios), it seems to contain two or more copies tandemly arrayed within the insertion site as indicated by failed TAIL-PCR experiments (data not shown) and Southern blotting (Fig. 2). Furthermore, unlike *LUCH,* which only harbors methylation at the 35S promoter (Fig. 1B), *LUCL* also has DNA methylation in the *LUC* coding region (Fig. 1B). Moreover, *LUCL* has a high amount of CG body methylation (Fig. 1F), thus making it the only existing reporter for body methylation. This is momentous as gene body methylation is observed in plants, vertebrates and invertebrates, yet not much is known about its function. Using *LUCL* as a reporter, through screening for chemicals that released LUC activity, we isolated two chemicals, MTX and CPT, which released LUC activity via the release of DNA methylation. As MTX was an obvious hit for our chemical screen, as it inhibits the production of SAM, the methyl donor, we focused our efforts on CPT, a well-known anti-cancer compound that targets *TOPIα*.

DNA topoisomerases are ubiquitous in nature and are pivotal in maintaining DNA topology during replication, recombination, and transcription. There are two main types of DNA topoisomerases but the evolution and diversity of the function and structure of

these enzymes highlight its diverse roles with regards to DNA topology in different species (Forterre et al., 2007). Although the main role of these enzymes is to regulate DNA topology, recent studies have shown that DNA topoisomerases play other cellular roles such as chromosome organization (Tadesse et al., 2005; Salceda et al., 2006)and epigenetic silencing (Huang et al., 2012). In addition, coincidentally, through a traditional forward genetic screen for genes involved in floral determinacy, we isolated and found that *TOPIα* contributes to floral stem cell homeostasis (Liu et al., in prep). Thus, while isolation of *TOPIα* in our cytosine methylation screen may have been counterintuitive with regards to its function in carcinogenesis, it coincides with *TOPIα*'s diverse functionality.

***TOPIα* is required for transposon silencing**

Analysis of siRNA and miRNA levels in the *topIα* mutant revealed that *TOPIα* does not play a major role in the biogenesis or accumulation of these small RNAs (Fig. 7A, 8). Instead, *TOPIα* is responsible for silencing transposons, as loss of *topIα* results in de-repression of siRNA target loci, such as transposable elements (Fig. 7B), and in genome-wide loss of CHH methylation at transposable elements. Being a general regulator of DNA topology, *TOPIα* has access to the entire genome, thus it is feasible for it to participate in maintaining genome integrity by repressing the expression of transposons.

Even though *TOPIα* is required for transposon repression, the pathway through which it functions is still unclear. There are two pathways by which *TOPIα* could act:

one, through RdDM in conjunction with Pol IV and/or Pol V, or two, through an independent pathway by regulating methylation status of the transposons directly. Support for the first hypothesis includes Real-Time RT-PCR experiments that show a slight, but consistent decrease in Pol V-dependent transcripts in *topIα* mutants (Fig. 7C). This slight reduction could be due to genetic redundancy masking the true effects of *topIα*, as loss of both *TOPIα* genes in *Arabidopsis* results in lethality (Takahashi et al., 2002). In addition, we are waiting for the genetic material to perform ChIP experiments to see whether loss of *TOPIα* would affect Pol V occupancy at RdDM loci. Furthermore, topoisomerases have been shown to interact with SMC-containing proteins for chromosome compaction (Maeshima and Laemmli, 2003; Tadesse et al., 2005), and DMS3, which is part of the RdDM machinery, contains a SMC domain, therefore, perhaps *TOPIα* may be an interacting partner of DMS3 in regulating RdDM.

In the second hypothesis, *TOPIα* silences transposons through a pathway independent of RdDM. Support for this hypothesis comes from genome-wide analysis in which *TOPIα* has a widespread effect on cytosine methylation. Loss of *TOPIα* results in both increased and decreased DMRs at transposable elements in terms of CHH methylation (Fig. 9). In addition, the hypothesis of a *TOPIα*-specific pathway is further supported by the fact that the majority of DMRs do not overlap with Pol IV and/or PolV loci (Fig. 11). Further, *TOPIα* may function to prevent the transcription of transposable elements through interactions with Pol II (Capranico et al., 2007) and/or nucleosome assembly/disassembly (Durand-Dubief et al., 2010; Durand-Dubief et al., 2011). The *topIα-2 nrpb2-3* double mutant is seedling lethal (data not shown; *nrpb2-3* is a weak Pol

II allele isolated by our lab), thus providing genetic evidence for the close relationship between *TOPIα* and Pol II.

TOPIα is important for the homeostasis of gene body methylation

Genome-wide analysis of *topIα* as compared to L*er* revealed a surprising role for *TOPIα* in gene body methylation. Gene body methylation is a controversial and poorly understood concept as evidence purports its importance to prevent spurious transcription and enhance accurate splicing of primary transcripts (Lorincz et al., 2004; Zilberman et al., 2007; Luco et al., 2010; Maunakea et al., 2010). However, since only minor effects on levels of expression were observed in genes that lost body methylation, and gene body methylation is highly polymorphic amongst different individuals (Vaughn et al., 2007; Zilberman et al., 2007; Zhang et al., 2008; Zemach et al., 2010), it may only be a by-product of transcription. Efforts have been made to differentiate between the two viewpoints (Takuno and Gaut, 2012). Thus, the discovery that *TOPIα* plays a role in gene body methylation may help shed light to these hypotheses.

We found that both increased and decreased DMRs had a propensity for genic regions in a CG context (Fig. 9B,C), thereby alluding to the importance of *TOPIα* in gene body methylation. All of our preliminary data (Figs. 17-20) suggest a link between *TOPIα*, transcription and gene body methylation. It would be tempting to postulate that *TOPIα* may be the "super-glue" that connects the two processes, but not enough direct evidence has been found to connect the three components. Studies have indicated a direct link between *TOPIα* and transcription. In 1993, Merino and colleagues used an *in*

vitro system to show that in the presence of an activator, Top1 could stimulate transcription from a circular DNA template at a higher rate, however in the absence of an activator, the enzyme could repress basal transcription levels (Merino et al., 1993). This study provided the first direct evidence that TopI acts on transcription.

Transcription can be divided in three components: initiation, elongation and termination. In *S. Pombe*, genome-wide analysis was performed on a *top1Δtop2-191*^{ts} mutant to determine the function of Top1 and Top2 in transcription (Durand-Dubief et al., 2010; Durand-Dubief et al., 2011). Using ChIP-chip, nucleosome occupancy was investigated and interestingly, they found that *Top1* contributes to transcription initiation by facilitating the eviction of nucleosomes through possible interactions with chromatin remodeling factors and the release of negative supercoils (Walfridsson et al., 2007; Durand-Dubief et al., 2010). Thus, *Top1* and *Top2* in yeast may function to maintain low histone density in promoters so Pol II can be efficiently recruited and high levels of transcription could ensue. Other studies have shown that nucleosomes are depleted at promoter regions and TSS, lending support to the role of TopI in transcription initiation (Lantermann et al., 2010). Studies in *Arabidopsis* lend support to this finding in that synergistic interactions were observed when *topIα* was crossed to genes in the Polycomb Repressive Complex 2 (PRC2) ((Graf et al., 2010); Liu et al., in prep). In addition, stochastic occurrence of ectopic AG::GUS expression in the *topIα* mutant suggests that it is needed for maintaining chromatin marks (Graf et al., 2010).

During elongation, as Pol II moves along the replication fork, positive supercoils are incurred in front of the replication fork. In yeast, it has been shown that *Top1* and

Top2 reduce the helical stress induced by Pol II. Moreover, in the *top1Δtop2-191^{ts}* double mutant, an accumulation of non-full-length transcripts was detected. Pol II ChIP showed that Pol II accumulates in the body of long genes in the *top1Δtop2-191^{ts},* which suggests that transcription elongation is compromised. This shows that topoisomerase activity is required for transcription elongation.

Top1 has a propensity to occupy 3' end of genes (Durand-Dubief et al., 2010) and TTS (Lantermann et al., 2010). Thus, it is feasible that topoisomerases play a role in transcription termination by evicting nucleosomes from that region so that termination can be processed properly. In *top1Δtop2-191^{ts}*, there was an increased amount of histone occupancy at the 3' end, regardless of the distance between adjacent genes (Durand-Dubief et al., 2010). Moreover, analysis of the *top1Δtop2-191^{ts}* transcriptome data revealed an increase in transcript levels at the 3' TTS region and an increase of Pol II occupancy was also observed (Durand-Dubief et al., 2010; Durand-Dubief et al., 2011). Taken together, these results highly implicate topoisomerase in promoting transcription termination.

In addition to the findings presented earlier, studies done in our lab reveal the intimate relationship between TOPIα and Pol II. In 2009, our lab isolated a Pol II weak allele in *Arabidopsis*, *nrpb2-3* (Zheng et al., 2009). We conducted expression analyses of *nrpb2-3* and *topIα-2* and found a surprising degree of overlap of differentially expressed genes (Liu et al., in prep). In *nrpb2-3*, 448 genes were reduced in expression. Of these 448 genes, 368 of them were also reduced in expression in *topIα* (there was a total of 716

genes reduced in expression in *topIα*). Furthermore, this intimate relationship is further purported by the seedling lethality of the *nrpb2-3 topIα* double mutant (data not shown).

Thus, taken together, there is a link between topoisomerases and transcription. Further, studies shown here reveal a role for *TOPIα* in gene body methylation. Is there a connection between the two processes? It may be highly provocative to say so, however, there is evidence to support that hypothesis. **First**, Top1 affects the production of long genes in transcription. Pol II accumulates on longer genes in *top1Δtop2-191*^{ts} (Durand-Dubief et al., 2010; Sperling et al., 2011). Our findings show that in the *topIα* DMR population, a high proportion of them are long genes (Fig. 18B). **Second**, Top1 mainly binds at gene promoter and 3' parts of the transcribed region (Durand-Dubief et al., 2010; Sperling et al., 2011). Our studies show that the *topIα* DMR population is enriched at the 3' end of a gene unit (Fig. 17B). **Third**, Top1 and Top2 are required for the proper expression of highly active genes. Our studies show that the *topIα* DMR population is enriched for moderately expressed genes (Fig. 18A). The sum of these results highlights an associative link between *TOPIα*, transcription and methylation. However, since all of these findings are associative, thus we are currently performing whole-genome mRNAseq analyses on *topIα* compared to L*er* in order to gain a better understanding of these three components and how they may be related.

Given the diversity of these enzymes, the newly discovered role of *TOPIα* in epigenetics is not surprising. DNA topoisomerases in general have been shown to play a role in epigenetic silencing (Huang et al., 2012), chromosome organization (Tadesse et al., 2005; Salceda et al., 2006), Pol II activity (Capranico et al., 2007; Sperling et al.,

2011), transcriptional regulation (Lotito et al., 2008), nucleosome assembly/disassembly (Durand-Dubief et al., 2010; Durand-Dubief et al., 2011; Sperling et al., 2011), transcription associated mutagenesis (Lippert et al., 2011; Takahashi et al., 2011) and transposon mobilization (Yang et al., 2012). Taken together, our findings shed light on our understanding of gene body methylation, a phenomenon that transcends the *Plantae* kingdom, and provide the groundwork for further studies in this field.

MATERIALS AND METHODS

Plant Growth Conditions and Luciferase Live Imaging

Arabidopsis thaliana seeds were surface-sterilized, planted on ½x MS-agar plates containing 1% sucrose (no sucrose was added in plates used for screening), and stratified at 4°C for two days. Seedlings were grown at 23°C under continuous light for ten days. All experiments were performed with ten-day old seedlings unless otherwise specified. For luciferase live imaging, 1 mM luciferin (Promega, Madison, WI) in 0.01% Triton X-100 was sprayed onto the seedlings and luciferase images were taken using a Stanford Photonics Onyx Luminescence Dark Box with a Roper Pixis 1024B camera.

Screening Conditions

220uL of ½x MS-agar media was placed into 96-well plates and two *LUCL*-containing seeds were placed in each well, equidistant apart. Plates were stratified at 4°C for two days and placed in the growth chamber (as described above). After 8-10 days, (when the two true leaves begin to emerge), 125uL of ½x MS media (no agar) was added to each well and the screening chemicals were added to each well. Luciferase imaging was done after three days of chemical addition (as described above). Secondary screens were performed in a 96-well format to titrate chemical concentrations. If the chemical produces consistent LUC activity, it was subjected to two different types of screening, addition of chemical to liquid media (at the screening concentration) or transferring

seedlings to chemical-laced solid media. With MTX (Sigma, St. Louis, MO) and CPT (Sigma, St. Louis, MO), both produced more consistent results when the chemical was added to the liquid media. MTX and CPT were obtained through a screen with the LATCA library of 3580 compounds (http://cutlerlab.blogspot.com/2008/05/latca.html).

Luciferase Activity Assays

LUCL, *LUCH*, *PM* and *PM* empty vector seeds were grown on selective media plates. *PM* is a miRNA reporter line from our collaborator Detlef Weigel. After nine days, seedlings were transferred one by one to 96-well 2 ml plates (Fisher Scientific, Waltham, MA) containing MS media and chemical. Luciferin substrate was added prior to loading into the Topcount NXT microplate luminescence counter (Perkin Elmer, Waltham, MA). During the next 48 hours, luciferase activity was read every 6-15 minutes for each plate, depending on the number of plates assayed. LUC levels were calculated via the Topcount Software Pack (Perkin Elmer, Waltham, MA).

RT-PCR and Real-Time RT-PCR

RNA was extracted with Tri-reagent (Trizol; Molecular Research Center) per manufacturer's instructions. Ten µg of total RNA was subjected to DNaseI treatment and reverse transcription. RT-PCR was performed on the cDNAs using primers specific for each locus of interest (Table 2). For detecting Pol V-dependent transcripts, the RT SuperScript III kit (Invitrogen, Carlsbad, CA) with gene specific primers was used and the RT reaction was performed per manufacturer's instructions. Real time RT-PCR was performed on the same cDNAs using a Biorad real time PCR SYBR Green system

(Biorad, Hercules, Ca). Three technical replicates were performed for each real time RT-PCR. Three biological replicates were obtained and Real Time RT-PCR was performed. Error bars represent the standard deviation from three technical replicates.

Small RNA Northern Blots

Total RNA was isolated as previously described. For miRNA detection, 5µg of total RNA was used, whereas 30-40µg was used for siRNA detection. Subsequent steps were performed as previously described (Zheng et al., 2009).

Small RNA-seq Library Generation

50ug total RNA was processed through a 15% denaturing polyacrylamide gel and a gel slice containing RNAs of 15 to 40 nucleotides (based on the O' Range Ruler 10 bp DNA ladder (Thermo Scientific, Glen Burnie, MD)) was extracted and transferred to 1.5 ml tube. The gel piece was ground, 500 µl of 0.4N NaCl (DEPC) was added to the ground mixture, and the tube containing the gel slice was shaken overnight at 4°C. Eluted RNAs were precipitated using ethanol and re-suspended in DEPC water. Gel-purified small RNAs were ligated sequentially to 3' and 5' adaptors followed by reverse transcription. PCR amplification of the cDNA resulted in the small RNA library, which is purified and subjected to high throughput sequencing. All of library construction steps were performed with the Illumina TruSeq-small RNA sample preparation kit per manufacturer's instructions (Illumina, San Diego, CA).

Sequencing Analysis

The raw data of small RNAs were processed by perl scripts built in house as described previously (Lertpanyasampatha, 2012). Briefly, the raw reads were screened with Illumina's quality control filter. The reads that passed the filter were separated into different bins according to their barcodes (indexes). The adaptor was trimmed for each read. Reads that match known rRNAs, tRNAs, snRNAs, and snoRNAs were removed. Reads of 20–24 nt were selected as the raw small RNA sequences. These reads were mapped to Arabidopsis genome with SOAP2 (Li et al., 2009).

Identification of DSRs (Differential Small RNA Accumulation Regions)

In order to identify the DSRs, every chromosome of Arabidopsis was divided into continuous 500bp windows. Small RNAs whose 5' nucleotide falls into a 500 bp window were counted as those belonging to this window. The number of reads in every window was recorded and serves as the abundance of sRNAs in this window. The edgeR (version 2.2.5) package (Robinson et al., 2010)in R was used to identify the windows showing differential small RNA accumulation between two genotypes. Fold change > 2 and adjusted p-value FDR < 0.05 were required.

DSR Distribution Across Genes, TEs and Intergenic Regions

The DSRs were processed as described below in DMR analysis. The DSRs overlapping with genes or TEs were counted once as gene-DSR or TE-DSR. The DSRs that did not overlap with genes or TEs were counted as intergenic-DSRs.

McrBC-based DNA Methylation Assays

Ten-day old seedlings were harvested and DNA was isolated via CTAB extraction. 500ng was subjected to McrBC (New England Biolabs, Ipswich, MA) treatment in the presence of GTP (New England Biolabs, Ipswich, MA) for two hours and 1µl was used for subsequent PCR reactions (for primer sequences, see Table 2).

Southern Blotting

Five or ten µg of CTAB-extracted DNA was subjected to *Hpa*II, *Hae*III or *Msp*I (New England Biolabs, Ipswich, MA) digestion overnight at 37°C and processed through a 1% gel for four hours. Subsequent processing, transfer, hybridization and analyses steps were performed as previously described (Won, 2012).

DNA Extraction and Bisulfite Conversion

For analysis of DNA methylation at a single locus (as in Fig. 1F), 2ug of RNAse-treated CTAB DNA was subjected to bisulfite conversion per manufacturer's instructions (Qiagen, Valencia, CA). The *35S* locus was amplified (for primers see Table 2) with Crimson Taq (New England Biolabs, Ipswich, MA) as follows: 94°C for 5 min.; 94°C for 30 sec, 56°C for 3 min, 68°C for 3 min, these steps are repeated an additional four times; 94°C for 30 sec, 56°C for 1min 30 sec, 68°C for 2 min, these steps are repeated an additional 39 times; final extension at 68°C for 5 min. The PCR product was then gel-purified (Zymo Research, Irvine, Ca) and subjected to standard cloning processes with

pGEM T-easy vector (Promega, Madison, WI). Positive clones were identified and 26 clones were sequenced for each genotype. Only unique clones were processed and analyzed with http://katahdin.mssm.edu/kismeth/revpage.pl.

MethylC-Seq Library Construction

For whole-genome bisulfite sequencing, approximately 1 µg of genomic DNA was sonicated to ~100 bp using the Covaris S2 System using the parameters: cycle number = 6, duty cycle = 20%, intensity = 5, cycles/burst = 200 and time = 60 seconds. Sonicated DNA was purified using Qiagen DNeasy minielute columns (Qiagen, Valencia, CA). Purified DNA was end repaired using End-It for 45 minutes at room temperature (Epicentre, Madison, WI). The end-repaired DNA was purified with a Qiagen DNeasy minielute column and A-tailed with dATP and Klenow 3'-5' exo- (New England Biolabs, Ipswich, MA) for 30 minute at 37°C and then purified with a Qiagen DNeasy minielute column. The purified DNA was ligated overnight at 16°C to genomic DNA adapters from Illumina Kit A (Illumina, San Diego, CA) with T4 DNA Ligase (New England Biolabs, Ipswich, MA). Ligation products were purified with AMPure XP beads (Beckman, Brea, CA). DNA (450 ng), was bisulfite treated using the MethylCode Kit (Invitrogen, Carlsbad, CA) following the manufacturer's guidelines and then PCR amplified using Pfu Cx Turbo (Agilent, Santa Clara, CA) using the following PCR conditions (2 minutes at 95°C, 4 cycles of 15 seconds at 98°C, 30 seconds at 60°C, 4 minutes at 72°C and 10 minutes at 72°C).

Sequencing

MethylC-Seq libraries were sequenced using the Illumina HiSeq 2000 (Illumina) as per manufacturer's instructions. Sequencing of libraries was performed up to 101 cycles. Image analysis and base calling were performed with the standard Illumina pipeline version RTA 1.13.48.

Sequencing Analysis

Fastq files were aligned to TAIR10 using Bowtie (Langmead et al, 2009) and custom algorithms were used for identification of mC sites as described previously (Lister et al, 2008). In brief, the raw data from Illumina were filtered with the following two steps. First, reads that failed in the quality control according to Illumina were removed. Second, ONLY reads that contained an adaptor were kept and the adaptor was subsequently trimmed.

Hereafter, the data were mapped to *Arabidopsis* genome as described (Lister et al., 2008). Briefly, first all the cytosine in the read was substituted by thymine. Second, two reference genome sequences were built, one in which cytosines that were converted to thymines, the other in which the guanines were converted to adenines. Only perfect match was allowed and used for further analyses. The reads mapped to more than one positions were disregarded.

Identification of Differentially Methylated Regions (DMRs)

DMR analysis was performed mostly as described in (Schmitz et al., 2011; Schmitz and Zhang, 2011) with the following modifications. Preliminary DMR tests were

performed for each type of methylation (CG, CHG and CHH) by scanning each chromosome requiring at least 10 methylated cytosines within a 200 bp window and neighboring DMRs were joined that occurred within 100 bp.

For the more detailed DMR analyses, first, 200bp windows with 10 mC differences between each sample were retained after the whole genome was scanned. Second, the window was reduced to the first and last cytosine. Third, the neighboring windows within 100bp were joined together. The methylation level was compared using a Kruskal-Wallis test. The joined windows with FDR < 0.01 were kept. The methylation density was computed as the number of mC divided by total C divided by length in base pairs of the window. The 8-fold difference in methylation density between the least and the most methylated sample was required. These DMRs were common for all samples.

For the DMRs of a single mutant, the methylation densities of the mutant and its corresponding WT were checked. The methylation density of the most methylated sample should be more than 0.01 and a 2-fold difference in methylation density between the two genotypes were required. For all of the analyses, the DMRs in the three contexts (CG, CHG and CHH) were counted separately.

DMR Distribution Along Chromosomes

Every chromosome was divided into continuous 300k bp bins. The DMRs of each methylation context in every bin were counted. The counts were plotted for each chromosome.

DMR Distribution Across Genes, TEs and Intergenic regions

The annotation of genes and TEs is according to TAIR10 (http://www.arabidopsis.org/). A DMR overlapping with a gene or a TE was counted once as a gene-DMR or a TE-DMR. The DMRs that did not overlap with genes or TEs were counted as intergenic-DMRs. The frequency of gene, TE and intergenic DMRs were computed as the number of DMRs divided by the total number of DMRs and shown in a pie chart.

Correlation of DMR with Gene Length

The lengths of genes with overlapping DMRs were examined. The lengths were divided into 11 bins (1, 2, 10kb and greater than 10kb). The percentage of genes of a certain length was calculated for total genes and genes harboring DMRs.

Correlation of DMR with the Levels of Gene Expression

Two microarray data sets of gene expression profile were downloaded from the GEO in NCBI (https://www.ncbi.nlm.nih.gov/geo): GSM8843 and GSM8844 (Schmid et al., 2003). The expression level estimated from Robust Multichip Analysis (RMA) was used. The RMA values were divided into 11 bins (4, 5, 13 and greater than 13). The expression levels of genes overlapping with DMRs were determined and the genes were binned accordingly. The distribution of genes along the expression spectrum was diagrammed for total genes as well as genes overlapping with DMRs.

DMR Distribution Along Genes

Gene-DMRs were mapped to the corresponding genes. The gene length was divided into 100 equal portions. The DMRs overlapping with each portion were counted as DMRs in that portion of the gene. The DMR distribution along genes was then plotted.

DMR Distribution Across UTRs and Coding Regions

The genes were divided into UTRs and coding regions according to TAIR10 annotation. Here, the different splicing isoforms of each gene were taken into account individually. A coding region-DMR, which overlapped with the UTR and coding region at the same time, was counted as an UTR-DMR once and a gene body DMR once. Otherwise, it was counted as an UTR-DMR or a coding region-DMR once.

DMR Distribution Across Exons and Introns

Here we also took every splicing isoform into account. Only internal exons were used. Internal exons were defined as the exons that are flanked by introns on both ends and do not contain any UTR as described (Feng et al., 2010). The upstream and downstream sequences of the same length as the exon were used as introns. The gene-DMRs were mapped to the intron-exon-intron model. The size of each model was normalized to the 300bp. The DMRs overlapping with each segment were counted.

Correlation of DMR with Alternative Splicing

All genes could be classified into 10 types according to the number of their alternative splicing isoforms from the annotation of TAIR10. The genes with *i* (*i*=1, 2, 10) isoforms are the *i*th type of genes. The DMR overlapping with i-type gene was

counted once as i-type-gene-DMR. A histogram was created to display the distribution of all as well as DMR-overlapping genes among the 10 gene classes.

REFERENCES

Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M. and Suzuki, M. (1998) 'A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA', *Embo J* 17(18): 5484-96.

Aravin, A. A. (2001) 'Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline', *Curr. Biol.* 11: 1017-1027.

Aravin, A. A. and Bourc'his, D. (2008) 'Small RNA guides for de novo DNA methylation in mammalian germ cells', *Genes Dev.* 22: 970-975.

Aravin, A. A., Hannon, G. J. and Brennecke, J. (2007a) 'The Piwi?뱎iRNA pathway provides an adaptive defense in the transposon arms race', *Science* 318: 761-764.

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. and Hannon, G. J. (2007b) 'Developmentally regulated piRNA clusters implicate MILI in transposon control', *Science* 316: 744-747.

Ball, M. P., Li, J. B., Gao, Y., Lee, J. H., LeProust, E. M., Park, I. H., Xie, B., Daley, G. Q. and Church, G. M. (2009) 'Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells', *Nat Biotechnol* 27(4): 361-8.

Baute, J. and Depicker, A. (2008) 'Base excision repair and its role in maintaining genome stability', *Crit. Rev. Biochem. Mol. Biol.* 43: 239-276.

Been, M. D., Burgess, R. R. and Champoux, J. J. (1984) 'Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA', *Nucleic Acids Res* 12(7): 3097-114.

Belova, G. I., Prasad, R., Nazimov, I. V., Wilson, S. H. and Slesarev, A. I. (2002) 'The domain organization and properties of individual domains of DNA topoisomerase V, a type 1B topoisomerase with DNA repair activities', *J Biol Chem* 277(7): 4959-65.

Beran-Steed, R. K. and Tse-Dinh, Y. C. (1989) 'The carboxyl terminal domain of Escherichia coli DNA topoisomerase I confers higher affinity to DNA', *Proteins* 6(3): 249-58.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P. C., Nicolas, A. and Forterre, P. (1997) 'An atypical topoisomerase II from Archaea with implications for meiotic recombination', *Nature* 386(6623): 414-7.

Bharti, A. K., Olson, M. O., Kufe, D. W. and Rubin, E. H. (1996) 'Identification of a nucleolin binding site in human topoisomerase I', *J Biol Chem* 271(4): 1993-7.

Bomblies, K., Dagenais, N. and Weigel, D. (1999) 'Redundant enhancers mediate transcriptional repression of *AGAMOUS* by *APETALA2*', *Dev Biol* 216(1): 260-4.

Bonven, B. J., Gocke, E. and Westergaard, O. (1985) 'A high affinity topoisomerase I binding sequence is clustered at DNAase I hypersensitive sites in Tetrahymena R-chromatin', *Cell* 41(2): 541-51.

Bostick, M. (2007) 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science* 317: 1760-1764.

Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (1989) 'Genes directing flower development in *Arabidopsis*.', *Plant Cell* 1: 37-52.
Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (1991) 'Genetic interactions among floral homeotic genes of *Arabidopsis*', *Development* 112: 1-20.
Breuer, C., Stacey, N. J., West, C. E., Zhao, Y., Chory, J., Tsukaya, H., Azumi, Y., Maxwell, A., Roberts, K. and Sugimoto-Shirasu, K. (2007) 'BIN4, a novel component of the plant DNA topoisomerase VI complex, is required for endoreduplication in Arabidopsis', *Plant Cell* 19(11): 3655-68.
Broitman-Maduro, G., Maduro, M. F. and Rothman, J. H. (2005) 'The noncanonical binding site of the MED-1 GATA factor defines differentially regulated target genes in the C. elegans mesendoderm', *Developmental cell* 8(3): 427-33.
Brown, P. O. and Cozzarelli, N. R. (1979) 'A sign inversion mechanism for enzymatic supercoiling of DNA', *Science* 206(4422): 1081-3.
Camilloni, G., Di Martino, E., Di Mauro, E. and Caserta, M. (1989) 'Regulation of the function of eukaryotic DNA topoisomerase I: topological conditions for inactivity', *Proc Natl Acad Sci U S A* 86(9): 3080-4.
Cao, X. (2003) 'Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation', *Curr. Biol.* 13: 2212-2217.
Capranico, G., Ferri, F., Fogli, M. V., Russo, A., Lotito, L. and Baranello, L. (2007) 'The effects of camptothecin on RNA polymerase II transcription: roles of DNA topoisomerase I', *Biochimie* 89(4): 482-9.
Cartolano, M., Castillo, R., Efremova, N., Kuckenberg, M., Zethof, J., Gerats, T., Schwarz-Sommer, Z. and Vandenbussche, M. (2007) 'A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity', *Nat Genet* 39(7): 901-5.
Champoux, J. J. (2001) 'DNA topoisomerases: structure, function, and mechanism', *Annu Rev Biochem* 70: 369-413.
Champoux, J. J. and Dulbecco, R. (1972) 'An activity from mammalian cells that untwists superhelical DNA--a possible swivel for DNA replication (polyoma-ethidium bromide-mouse-embryo cells-dye binding assay)', *Proc Natl Acad Sci U S A* 69(1): 143-6.
Chan, S. W., Henderson, I. R. and Jacobsen, S. E. (2005) 'Gardening the genome: DNA methylation in Arabidopsis thaliana', *Nature Rev. Genet.* 6: 351-360.
Chappell, C., Hanakahi, L. A., Karimi-Busheri, F., Weinfeld, M. and West, S. C. (2002) 'Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining', *EMBO J* 21(11): 2827-32.
Chen, X. (2004) 'A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development', *Science* 303(5666): 2022-5.
Choi, Y. (2002) 'DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis', *Cell* 110: 33-42.
Ciccone, D. N. and Chen, T. (2009) 'Histone lysine methylation in genomic imprinting', *Epigenetics* 4: 216-220.
Cokus, S. J. (2008) 'Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning', *Nature* 452: 215-219.

Corbett, K. D. and Berger, J. M. (2004) 'Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases', *Annu Rev Biophys Biomol Struct* 33: 95-118.
Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D. et al. (2011) 'Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair', *Cell* 146(1): 67-79.
Creemers, G. J., Lund, B. and Verweij, J. (1994) 'Topoisomerase I inhibitors: topotecan and irenotecan', *Cancer Treat Rev* 20(1): 73-96.
Cronstein, B. N. (2005a) 'Folic acid and folinic acid supplements and methotrexate therapy: comment on the article by Morgan et al', *Arthritis Rheum* 52(4): 1338-9; author reply 1339-40.
Cronstein, B. N. (2005b) 'Low-dose methotrexate: a mainstay in the treatment of rheumatoid arthritis', *Pharmacol Rev* 57(2): 163-72.
Dalmay, T., Hamilton, A., Rudd, S., Angell, S. and Baulcombe, D. C. (2000) 'An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus', *Cell* 101(5): 543-53.
De Rybel, B., Audenaert, D., Vert, G., Rozhon, W., Mayerhofer, J., Peelman, F., Coutuer, S., Denayer, T., Jansen, L., Nguyen, L. et al. (2009) 'Chemical inhibition of a subset of Arabidopsis thaliana GSK3-like kinases activates brassinosteroid signaling', *Chem Biol* 16(6): 594-604.
De Silva, E. K., Gehrke, A. R., Olszewski, K., Leon, I., Chahal, J. S., Bulyk, M. L. and Llinas, M. (2008) 'Specific DNA-binding by apicomplexan AP2 transcription factors', *Proceedings of the National Academy of Sciences of the United States of America* 105(24): 8393-8.
Delaval, K. (2007) 'Differential histone modifications mark mouse imprinting control regions during spermatogenesis', *EMBO J.* 26: 720-729.
Dennis, K., Fan, T., Geiman, T., Yan, Q. and Muegge, K. (2001) 'Lsh, a member of the SNF2 family, is required for genome-wide methylation', *Genes Dev.* 15: 2940-2944.
Deyholos, M. K. and Sieburth, L. E. (2000) 'Separable whorl-specific expression and negative regulation by enhancer elements within the *AGAMOUS* second intron', *Plant Cell* 12(10): 1799-810.
Drews, G. N., Bowman, J. L. and Meyerowitz, E. M. (1991) 'Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product', *Cell* 65(6): 991-1002.
Durand-Dubief, M., Persson, J., Norman, U., Hartsuiker, E. and Ekwall, K. (2010) 'Topoisomerase I regulates open chromatin and controls gene expression in vivo', *EMBO J* 29(13): 2126-34.
Durand-Dubief, M., Svensson, J. P., Persson, J. and Ekwall, K. (2011) 'Topoisomerases, chromatin and transcription termination', *Transcription* 2(2): 66-70.
Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E. et al. (2010) 'Conservation and divergence of methylation patterning in plants and animals', *Proc Natl Acad Sci U S A* 107(19): 8689-94.

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., Marques, C. J., Andrews, S. and Reik, W. (2011) 'Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation', *Nature* 473(7347): 398-402.
Forterre, P. (2006) 'DNA topoisomerase V: a new fold of mysterious origin', *Trends Biotechnol* 24(6): 245-7.
Forterre, P., Gribaldo, S., Gadelle, D. and Serre, M. C. (2007) 'Origin and evolution of DNA topoisomerases', *Biochimie* 89(4): 427-46.
Fournier, C. (2002) 'Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes', *EMBO J.* 21: 6560-6570.
Franks, R. G., Wang, C., Levin, J. Z. and Liu, Z. (2002) 'SEUSS, a member of a novel family of plant regulatory proteins, represses floral homeotic gene expression with LEUNIG', *Development* 129(1): 253-263.
Fromme, J. C. and Verdine, G. L. (2004) 'Base excision repair', *Adv Protein Chem* 69: 1-41.
Galat, A., Lane, W. S., Standaert, R. F. and Schreiber, S. L. (1992) 'A rapamycin-selective 25-kDa immunophilin', *Biochemistry* 31(8): 2427-34.
Gehring, M., Bubb, K. L. and Henikoff, S. (2009) 'Extensive demethylation of repetitive elements during seed development underlies gene imprinting', *Science* 324: 1447-1451.
Gellert, M., Mizuuchi, K., O'Dea, M. H. and Nash, H. A. (1976) 'DNA gyrase: an enzyme that introduces superhelical turns into DNA', *Proc Natl Acad Sci U S A* 73(11): 3872-6.
Goll, M. G. and Bestor, T. H. (2005) 'Eukaryotic cytosine methyltransferases', *Annu. Rev. Biochem.* 74: 481-514.
Gomez-Mena, C., de Folter, S., Costa, M. M., Angenent, G. C. and Sablowski, R. (2005) 'Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis', *Development* 132(3): 429-38.
Gong, Z. (2002) 'ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase', *Cell* 111: 803-814.
Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E. M. and Coupland, G. (1997) 'A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*', *Nature* 386(6620): 44-51.
Goto, T. and Wang, J. C. (1982) 'Yeast DNA topoisomerase II. An ATP-dependent type II topoisomerase that catalyzes the catenation, decatenation, unknotting, and relaxation of double-stranded DNA rings', *J Biol Chem* 257(10): 5866-72.
Graf, P., Dolzblasz, A., Wurschum, T., Lenhard, M., Pfreundt, U. and Laux, T. (2010) 'MGOUN1 encodes an Arabidopsis type IB DNA topoisomerase required in stem cell regulation and to maintain developmentally regulated gene silencing', *Plant Cell* 22(3): 716-28.
Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. and Young, R. A. (2007) 'A chromatin landmark and transcription initiation at most promoters in human cells', *Cell* 130: 77-88.
Gustafson-Brown, C., Savidge, B. and Yanofsky, M. F. (1994) 'Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*', *Cell* 76: 131-143.
Hanai, R., Caron, P. R. and Wang, J. C. (1996) 'Human TOP3: a single-copy gene encoding DNA topoisomerase III', *Proc Natl Acad Sci U S A* 93(8): 3653-7.

Hao, D., Ohme-Takagi, M. and Sarai, A. (1998) 'Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant', *J Biol Chem* 273(41): 26857-61.
Hartung, F., Wurz-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S. and Puchta, H. (2007) 'The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in Arabidopsis', *Plant Cell* 19(10): 3090-9.
He, X. J. (2009) 'A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development', *Genes Dev.* 23: 2717-2722.
He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. et al. (2011) 'Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA', *Science* 333(6047): 1303-7.
Henderson, I. R. and Jacobsen, S. E. (2007) 'Epigenetic inheritance in plants', *Nature* 447: 418-424.
Hiasa, H. and Marians, K. J. (1996) 'Two distinct modes of strand unlinking during theta-type DNA replication', *J Biol Chem* 271(35): 21529-35.
Hirochika, H., Okamoto, H. and Kakutani, T. (2000) 'Silencing of retrotransposons in Arabidopsis and reactivation by the ddm1 mutation', *Plant Cell* 12: 357-369.
Hodges, E., Smith, A. D., Kendall, J., Xuan, Z., Ravi, K., Rooks, M., Zhang, M. Q., Ye, K., Bhattacharjee, A., Brizuela, L. et al. (2009) 'High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing', *Genome Res* 19(9): 1593-605.
Hong, R. L., Hamaguchi, L., Busch, M. A. and Weigel, D. (2003) 'Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing', *Plant Cell* 15(6): 1296-309.
Hsieh, T. and Brutlag, D. (1980) 'ATP-dependent DNA topoisonmerase from D. melanogaster reversibly catenates duplex DNA rings', *Cell* 21(1): 115-25.
Huang, H. S., Allen, J. A., Mabb, A. M., King, I. F., Miriyala, J., Taylor-Blake, B., Sciaky, N., Dutton, J. W., Jr., Lee, H. M., Chen, X. et al. (2012) 'Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons', *Nature* 481(7380): 185-9.
Huang, J. (2004) 'Lsh, an epigenetic guardian of repetitive elements', *Nucleic Acids Res.* 32: 5019-5028.
Huber, L. C., Stanczyk, J., Jungel, A. and Gay, S. (2007) 'Epigenetics in inflammatory rheumatic diseases', *Arthritis Rheum* 56(11): 3523-31.
Husbands, A., Bell, E. M., Shuai, B., Smith, H. M. and Springer, P. S. (2007) 'LATERAL ORGAN BOUNDARIES defines a new family of DNA-binding transcription factors and can interact with specific bHLH proteins', *Nucleic Acids Res* 35(19): 6663-71.
Hutvagner, G. and Simard, M. J. (2008) 'Argonaute proteins: key players in RNA silencing', *Nature Rev. Mol. Cell Biol.* 9: 22-32.
Inoue, A. and Zhang, Y. (2011) 'Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos', *Science* 334(6053): 194.
Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. and Zhang, Y. (2011) 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science* 333(6047): 1300-3.

Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J. L. and Meyerowitz, E. M. (2004) 'The homeotic protein *AGAMOUS* controls microsporogenesis by regulation of *SPOROCYTELESS*', *Nature* 430(6997): 356-60.
Jackson, J. P., Lindroth, A. M., Cao, X. and Jacobsen, S. E. (2002) 'Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase', *Nature* 416: 556-560.
Jaxel, C., Capranico, G., Kerrigan, D., Kohn, K. W. and Pommier, Y. (1991) 'Effect of local DNA sequence on topoisomerase I cleavage in the presence or absence of camptothecin', *J Biol Chem* 266(30): 20418-23.
Jeltsch, A. (2010) 'Molecular biology. Phylogeny of methylomes', *Science* 328(5980): 837-8.
Jofuku, K. D., den Boer, B. G., Van Montagu, M. and Okamuro, J. K. (1994) 'Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*', *Plant Cell* 6(9): 1211-25.
Jofuku, K. D., Omidyar, P. K., Gee, Z. and Okamuro, J. K. (2005) 'Control of seed mass and seed yield by the floral homeotic gene *APETALA2*', *Proc Natl Acad Sci U S A* 102(8): 3117-22.
Johnson, L. M. (2007) 'The SRA methyl-cytosine-binding domain links DNA and histone methylation', *Curr. Biol.* 17: 379-384.
Johnson, L. M., Law, J. A., Khattar, A., Henderson, I. R. and Jacobsen, S. E. (2008) 'SRA-domain proteins required for DRM2-mediated de novo DNA methylation', *PLoS Genet.* 4: e1000280.
Kagaya, Y., Ohmiya, K. and Hattori, T. (1999) 'RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants', *Nucleic Acids Res* 27(2): 470-8.
Kanno, T. (2004) 'Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation', *Curr. Biol.* 14: 801-805.
Kanno, T. (2008) 'A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation', *Nature Genet.* 40: 670-675.
Kanno, T. (2009) 'RNA-directed DNA methylation and plant development require an IWR1-type transcription factor', *EMBO Rep.* 11: 65-71.
Kanoh, N., Kumashiro, S., Simizu, S., Kondoh, Y., Hatakeyama, S., Tashiro, H. and Osada, H. (2003) 'Immobilization of natural products on glass slides by using a photoaffinity reaction and the detection of protein-small-molecule interactions', *Angew Chem Int Ed Engl* 42(45): 5584-7.
Keck, E., McSteen, P., Carpenter, R. and Coen, E. (2003) 'Separation of genetic functions controlling organ identity in flowers', *Embo J* 22(5): 1058-66.
Khobta, A., Ferri, F., Lotito, L., Montecucco, A., Rossi, R. and Capranico, G. (2006) 'Early effects of topoisomerase I inhibition on RNA polymerase II along transcribed genes in human cells', *J Mol Biol* 357(1): 127-38.
Kieber, J. J., Tissier, A. F. and Signer, E. R. (1992) 'Cloning and Characterization of an Arabidopsis thaliana Topoisomerase I Gene', *Plant Physiol* 99(4): 1493-501.
Kim, J. K., Samaranayake, M. and Pradhan, S. (2009) 'Epigenetic mechanisms in mammals', *Cell Mol Life Sci* 66(4): 596-612.

Kim, R. A. and Wang, J. C. (1989) 'Function of DNA topoisomerases as replication swivels in Saccharomyces cerevisiae', *J Mol Biol* 208(2): 257-67.
Kirik, V., Schrader, A., Uhrig, J. F. and Hulskamp, M. (2007) 'MIDGET unravels functions of the Arabidopsis topoisomerase VI complex in DNA endoreduplication, chromatin condensation, and transcriptional silencing', *Plant Cell* 19(10): 3100-10.
Kitagawa, M., Ikeda, S., Tashiro, E., Soga, T. and Imoto, M. (2010) 'Metabolomic identification of the target of the filopodia protrusion inhibitor glucopiericidin A', *Chem Biol* 17(9): 989-98.
Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S. and Ahringer, J. (2009) 'Differential chromatin marking of introns and expressed exons by H3K36me3', *Nat Genet* 41(3): 376-81.
Koster, D. A., Croquette, V., Dekker, C., Shuman, S. and Dekker, N. H. (2005) 'Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB', *Nature* 434(7033): 671-4.
Krah, R., Kozyavkin, S. A., Slesarev, A. I. and Gellert, M. (1996) 'A two-subunit type I DNA topoisomerase (reverse gyrase) from an extreme hyperthermophile', *Proc Natl Acad Sci U S A* 93(1): 106-10.
Krizek, B. A. (2003) 'AINTEGUMENTA utilizes a mode of DNA recognition distinct from that used by proteins containing a single AP2 domain', *Nucleic Acids Res* 31(7): 1859-68.
Krizek, B. A., Lewis, M. W. and Fletcher, J. C. (2006) '*RABBIT EARS* is a second-whorl repressor of *AGAMOUS* that maintains spatial boundaries in *Arabidopsis* flowers', *Plant J* 45(3): 369-83.
Krogh, B. O. and Shuman, S. (2002) 'Proton relay mechanism of general acid catalysis by DNA topoisomerase IB', *J Biol Chem* 277(8): 5711-4.
Kunst, L., Klenz, J. E., Martinez-Zapater, J. and Haughn, G. W. (1989) '*AP2* gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana*', *Plant Cell* 1(12): 1195-1208.
Kuramochi-Miyagawa, S. (2008) 'DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes', *Genes Dev.* 22: 908-917.
Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N. et al. (2006) 'The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease', *Science* 313(5795): 1929-35.
Lantermann, A. B., Straub, T., Stralfors, A., Yuan, G. C., Ekwall, K. and Korber, P. (2010) 'Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae', *Nat Struct Mol Biol* 17(2): 251-7.
Law, J. A. and Jacobsen, S. E. (2010) 'Establishing, maintaining and modifying DNA methylation patterns in plants and animals', *Nat Rev Genet* 11(3): 204-220.
Law, J. A., Vashisht, A. A., Wohlschlegel, J. A. and Jacobsen, S. E. (2011) 'SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and

chromatin remodeling factors, associate with RNA polymerase IV', *PLoS Genet* 7(7): e1002195.
Lertpanyasampatha, M., Gao, L., Kongsawaworaku, P., Viboonjun, U., Chrestin, H., Liu, R., Chen, X., Narangajavana, J. (2012) 'Genome-wide analysis of microRNAs in rubber tree (*Hevea brasiliensis L.*) using high-throughput sequencing.', *Planta*.
Levin, N. A., Bjornsti, M. A. and Fink, G. R. (1993) 'A novel mutation in DNA topoisomerase I of yeast causes DNA damage and RAD9-dependent cell cycle arrest', *Genetics* 133(4): 799-814.
Li, L., Thomas, S. A., Klein, L. L., Yeung, C. M., Maring, C. J., Grampovnik, D. J., Lartey, P. A. and Plattner, J. J. (1994) 'Synthesis and biological evaluation of C-3'-modified analogs of 9(R)-dihydrotaxol', *J Med Chem* 37(17): 2655-63.
Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K. and Wang, J. (2009) 'SOAP2: an improved ultrafast tool for short read alignment', *Bioinformatics* 25(15): 1966-7.
Lippert, M. J., Kim, N., Cho, J. E., Larson, R. P., Schoenly, N. E., O'Shea, S. H. and Jinks-Robertson, S. (2011) 'Role for topoisomerase 1 in transcription-associated mutagenesis in yeast', *Proc Natl Acad Sci U S A* 108(2): 698-703.
Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H. and Ecker, J. R. (2008) 'Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis', *Cell* 133(3): 523-536.
Litt, A. (2007) 'An evaluation of A-funciton: evidence from the *APETALA1* and *APETALA2* gene lineages', *Int. J. Plant Sci.* 168: 73-91.
L iu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1998) 'Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*', *Plant Cell* 10(8): 1391-406.
Lorincz, M. C., Dickerson, D. R., Schmitt, M. and Groudine, M. (2004) 'Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells', *Nat Struct Mol Biol* 11(11): 1068-75.
Lorsbach, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T. and Downing, J. R. (2003) 'TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23)', *Leukemia* 17(3): 637-41.
Lotito, L., Russo, A., Chillemi, G., Bueno, S., Cavalieri, D. and Capranico, G. (2008) 'Global transcription regulation by DNA topoisomerase I in exponentially growing Saccharomyces cerevisiae cells: activation of telomere-proximal genes by TOP1 deletion', *J Mol Biol* 377(2): 311-22.
Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M. and Misteli, T. (2010) 'Regulation of alternative splicing by histone modifications', *Science* 327(5968): 996-1000.
Maes, T., Van de Steene, N., Zethof, J., Karimi, M., D'Hauw, M., Mares, G., Van Montagu, M. and Gerats, T. (2001) 'Petunia *Ap2*-like genes and their role in flower and seed development', *Plant Cell* 13(2): 229-44.

Maeshima, K. and Laemmli, U. K. (2003) 'A two-step scaffolding model for mitotic chromosome assembly', *Dev Cell* 4(4): 467-80.
Maiti, A. and Drohat, A. C. (2011a) 'Dependence of substrate binding and catalysis on pH, ionic strength, and temperature for thymine DNA glycosylase: Insights into recognition and processing of G.T mispairs', *DNA Repair (Amst)* 10(5): 545-53.
Maiti, A. and Drohat, A. C. (2011b) 'Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites', *J Biol Chem* 286(41): 35334-8.
Makarevitch, I. and Somers, D. A. (2006) 'Association of Arabidopsis topoisomerase IIA cleavage sites with functional genomic elements and T-DNA loci', *Plant J* 48(5): 697-709.
Malik, S. B., Ramesh, M. A., Hulstrand, A. M. and Logsdon, J. M., Jr. (2007) 'Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss', *Mol Biol Evol* 24(12): 2827-41.
Martinez-Macias, M. I., Qian, W., Miki, D., Pontes, O., Liu, Y., Tang, K., Liu, R., Morales-Ruiz, T., Ariza, R. R., Roldan-Arjona, T. et al. (2012) 'A DNA 3' phosphatase functions in active DNA demethylation in Arabidopsis', *Mol Cell* 45(3): 357-70.
Mathieu, J., Yant, L. J., Murdter, F., Kuttner, F. and Schmid, M. (2009) 'Repression of flowering by the miR172 target SMZ', *PLoS biology* 7(7): e1000148.
Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y. et al. (2010) 'Conserved role of intragenic DNA methylation in regulating alternative promoters', *Nature* 466(7303): 253-7.
Merino, A., Madden, K. R., Lane, W. S., Champoux, J. J. and Reinberg, D. (1993) 'DNA topoisomerase I is involved in both repression and activation of transcription', *Nature* 365(6443): 227-32.
Mizuuchi, K., Fisher, L. M., O'Dea, M. H. and Gellert, M. (1980) 'DNA gyrase action involves the introduction of transient double-strand breaks into DNA', *Proc Natl Acad Sci U S A* 77(4): 1847-51.
Mlotshwa, S., Yang, Z., Kim, Y. and Chen, X. (2006) 'Floral patterning defects induced by *Arabidopsis APETALA2* and microRNA172 expression in *Nicotiana benthamiana*', *Plant Mol Biol* 61(4-5): 781-93.
Monk, M., Boubelik, M. and Lehnert, S. (1987) 'Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development', *Development* 99: 371-382.
Morgan, H. D., Dean, W., Coker, H. A., Reik, W. and Petersen-Mahrt, S. K. (2004) 'Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming', *J. Biol. Chem.* 279: 52353-52360.
Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pages, H. and Gentleman, R. (2009) 'ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data', *Bioinformatics* 25(19): 2607-8.
Morgenstern, B. (2004) 'DIALIGN: multiple DNA and protein sequence alignment at BiBiServ', *Nucleic Acids Res* 32(Web Server issue): W33-6.

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J. B., Jouette, D., Lacombe, A. M., Nikic, S., Picault, N. et al. (2000) '*Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance', *Cell* 101(5): 533-42.
Muroi, M., Kazami, S., Noda, K., Kondo, H., Takayama, H., Kawatani, M., Usui, T. and Osada, H. (2010) 'Application of proteomic profiling based on 2D-DIGE for classification of compounds according to the mechanism of action', *Chem Biol* 17(5): 460-70.
Nole-Wilson, S. and Krizek, B. A. (2000) 'DNA binding properties of the *Arabidopsis* floral development protein AINTEGUMENTA', *Nucleic Acids Res* 28(21): 4076-82.
Ohme-Takagi, M. and Shinshi, H. (1995) 'Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element', *Plant Cell* 7(2): 173-82.
Ohto, M. A., Floyd, S. K., Fischer, R. L., Goldberg, R. B. and Harada, J. J. (2009) 'Effects of *APETALA2* on embryo, endosperm, and seed coat development determine seed size in *Arabidopsis*', *Sex Plant Reprod* 22(4): 277-89.
Okamuro, J. K., Caster, B., Villarroel, R., Van Montagu, M. and Jofuku, K. D. (1997) 'The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*', *Proc Natl Acad Sci U S A* 94(13): 7076-81.
Okitsu, C. Y. and Hsieh, C. L. (2007) 'DNA methylation dictates histone H3K4 methylation', *Mol. Cell. Biol.* 27: 2746-2757.
Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H. and Hayashi, Y. (2002) 'LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23)', *Cancer Res* 62(14): 4075-80.
Ooi, S. K. (2007) 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature* 448: 714-717.
Otani, J. (2009) 'Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX?밒NMT3?밒NMT3L domain', *EMBO Rep.* 10: 1235-1241.
Park, S. Y., Fung, P., Nishimura, N., Jensen, D. R., Fujii, H., Zhao, Y., Lumba, S., Santiago, J., Rodrigues, A., Chow, T. F. et al. (2009) 'Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins', *Science* 324(5930): 1068-71.
Parsons, A. B., Brost, R. L., Ding, H., Li, Z., Zhang, C., Sheikh, B., Brown, G. W., Kane, P. M., Hughes, T. R. and Boone, C. (2004) 'Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways', *Nat Biotechnol* 22(1): 62-9.
Parsons, A. B., Lopez, A., Givoni, I. E., Williams, D. E., Gray, C. A., Porter, J., Chua, G., Sopko, R., Brost, R. L., Ho, C. H. et al. (2006) 'Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast', *Cell* 126(3): 611-25.
Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., McLoughlin, E. M., Brudno, Y., Mahapatra, S., Kapranov, P. et al. (2011) 'Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells', *Nature* 473(7347): 394-7.
Penterman, J. (2007) 'DNA demethylation in the Arabidopsis genome', *Proc. Natl Acad. Sci. USA* 104: 6752-6757.

Penterman, J., Uzawa, R. and Fischer, R. L. (2007) 'Genetic interactions between DNA demethylation and methylation in Arabidopsis', *Plant Physiol.* 145: 1549-1557.
Petrucco, S., Volpi, G., Bolchi, A., Rivetti, C. and Ottonello, S. (2002) 'A nick-sensing DNA 3'-repair enzyme from Arabidopsis', *J Biol Chem* 277(26): 23675-83.
Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E. and Reik, W. (2010) 'Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency', *Nature* 463(7284): 1101-5.
Rai, K. (2008) 'DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and Gadd45', *Cell* 135: 1201-1212.
Rajagopalan, P. T., Zhang, Z., McCourt, L., Dwyer, M., Benkovic, S. J. and Hammes, G. G. (2002) 'Interaction of dihydrofolate reductase with methotrexate: ensemble and single-molecule kinetics', *Proc Natl Acad Sci U S A* 99(21): 13481-6.
Rakic, B., Clarke, J., Tremblay, T. L., Taylor, J., Schreiber, K., Nelson, K. M., Abrams, S. R. and Pezacki, J. P. (2006) 'A small-molecule probe for hepatitis C virus replication that blocks protein folding', *Chem Biol* 13(10): 1051-60.
Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J. and Hol, W. G. (1998) 'Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA', *Science* 279(5356): 1504-13.
Reik, W. (2007) 'Stability and flexibility of epigenetic gene regulation in mammalian development', *Nature* 447: 425-432.
Riechmann, J. L. and Meyerowitz, E. M. (1998) 'The AP2/EREBP family of plant transcription factors', *Biol Chem* 379(6): 633-46.
Ripoll, J. J., Roeder, A. H., Ditta, G. S. and Yanofsky, M. F. (2011) 'A novel role for the floral homeotic gene APETALA2 during Arabidopsis fruit development', *Development* 138(23): 5167-76.
Robert, S., Chary, S. N., Drakakaki, G., Li, S., Yang, Z., Raikhel, N. V. and Hicks, G. R. (2008) 'Endosidin1 defines a compartment involved in endocytosis of the brassinosteroid receptor BRI1 and the auxin transporters PIN2 and AUX1', *Proc Natl Acad Sci U S A* 105(24): 8464-9.
Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics* 26(1): 139-40.
Roudier, F., Teixeira, F. K. and Colot, V. (2009) 'Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more', *Trends Genet* 25(11): 511-7.
Sablowski, R. W. and Meyerowitz, E. M. (1998) 'A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3*/*PISTILLATA*', *Cell* 92(1): 93-103.
Sakamoto, S., Kabe, Y., Hatakeyama, M., Yamaguchi, Y. and Handa, H. (2009) 'Development and application of high-performance affinity beads: toward chemical biology and drug discovery', *Chem Rec* 9(1): 66-85.
Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2002) 'DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression', *Biochem Biophys Res Commun* 290(3): 998-1009.

Salceda, J., Fernandez, X. and Roca, J. (2006) 'Topoisomerase II, not topoisomerase I, is the proficient relaxase of nucleosomal DNA', *EMBO J* 25(11): 2575-83.
Sasaki, H. and Matsui, Y. (2008) 'Epigenetic events in mammalian germ-cell development: reprogramming and beyond', *Nature Rev. Genet.* 9: 129-140.
Schmid, M., Uhlenhaut, N. H., Godard, F., Demar, M., Bressan, R., Weigel, D. and Lohmann, J. U. (2003) 'Dissection of floral induction pathways using global expression analysis', *Development* 130(24): 6001-12.
Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger, O., Schork, N. J. and Ecker, J. R. (2011) 'Transgenerational epigenetic instability is a source of novel methylation variants', *Science* 334(6054): 369-73.
Schmitz, R. J. and Zhang, X. (2011) 'High-throughput approaches for plant epigenomic studies', *Curr Opin Plant Biol* 14(2): 130-6.
Schultz, E. A. and Haughn, G. W. (1993) 'Genetic analysis of the floral initiation process (FLIP) in *Arabidopsis*', *Development* 119: 745-765.
Schwartz, S., Meshorer, E. and Ast, G. (2009) 'Chromatin organization marks exon-intron structure', *Nat Struct Mol Biol* 16(9): 990-5.
Shannon, S. and Meeks-Wagner, D. R. (1993) 'Genetic interactions that regulate inflorescence development in *Arabidopsis*', *Plant Cell* 5: 639-655.
Sharif, J. (2007) 'The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA', *Nature* 450: 908-912.
Sieburth, L. E. and Meyerowitz, E. M. (1997) 'Molecular Dissection of the AGAMOUS Control Region Shows That cis Elements for Spatial Regulation Are Located Intragenically', *Plant Cell* 9(3): 355-365.
Slesarev, A. I., Stetter, K. O., Lake, J. A., Gellert, M., Krah, R. and Kozyavkin, S. A. (1993) 'DNA topoisomerase V is a relative of eukaryotic topoisomerase I from a hyperthermophilic prokaryote', *Nature* 364(6439): 735-7.
Smith, H. M., Boschke, I. and Hake, S. (2002) 'Selective interaction of plant homeodomain proteins mediates high DNA-binding affinity', *Proceedings of the National Academy of Sciences of the United States of America* 99(14): 9579-84.
Smith, L. M. (2007) 'An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis', *Plant Cell* 19: 1507-1521.
Soret, J., Gabut, M., Dupon, C., Kohlhagen, G., Stevenin, J., Pommier, Y. and Tazi, J. (2003) 'Altered serine/arginine-rich protein phosphorylation and exonic enhancer-dependent splicing in Mammalian cells lacking topoisomerase I', *Cancer Res* 63(23): 8203-11.
Sperling, A. S., Jeong, K. S., Kitada, T. and Grunstein, M. (2011) 'Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast', *Proc Natl Acad Sci U S A* 108(31): 12693-8.
Staker, B. L., Feese, M. D., Cushman, M., Pommier, Y., Zembower, D., Stewart, L. and Burgin, A. B. (2005) 'Structures of three classes of anticancer agents bound to the human topoisomerase I-DNA covalent complex', *J Med Chem* 48(7): 2336-45.
Staker, B. L., Hjerrild, K., Feese, M. D., Behnke, C. A., Burgin, A. B., Jr. and Stewart, L. (2002) 'The mechanism of topoisomerase I poisoning by a camptothecin analog', *Proc Natl Acad Sci U S A* 99(24): 15387-92.

Stewart, L., Ireton, G. C. and Champoux, J. J. (1996a) 'The domain organization of human topoisomerase I', *J Biol Chem* 271(13): 7602-8.
Stewart, L., Ireton, G. C., Parker, L. H., Madden, K. R. and Champoux, J. J. (1996b) 'Biochemical and biophysical analyses of recombinant forms of human topoisomerase I', *J Biol Chem* 271(13): 7593-601.
Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G. and Champoux, J. J. (1998) 'A model for the mechanism of human topoisomerase I', *Science* 279(5356): 1534-41.
Stockinger, E. J., Gilmour, S. J. and Thomashow, M. F. (1997) 'Arabidopsis thaliana *CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit', *Proc Natl Acad Sci U S A* 94(3): 1035-40.
Sugimoto-Shirasu, K., Stacey, N. J., Corsar, J., Roberts, K. and McCann, M. C. (2002) 'DNA topoisomerase VI is essential for endoreduplication in Arabidopsis', *Curr Biol* 12(20): 1782-6.
Suzuki, M. M. and Bird, A. (2008a) 'DNA methylation landscapes: provocative insights from epigenomics', *Nature Rev. Genet.* 9: 465-476.
Suzuki, M. M. and Bird, A. (2008b) 'DNA methylation landscapes: provocative insights from epigenomics', *Nat Rev Genet* 9(6): 465-76.
Tadesse, S., Mascarenhas, J., Kosters, B., Hasilik, A. and Graumann, P. L. (2005) 'Genetic interaction of the SMC complex with topoisomerase IV in Bacillus subtilis', *Microbiology* 151(Pt 11): 3729-37.
Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. et al. (2009) 'Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1', *Science* 324(5929): 930-5.
Takahashi, T., Burguiere-Slezak, G., Van der Kemp, P. A. and Boiteux, S. (2011) 'Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in Saccharomyces cerevisiae', *Proc Natl Acad Sci U S A* 108(2): 692-7.
Takahashi, T., Matsuhara, S., Abe, M. and Komeda, Y. (2002) 'Disruption of a DNA topoisomerase I gene affects morphogenesis in Arabidopsis', *Plant Cell* 14(9): 2085-93.
Takuno, S. and Gaut, B. S. (2012) 'Body-methylated genes in Arabidopsis thaliana are functionally important and evolve slowly', *Mol Biol Evol* 29(1): 219-27.
Taneja, B., Patel, A., Slesarev, A. and Mondragon, A. (2006) 'Structure of the N-terminal fragment of topoisomerase V reveals a new family of topoisomerases', *EMBO J* 25(2): 398-408.
Tanizawa, A., Kohn, K. W. and Pommier, Y. (1993) 'Induction of cleavage in topoisomerase I c-DNA by topoisomerase I enzymes from calf thymus and wheat germ in the presence and absence of camptothecin', *Nucleic Acids Res* 21(22): 5157-66.
Team, R. D. C. (2010) R: a language and environment for statistical computing.
Teixeira, F. K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M. F. et al. (2009) 'A role for RNAi in the selective correction of DNA methylation defects', *Science* 323(5921): 1600-4.
Theissen, G. and Saedler, H. (2001) 'Floral quartets', *Nature* 409(6819): 469-71.

Tilly, J. J., Allen, D. W. and Jack, T. (1998) 'The CArG boxes in the promoter of the *Arabidopsis* floral organ identity gene *APETALA3* mediate diverse regulatory effects', *Development* 125(9): 1647-57.
Tse, Y. C., Kirkegaard, K. and Wang, J. C. (1980) 'Covalent bonds between protein and DNA. Formation of phosphotyrosine linkage between certain DNA topoisomerases and DNA', *J Biol Chem* 255(12): 5560-5.
Tse-Dinh, Y. C. (1991) 'Zinc (II) coordination in Escherichia coli DNA topoisomerase I is required for cleavable complex formation with DNA', *J Biol Chem* 266(22): 14317-20.
Vaughn, M. W., Tanurdzic, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P. D., Dedhia, N., McCombie, W. R., Agier, N., Bulski, A. et al. (2007) 'Epigenetic natural variation in Arabidopsis thaliana', *PLoS Biol* 5(7): e174.
Vongs, A., Kakutani, T., Martienssen, R. A. and Richards, E. J. (1993) 'Arabidopsis thaliana DNA methylation mutants', *Science* 260: 1926-1928.
Vos, S. M., Tretter, E. M., Schmidt, B. H. and Berger, J. M. (2011) 'All tangled up: how cells direct, manage and exploit topoisomerase function', *Nat Rev Mol Cell Biol* 12(12): 827-41.
Wagner, D., Sablowski, R. W. and Meyerowitz, E. M. (1999) 'Transcriptional activation of *APETALA1* by *LEAFY*', *Science* 285(5427): 582-4.
Wakem, M. P. and Kohalmi, S. E. (2003) 'Mutation in the ap2-6 allele causes recognition of a cryptic splice site', *Journal of experimental botany* 54(393): 2655-60.
Walfridsson, J., Khorosjutina, O., Matikainen, P., Gustafsson, C. M. and Ekwall, K. (2007) 'A genome-wide role for CHD remodelling factors and Nap1 in nucleosome disassembly', *EMBO J* 26(12): 2868-79.
Wall, M. E., Wani, M.C., Cook, C.E., Palmer, K.H., McPhail, A.T., Sim, G.A. (1966) 'Plant anti-tumor agents: the isolation and structure of campothecin, a novel alkaloidal leukemia and tumor inhibitor from *Camptotheca acuminata*.', *J Am Chem Soc* 88.
Wang, Z. and Droge, P. (1996) 'Differential control of transcription-induced and overall DNA supercoiling by eukaryotic topoisomerases in vitro', *EMBO J* 15(3): 581-9.
Wassenegger, M., Heimes, S., Riedel, L. and Sanger, H. L. (1994) 'RNA-directed de novo methylation of genomic sequences in plants', *Cell* 76: 567-576.
Weigel, D. (1995) 'The APETALA2 domain is related to a novel type of DNA binding domain', *Plant Cell* 7(4): 388-9.
Whitehouse, C. J., Taylor, R. M., Thistlethwaite, A., Zhang, H., Karimi-Busheri, F., Lasko, D. D., Weinfeld, M. and Caldecott, K. W. (2001) 'XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair', *Cell* 104(1): 107-17.
Wierzbicki, A. T., Haag, J. R. and Pikaard, C. S. (2008a) 'Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes', *Cell* 135(4): 635-48.
Wierzbicki, A. T., Haag, J. R. and Pikaard, C. S. (2008b) 'Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes', *Cell* 135: 635-648.

William, D. A., Su, Y., Smith, M. R., Lu, M., Baldwin, D. A. and Wagner, D. (2004) 'Genomic identification of direct target genes of *LEAFY*', *Proc Natl Acad Sci U S A* 101(6): 1775-80.
Williams, K., Christensen, J. and Helin, K. (2012) 'DNA methylation: TET proteins-guardians of CpG islands?', *EMBO Rep* 13(1): 28-35.
Wilson, T. M., Chen, A. D. and Hsieh, T. (2000) 'Cloning and characterization of Drosophila topoisomerase IIIbeta. Relaxation of hypernegatively supercoiled DNA', *J Biol Chem* 275(3): 1533-40.
Won, S. Y., Li, S., Zheng, B., Zhao, Y.Y., Li, D., Zhao, X., Yi, H., Gao, L., Dinh, T.T. and Chen, X. (2012) 'Development of a luciferase-based reporter of transcriptional gene silencing that enables bidirectional mutant screening in *Arabidopsis thaliana*', *Silence*.
Woo, H. R., Dittmer, T. A. and Richards, E. J. (2008) 'Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis', *PLoS Genet.* 4: e1000156.
Woo, H. R., Pontes, O., Pikaard, C. S. and Richards, E. J. (2007) 'VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization', *Genes Dev.* 21: 267-277.
Wood, J. G., Rogina, B., Lavu, S., Howitz, K., Helfand, S. L., Tatar, M. and Sinclair, D. (2004) 'Sirtuin activators mimic caloric restriction and delay ageing in metazoans', *Nature* 430(7000): 686-9.
Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E. and Zhang, Y. (2011a) 'Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells', *Genes Dev* 25(7): 679-84.
Wu, H., D'Alessio, A. C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y. E. and Zhang, Y. (2011b) 'Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells', *Nature* 473(7347): 389-93.
Wu, H. and Zhang, Y. (2011) 'Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation', *Genes Dev* 25(23): 2436-52.
Würschum, T., Gross-Hardt, R. and Laux, T. (2006) '*APETALA2* regulates the stem cell niche in the *Arabidopsis* shoot meristem', *Plant Cell* 18(2): 295-307.
Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A. J., Zheng, L., Zhang, H., Huang, S. et al. (2011) 'Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells', *Mol Cell* 42(4): 451-64.
Yang, J., Shamji, A., Matchacheep, S. and Schreiber, S. L. (2007) 'Identification of a small-molecule inhibitor of class Ia PI3Ks with cell-based screening', *Chem Biol* 14(4): 371-7.
Yang, X., Yu, Y., Jiang, L., Lin, X., Zhang, C., Ou, X., Osabe, K. and Liu, B. (2012) 'Changes in DNA methylation and transgenerational mobilization of a transposable element (mPing) by the topoisomerase II inhibitor, etoposide, in rice', *BMC Plant Biol* 12(1): 48.
Yant, L., Mathieu, J., Dinh, T. T., Ott, F., Lanz, C., Wollmann, H., Chen, X. and Schmid, M. (2010) 'Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2', *The Plant cell* 22(7): 2156-70.

Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G. R., Janse, C. J., Waters, A. P., Kato, T. and Kaneko, I. (2009) 'Identification of a transcription factor in the mosquito-invasive stage of malaria parasites', *Mol Microbiol* 71(6): 1402-14.
Zabotina, O., Malm, E., Drakakaki, G., Bulone, V. and Raikhel, N. (2008) 'Identification and preliminary characterization of a new chemical affecting glucosyltransferase activities involved in plant cell wall biosynthesis', *Mol Plant* 1(6): 977-89.
Zechiedrich, E. L. and Osheroff, N. (1990) 'Eukaryotic topoisomerases recognize nucleic acid topology by preferentially interacting with DNA crossovers', *EMBO J* 9(13): 4555-62.
Zemach, A., McDaniel, I. E., Silva, P. and Zilberman, D. (2010) 'Genome-wide evolutionary analysis of eukaryotic DNA methylation', *Science* 328(5980): 916-9.
Zhang, H., Deng, X., Miki, D., Cutler, S., La, H., Hou, Y. J., Oh, J. and Zhu, J. K. (2012) 'Sulfamethazine suppresses epigenetic silencing in Arabidopsis by impairing folate synthesis', *Plant Cell* 24(3): 1230-41.
Zhang, H., Meng, L. H., Zimonjic, D. B., Popescu, N. C. and Pommier, Y. (2004) 'Thirteen-exon-motif signature for vertebrate nuclear and mitochondrial type IB topoisomerases', *Nucleic Acids Res* 32(7): 2087-92.
Zhang, X. (2006) 'Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis', *Cell* 126: 1189-1201.
Zhang, X., Shiu, S. H., Cal, A. and Borevitz, J. O. (2008) 'Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays', *PLoS Genet* 4(3): e1000032.
Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E. et al. (2006) 'Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis', *Cell* 126(6): 1189-201.
Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J. Y. and Chen, X. (2009) 'Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis', *Genes Dev* 23(24): 2850-60.
Zhu, C. X., Qi, H. Y. and Tse-Dinh, Y. C. (1995) 'Mutation in Cys662 of Escherichia coli DNA topoisomerase I confers temperature sensitivity and change in DNA cleavage selectivity', *J Mol Biol* 250(5): 609-16.
Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. and Henikoff, S. (2007) 'Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription', *Nature Genet.* 39: 61-69.
Zunino, F., Dallavalleb, S., Laccabuea, D., Berettaa, G., Merlinib, L. and Pratesi, G. (2002) 'Current status and perspectives in the development of camptothecins', *Curr Pharm Des* 8(27): 2505-20.

Figure 2.1 *LUCL* is silenced by DNA methylation

(A) Diagram of the reporter construct

(B) Basal LUC activity levels of *LUCL*, *LUCH*. Col-0 was present as the negative control. Seedlings of the indicated transgenes were plated on MS agar plates and placed in the growth chamber for 10 days under continuous light. The seedlings were then sprayed with luciferin substrate imaged with a CCD camera to determine qualitative LUC activity levels. *LUCH* has much higher LUC activity than *LUCL* as indicated by the visible luminescence pictured. *LUCL* basal LUC activity is practically non-existent, as indicated by the lack of luminescence observed.

(C) McrBC PCR-based methylation assay. The (+) gels are DNA treated with McrBC. The (-) gels are DNA treated in the same manner as the (+) gels except no McrBC enzyme was added. *At2g19920* was used as an unmethylated internal control.

(D) Ten day old *LUCL*, *LUCH*, and Col-0 seedlings treated with 5-aza-2'-deoxycytidine, then luminescence was observed.

(E) RT-PCR of mock-treated and 5-aza-2'-deoxycytidine-treated *LUCL* and *LUCH*. The *LUC* and *NPTII* genes were examined. *UBQ* served as an internal loading control.

(F) Levels of DNA methylation of *LUCH* and *LUCL* at the 35S promoter region as determined by bisulfite sequencing.

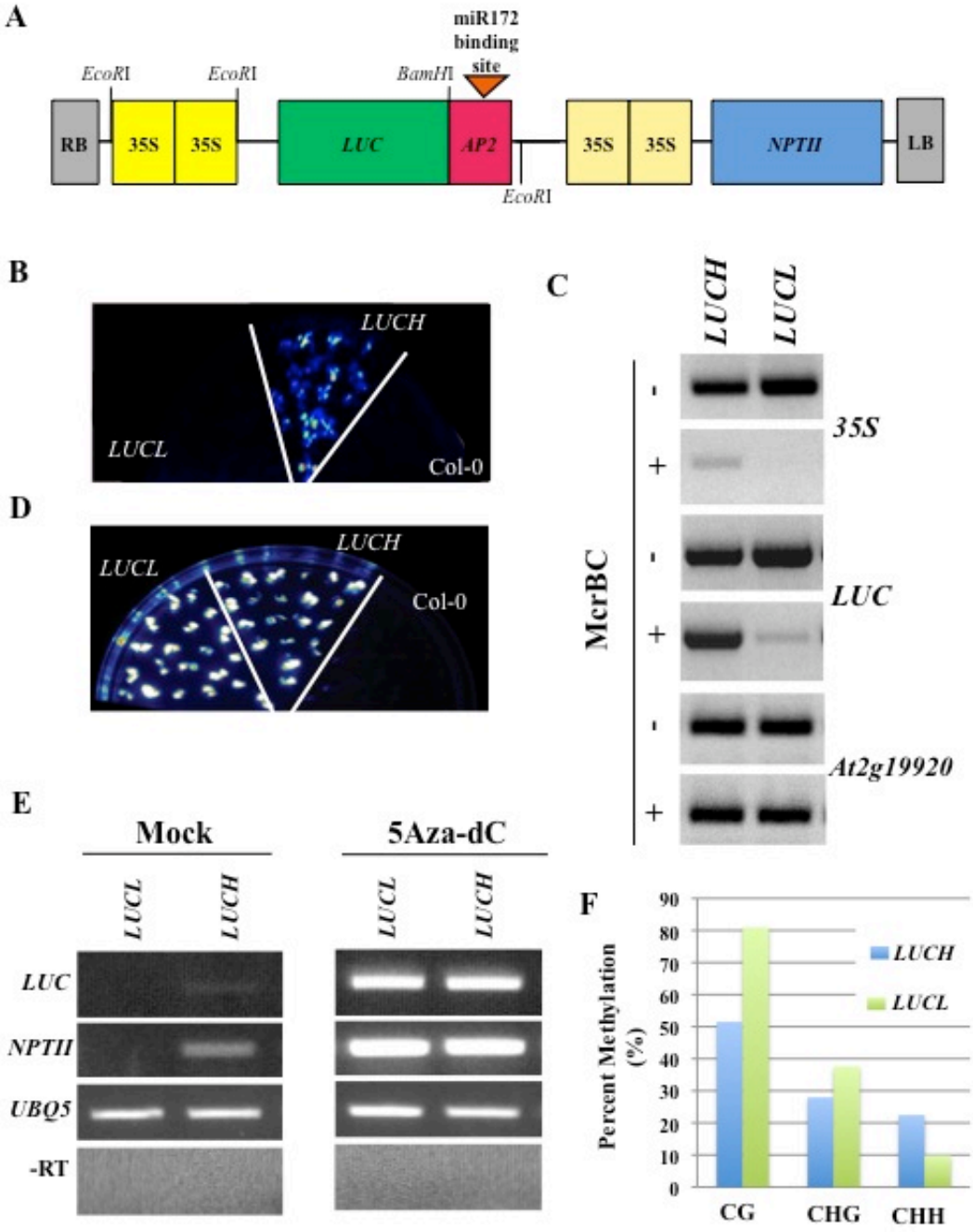

A
miR172 binding site
EcoRI
EcoRI
BamHI
EcoRI
RB
35S
35S
LUC
AP2
35S
35S
NPTII
LB
B
LUCH
LUCL
Col-0
C
LUCH
LUCL
-
+
35S
-
+
LUC
-
+
At2g19920
McrBC
D
LUCH
LUCL
Col-0
E
Mock
5Aza-dC
LUCL
LUCH
LUCL
LUCH
LUC
NPTII
UBQ5
-RT
F
Percent Methylation (%)
90
80
70
60
50
40
30
20
10
0
LUCH
LUCL
CG
CHG
CHH

Figure 2.2 Conversion rates of the bisulfite sequencing experiments in Fig 2.1F

(A) Using chloroplast DNA-specific primers (Table 2), a bisulfite sequencing reaction was performed to determine the conversion rate of our samples, as chloroplast DNA is unmethylated. The graph depicts the bisulfite conversion rate of *LUCH* and *LUCL* in the different cytosine methylation contexts.

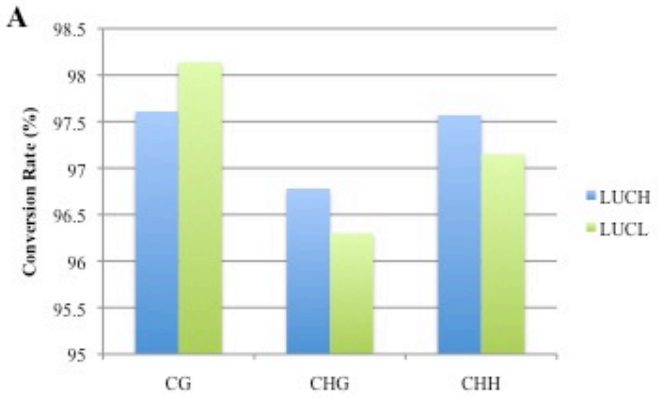

A
98.5
98
97.5
97
96.5
96
95.5
95
Conversion Rate (%)
CG
CHG
CHH
LUCH
LUCL

Figure 2.3 *LUCL* is a multi-copy, single insertion transgene

(A) Diagram of *LUCL* as a multi-copy transgene. Restriction sites and distances between sites are noted.

(B) Southern blots of *LUCL*, Col-0 and *LUCH* (increasing genomic amounts; first lane is with equal to *LUCL* whereas the second lane contains 1.5 times than that of *LUCL* (gray triangle)) with *BamH*I. 50ng of probe was made with primers specific to *LUC* (red line). The 5.5 kb band present in *LUCL* (arrow) represents the possibility of a multi-copy transgene as the distance between the two *BamH*I sites is 5.427 kb (we do not know the distance between the LB and RB (question mark).

(C) Southern blots of *LUCL*, Col-0 and *LUCH* (increasing genomic amounts; first lane is with equal to *LUCL* whereas the second lane contains 1.5 times than that of *LUCL* (gray triangle)) with *EcoR*I. 50ng of probe was made with primers specific to *LUC*. The 2.1 kb band is highlighted with a red arrow. The intensity of the 2.1 kb band in *LUCL* is much higher than in *LUCH*.

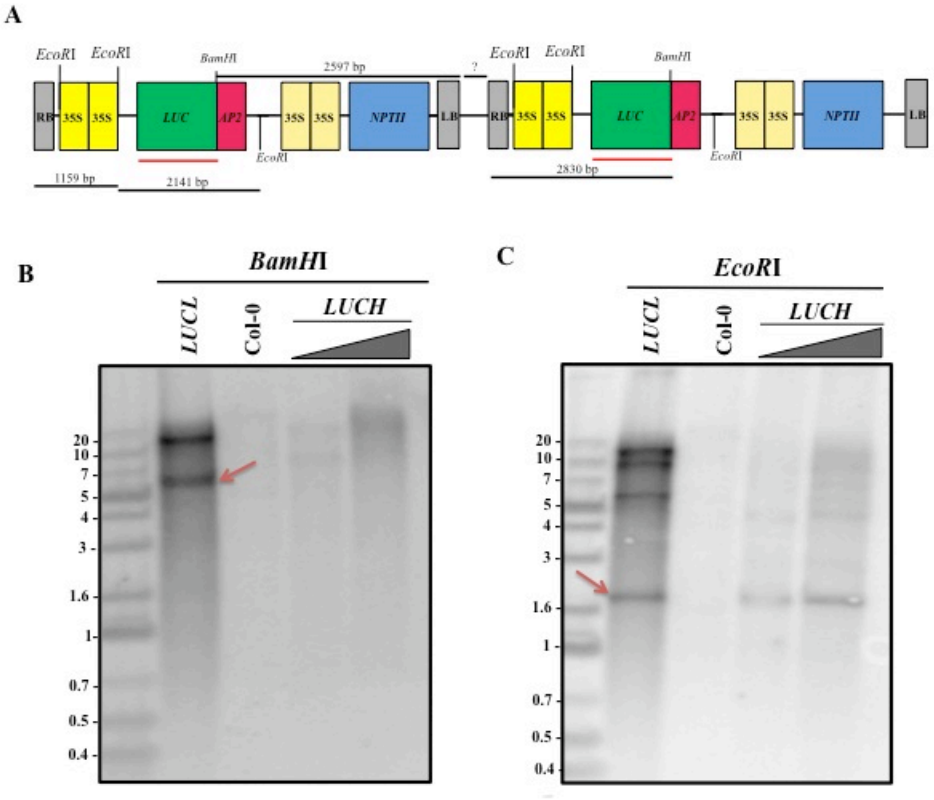

A
EcoRI
EcoRI
BamHI
2597 bp
?
EcoRI
EcoRI
BamHI
RB
35S
35S
LUC
AP2
35S
35S
NPTII
LB
RB
35S
35S
LUC
AP2
35S
35S
NPTII
LB
EcoRI
EcoRI
1159 bp
2141 bp
2830 bp
B
BamHI
LUCL
Col-0
LUCH
20
10
7
5
4
3
1.6
1
0.7
0.5
0.4
C
EcoRI
LUCL
Col-0
LUCH
20
10
7
5
4
3
1.6
1
0.7
0.5
0.4

Figure 2.4 MTX releases methlyation of *LUCL*

(A) DMSO –treated *LUCL* seedlings.

(B-H) The less-active form of MTX was added to *LUCL*. The concentration added is indicated on the bottom left hand corner of each figure.

(I-P) The racemic mixture of less-active and active forms of MTX was added to *LUCL*. The concentration added is indicated on the bottom left hand corner of each figure.

(Q-X) The active form of MTX was added to *LUCL*. The concentration added is indicated on the bottom left hand corner of each figure.

(Y) Chemical structures of the active (top) and less-active (bottom) forms of MTX. The red circle indicates the chirality of the molecule between the two forms.

(Z) McrBC-PCR-based methylation assay of *LUCL* seedlings treated with the less-active MTX. DC=DMSO-treated Col-O control, D= DMSO-treated *LUCL*. The gray triangle represents increasing concentrations of MTX added from left to right.

(AA) MTX inhibits S-adenosyl methionine (SAM) biosynthesis to indirectly affect gene silencing via DNA methylation. MTX inhibits the conversion of DHF to THF. Under normal circumstances, the energy given off by the conversion of THF to 5-methyl THF catalyzes the reaction of homocysteine to methionine in the presence of Vitamin B12. SAM is used to deposit a methyl group at the 5-position of cytosine to generate 5-Methylcytosine, which subsequently promotes gene silencing.

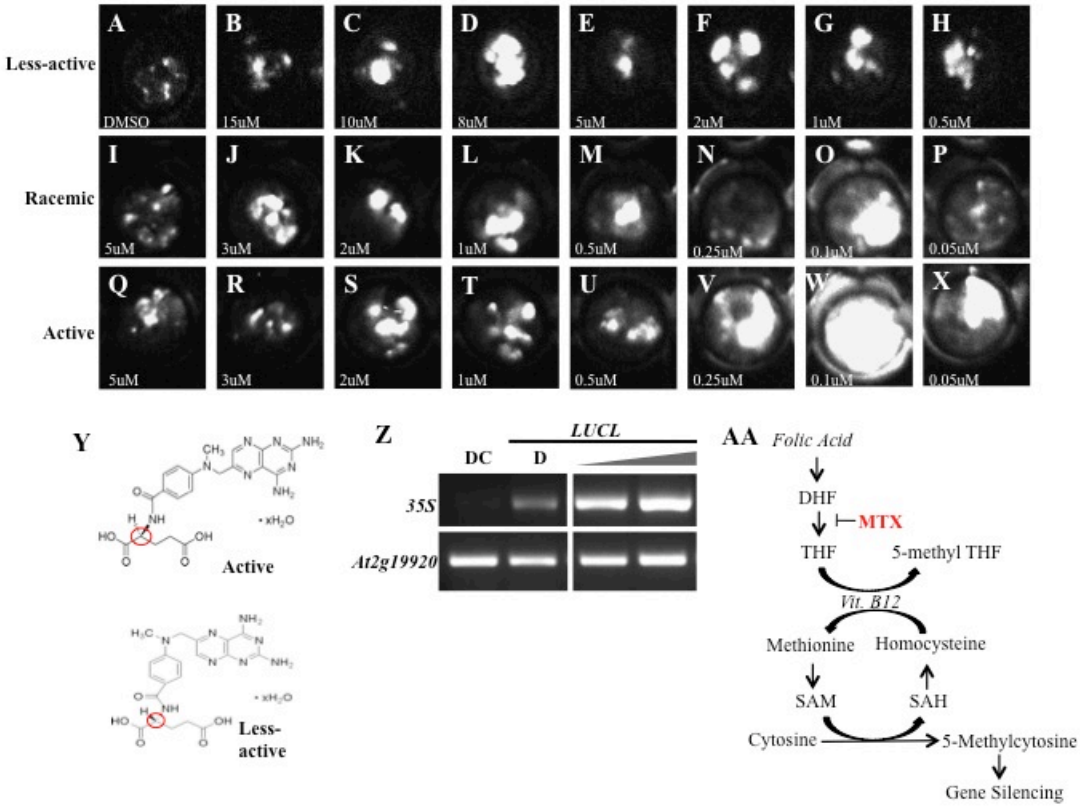

Less-active
A
DMSO
B
15uM
C
10uM
D
8uM
E
5uM
F
2uM
G
1uM
H
0.5uM
Racemic
I
5uM
J
3uM
K
2uM
L
1uM
M
0.5uM
N
0.25uM
O
0.1uM
P
0.05uM
Active
Q
5uM
R
3uM
S
2uM
T
1uM
U
0.5uM
V
0.25uM
W
0.1uM
X
0.05uM
Y
Active
Less-active
Z
LUCL
DC
D
35S
At2g19920
AA
Folic Acid
DHF
MTX
THF
5-methyl THF
Vit. B12
Methionine
Homocysteine
SAM
SAH
Cytosine
5-Methylcytosine
Gene Silencing

Figure 2.5 CPT releases methylation of *LUCL*

(A) CPT affects *LUCL* luciferase luminescence in a concentration-dependent manner.

(B) The chemical structure of CPT.

(C) RT-PCR analysis of CPT-treated *LUCL* seedlings showed an increase of *LUC* expression upon chemical addition. *115-1H* is an *ago4* allele in the *LUCH* background and serves as a positive control (Won, 2012). *UBQ* was used as an internal control for quantification purposes. 30 and 25 cycles of *LUC* and *UBQ,* respectively, were performed.

(D) Real-Time RT-PCR results of the samples used in (C). Three biological replicates were done, all of which showed the same trend. A representative image from one biological replicate is shown. Error bars represent the standard deviation from three technical replicates.

(E) Measurement of LUC activity levels over time using an automated LUC counter machine. CPT specifically affects the activity of *LUCL* but not *LUCH*, or *PM*, a miRNA reporter line and *PM-empty vector* (the vector used in *PM* without the miRNA binding site). The numbers on the x-axis indicate hours: minutes.

(F) Analysis of time- and concentration- dependent de-repression of *LUCL* by CPT. The left hand side shows the time in which the seedling was incubated with CPT. H= hour, D= Day. For the top (of the figure), D= DMSO-treated *LUCL*, the blue triangle indicates the concentration added from greatest to least (left to right, respectively) and DC= DMSO-treated Col-0 control.

(G) McrBC-PCR-based methylation analysis of the *35S* promoter in *LUCL* seedlings treated with CPT. *At2g19920* served as an internal unmethylated control. D= DMSO-treated *LUCL* seedlings and the blue triangle indicates the concentration added from greatest to least (left to right, respectively). DC= DMSO-treated Col-0 control.

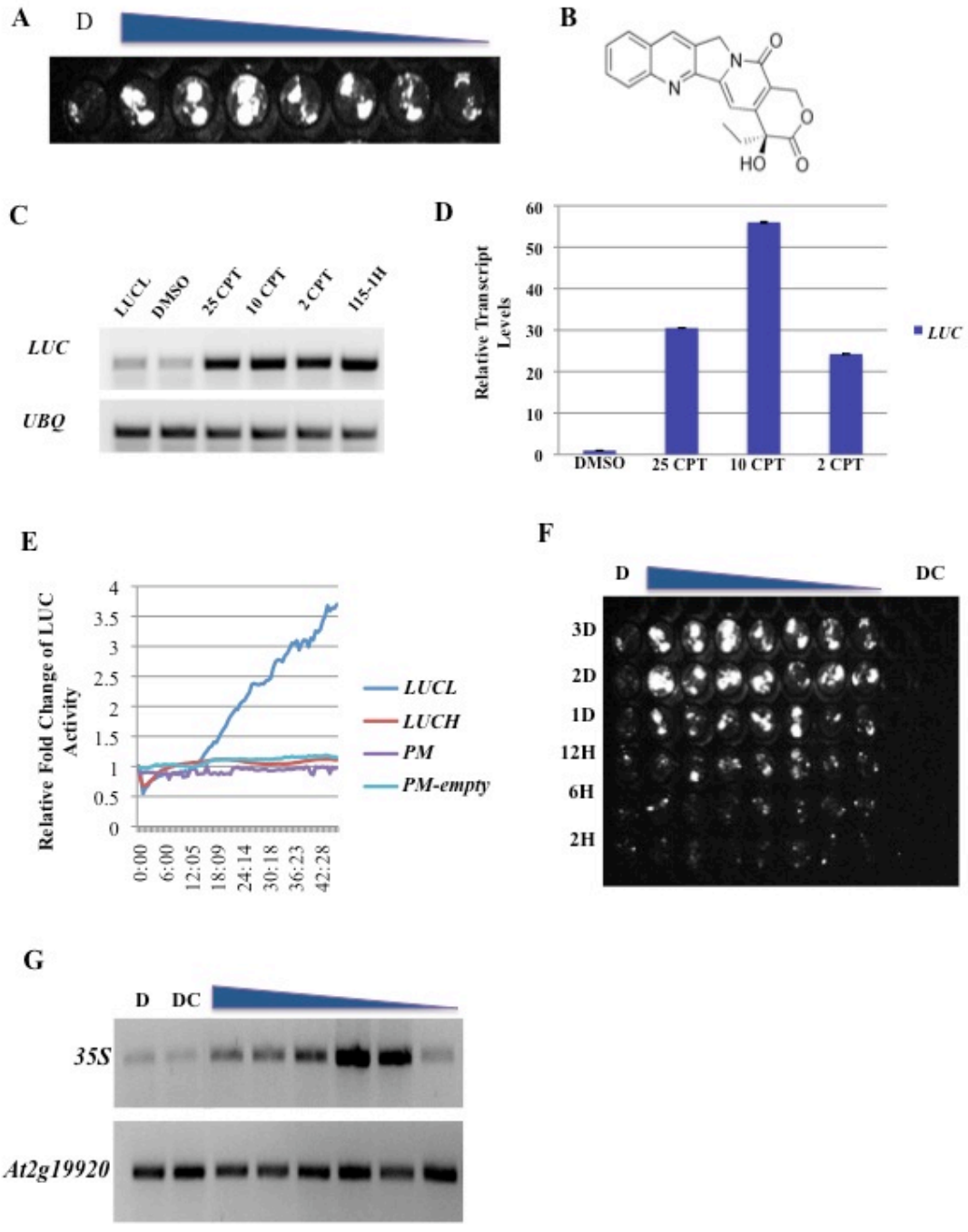

A
D
B
C
LUCL
DMSO
25 CPT
10 CPT
2 CPT
115-1H
LUC
UBQ
D
Relative Transcript Levels
60
50
40
30
20
10
0
DMSO
25 CPT
10 CPT
2 CPT
LUC
E
Relative Fold Change of LUC Activity
4
3.5
3
2.5
2
1.5
1
0.5
0
0:00
6:00
12:05
18:09
24:14
30:18
36:23
42:28
LUCL
LUCH
PM
PM-empty
F
D
DC
3D
2D
1D
12H
6H
2H
G
D
DC
35S
At2g19920

Figure 2.6 *topIα-2* does not release LUC activity in *LUCL*

topIα-2^{Col} was crossed to *LUCL*. *topIα-2*^{Col} is the same mutant allele as *topIα-2* except that it has been back-crossed with Col-0 five times to make it suitable for crossing into *LUCL*. No substantial de-repression of LUC activity was observed in the mutant containing the transgene (*topIα-2*^{Col} *LUCL*) as compared to *LUCL*. *LUCL*/+ indicates that the transgene is hemizygous whereas "*LUCL*" indicates that the transgene is homozygous. The *ago4* (*LUCH*) serves as a "high-line" control.

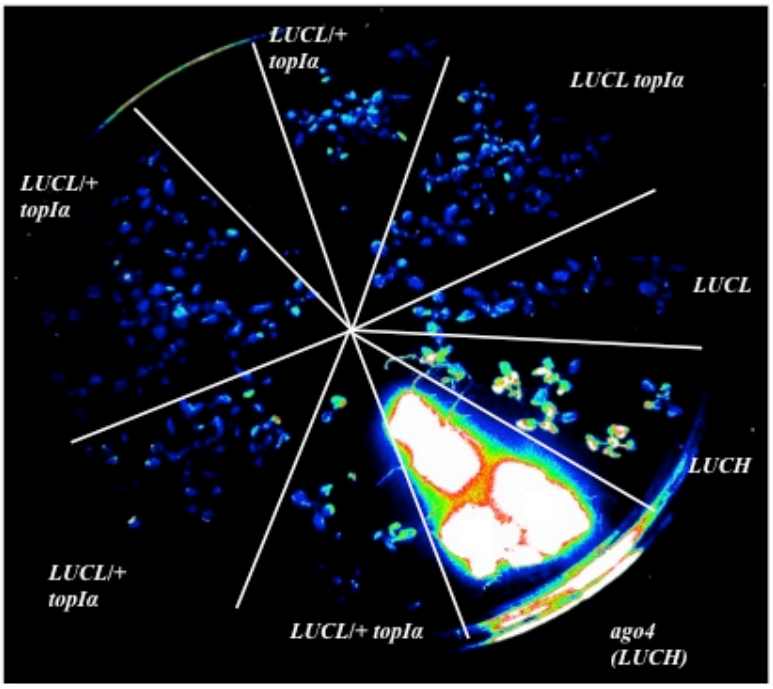

LUCL/+
topIα
LUCL topIα
LUCL/+
topIα
LUCL
LUCH
LUCL/+
topIα
LUCL/+ topIα
ago4
(LUCH)

Figure 2.7 Loss of *TOPIα* does not affect siRNA or miRNA accumulation levels, but does de-repress transposon silencing and slightly affects Pol V dependent transcript levels in ten-day old seedlings.

(A) siRNA and miRNA levels are not affected in the *topIα* mutant. Loss of *TOPIα* did not significantly change siRNA (*cluster4, SoloLTR, sir1003*) and miRNA (*miR173*) levels. *U6* was used as an internal loading control. The numbers indicate the relative abundance of the small RNAs in the mutant (with that in the wild type set to 1.0).

(B) Real-Time RT-PCR analysis indicates that loss of *TOPIα* or addition of CPT results in siRNA target loci de-repression. The loci tested are *Cluster4, AtCopia, IG/LINE* and *AtMuI*. At least four different biological replicates were done, all of which showed the same trend. A representative image is shown. Error bars represent the standard deviation from three technical replicates in one biological replicate.

(C) Real-Time RT-PCR analysis indicates that *TOPIα* slightly contributes to the production of Pol V-dependent transcripts at *MEA-ISR*. *nrpe* is a Pol V mutant and was used as a positive control to be compared to Col-0, whereas *topIα* should be compared to L*er*. Error bars represent the standard deviation from three technical replicates.

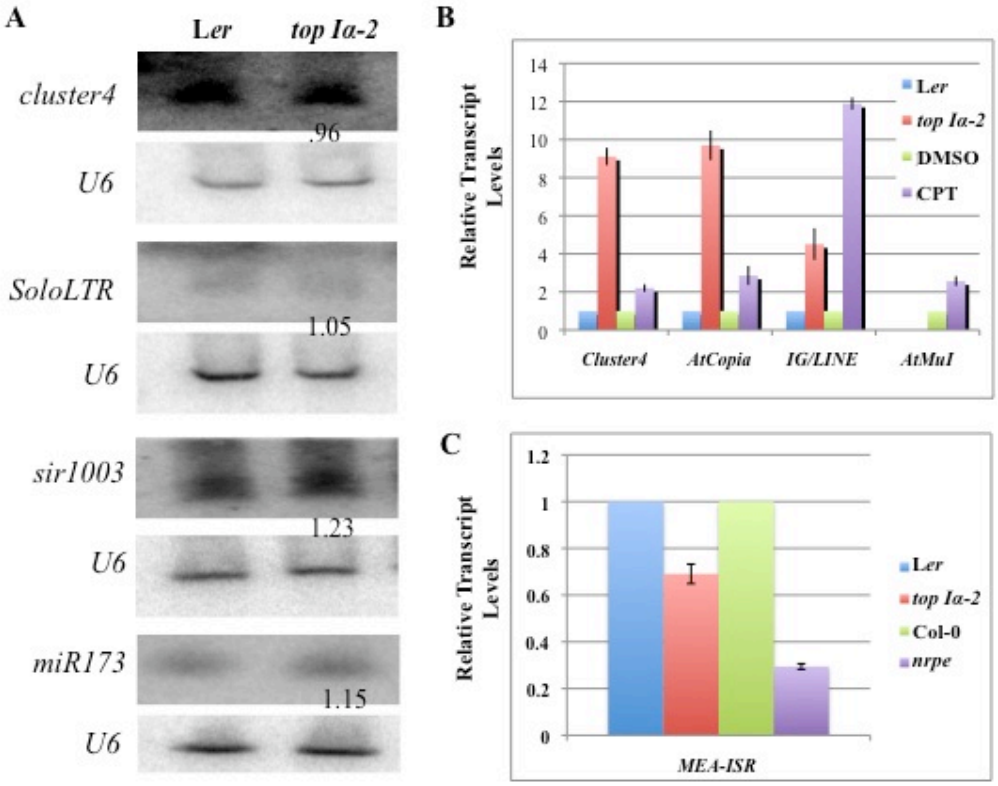

A
Ler
top Iα-2
cluster4
.96
U6
SoloLTR
1.05
U6
sir1003
1.23
U6
miR173
1.15
U6
B
Relative Transcript Levels
14
12
10
8
6
4
2
0
Cluster4
AtCopia
IG/LINE
AtMu1
Ler
top Iα-2
DMSO
CPT
C
Relative Transcript Levels
1.2
1
0.8
0.6
0.4
0.2
0
MEA-ISR
Ler
top Iα-2
Col-0
nrpe

Figure 2.8 Genome-wide analysis of small RNA accumulation in *topIα*

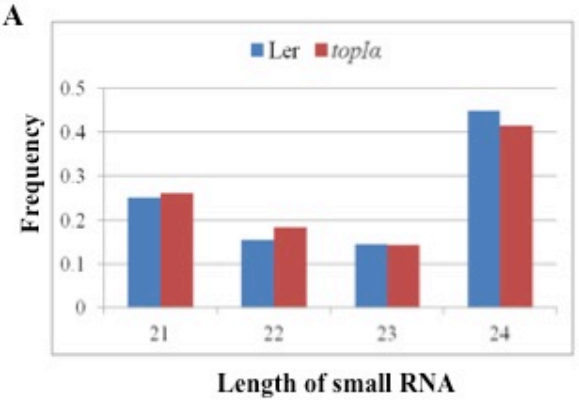
(A) The size distribution of total small RNA reads in L*er* and *topIα* is largely identical.

(B) Analysis of DSRs (Differentially Accumulated small RNAs) in L*er* and *topIα*. .The table gives a brief overview and breakdown of the small number of DSRs (37 in total). TE= transposable element; L24 or L21= Length of small RNA (24 nt or 21 nt).

(C) Pie charts of DSR distribution for the increased DSRs in *topIα*.

(D) Pie charts of DSR distribution for the decreased DSRs in *topIα*.

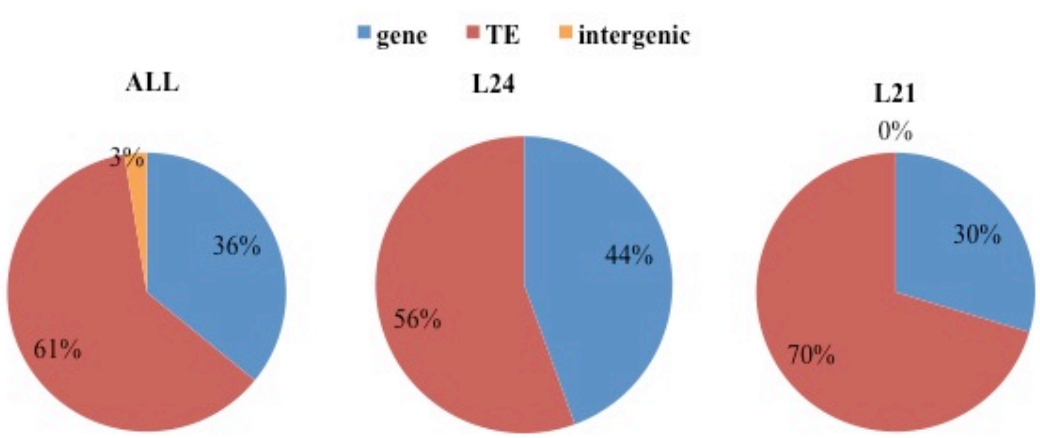
A



B

		All	L24	L21
DSR		37	10	25
gene+TE	all	38	10	27
	unique	36	10	25
	overlapping	2	0	2
gene		14	4	8
TE		24	5	19
intergenic		1	0	0
total		39	9	27

C



D

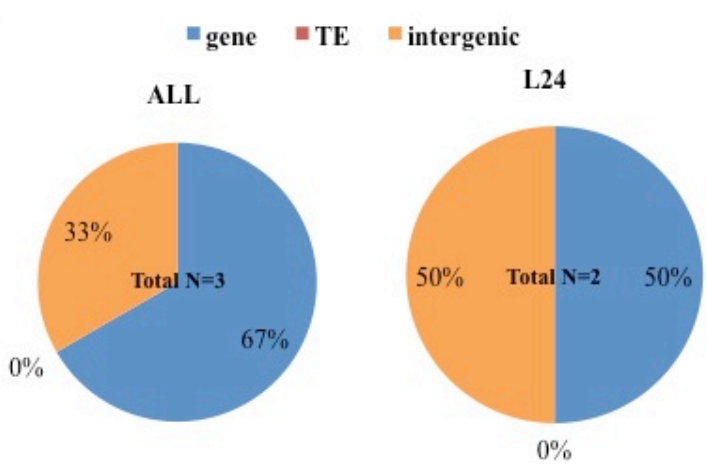


Figure 2.9 The *topIα* mutation affects DNA methylation at thousands of genomic loci

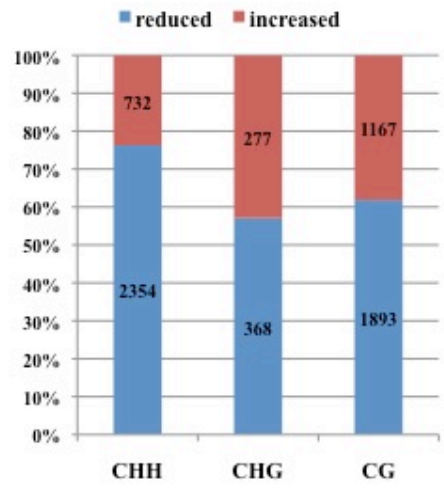
(A) The number of DMRs (Differentially Methylated Regions) between wild type and *topIα* in the three contexts. Whole-genome bisulfite sequencing was performed using ten-day old seedling tissue. Blue bars represent the number of loci with reduced DNA methylation in *topIα*, whereas the red bars represent the number of loci with "increased" DNA methylation in *topIα*.

(B) The region-specific breakdown of DMRs with increased DNA methylation in *topIα* in the three contexts (CHH, CHG and CG).

(C) The region-specific breakdown of DMRs with reduced DNA methylation in *topIα* in the three contexts (CHH, CHG and CG).

(D) Top panel: the pie chart of the region-specific breakdown in (B). Bottom panel: the pie chart of the region-specific breakdown in (C). blue= gene region, red= transposable element, orange=intergenic region.

A



B

		CHH	CHG	CG
DMR		732	277	1167
gene+TE	all	630	213	1001
	unique	619	203	983
	overlapping	11	10	18
gene		50	56	595
TE		580	157	406
intergenic		113	74	184
total		743	287	1185

C

		CHH	CHG	CG
DMR		2354	368	1893
gene+TE	all	2130	284	1723
	unique	2053	269	1678
	overlapping	77	15	45
gene		211	85	1185
TE		1919	199	538
intergenic		301	99	215
total		2431	383	1938

D

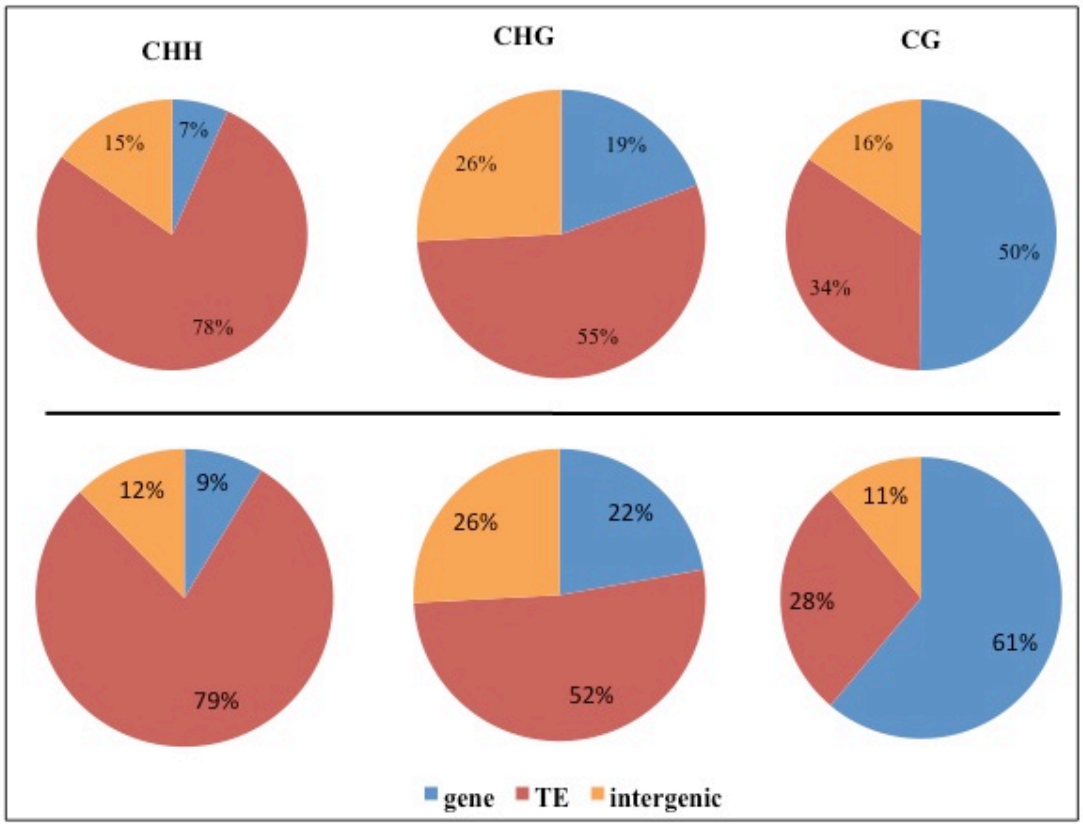


Figure 2.10 Box plots of levels of DNA methylation at the DMRs between L*er* and *topIα*

Top panel: DNA methylation density for DMRs with increased DNA methylation in *topIα* as compared to L*er* for the three cystosine methylation contexts. Bottom panel: Levels of DNA methylation DMRs with reduced DNA methylation in *topIα* as compared to L*er* for the three cystosine methylation contexts for.

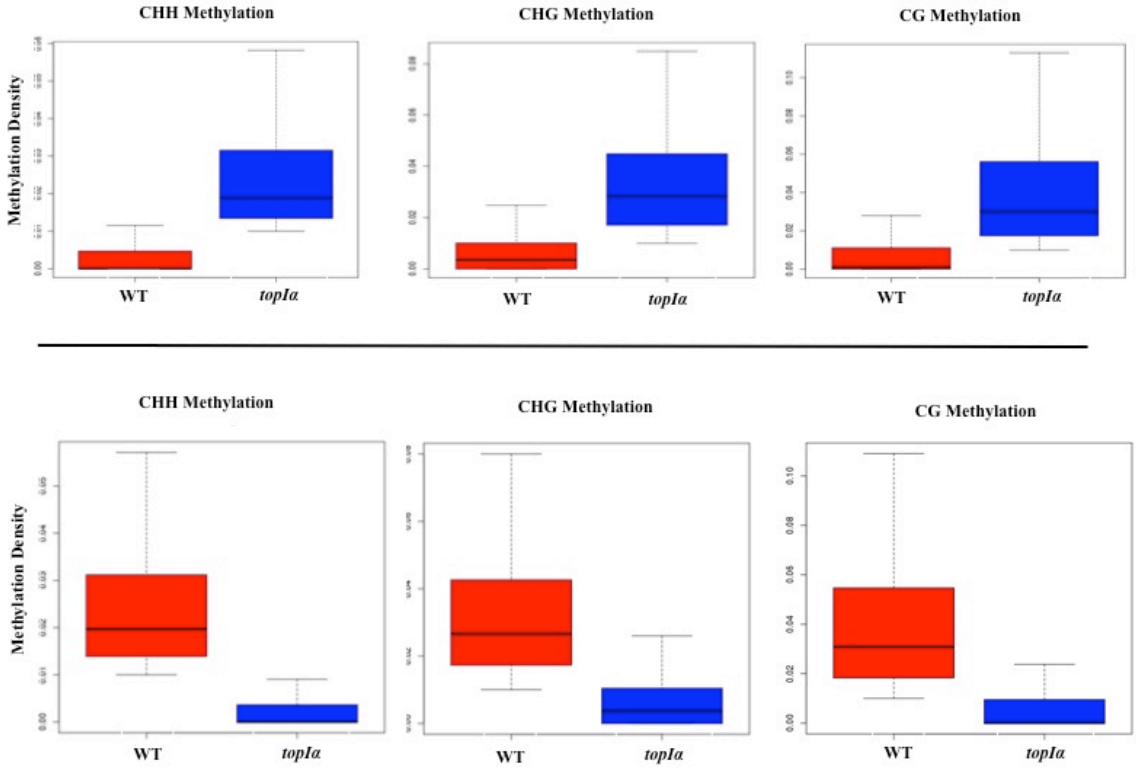

CHH Methylation
CHG Methylation
CG Methylation
Methylation Density
WT
topIα
WT
topIα
WT
topIα
CHH Methylation
CHG Methylation
CG Methylation
Methylation Density
WT
topIα
WT
topIα
WT
topIα

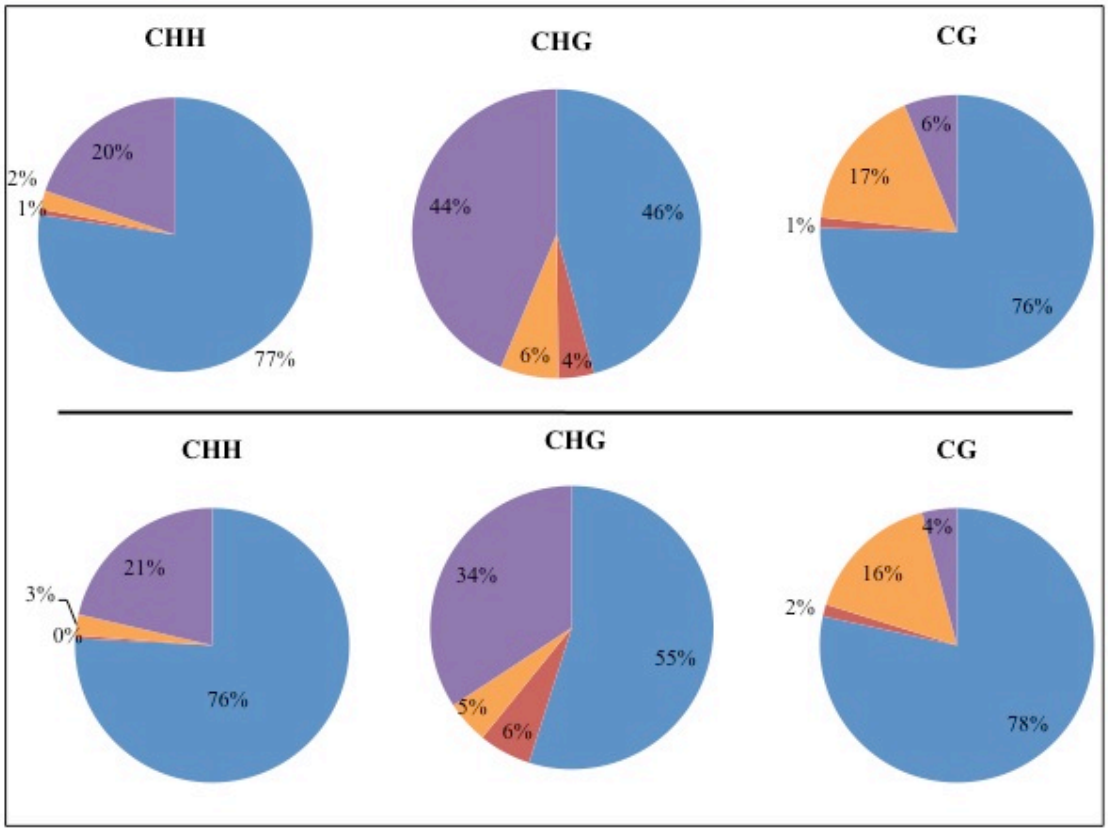
Figure 2.11 *topIα* DMRs do not significantly overlap with DMRs from Pol IV or Pol V mutants

topIα DMRs (either increased or decreased in DNA methylation relative to L*er*) were compared to *sde4-3* or *nrpe1-1* DMRs. *sde4-3* is a Pol IV mutant whereas *nrpe1-1* is a Pol V mutant. Their respective DMRs are as compared to Col-0 and are reduced in DNA methylation relative to Col-0.

(A) Top panel: Pie-chart of the percentages of *topIα* DMRs (with increased DNA methylation in the mutant) that did or did not overlap with Pol IV and/or Pol V DMRs in the three cystosine contexts. Bottom panel: Pie chart of the percentages of *topIα* DMRs (with reduced DNA methylation in the mutant) that did or did not overlap with Pol-IV and/or Pol-V in the three cystosine contexts.

(B) Left table: breakdown of the raw numbers for *topIα* DMRs with increased DNA methylation in the mutant. These are the numbers by which the pie chart (A, top panel) is derived. Right table: breakdown of the raw numbers for *topIα* DMRs with reduced DNA methylation in the mutant. These are the numbers by which the pie chart (A, bottom panel) is derived.

A



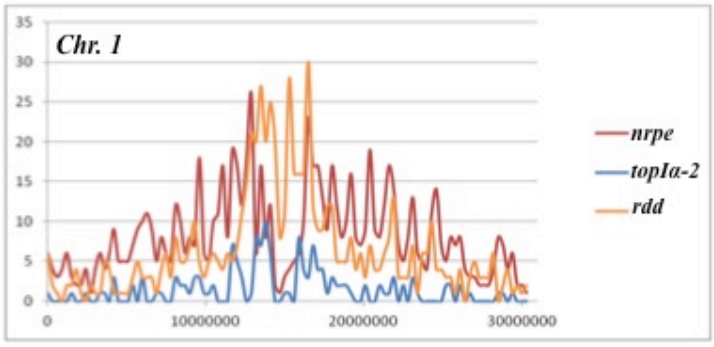
B

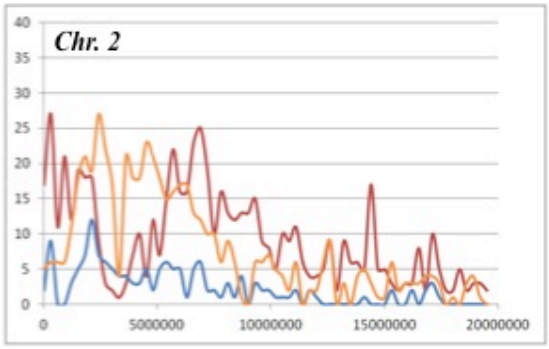
	CHH	CHG	CG
top Iα	566	127	881
top Iα & sde4-3	5	11	14
top Iα & nrpe1-1	16	18	199
top Iα & sde4-3 & nrpe1-1	145	121	73
total	732	277	1167

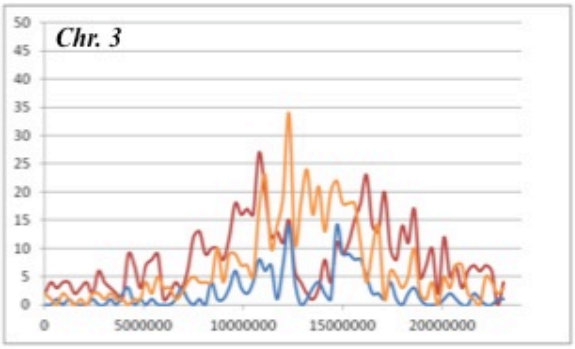
	CHH	CHG	CG
top Iα	1785	202	1484
top Iα & sde4-3	8	22	27
top Iα & nrpe1-1	58	18	304
top Iα & sde4-3 & nrpe1-1	503	126	78
total	2354	368	1893

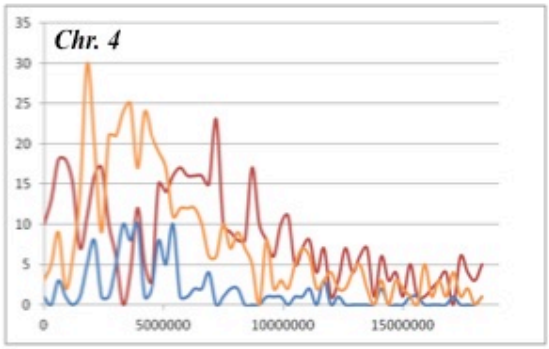
Figure 2.12 The distribution of CHH DMR density along the five chromosomes

The five boxes represent the distribution of DMRs increased in methylation in *topIα* (blue line) along the five chromosomes (as indicated on the top left corner). *nrpe* is a Pol V mutant and its DMR (decreased in DNA methylation) distribution is also shown (red line). *rdd* is the *ros1 dml2 dml3* triple mutant and its DMR (increased in DNA methylation) distribution is shown (orange line).


Chr. 1
nrpe
topIα-2
rdd
0
10000000
20000000
30000000


Chr. 2
0
5000000
10000000
15000000
20000000


Chr. 3
0
5000000
10000000
15000000
20000000


Chr. 4
0
5000000
10000000
15000000

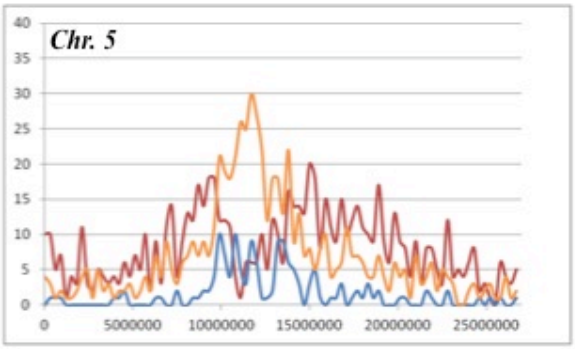

Chr. 5
0
5000000
10000000
15000000
20000000
25000000

Figure 2.13 The distribution of CHH DMR density along the five chromosomes

The five boxes represent the distribution of DMRs reduced in methylation in *topIα* (blue line) along the five chromosomes (as indicated on the top left corner). *nrpe* is a Pol V mutant and its DMR (decreased in DNA methylation) distribution is also shown (red line). *rdd* is the *ros1 dml2 dml3* triple mutant and its DMR (increased in DNA methylation) distribution is shown (orange line).

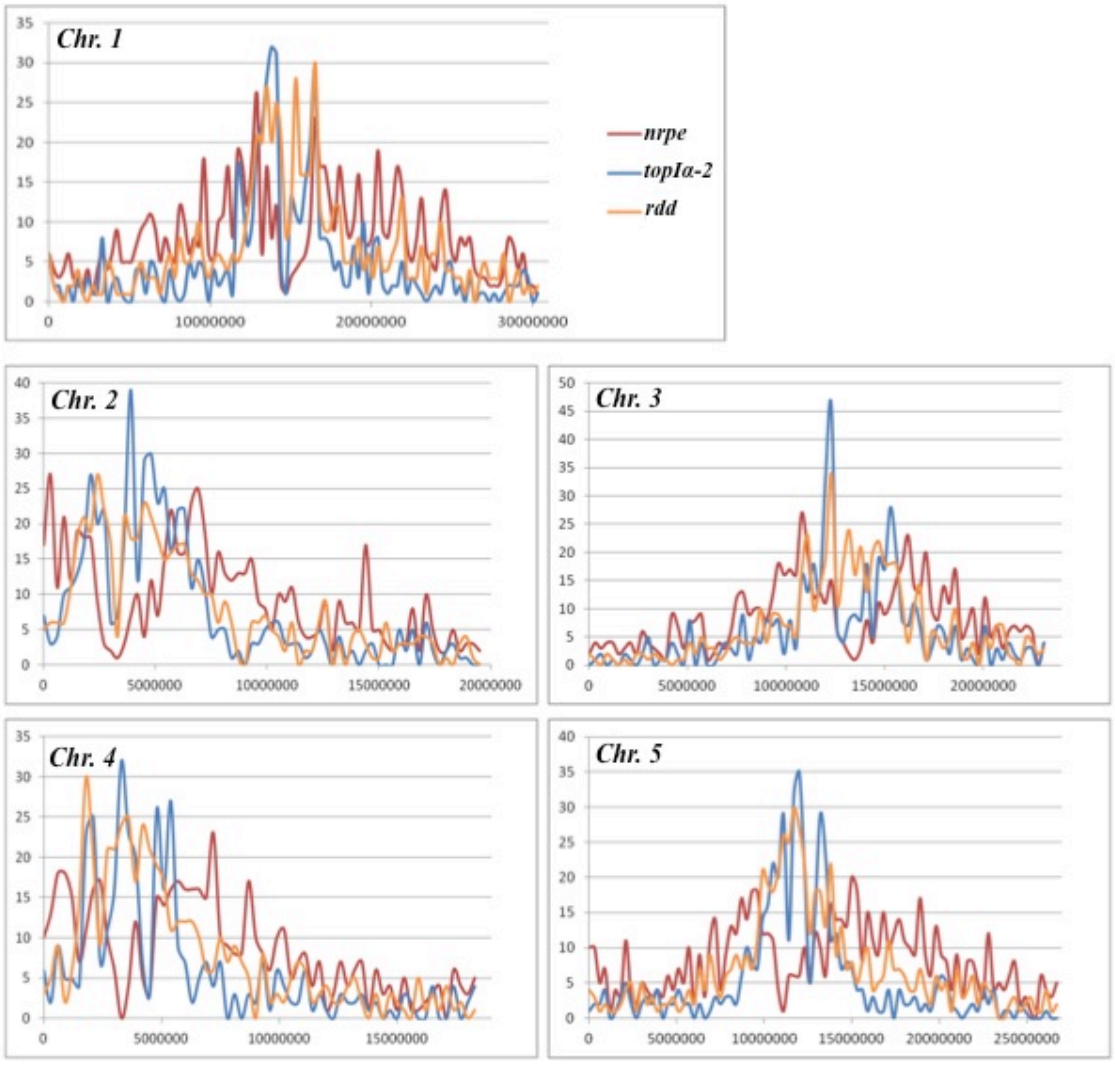

Chr. 1
nrpe
topIα-2
rdd
Chr. 2
Chr. 3
Chr. 4
Chr. 5

Figure 2.14 *topIα* DMRs with increased DNA methylation do not overlap with those of *rdd*. *topIα* DMRs with increased DNA methylation were compared with DMRs with increased DNA methylation in the *rdd* mutant to determine the extent of overlap.

(A) Percentage of *topIα*- specific DMRs (*topIα*) and DMRs common to *topIα* and *rdd* (*topIα & rdd*) among all *topIα* DMRs. The percentage of *topIα*-specific DMRs was much higher than that of the common DMRs.

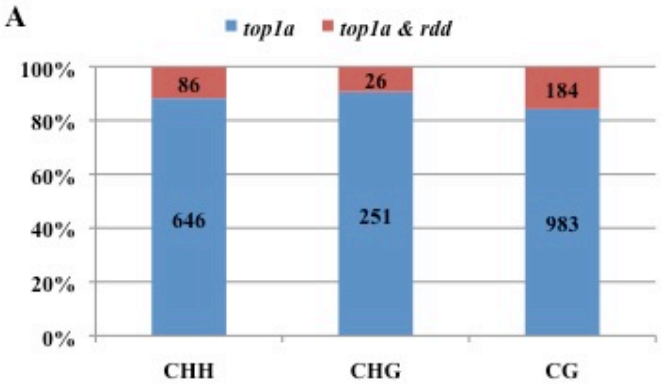

A
top1a
top1a & rdd
100%
80%
60%
40%
20%
0%
86
646
26
251
184
983
CHH
CHG
CG

Figure 2.15 *TOPIα* promotes CG methylation at 5S loci

Loss of *TOPIα* results in reduced CG-specific DNA methylation. 5μg of genomic DNA was digested with *Hpa*II. *Hpa*II cuts unmethylated DNA in a CG context. Thus, if the DNA is unmethylated we expect to see a higher intensity of bands lower down the gel as in *nrpe1-11*. *nrpe1-11* is a Pol V mutant. In the CPT-treated sample, 25 μM of CPT was used. *topIα-7* (also known as *mgo1-7* (Graf et al., 2010)) is a SALK line (S112625) and a weak allele (Liu et al., in prep). Thus, only a slight increase in band intensity lower down the gel is observed. *topIβ-1* is also a SALK line (S069847C) (no morphological phenotypes were observed in this mutant as compared to WT) (Liu et al., in prep). Thus, no change in band intensity is observed for *topIβ-1*.

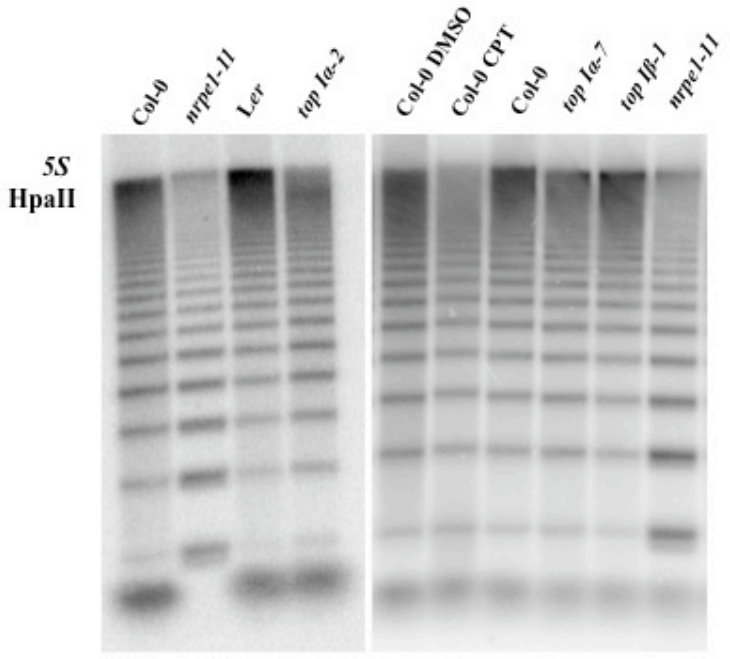

Col-0
nrpe1-11
Ler
top 1α-2
Col-0 DMSO
Col-0 CPT
Col-0
top 1α-7
top 1β-1
nrpe1-11
5S
HpaII

Figure 2.16 DNA methylation levels are unchanged in terms of CHH methylation at *MEA-ISR*, *180S*, and *5S*

(A) 7.5 μg of genomic DNA from ten-day old seedling were digested with *Msp*I at 37°C overnight and hybridized with a probe corresponding to the *MEA-ISR* locus. *MspI* cuts unmethylated DNA in a CHG context. The upper band is methylated DNA, whereas the lower band is unmethylated DNA. *nrpe1-11* is a Pol-V mutant. No change is observed between *topIα-2* and L*er*.

(B) 5 μg of genomic DNA from ten-day old seedlings were digested with *Msp*I at 37°C overnight and hybridized with a probe corresponding to the *180S* loci. c*mt3-7* is the control with reduced CHG methylation. No change is observed between *topIα-2* and L*er*.

(C) 5 μg of genomic DNA from ten-day old seedlings were digested with *Hae*III at 37°C overnight and hybridized with a probe corresponding to the *5S* loci. *sde4-3* is a Pol IV mutant and *nrpd1e -11* is a Pol V mutant. Both serve as controls with reduced CHH methylation. No change is observed between *topIα-2* and L*er*.

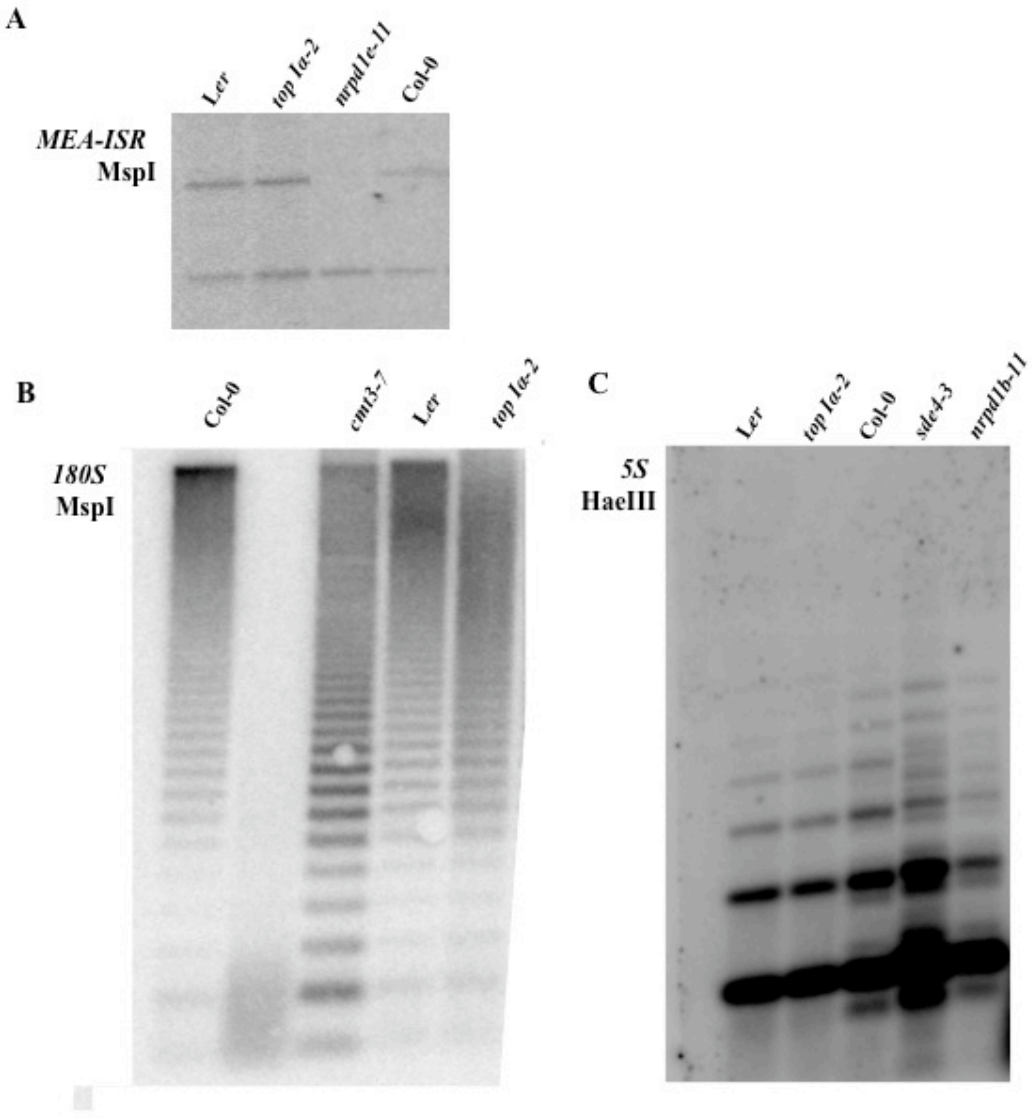

A
Ler
top 1α-2
nrpd1e-11
Col-0
MEA-ISR
MspI
B
Col-0
cmt3-7
Ler
top 1α-2
180S
MspI
C
Ler
top 1α-2
Col-0
sde4-3
nrpd1b-11
5S
HaeIII

Figure 2.17 Distribution of *topIα* DMRs in genic regions

(A) The raw number break down of *topIα* DMRs mapping to different types of genic regions. DMRs with increased or reduced DNA methylation in *topIα* were considered separately. *topIα* DMRs are mostly in protein-coding genes.

(B) Distribution of DMRs along the gene transcription unit. The Y-axis represents the number of DMRs and the x-axis represents the relative distance from the TSS. The relative distance from the TSS was calculated as follows: Relative Distance= [(DS-TSS)/(TTS-TSS)]*100 such that DS is the DMR site. *topIα* DMRs (with reduced methylation; red) are highly enriched at a further distance from the TSS (Transcription Start Site). *topIα* DMRs (with increased methylation; blue) are mildly enriched at a further distance from the TSS.

A

		reduced	increase
total_gene		1185	595
UTR	5UTR	28	12
	3UTR	164	68
	UTR	192	80
protein	Coding region	1509	743
RNA(miRNA,ncRNA,tRNA,etc)		32	10
pseudogenic_transcripts		26	15
total		1759	848

B

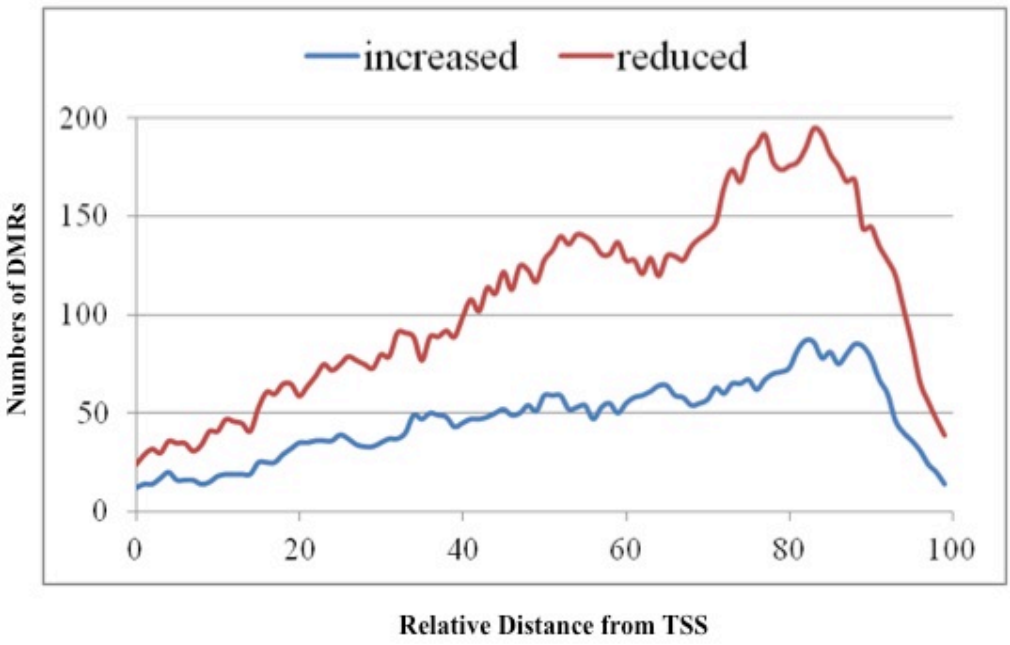
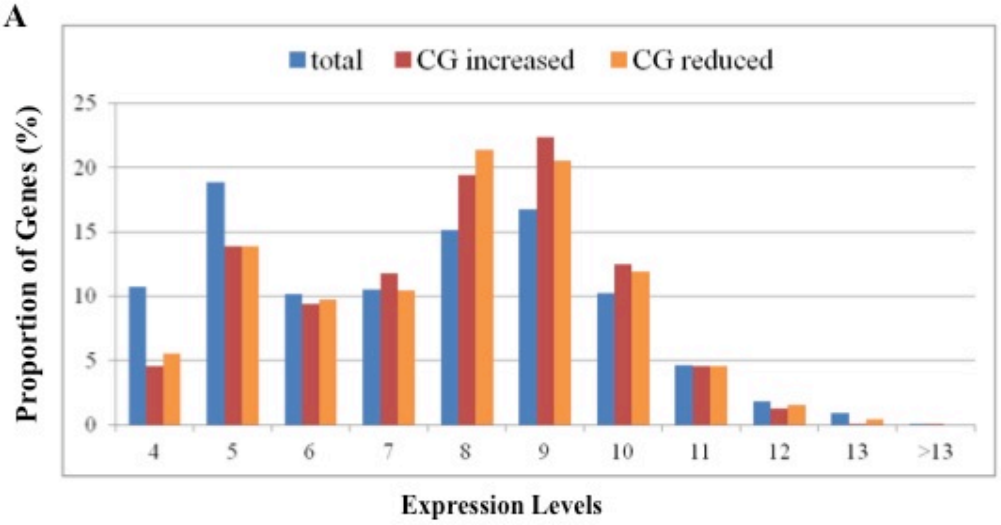


Figure 2.18 The *topIα* CG DMRs are enriched in long genes with moderate levels of expression

(A) *topIα* CG DMRs are enriched in genes with moderate expression levels. The frequency of genes (y-axis) according to transcriptional level (x-axis) was plotted. The blue bars represent all genes in WT (wild type), the red bars represent genes containing CG DMRs with increased methylation in *topIα*, and the orange bars represent genes containing CG DMRs with reduced methylation in *topIα*.

(B) *topIα* CG DMRs are enriched at longer (3 kb+) genes. The frequency of genes (y-axis) according to gene length in base pairs (x-axis) was plotted. The blue bars represent all genes in WT (wild type), the red bars represent genes containing CG DMRs with increased methylation in *topIα*, and the orange bars represent genes containing CG DMRs with reduced methylation in *topIα*.

A


total
CG increased
CG reduced
Proportion of Genes (%)
25
20
15
10
5
0
4
5
6
7
8
9
10
11
12
13
>13
Expression Levels

B

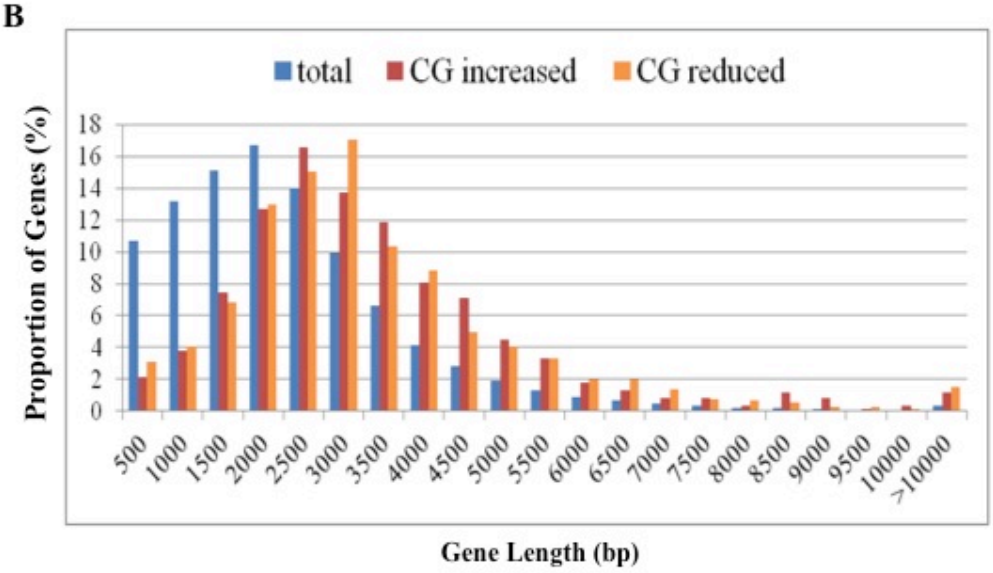
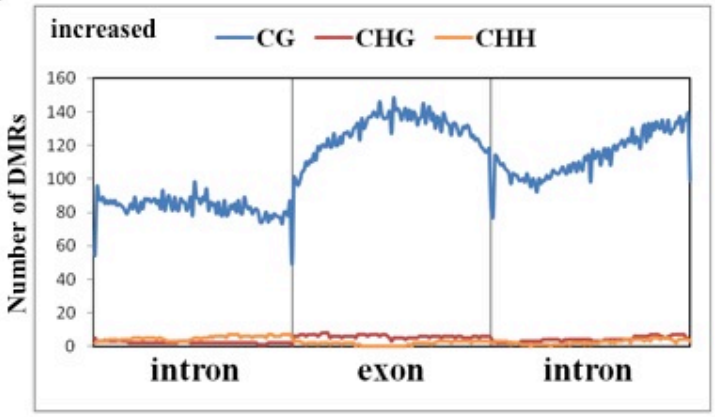

total
CG increased
CG reduced
Proportion of Genes (%)
18
16
14
12
10
8
6
4
2
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
10000
>10000
Gene Length (bp)

Figure 2.19 Partition of *topIα* DMRs among exons than introns.

(A) CG DMRs (blue line) with increased methylation in *topIα* are enriched in exons as compared to flanking introns. There are few CHG (red line) or CHH (orange line) DMRs mapping to exons or introns of genes. The Y-axis represents the numbers of DMRs whereas the x-axis indicates the region.

(B) CG DMRs (blue line) with reduced methylation in *topIα* are enriched in exons as compared to flanking introns. There are few CHG (red line) or CHH (orange line) DMRs mapping to exons or introns of genes. The Y-axis represents the numbers of DMRs whereas the x-axis indicates the region.

A


increased
CG
CHG
CHH
Number of DMRs
160
140
120
100
80
60
40
20
0
intron
exon
intron

B

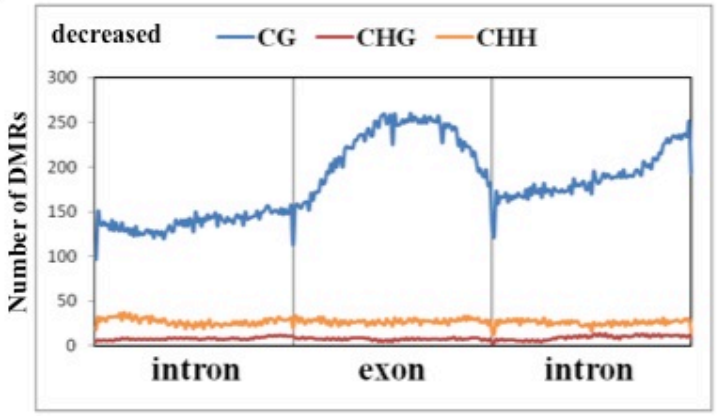

decreased
CG
CHG
CHH
Number of DMRs
300
250
200
150
100
50
0
intron
exon
intron

Figure 2.20 Genes with *topIα* DMRs tend to have more spliced isoforms.

(A, C, E) Data on *topIα* DMRs with increased DNA methylation.

(B, D, F) Data on *topIα* DMRs with reduced DNA methylation.

(A) Distribution of genes according to their transcript isoforms. *topIα* DMRs with increased methylation are slightly enriched in genes with 3 splice isoforms.

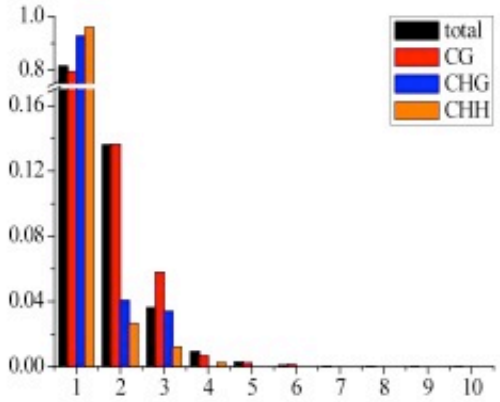
(B) Distribution of genes according to their transcript isoforms. *topIα* DMRs with decreased methylation are slightly enriched in genes with 2-4 splice isoforms.

For (A) and (B), the Y-axis represents the percentage of genes (multiply by 100 to get percentage) and the X-axis represents the number of spliced isoforms. The black bars represent the total genes in wild type. The red bars represent the genes with CG DMRs, the blue bars with CHG DMRs and the orange with CHH DMRs.

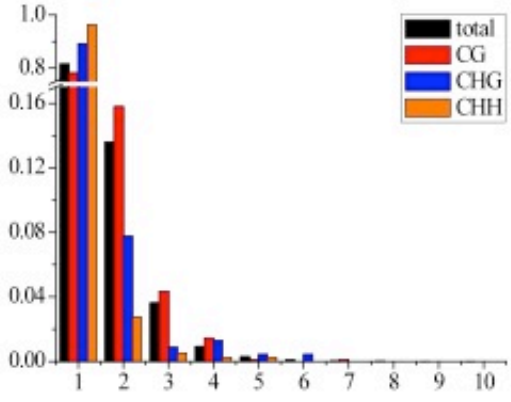
(C and D) The raw numbers by which (A) and (B), respectively, were derived from.

(E and F) The raw percentages by which (A) and (B), respectively, were derived from.

A



B



C

increased	total	CG	CHG	CHH
1	25507	700	137	394
2	4259	120	6	11
3	1133	51	5	5
4	291	6	0	1
5	89	2	0	0
6	26	1	0	0
7	7	0	0	0
8	5	0	0	0
9	1	0	0	0
10	1	0	0	0

D

reduced	total	CG	CHG	CHH
1	25507	1238	207	1266
2	4259	250	18	36
3	1133	69	2	7
4	291	23	3	3
5	89	2	1	3
6	26	0	1	0
7	7	2	0	0
8	5	0	0	0
9	1	0	0	0
10	1	0	0	0

E

increased	total	CG	CHG	CHH
1	0.8144	0.79545	0.92568	0.95864
2	0.136	0.13636	0.04054	0.02676
3	0.0362	0.05795	0.03378	0.01217
4	0.0093	0.00682	0	0.00243
5	0.0028	0.00227	0	0
6	0.0008	0.00114	0	0
7	0.0002	0	0	0
8	0.0002	0	0	0
9	3E-05	0	0	0
10	3E-05	0	0	0

F

reduced	total	CG	CHG	CHH
1	0.8144	0.78157	0.89224	0.96274
2	0.136	0.15783	0.07759	0.02738
3	0.0362	0.04356	0.00862	0.00532
4	0.0093	0.01452	0.01293	0.00228
5	0.0028	0.00126	0.00431	0.00228
6	0.0008	0	0.00431	0
7	0.0002	0.00126	0	0
8	0.0002	0	0	0
9	3E-05	0	0	0
10	3E-05	0	0	0

Table 2.1 Small RNA Levels are not changed in *topIα*

(A) Only a small number of small RNA-producing regions are affected by loss of *TOPIα*. The raw reads and loci affected in *topIα*. *topIα* was compared to L*er* whereas *sde4-3* (a Pol IV mutant), *nrpb2-3* (a Pol II mutant) and *nrpe1* (a Pol V mutant) were compared to Col-0, *LUCH* and *YJ11-3F* (another luciferase-based reporter line) (the three were considered together as wild type).

(B) The maximum and minimum small RNA fold-change in L*er* vs. *topIα* amongst all small RNAs or small RNAs with a length of 24 nt (L24) or 21 nt (L21).

(C) The overlap of *topIα* DSRs with that of *sde4-3*, *nrpb2-3*, or *nrpe1*. The + and/or – signs in the last column represents increased or reduced DSRs, respectively. The first sign is for *topIα*, the second represents the other mutants.

A

average		All		L24		L21	
total		56134		43218		12520	
		increased	reduced	increased	reduced	increased	reduced
Col	*sde4-3*	756	31246	835	832	36	1518
LUCH	*nrpb2-3*	3132	23	1350	19	1216	0
YJ11-3F	*nrpe/pol5*	1186	3620	4069	3686	68	74
Ler	*topIα*	37	3	10	2	25	0

B

average		All		L24		L21	
		increased	reduced	increased	reduced	increased	reduced
Fold Change	maximum	192.829	227.959	200.632	62.9853	234.769	-
	minimum	8.46983	55.6072	9.3128	48.3877	7.8847	-

C

ALL	*sde4-3*	*nrpb2-3*	*nrpe*	
topIα	2	0	2	--
	3	13	1	++
	26	0	0	+-
	0	0	0	-+

L24	*sde4-3*	*nrpb2-3*	*nrpe*	
topIα	1	0	2	--
	0	4	0	++
	0	0	0	+-
	0	0	0	-+

L21	*sde4-3*	*nrpb2-3*	*nrpe*	
topIα	0	0	0	--
	0	9	0	++
	7	0	0	+-
	0	0	0	-+

Table 2.2 Oligonucleotide Sequences

Name	Oligonucleotide Sequence	Application
AP2p26	5'-CCGGTTTGATGGTCGGGCCTCGAC -3'	partial AP2 amplification
AP2p28	5'- GTTTTTTTTAATTACCTTTAGAAAAAGGGA -3'	partial AP2 amplification
Rlucp1	5'-AGGGGATCCACCATGGCTTCGAAAGTTTATG ATCCAGAAC-3'	LUC amplification
lucp6	GCACCCGGGGAAGACGCCAAAAACATAAAA GAAA	McrBC-PCR, southern blot
lucp7	GGACCCGGGTGCGATCTTTCCGCCCTTCTTGG CCT	McrBC-PCR, southern blot
Actin1-F	CCAAGCAGCATGAAGATCAA	McrBC-PCR
Actin1-R	TGAACAATCGATGGACCTGA	McrBC-PCR
35Sf	CAAAGCAAGTGGATTGATGTGA	McrBC-PCR, southern blot
35Sr	TTTCCACGATGCTCCTCGT	Southern blot
LUC 0.13k R	TATGTGCATCTGTAAAAGCAA	McrBC-PCR
YZ 35S Bis F	AttAtTGTyGGtAGAGGtATtTTGAAyGATAGtt	Bisulfite sequencing
YZ LUC Bis R	CATCTaTAAAAaCAATTaTTCCAaaAACCAaa	Bisulfite sequencing
N_UBQ5	GGTGCTAAGAAGAGGAAGAAT	RT-PCR, loading control
C_UBQ5	CTCCTTCTTTCTGGTAAACGT	RT-PCR, loading control

LUCmF5	CTCCCCTCTCTAAGGAAGTCG	RT-PCR for LUC
LUCmR5	CCAGAATGTAGCCATCCATC	RT-PCR for LUC
Kan-RT-F	AGGTTCCATCTGCCAGGTATCA	RT-PCR for NPTII
Kan-RT-R	CCCGGTATCCAGATCCACAA	RT-PCR for NPTII
At2g19990-F	5'- TCACCCGAACAGTTGGAAGAA-3'	McrBC
At2g19990-R	5'-GTGAGGAACCGGTCCATTATTGCT-3'	McrBC
Cluster4-F1	5'- CGTCCTCAAAGTTCCAGAGAT -3'	Real-time RT-PCR
Cluster4-R1	5'- CGGTATTCTCCATCCCAAAG -3'	Real-time RT-PCR
AtCopia2-F1	5'- TTGCCCCAACAACAAAAA -3'	Real-time RT-PCR
AtCopia2-R1	5'- CAGAGAAAGAGATAGAAGAAATGA -3'	Real-time RT-PCR
AtMuI-F1	5'- GGCAGTCGGTTTGTCATTCT -3'	Real-time RT-PCR
AtMuI-R1	5'- CCTTCTTGGCATGGTTCTTC -3'	Real-time RT-PCR
MEA ISR-F1	5'- cgcgaacgactattgctaaa -3'	Real-time PCR B region
MEA ISR-R1	5'-acgattccacaaatccaaca-3'	Real-time PCR B region
MEA-RT-R	5'- TGAAATCTAACCGGATTTTGG -3'	Gene specific primer for RT B region
LNA-siR1003	5'- A+TGC+CAA+GTT+TGG+CCT+CAC+CGT+C-3'	Probe
LNA-cluster4	5'-AA+GATC+AAAC+ATCA+GCA+GCGTC+AG+AGG+CTT-3'	Probe
SoloLTR	5'-GGATTCACGATTAGAGAACGTAGA-3'	Probe
LNA-miR173	5'-GT+GAT+TTC+TCT+CTG+TAA+GCG+AA-3'	Probe
ACTIN60-F	5'-ATCCCTCAGCACCTTCCAAC-3'	Real-time RT-

		PCR; control for B region
ACTIN60-R	5'-AAAATCCACATAACAACAGATAGTTCA-3'	Real-time RT-PCR; control for B region
Chloro_up_bis F	5'-TATGGTGAGYTACAATAATGGTTAAAGAG-3'	Bisulfite sequencing
Chloro_up_bis R	5'-TATCTTTACCRATTAACCAATTTCTAAAC-3'	Bisulfite sequencing
Chloro_up_bis R	5'-TATCTTTACCRATTAACCAATTTCTAAAC-3'	Bisulfite sequencing

CONCLUSION

In my thesis research, I have mainly focused on uncovering factors and molecular mechanisms in transcriptional regulation in *Arabidopsis thaliana*. Transcriptional regulation through a myriad of pathways affords the organism plasticity and ultimately, contributes to overall fitness. Through the identification of the elusive APETALA2 (AP2) target binding sequence, I provide a direct molecular link between *AP2* and *AGAMOUS* (*AG*), thereby answering a long-standing question in the field of floral development. Another goal was to identify genes that act in transcriptional gene silencing, specifically DNA methylation. I identified one gene, *DNA Topoisomerase Iα* and found that it is involved in RNA-directed DNA methylation and gene body methylation. The studies of this gene extend our understanding of DNA methylation and function.

The identification of the AP2 target binding sequence and its biological relevance

I identified the AP2 target binding sequence, TTTGTT and/or AACAA through an *in vitro* DNA-binding assay with AP2R2 (the second DNA binding domain of AP2). Using this novel binding sequence, I found that AP2 directly binds *AG*, *in vitro* and *in vivo*. In addition, using a GUS reporter system, I demonstrated that the presence of this sequence in the *AG* 2nd intron is important for restricting *AG* expression *in vivo*.

Computational analyses revealed that this binding site is highly conserved within the 2nd intron of the *Brasscicaceae* family. The sum of my findings not only solved the missing molecular link between AP2 and *AG*, it also set a foundation for understanding the broad biological roles of *AP2* and its orthologs in other plants.

Identification of a positive factor in DNA methylation

Campothecin (CPT) was identified as a negative factor in DNA methylation through a forward chemical genetics screen of a transcriptionally silenced reporter. CPT is a well-known compound with proliferative anti-cancer properties and its target *DNA Topoisomerase*, is regulates DNA topology. Fortuitously, the mutant (*topIα-2*) was isolated in an independent genetic screen for genes involved in floral determinacy by Dr. Xigang Liu. Using this mutant for mechanistic studies, we found that *TOPIα* is required for silencing transposons at several loci and promoting methylation at 5S repeats.. Genome-wide analyses showed that *TOPIα* affects DNA methylation in two different contexts: CHH and CG. In the CHH context, we found that *TOPIα* is required for CHH methylation at transposons on a genome-wide scale. Furthermore, we found that although *TOPIα* does not play a role in siRNA production, it may play a role in the production of Pol V dependent transcripts. In the CG context, we found that *TOPIα* is important for gene body methylation, an ancient, yet very poorly understood phenomenon. Further, we observed that the DMRs with reduced CG methylation in

topIα are enriched at distances further from the transcription start site. Moreover, *topIα* DMRS are derived from genes with moderate levels of expression, from longer genes, tend to be enriched at exons and have more spliced isoforms. Taken together, these results bear tremendous similarities to the documented role in transcription for *Top1* and *Top2* in *S. Pombe*. These findings reveal a possibility for a link between transcription and body methylation through *TOPIα*. The sum of our findings goes beyond the plant model species and lays groundwork for future studies in DNA methylation and carcinogenesis. Further, it highlights the intricate role *TOPIα* in diverse facets of transcriptional regulation.

Appendix A: Direct Targets of AP2

Although *APETALA2* (AP2) has been shown to play a role in diverse processes, it was not until recently that direct targets were identified for this transcription factor (Yant et al., 2010; Dinh et al., 2012). Thus, at the start of my graduate career, in conjunction with identifying the AP2 binding site, I also generated two inducible lines AP2m3-pER and AP2m3-GR in *ap2*/+ plants. The pER8 vector was a gift from Dr. Nam-Hai Chua from Rockefeller University and the GR vector was a gift from Dr. Patricia Springer from UC Riverside. The primers used to generate these vectors are as listed below (Table 3). Both transgenic lines (AP2m3-ER and AP2m3GR) displayed the AP2 overexpression phenotype (phenocopied AP2m3) upon induction with 10uM estradiol or dexamethasone, respectively, in *ap2*/+ plants (Fig. 1). Since only the AP2m3-translational fusion protein was able to allow us to identify direct targets, we opted to use this line for microarray experiments on inflorescence tissue comparing the gene expression data between GR-induced versus non GR-induced plants. Since I performed one biological replicate first, then later performed two more biological replicates, bioinformatics analyses revealed that the discrepancy between those two sets of experiments were very high. Thus, from the combined microarray experiments, we were only able to obtain ten direct targets (Table 1). Due to the discrepancy issues, this expression data were not suitable for publication. However, I went ahead and

confirmed these genes using RT-PCR of AP2m3GR *ap2*/+ or AP2m3GR *ap2-2* uninduced or induced tissue and L*er* versus *ap2-2* inflorescence tissue. I was able to confirm half of the genes in the induction experiments and all the genes in the *ap2* mutant as compared to L*er*. All genes except *At1g47670*, the Lys/His transporter, showed a consistent molecular phenotype as compared to the results obtained from the microarray.

In parallel, our collaborator also performed a genome-wide deep sequencing experiment in order to identify genes bound by AP2 at the stage of floral transition (he used just bolted inflorescences), using the AP2 antibody that I generated. From his ChIP-seq data, he found a subset of genes that he was interested in (Fig. 3), and I performed the induction experiments to determine whether those genes were indeed direct targets (Fig. 3). A few genes were selected for Real-Time RT-PCR analysis (Fig. 4) and this data were published (except for the Real-Time RT-PCR data for *ARF3*, this will go into another paper) (Yant et al., 2010). In addition, as the collaborator wanted to test his model of an AP2-mir156e-mir172 feedback loop, I also performed small RNA Northern blotting for these two microRNAs in L*er* versus the *ap2-2* mutant Fig. 5; Yant et al., 2010).

As our collaborator had the ChIP-seq data, I asked him to see whether AP2 directly binds the genes that I had obtained from the microarray experiments (Table 1). Indeed, all the genes were bound by AP2. In addition, when I parsed through the AP2-bound regions of each gene, the AP2 binding site that I identified (Dinh et al.,

2012) was present (I have printed out the sequence of every target, marked the region bound by AP2, underlined or highlighted the AP2 binding site and collated this information in a binder marked, "AP2 direct targets"). Thus, these are bona fide direct targets of AP2.

I have also ordered the SALK or SAIL lines of each gene except for *At3G02140* (ABI-Five binding protein, a gene that functions in the ABA pathway), in which the SALK line was not available (Table 2). I have contacted the lab in which this mutant was identified several times (multiple times to the PI as well the first author of the paper in which this mutant was characterized) in order to obtain seeds, however, no response was obtained. All the SALK lines have been genotyped and homozygous lines have been identified (for primer sequences see Table 3). The seeds are stored in a white box labeled, "AP2 direct targets."

Figure A.1 The two transgenes are functional in *ap2/+*

Daily application (?) of dexamethasone and β-estradiol, respectively, to AP2m3GR *ap2-2*/+ and AP2m3-ER *ap2-2*/+ inflorescences led to an AP2m3 phenotype (Chen et al., 2004) in both lines. This shows that both transgenes are functional.


Ler
AP2m3-GR
AP2m3-ER

Figure 2 Transcript levels of the AP2 targets obtained from the microarray experiments show a consistent molecular phenotype.

AP2m3GR *ap2-2/+*or AP2m3GR *ap2-2 i*nflorescences were subjected to 4 hours of cyclohexamide or dexamethasone and cyclohexamide treatments. RT-PCR analysis on these lines along with L*er* and *ap2-2* was performed. The genes show a consistent molecular phenotype as the results obtained from the microarray. The genes also show the expected opposite changes in expression in *ap2-2* vs. *AP2* induction.

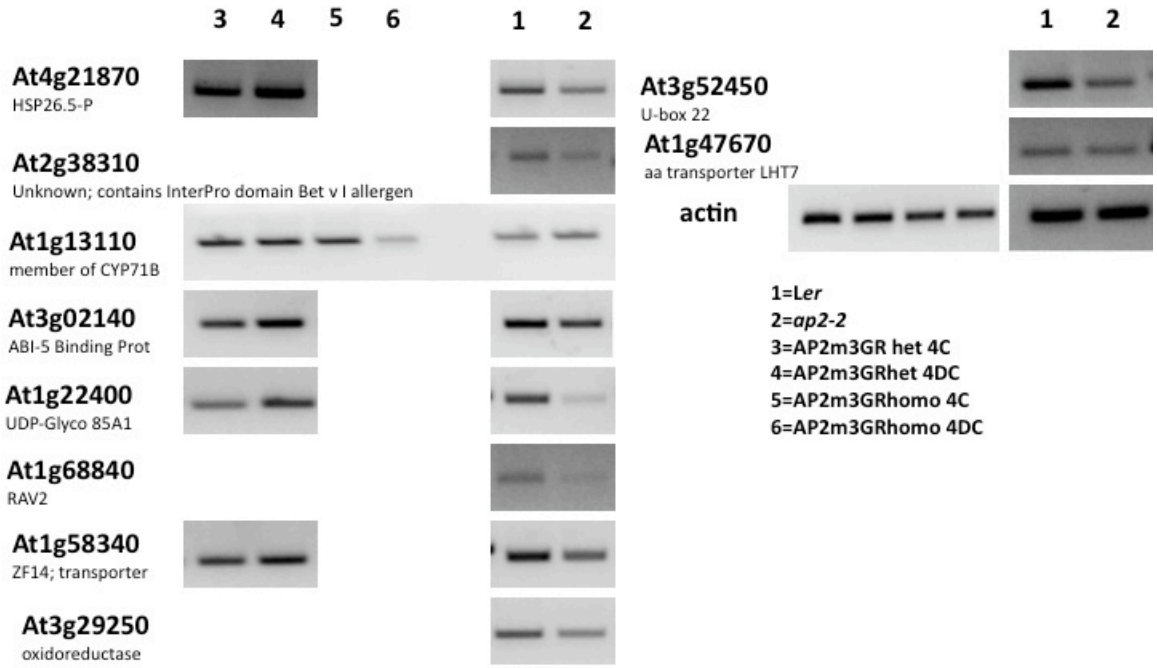

3 4 5 6
1 2
1 2
At4g21870
HSP26.5-P
At2g38310
Unknown; contains InterPro domain Bet v I allergen
At1g13110
member of CYP71B
At3g02140
ABI-5 Binding Prot
At1g22400
UDP-Glyco 85A1
At1g68840
RAV2
At1g58340
ZF14; transporter
At3g29250
oxidoreductase
At3g52450
U-box 22
At1g47670
aa transporter LHT7
actin
1=Ler
2=ap2-2
3=AP2m3GR het 4C
4=AP2m3GRhet 4DC
5=AP2m3GRhomo 4C
6=AP2m3GRhomo 4DC

Figure 3 RT-PCR analyses of genes bound by AP2

From the Chip-seq data performed by our collaborator (Yant et al., 2010), I performed RT-PCR analyses at several loci on AP2m3GR *ap2-2/+* or *ap2-2* lines induced with cyclohexamide or dexamethasone and cyclohexamide for six hours.

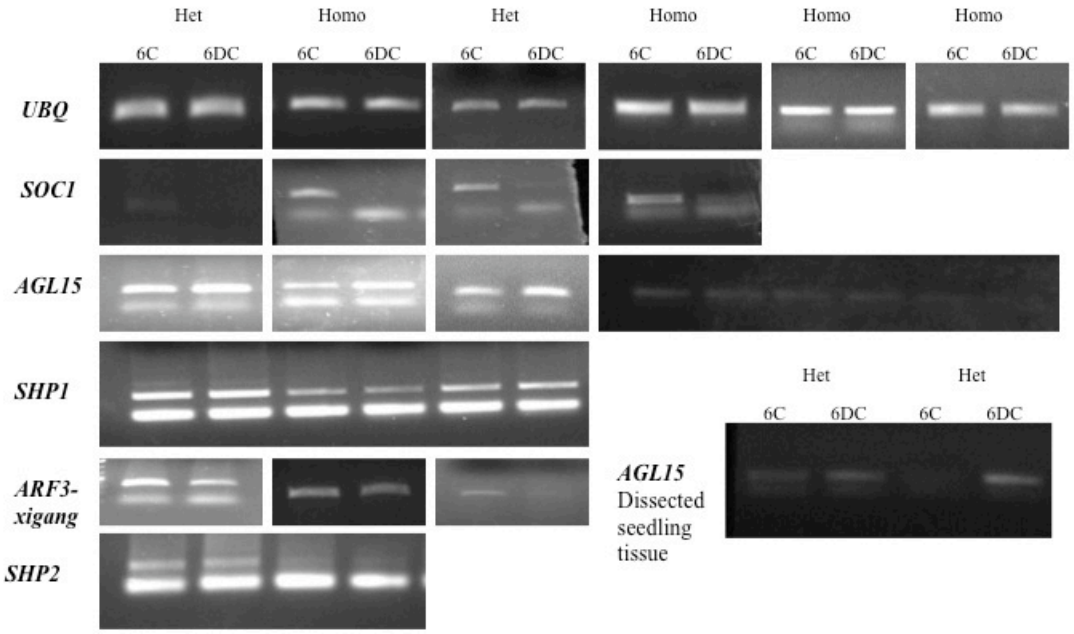

Het
Homo
Het
Homo
Homo
Homo
6C
6DC
UBQ
SOC1
AGL15
SHP1
ARF3-xigang
SHP2
Het
Het
AGL15
Dissected seedling tissue

Figure 4 Real-Time RT-PCR analyses of genes bound by AP2

Real-Time RT-PCR analyses were performed on a few selected loci from the RT-PCR experiment (Fig. 3).

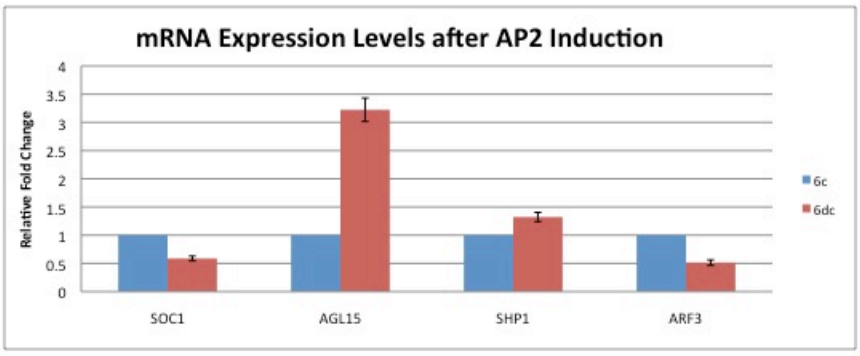

mRNA Expression Levels after AP2 Induction
Relative Fold Change
4
3.5
3
2.5
2
1.5
1
0.5
0
SOC1
AGL15
SHP1
ARF3
6c
6dc

Figure 5 AP2-mir156e-mir172 feedback loop

Mir156e and mir172 were found to be direct targets of AP2 (Yant et al, 2010). Thus, I performed Northern Blot analysis to determine the levels of these two microRNAs in the *ap2-2* mutant.

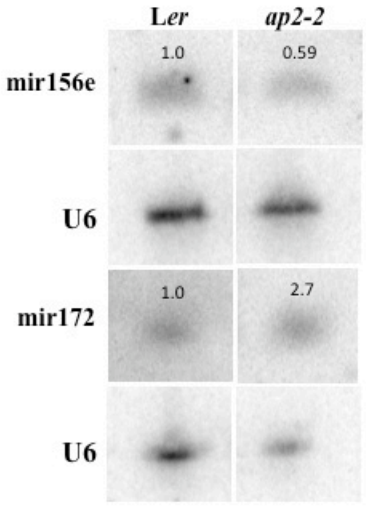

L*er*
ap2-2
1.0
0.59
mir156e
U6
1.0
2.7
mir172
U6

Table A.1 Differentially expressed genes upon induction of AP2

Gene Name/Description	Fold Change
Heat Shock protein	4
UDP-glucosyl transferase 85A1	3
CYP71B7	3.5
ZF14 (MATE efflux)	3
ABI Five binding protein	2
oxidoreductase	3
Plant U box	2.8
Lys/His transporter	2.8
unknown	2
RAV2	2

Table A.2 SALK numbers of the AP2 direct targets

Gene	Acession #	SALK lines
RAV2	At1G68840	S_131565, S_070847C
ABI-Five binding protein	At3G02140	Not availble
HSP26.5	At4g21870	S_2698, S_2700, S_2704
ZF14	At1g58340	S_64288, S_67666, S_67667, S_98386, S_98387, S_98483, S_136101, cs26768, cs859496
UDP-Glucotransferase	At1g22400	S_135383, S_85809C, S_146306C
Unknown protein	At2g38310	S_46674, S_S078338, cs859824
LHT7 amino acid transporter	AT1g47670	Cs859439
Plant Ubox	At3g52450	*pub22-1*
Oxidoreductase	At3g29250	S082262
Cyp71B	At1g13110	S_127944, S_137911, S_12075, S_2841C, S_119574

Table A.3 Oligonucleotide sequences used in this study

Name	Sequence	Purpose
AP2-F	5'-CACCTTAGGCCCGACCTATCGTC-3'	Cloning AP2 into pENTRY vector
AP2-R	5'-AGAAGGTCTCATGAGAGGAGGTTGG-3'	Cloning AP2 into pENTRY vector
mir172-AS	CAGCATCATCAAGATTCT	Northern Blot
mir156-AS	GTGCTCAGTCTCTTCTGTCA	Northern Blot
U6-AS	AGGGGCCATGCTAATCTTCTC	Northern Blot
LBa1	5’-TGGTTCACGTAGTGGGCCATCG-3’	Genotyping
LBb1	5’-GCGTGGACCGCTTGCTGCAACT-3’	Genotyping
RAV2S131565-RP	5’-GAGAAACACTTTCCGTTACCGTCACC-3’	genotyping
RAV2S131565-LP	5’-tgaatggtgccttgacaatgtgggct-3’	genotyping
RAV2S070847C-RP	5’-GAGAAAACACACTTACGCCGACGAGC-3’	genotyping
RAV2S070847C-LP:	5’-taaaaggttgctaacttgctatgtg-3’	genotyping
HSPS002698-RP	5’-CAGGAAGATCAACAGAGAAAG-3’	genotyping
HSP002704-LP:	5’-AGCTTTTCTTCGCCTTTAAAGCCAT-3’	genotyping

ZF14S098387-RP	5’-gccgacataaccgaaagagttcaat-3’	genotyping
ZF14S067667-LP	5’-aagaggactacaaggttgagatttg-3’	genotyping
UDPS135383-RP	5’-agccattaggtgatgaagtgaac-3’	genotyping
UDPS085809C-LP	5’-tggaatgatgagttcgaatttatg-3’	genotyping
UnknownS0778338-RP	5’-gttcctccttctgcttgttcatc-3’	genotyping
unknownS046674-LP	5’-gaaagtgtttgtaagcttgag-3’	genotyping
unknowncs859824-RP	5’-gaaactcttaaagcgtattcaat-3’	genotyping
unknowncs859824-LP	5’-tcacagagacatcttcttctt-3’	genotyping
LHT7cs859439-RP	5’-ccgtatctgctcttatacaaat-3’	genotyping
LHT7cs859439-LP	5’-ggaaccttgtaagttc-3’	genotyping
uboxS072621-RP	5’-ttaaatggtcagaaacgttaca-3’	genotyping
uboxS072621-LP	5’-acgacttcgtcaaatagtatc-3’	genotyping
cyp71BS127944-RP	5’-tataacgcaaccaaaccggtcc-3’	genotyping
CYP71B12075-LP	5’-cggaggaacgttccattttgag-3’	genotyping
Cyp71BS002841-RP	5’-TTGGCTTGTAGACCGAATCTCA-3’	genotyping
cyp71BS119574-LP	5’-tcatagccgaaatcacagcgta-3’	genotyping
oxidoS082262-RP	5’-caccaaatacgtcattacggag-3’	genotyping
oxidoS082262-LP	5’-cctcgccgtttcgatagggc-3’	genotyping

Appendix B: Map-based cloning of S.6.3.2.1

Xuemei performed an EMS mutagenesis screen to look for enhancers of the *hua-1 ag-10* phenotype. Map based cloning was initiated by crossing the *hua-1 ag-10 S6.3.2/+* plants to wild-type plants of the Columbia ecotype. F2 seeds were harvested from single F1 plants. The F2 families that segregated *S6.3.2* were identified by the presence of the "pointy leaf" vegetative and the aberrant floral phenotypes (Fig. 1). In these F2 families, 566 F2 plants that were homozygous for *S6.3.2.1* were identified with the vegetative phenotypes. Genomic DNA was isolated from each of the 566 *S6.3.21* plants. Initial mapping was completed with 36 such plants using simple sequence length polymorphism (SSLP) markers on chromosome V such as ciw8 gave 100% linkage and therefore two markers, nga106 (33 cM) and nga139 (50 cM) were chosen as flanking markers (Fig. 2). Fine mapping was further performed using RFLPs between Col-0 and L*er* (Fig. 3) and the mapping region was ultimately narrowed down to a 300 kb region(for all the mapping data, primers, gel pictures, please see binder labeled, "S6.3.2 mapping data"). This region included *TOUSLED* (*TSL*) as a gene of interest. The phenotype of *S.6.3.2* resembled that of *tsl* mutants.

As the mutant was already published before and no obvious, consistent changes were seen in terms of AP2 and AGO protein levels (Fig. 4A and B), *AP2* or

AGO transcript levels (Fig. 4C and 5), and microRNA levels were unchanged (Fig. 6), this project was no longer pursued.

Figure B.1 Phenotypic analyses of *S.6.3.2*

S.6.3.2 was compared against L*er*. The top pictures are all L*er* and the bottom pictures are *S.6.3.2.*


Ler
S6.3.2.1

Figure B.2 *S6.3.2* linkage mapping data

S.6.3.2 was linked to ciw8 in the rough mapping. The mapping population was expanded (to 2000 plants) and flanking markers, nga106 and nga139, were used to test for recombinants.

Chr. 5: 139 cM; 26170 kb

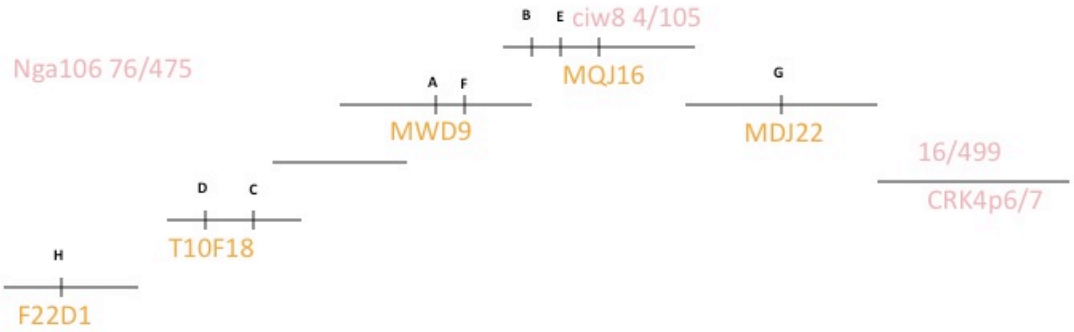
Marker	Rec.	#	%
nga106	76	475	16
mva3	77	2000	3.8
ciw8	52	2000	2.6
CRK4p6/7	16	499	3

Table 1.1 Percent of recombinants out of total sample size

* All based on AGI Map unless noted otherwise

Figure B.3 Fine mapping of *S6.3.2*

A list of the BAC clones near the site of mutation in *S.6.3.2.* RFLP primers were generated and recombinants were tested.



ciw8: 7485588 bp
MDJ spans from 7446495-751325

A) 7369601 0/76
B) 7456801 0/76
C) 7134245
D) 7134247 **Does not work**
E) 7464553 **No recomb**
F) 7417254 **Does not work**
G) 7514969
H) 7062136 **No recomb**

Figure B.4 *S.6.3.2* does not show consistent changes in AP2 and AGO1 protein levels, and *AP2* transcript levels are unchanged

(A) This western shows an increase in AP2 protein levels, however this data was not consistently reproducible.

(B) This western shows no change in AGO protein levels, however this data was not consistently reproducible.

(C) *AP2* transcript levels are unchanged in *S.6.3.2.*

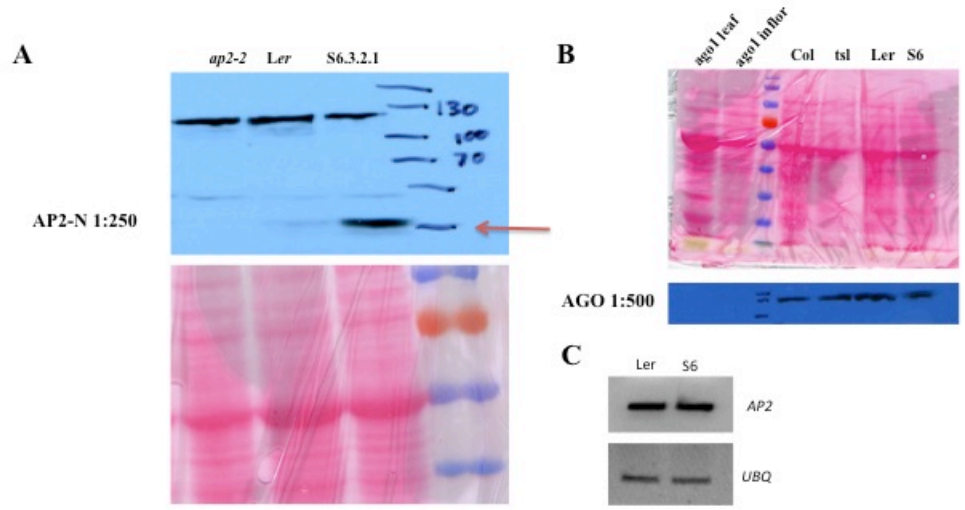

A
ap2-2
Ler
S6.3.2.1
130
100
70
AP2-N 1:250
B
ago1 leaf
ago1 inflor
Col
tsl
Ler
S6
AGO 1:500
C
Ler
S6
AP2
UBQ

Figure B.5 AGO1 transcript levels are unchanged in *S.6.3.2*

There are no changes in AGO1 transcript levels in S.6.3.2.

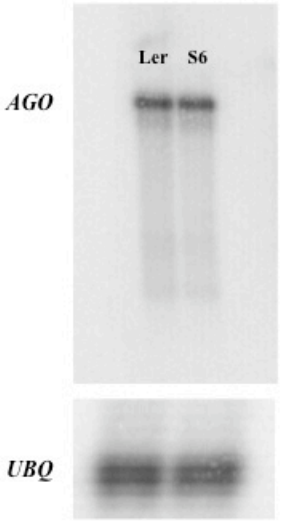
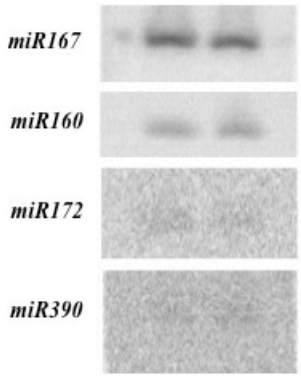

Ler
S6
AGO
UBQ

Figure B.6 *S.6.3.2* does not alter miRNA accumulation levels

S.6.3.2 does not have any changes in miRNA levels in the loci tested as compared to L*er.*


miR167
miR160
miR172
miR390

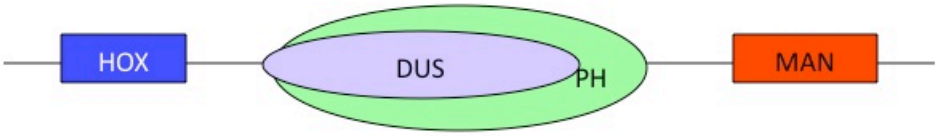
Appendix C: A region in the WUS C-terminal domain can bind lipids

When I first started working Dr. Xuemei Chen's lab, one of my rotation projects was to analyze the WUS protein with Dr. Binglian Zheng, a post-doc in the lab. She had found that WUS had a putative lipid-binding domain (Fig. 1) and so we fused this region of WUS to a GST-tag (pGEX-2TK) and His-tag (prSETA and pet15b). Using PIP strips, we found that this domain bound to PI3P, PI4P, PI4,5P and PI3,4,5P (Fig. 2). Binglian went on to identify several conserved residues within this region (Fig. 3), and performed site directed mutagenesis of these residues (Fig. 3) but the protein still bound lipids (Fig. 4). So, the project was dropped. Several years later, Dr. Brett Tyler's group identified that lipids may be important for the cell-to-cell movement and found that RXLR motifs that contributed to its movement. Parsing the WUS putative lipid-binding domain, with the aid of Dr. Tyler, we found several of these motifs but they were present on both the N- and C-terminal of this putative lipid-binding domain. HHLYH was a strong candidate, HPLLH, KLNQD, KPYP, HVYG and HHYS were weak candidates. Thus, I divided the region in half and made GST- (pGEX2TK) and His- (pET21A) constructs. Region 1A did not bind and region 1B bound lipids at a very low level, especially when compared to the control (Fig. 5). Furthermore, when I quantified WUS to the Avr protein (used as a positive control), I found that the lipid binding capacity is way lower than that of Avr (Fig. 6). Moreover, the SDM constructs that I generated still bound lipids (Per Dr. Tyler's

suggestion, I mutated HHLYH→ qqLYH (Fig. 7). Thus, I did not pursue this project anymore. All primers used are listed in Table 1.

Figure C.1 Schematic diagram of WUS C-terminal domain

Representation of the different putative domains at the C-terminal region of WUS.



Total: 292 AA

HOX (32-100 AA): homeodomain, DNA binding activity

PH (108-205 AA): homology to lipid binding proteins ~21%

DUS (104-178 AA): homology to NADH-dehydrogenases ~24%

MAN (233-286 AA): mannanase, hydrolyzes mannan ~22%

Figure C.2 The putative WUS lipid-binding domain binds lipids *in vitro*

Purified WUS containing the lipid-binding domain was purified via its GST-tag and subjected to lipid binding assay with PIP strips. This domain binds lipids.

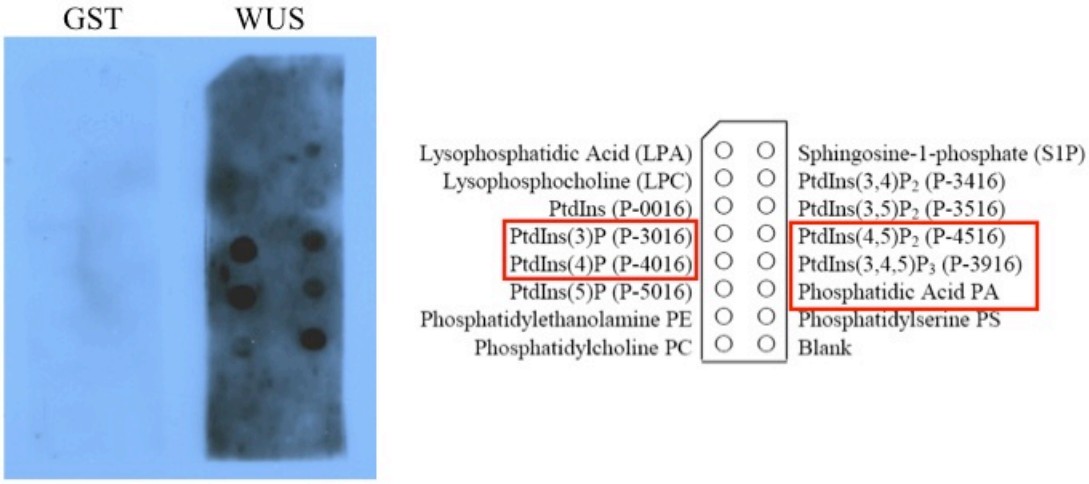

GST
WUS
Lysophosphatidic Acid (LPA)
Lysophosphocholine (LPC)
PtdIns (P-0016)
PtdIns(3)P (P-3016)
PtdIns(4)P (P-4016)
PtdIns(5)P (P-5016)
Phosphatidylethanolamine PE
Phosphatidylcholine PC
Sphingosine-1-phosphate (S1P)
PtdIns(3,4)P2 (P-3416)
PtdIns(3,5)P2 (P-3516)
PtdIns(4,5)P2 (P-4516)
PtdIns(3,4,5)P3 (P-3916)
Phosphatidic Acid PA
Phosphatidylserine PS
Blank

Figure C.3 Critical amino acids in the WUS lipid-binding domain

The sequence of the WUS lipid-binding domain. The red box highlights the conserved amino acids. Site-directed mutagenesis of these amino acids was performed.

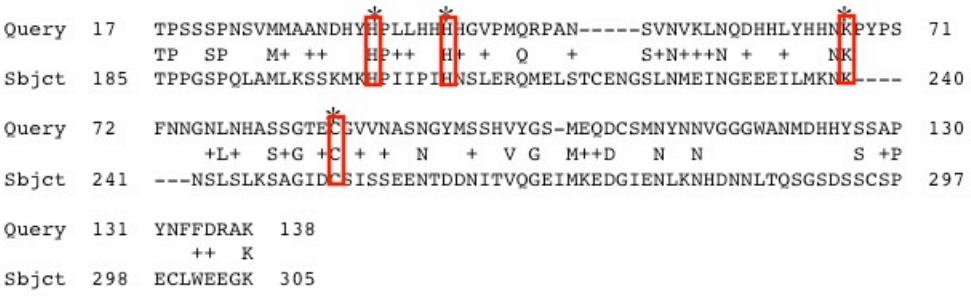
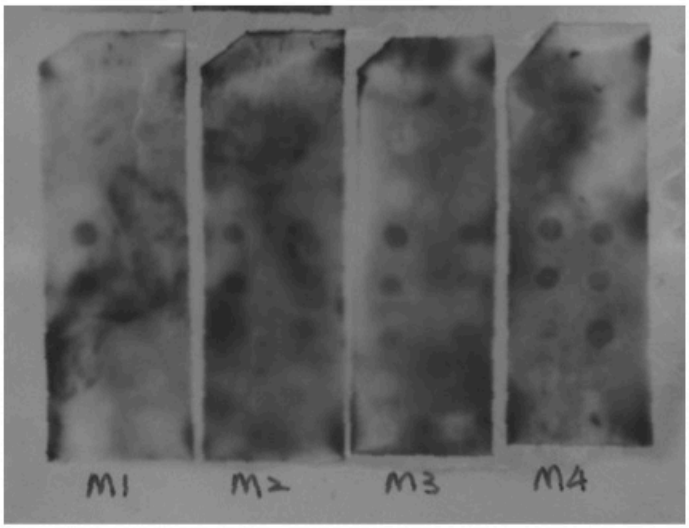

Query 17 TPSSSPNSVMMAANDHYHPLLHHHHGVPMQRPAN-----SVNVKLNQDHHLYHHNKPYPS 71
TP SP M+ ++ HP++ H+ + Q + S+N+++N + + NK
Sbjct 185 TPPGSPQLAMLKSSKMKHPIIPIHNSLERQMELSTCENGSLNMEINGEEEILMKNK---- 240
Query 72 FNNGNLNHASSGTECGVVNASNGYMSSHVYGS-MEQDCSMNYNNVGGGWANMDHHYSSAP 130
+L+ S+G +C + + N + V G M++D N N S +P
Sbjct 241 ---NSLSLKSAGIDCSISSEENTDDNITVQGEIMKEDGIENLKNHDNNLTQSGSDSSCSP 297
Query 131 YNFFDRAK 138
++ K
Sbjct 298 ECLWEEGK 305

Figure C.4 Mutation of the amino acids did not abolish binding

After site-directed mutagenesis (SDM) was performed and the purified protein was obtained, they were subjected to PIP binding and all mutations did not abolish binding. Thus, the amino acids chosen are not critical for lipid binding.



M1(125-126th): HP---AA

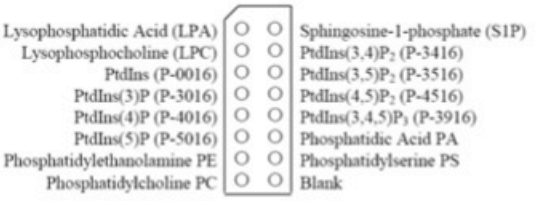
M2(131th): H---A

M3(157-158th): NK---AA

M4(177th): C---A

Figure C.5 Truncations of the WUS lipid-binding domain decreases binding affinity

The WUS lipid-binding domain was divided in half (A and B) and purified via its His tag. The A region no longer bound lipids, however, binding of the B region was very low (as compared to the positive control, Avr). Plus, lipid binding was mostly lost with the truncated protein (B).


Lysophosphatidic Acid (LPA)
Lysophosphocholine (LPC)
PtdIns (P-0016)
PtdIns(3)P (P-3016)
PtdIns(4)P (P-4016)
PtdIns(5)P (P-5016)
Phosphatidylethanolamine PE
Phosphatidylcholine PC
Sphingosine-1-phosphate (S1P)
PtdIns(3,4)P2 (P-3416)
PtdIns(3,5)P2 (P-3516)
PtdIns(4,5)P2 (P-4516)
PtdIns(3,4,5)P3 (P-3916)
Phosphatidic Acid PA
Phosphatidylserine PS
Blank

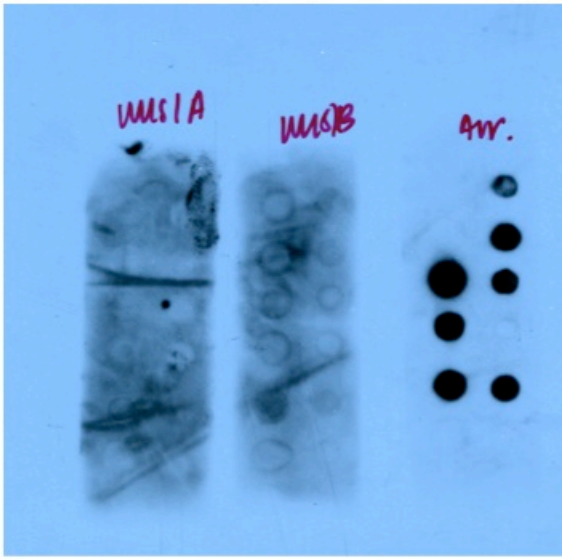

WT S1A
WT S1B
Arr.

Figure C.6 The WUS lipid-binding domain binds lipids at a very low level

As compared to the Avr + control, the lipid binding capacity of WUS was very low, even when 5 times the amount of protein was added compared to Avr.

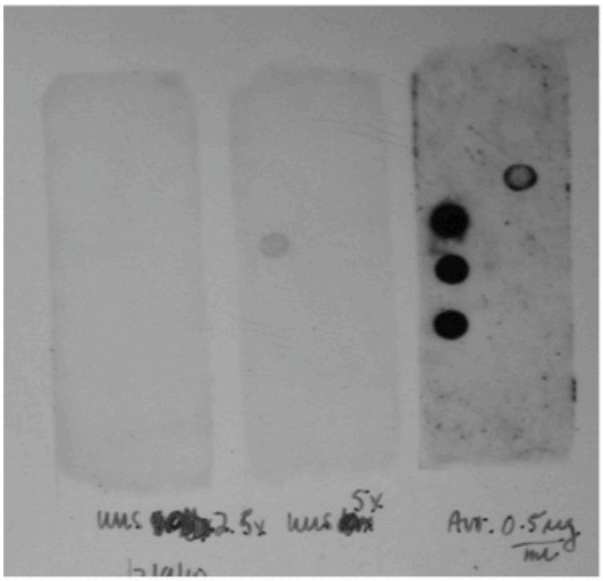

5x
Avr. 0.5

Figure C.7 SDM of the DEER motif did not abolish PI3P binding

Site-directed mutagenesis (SDM) of the DEER motifs result in decreased binding. Only binding was observed for PI3P. However, binding capacity is tremendously lower than Avr.

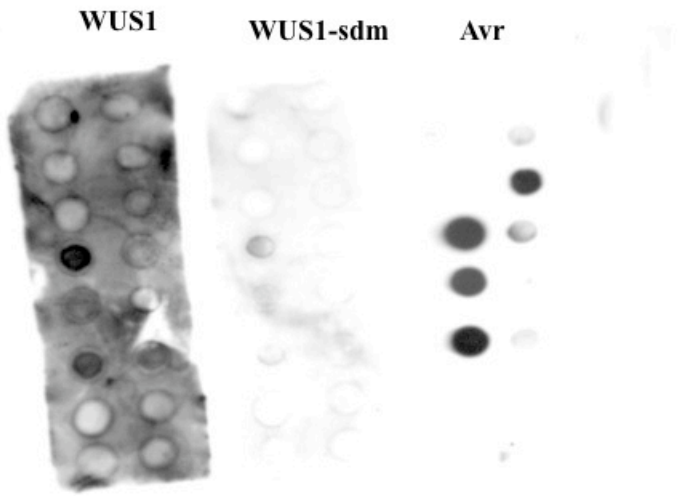

WUS1
WUS1-sdm
Avr

Table 1 Oligonucleotide sequences

Name	Sequence	Purpose
WUS-R10	5'-CCACATTCAGTACCTGAGCTTGCATG-3'	To clone wus1 (the lipid domain) from pGEX2TK to pET21A. I used the same forward primer as the cloning Binglian used to generate the GST-tag.
WUS-F11SDM	5'-CTTAACCAAGACCAgCAgCTCTATCATCATAACAAGCC-3'	SDM primers used for both wus1pGEX2tk and pET21A.
WUS-R11SDM	5'-GGCTTGTTATGATGATAGAGcTGcTGGTCTTGGTTAAG-3'	SDM primers used for both wus1pGEX2tk and pET21A.
WUS-F12	5'-cgcggatccgcgAGCTTCAATAACGGGAATTTA-3' position: 484 at CDS; has BamHI site	For cloning wus1B-pGEXT2TK, used wus12F & wus R1 (Binglian); For cloning wus1B-pET21A, used wusF12 & wus R3 (Binglian)
WUS-R12	5'-ccggaattccggGGGATATGGCTTGTTATGATG-3'	For cloning

	position: 483 at CDS; has ECoR1 site	wus1A-pGEXT2TK, used wus1(F) (Binglian) & wus R12
WUS-R13	5'-ccggaattccgGGGATATGGCTTGTTATGATG-3' position: 484 at CDS; has EcoRI site	For cloning wus1A-pET21A, used wus1(F) (Binglian) & wus R13
WUSFLpSDK1-F	5'-atgcACAAGTTTGTACAAAAAAGCAGGCTccA TGGAGCCGCCACAGCATCAGCA-3'	For cloning into the pSDK vector given to us by Dr. Tyler; did not clone yet.
WUSFLpSDK1-R	5'-GCATACCACTTTGTACAAGAAAGCTGGGTG TTCAGACGTAGCTCAAGAG-3'	For cloning into the pSDK vector given to us by Dr. Tyler; did not clone yet.