Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Stefan L. Farrugia March 14, 2019

Investigation of MBD9's Role in H2A.Z Deposition in *Arabidopsis thaliana*

by

Stefan L. Farrugia

Dr. Roger Deal
Adviser

Department of Biology

Dr. Roger Deal

Adviser

Dr. Megan Cole

Committee Member

Dr. Adriana Chira

Committee Member

2019

Investigation of MBD9's Role in H2A.Z Deposition in *Arabidopsis thaliana*

By

Stefan L. Farrugia

Dr. Roger Deal

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2019

Abstract

Investigation of MBD9's Role in H2A.Z Deposition in *Arabidopsis thaliana*
By Stefan L. Farrugia

Gene expression regulation is vital to the proper functioning of an organism. This is accomplished in a variety of ways, one of which is with nucleosomes. Nucleosomes consist of DNA wrapped around proteins called histones, which can conceal or expose DNA to transcription factors. Besides the four canonical histones H2A, H2B, H3, and H4, there are histone variants like H2A.Z which seem to provide more specialized roles. The plant *Arabidopsis thaliana* is a great model system to study H2A.Z as it is viable without any of its three copies of the *H2A.Z* gene, it has easy methods to introduce new genetic material, and it has several visible phenotypes for *H2A.Z* mutants. The protein complex that is responsible for depositing H2A.Z in *A. thaliana* is called SWR1, and only recently has work been done to establish the identity of the SWR1's subunits, one of which is MBD9. Not much is understood about MBD9, including its status as a stable subunit of SWR1 and its role in the deposition of H2A.Z. A proposed hypothesis is that MBD9 recognizes histone tail modifications and also interacts with SWR1, thereby guiding the complex to specific genomic regions for H2A.Z deposition. In order to test this, molecular tools for assays like histone peptide microarrays need to be developed. Here, a MBD9-MYC tag fusion protein was constructed and successfully translated. Additionally, a system to create MBD9 joined with a variety of tags was initiated for the purpose of protein interaction assays. Finally, phenotypic analysis of *A. thaliana H2A.Z* CRISPR mutant lines was accomplished. These novel *H2A.Z* mutants will be utilized to test protein interactions of MBD9-tag fusion proteins *in vivo*. Together, these results provide the groundwork for further experiments in understanding MBD9 and H2A.Z in *A. thaliana*.

Investigation of MBD9's Role in H2A.Z Deposition in *Arabidopsis thaliana*

By

Stefan L. Farrugia

Dr. Roger Deal

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2019

Acknowledgements

I would like to thank Dr. Roger Deal for allowing me to work his lab. I also would like to thank all members of Dr. Deal's lab for their guidance and support. I especially would like to recognize Dr. Paja Sijacic for his role as a mentor during my research. I want to thank my committee members Dr. Roger Deal, Dr. Megan Cole, and Dr. Adriana Chira for their help and input during the thesis process as well as during my academic career at Emory. Finally, I want to thank my family, Dr. Gianrico Farrugia, Geraldine Farrugia, and Luca Farrugia for their continual support and encouragement.

Table of Contents

List of Figures

Introduction

Histones as regulators of gene expression

Regulation of gene expression is one of the most vital processes in an organism. For example, the precise amount and timing of gene expression is critical to proper embryonic development (Hasegawa et al., 2015). One of the many ways that organisms regulate gene expression is through nucleosomes, which consist of roughly 147 base pairs of DNA wrapped around proteins called histones (Cutter & Hayes, 2015). Nucleosomes influence the expression of genes at the transcriptional level physically by exposing or blocking key segments of DNA, such as promoters or regulatory sequences (Luger, Dechassa, & Tremethick, 2012). Generally, nucleosomes have two copies of each of the four core canonical histones, H2A, H2B, H3, and H4, which are paired in dimers consisting of H2A with H2B, and H3 with H4 (Luger et al., 1997). However, individual histone proteins can be swapped for histone variants, which are similar in amino acid sequence to the canonical histones but are divergent in the amino-terminal and carboxy-terminal tails (Ausio and Abbott, 2002). These divergences could be responsible for more specialized roles that histone variants accomplish, like the H3 variant CenH3 which is involved in centromere identity (Palmer et al., 1991).

H2A.Z's significance and deposition

H2A.Z is one of the variants of the core canonical histone H2A that is found in all known eukaryotes, including humans (Thatcher & Gorovsky, 1994). In humans, there are two H2A.Z genes called *H2A.Z.1 and H2A.Z.2*. These *H2A.Z* genes are involved in essential processes like embryonic stem cell differentiation and T-cell activation (Creyghton et al., 2008; Sutcliffe et al., 2009). Additionally, abnormal expression of H2A.Z has been linked with prostate and breast cancer (Dryhurst et al., 2011; Rangasamy, 2010). The specific ways that H2A.Z is influencing

these processes and diseases is unknown. Part of the mystery is that H2A.Z has a context-dependent influence on gene expression; it can increase or decrease transcription depending on the location it is placed in the genome and likely the timing of the placement (Soboleva et al., 2014). H2A.Z is deposited in the genome by chromatin-remodeling complexes that exchange the canonical H2A-H2B dimer for H2A.Z-H2B (Krogan et al., 2003). In humans, this is the SNF2-related CREB-binding protein activator protein complex, known as the SRCAP complex, which has a catalytic ATPase subunit (Cai et al., 2005; Adam et al., 2001). Further understanding of chromatin-remodeling complexes that incorporate H2A.Z into chromatin is necessary to fully understand the role H2A.Z plays in gene regulation in humans and other organisms.

Arabidopsis thaliana as a model organism for studying H2A.Z

Arabidopsis thaliana, also known as the thale cress plant, has several distinct advantages as a model organism for studying H2A.Z. *A. thaliana* has 13 H2A-encoding genes, including three with high homology to human H2A.Z: *HTA8*, *HTA9*, and *HTA11* (March-Diaz and Reyes, 2009). While animal model organisms are not viable when missing H2A.Z, *A. thaliana* is viable as a single, double, and even triple mutant for *H2A.Z* genes (Coleman-Derr and Zilberman, 2012). These genes show some level of redundancy, as knockouts of any one of these three genes solely has no obvious phenotypic effect (Choi et al., 2007). However, it is important to note that when double mutants of *HTA9* and *HTA11* are grown, they do exhibit an altered phenotype, indicating their redundancy and showing that the other *HTA* genes, including *HTA8*, are not able to fulfill the specialized role that these particular genes are responsible for (March-Diaz and Reyes, 2009).

Another advantage of *A. thaliana* is the presence of easily detectable phenotypes for *H2A.Z* mutants. These include serrated rosette leaves, reduction in serrated leaf number, paler coloration in rosette leaves, abnormal petal numbers on flowers, shorter plants, lowered fertility, curved siliques, and loss of apical dominance (Coleman-Derr and Zilberman, 2012). There are several ways to generate mutant lines in *A. thaliana*. Specific to this model organism is T-DNA mutagenesis (Alonso and Stepanova, 2003). This method involves the use of the proteobacteria *Agrobacterium tumefaciens*. *A. tumefaciens* carries a circular plasmid, and when it is exposed to *A. thaliana*, the circular plasmid is incorporated into the plant genome, creating an insertion that can disrupt genes (Gelvin, 2003). As is possible for most model organisms, CRISPR mutant lines can also be generated in *A. thaliana* (Cong et al., 2013).

In comparison to the SRCAP complex found in mammals, *A. thaliana* has the Swi2/Snf2-related complex I, called SWR1. SWR1 contains many homologous protein subunits found in SRCAP, like ARP6 (Actin Related Protein 6) and PIE1 (Photoperiod-Independent Early flowering 1), which are important in the complex's ability to incorporate H2A.Z (March-Diaz and Reyes, 2009). The identity of all of the subunits of the SWR1 complex has only recently been accomplished by preliminary work in Dr. Roger Deal's laboratory. However, for identified subunits, their precise role in incorporation is still being investigated.

MBD9 is required for H2A.Z deposition

One potential subunit of the SWR1 complex is MBD9 (methyl-CpG-binding domain 9). It has been shown that *MBD9* mutants share some phenotypes with mutants of proteins found in SWR1, like *ARP6* (Peng et al., 2006; Deal et al., 2005). Previous Tandem Affinity Purification (TAP) experiments in Dr. Roger Deal's laboratory have shown that MBD9 interacts with ARP6.

TAP experiments involve two rounds of affinity purifications to beads using a tag fused to a protein of interest in order to test for protein interactions (Van Leene et al., 2015). ARP6 was used as bait to identify the proteins it interacts with by creating a fusion protein of ARP6 with a GS-rhino TAP tag. Previous experiments using this fusion protein generated a list of non-specific binding proteins to the tag, allowing them to be excluded as possible interacting proteins with ARP6 (Van Leene et al., 2015). N-terminal and C-terminal fusions of the tag to ARP6 were created to ensure that interference from the tag itself to protein interactions was not a factor. To ensure that the fusion protein was able to perform its function, it was introduced into the *arp6* mutant *A. thaliana*. These rescue experiments produced plants identical to wild type plants, showing that the ARP6-TAP tag fusion protein is operating correctly (submitted, Sijacic et al., 2019).

Chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) was performed using H2A.Z antibody on *mbd9* mutant plants. Lower levels of H2A.Z were found genome wide in comparison to wild type plants in the *mbd9* mutants. Next, to see if MBD9 is not only interacting with ARP6, but is also a core subunit of SWR1, size-exclusion chromatography was performed on both wild type and *mbd9* mutant SWR1 complexes. If MBD9 (around 240 kilodaltons) is a stable subunit of SWR1, it is expected that there would be a noticeable difference in weight between the genotypes. However, similar weights were found, indicating that MBD9 is not a stable subunit of SWR1 (submitted, Sijacic et al., 2019). Further investigation in MBD9's involvement with SWR1 and in H2A.Z incorporation is needed to expand on these findings.

MBD9 contains a bromodomain which has been shown to recognize histone modifications like acetylation (Yang 2004). The presence of bromodomains has also been

verified in chromatin-remodeling complexes (Fry and Peterson, 2001). So, one possible role of MBD9 is to recognize various histone modifications to inform SWR1 when and where to deposit H2A.Z. Histones can be modified by the post-translational addition of chemical modifications including phosphorylation, acetylation, and methylation to the N-terminal and C-terminal histone tails (Strahl & Allis, 2000). MBD9 may be "reading" where to deposit H2A.Z in the genome through these modifications. A way to test which particular modifications MBD9 could be interacting with is with a histone peptide microarray assay (Figure 1). The array consists of hundreds of different combinations of histone tail modifications. After the protein of interest interacts with the modifications, all non-interacting protein is washed away. A tag attached to the remaining interacting proteins is recognized by antibodies, which lead to the emittance of a fluorescent signal. A key can then be used to determine what histone tail modifications are represented by each signal location (Rothbart et al., 2012). In order to do this assay, translation of a functional version of MBD9 fused with a tag is needed as an effective MBD9 antibody is not available.

Objectives

Given the current understanding of H2A.Z and MBD9 in *A. thaliana*, the objectives listed below were selected to further progression in this field:

1. *In vitro* translation of MBD9 fused with a MYC tag for the purpose of a histone peptide microarray assay.
2. Creation of several MBD9-tag constructs that can be used for *in vivo* protein interaction assays.
3. Verification of CRISPR *H2A.Z* mutant plants through phenotypic analysis.

Methods

Plasmid and Primer Design

The program "A plasmid Editor" (ApE) created by M. Wayne Davis of the University of Utah was used to create plasmid and DNA constructs. Primers were designed in the software AmplifX to assess primer quality based on primer length and GC% content. Additionally, predicted annealing temperatures from AmplifX were used to make primer pairs within 4°C from each other. The NEB Tm Calculator was used to determine annealing temperatures of primer pairs. MYC tag, MBD9 coding sequence DNA (cMBD9), and plasmid pT7CFE1 (Thermo Scientific, Catalog #88860) sequence files were obtained from preliminary lab work. MBD9 genomic DNA (gMBD9) sequence information was obtained from The Arabidopsis Information Resource (TAIR).

For the cMBD9-MYC fusion protein, primers were designed to overlap with plasmid pT7CFE1 on the C-terminal end. The MYC tag was added to the N-terminal end of MBD9, so primers were designed to encompass the MYC tag and cover portions of MBD9 and pT7CFE1 on the N-terminal side of MBD9. Primers were designed in such a manner to proceed with Gibson Assembly, to combine the cMBD9-MYC PCR product with the plasmid by inserting the sequence into the plasmid's Multiple Cloning Site (MCS). The plasmid pT7CFE1 features a 6x His tag at the C-terminus for protein purification and an Ampicillin-resistance gene for bacterial selection (Figure 2).

For the protein-interaction constructs, primers were designed for Gateway BP Recombination Reaction (Invitrogen, Catalog #11789100). Primers were made to introduce *att*B flanking sequences on the 5' and 3' ends of MBD9 (Figure 3).

SPRI Bead DNA Purification

Purification of DNA products was accomplished using Solid Phase Reversible Immobilization Beads. SPRI Beads were kept at 4°C. Before use, an aliquot of beads was brought to room temperature and 1.8 volumes of beads in comparison to the volume being purified was added to the DNA samples. The mixture rested for 5 minutes at room temperature (around 22°C) before being placed on a magnetic rack to aspirate the supernatant. Beads were then cleaned with two rounds of 150 µL of 80% ethanol. After removal of all ethanol, DNA was collected in elution buffer (10 mM Tris, pH of 8). Following all DNA purifications, DNA concentrations were measured with a DNA spectrophotometer.

Gibson Assembly

New England BioLabs Protocol E2611 was followed for Gibson Assembly of PCR products with pT7CFE1 plasmids. Around 200 nanograms of cMBD9-MYC, 60 nanograms of pT7CFE1, and 10 µL of Gibson Assembly Master Mix were used to achieve the ideal molar ratio of plasmid to PCR product. A negative control with nuclease-free water instead of Gibson Assembly Master Mix was performed in parallel. The mixture was then incubated at 50°C for 60 minutes, followed by being held at 4°C for future use.

Transformation

Plasmids were transformed into NEB 5-alpha Competent *E. coli* cells. After adding plasmid to the thawed cells, they were then kept on ice for 30 minutes. Next, the sample was placed in a 42 °C water bath for precisely 30 seconds before being transferred back to ice for 2 minutes. Following this, 950 µL of Lysogeny Broth (LB) was added and the mixture was moved

to an incubated shaker for 60 minutes at 37°C. Then, 100-200 µL of the liquid was spread on Ampicillin or Kanamycin-containing plates based on the plasmid's selection gene and plates were kept in a 37°C incubator for approximately 18 hours.

Inoculation

For every colony being inoculated, 250 µg of ampicillin or kanamycin was added to 5 µL of LB solution. An individual, transformed *E. coli* colony containing the plasmid of interest was then added to the LB solution. The solution was then placed in an incubated shaker at 37°C for approximately 18 hours.

Miniprep Plasmid Preparation

Plasmids were isolated using the Qiagen QIAprep Spin Miniprep Kit 250 (Catalog #27106). Before starting the protocol, glycerol stocks were made by adding 750 µL of inoculated bacteria to 750 µL of a 50% glycerol, 50% LB solution in cryotubes to be kept at -80°C. Per the protocol, *E. coli* cells are lysed with a series of buffers, washed with an ethanol-based buffer, and then DNA was eluted in 50 µL of 10 mM TrisCl, pH 8.5.

In Vitro Transcription and Translation

Thermo Scientific 1-Step Human Coupled IVT (88881) was used for transcription of plasmid and translation of protein. Vector DNA was cleaned using a 3M sodium acetate solution pH 5.5 and 100% ethanol following Gibson Assembly as per protocol. HeLa lysate, accessory proteins, reaction mix, plasmid DNA, and nuclease-free water were mixed and incubated at 30°C for 6 hours. A GFP positive control was included in translation. The presence of GFP was

verified with a fluorescent microscope. Visualization of translation products with Western Blot was performed by Dr. Paja Sijacic.

Gateway Cloning Recombination

The Gateway Cloning Recombination system was chosen to introduce gMBD9 and cMBD9 fused with a variety of tags into *A. thaliana* (Invitrogen). The principle behind the concept is to move MBD9 into an entry vector, which then can be used to transfer MBD9 into a destination vector that introduces the tag of choice. This construct can then be transformed into *A. thaliana*. The first transfer, called BP Recombination (Invitrogen, Catalog #12536-017), into the entry vector pDONR221 is accomplished by the addition of *att*B sequences to both ends of the MBD9 PCR product. The *att*B sequences match the *att*P sequences located in the entry vector, allowing recombination to take place. Next, MBD9 in the pDONR221 vector is transferred to the pEarleyGate 200 and 300 series destination vectors with LR Recombination (Invitrogen, Catalog #11791-019). The same recombination principle of BP Recombination is used with *att*R sequences in the destination vector. In order to achieve ideal ratios of vector and insert, 200 to 250 ng of MBD9 with *att*B flanking sequences and 130 ng of pDONR221 were used. In addition to these components, TE pH 8.0 buffer was added to bring the total volume to 8 µL. 2 µL of Gateway BP Clonase II was then added and the mixture was incubated at room temperature (around 22°C) for 18 hours. 1 µL of Proteinase K was then mixed in and the reaction was incubated at 37°C for 10 minutes.

Assessment of A. thaliana Phenotypes

Plants were grown in growth chambers at 20°C with long-day cycles (16 hours of light, 8 hours of dark). Genotypes grown were Wild Type Col-0 ecotype, and several *H2A.Z* mutants:

hta9-1; hta11-2 T-DNA mutant, *hta9-1; hta11-1* T-DNA mutant, CRISPR *6-3* mutant, CRISPR *21-6* mutant, and a triple *H2A.Z* mutant. Genotypes were developed by preliminary lab work. Either 6 or 7 plants were grown for each genotype. Counting of rosette leaves was done at the emergence of the shoot when plants switched from leaf production to flower production (time of flowering). Rosette leaves are located at the base of the plant, and cotyledons were counted as rosette leaves. Flower petal counting was done at the emergence of distinguishable flower petals. Preliminary data shows that the normal number of petals per flower is 4, so flowers with more than 4 petals were counted as abnormal. 10 flowers were counted per plant and the proportion of flowers with abnormal petal numbers was calculated. Two-tailed, independent means t-tests were calculated between wild type averages and mutant plant averages for both phenotype characteristics being observed.

Results and Discussion

Expression of cMBD9-MYC Protein

Successful PCRs of the coding sequence of MBD9-MYC (cMBD9-MYC) and pT7CFE1 were verified with gel electrophoresis. Multiple cDNA templates were used; two PCR products were chosen to proceed into Gibson Assembly due to having adequate concentrations following SPRI bead purification (greater than 20 ng/μL). After bacterial transformation on Ampicillin-containing plates, six colonies were selected, inoculated, and miniprepped. Four of the six selected colonies showed successful Gibson Assembly (Samples 1, 3, 4, and 5). Samples were digested with AvrII restriction enzyme and checked with gel electrophoresis for correct length of product. The two samples negative for Gibson Assembly (Samples 2 and 6) may have had self-ligation of the plasmid without the insert. Presence of cMBD9-MYC as the insert was tested with

ClaI digestion, a restriction enzyme that cuts in the insert but not in the plasmid. The four successful samples from Gibson Assembly showed linearized DNA with gel electrophoresis, confirming the presence of cMBD9-MYC (Figure 4).

Initially, Sample #5 was chosen to proceed with *in vitro* translation due to having the highest DNA concentration. However, Western Blot analysis after translation showed a clear band around 130 kDa while the cMBD9-MYC product should be around 240 kDa. The sample was sent for sequencing to determine the cause of this discrepancy in size, and the results were aligned with the genomic MBD9 ApE file. An additional sequence in Sample #5 was discovered, and when compared to MBD9's annotated sequence from TAIR, was identified as the fifth intron of MBD9 which introduces a premature stop codon. To confirm this, primers were designed on either side of the intron so that the PCR product would be approximately 700 bp if the intron is present, while it would only be 500 bp if the intron was missing. All four samples were tested, and Sample #5 was confirmed to have the intron while Samples #1, #3, and #4 did not show presence of the intron (Figure 5). Sample #4 was then chosen for *in vitro* translation, so in order to fully confirm that no intron is present, the sample was sent for sequencing. Sequencing results were confirmed through alignment to the reference sequence. Although not in the scope of this study, the presence of intron 5 in MBD9-MYC Sample #5 could suggest a functional, truncated version of MBD9 that serves an alternate or specialized function.

Sample #4 was prepared with Vector DNA Clean Up as part of the 1-Step Human Coupled IVT protocol. Alongside Sample #4, a GFP positive control was included in translation. After the reaction, GFP was observed with a fluorescent microscope to confirm that translation worked. Translated cMBD9-MYC protein, along with *E. coli* MYC protein as a positive control and *in vitro* translated GFP as a negative control, were used in Western Blot analysis with MYC

antibody to test whether cMBD9-MYC protein can be detected (Figure 6). Figure 6 shows successful detection of the cMBD9-MYC protein at the expected size of 240 kDA. The ability to make a MBD9-MYC protein is useful for future studies such as a histone peptide microarray. Development of this successful protocol is advantageous for multiple applications of MBD9, as the MYC tag is a commonly used component of many assays.

Establishment of a MBD9-Tag Construct System

In order to create a system where multiple tags can be added to MBD9, the Gateway Cloning system of recombination was chosen. To set up this system, BP recombination of MBD9 into the pDONR221 entry vector is needed, so that LR recombination can move MBD9 from pDONR221 into a destination vector. The destination vectors pEarleyGate 200 and 300 series were chosen to introduce HA, FLAG, and MYC tags to to either the N-terminus or C-terminus of MBD9 in a format that can be transformed into *A. thaliana* for *in vivo* studies (Earley et al., 2006). This will provide the opportunity to perform protein interaction assays to see which proteins MBD9 interacts with *in vivo*. The main challenge was designing effective primers to introduce the *att*B sequences needed for recombination. Additionally, when considering genomic MBD9, the inclusion of promoter and terminator needs to be considered depending on whether a N-terminal or C-terminal tag is being added. Initially, PCR reactions using gDNA or cDNA were unsuccessful, so vector templates containing either gMBD9 or cMBD9 were utilized for successful PCR amplifications. So far, the optimization of PCR and analysis of PCR products through gel electrophoresis were successfully completed for gMBD9 C-fusion tag and cMBD9 N-fusion tag (Figure 7). These products can now be used in BP reaction to clone into pDONR221 entry vector.

Analysis of CRISPR H2A.Z Mutant Phenotypes

Two CRISPR *H2A.Z* mutant lines were recently created in the Dr. Deal lab. Phenotypic analysis of previously described *H2A.Z* mutants (*hta9-1; hta11-2* T-DNA mutant, *hta9-1; hta11-1* T-DNA mutant, and a triple *H2A.Z* mutant) and wild type Col-0 ecotype plants was performed in order to confirm that the two CRISPR lines (CRISPR *6-3* and CRISPR *21-6*) are indeed *H2A.Z* mutants. One of the most visible phenotypes of *A. thaliana H2A.Z* mutants is a decrease in the number of rosette leaves at the time of flowering. Two-tailed, independent means t-tests showed a significant difference (p-value <0.005) between the average number of rosette leaves of the wild type plants and the average number of rosette leaves for all five *H2A.Z* mutant genotypes (Figure 8). As expected, the greatest difference was found with the triple *H2A.Z* mutant genotype as all three H2A.Z genes are non-functional in comparison to the other genotypes in which two of the three genes were knocked out. Additionally, rosette leaves of all mutant plants displayed visible serration while wild type plants had smooth edges of rosette leaves (Figure 9). This confirms that the CRISPR lines are *H2A.Z* mutants.

Assessment of the petal number of the *H2A.Z* mutant plants was an additional way to confirm the CRISPR genotypes. Generally, *H2A.Z* mutants have a noticeably greater number of flowers with more than 4 petals, with 4 petals considered normal (Figure 10). The average number of flowers with abnormal petal number out of ten counted flowers was 1.33 for the wild type genotype and ranged from 3.50 to 6.67 for the *H2A.Z* mutants (Figure 11). Two-tailed, independent means t-tests showed a significant difference between all *H2A.Z* mutants and the wild type plants. For the T-DNA *hta9-1*; *hta11-2* genotype the p-value was less than 0.05, while all other mutant genotypes had p-values less than 0.005 including the CRISPR lines (Figure 11).

Conclusion

The study of the histone variant H2A.Z is an important field of work for understanding complex processes and diseases like embryonic stem cell differentiation and breast cancer. However, little is known about the mechanisms that deposit H2A.Z, including where and when H2A.Z needs to be placed in the genome. The use of the model organism *Arabidopsis thaliana* is an incredibly powerful tool in efforts to uncover these mysteries due to its easily genetically transformable nature, viability without H2A.Z, and several detectible *H2A.Z* mutant phenotypes. MBD9 seems to play a role in the deposition of H2A.Z, however, this role is mostly unexplained and poorly understood. One hypothesis is that MBD9 recognizes histone modifications to inform the SWR1 complex of the local epigenetic conditions. In order to test this, molecular tools needed to be created. Here, the successful production of a cMBD9-MYC fusion protein was shown which is a powerful asset for future experiments. Additionally, the initial constructs for multiple N-terminal and C-terminal tags fused to MBD9 were developed. These are beneficial tools for studying MBD9 protein interactions *in vivo*. While much more work is needed to uncover MBD9's function in H2A.Z deposition, these results provide a good foundation to continue building on. Lastly, the confirmation of CRISPR *H2A.Z* mutant plants was achieved through phenotypic analysis and comparison to other verified lines. CRISPR *H2A.Z* mutants were shown to have an abnormal number of flower petals and rosette leaves when compared to wild type plants. These CRISPR lines can be used for *in vivo* experiments.

Figures

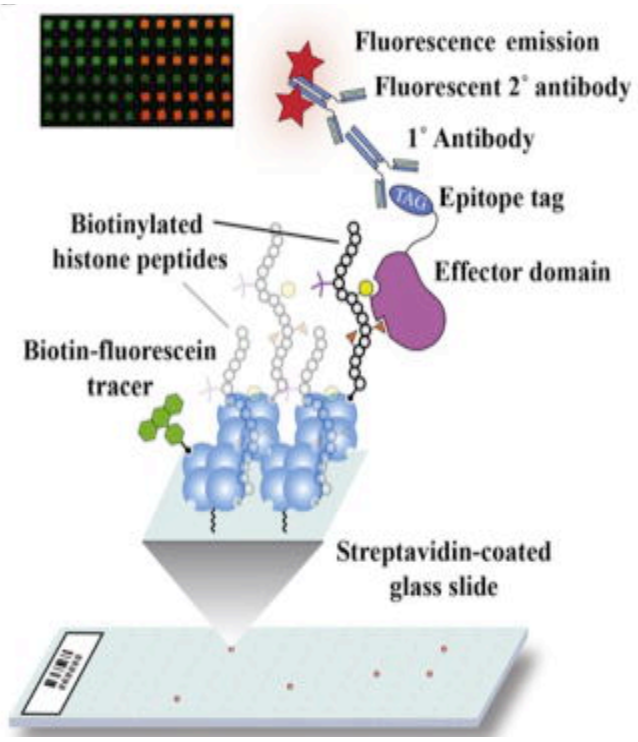

Figure 1. *Illustration of a histone peptide microarray assay*. The protein of interest (shown in purple) interacts with the histone peptides. The tag on the protein is recognized by a primary antibody, which is in turn recognized by a secondary antibody which causes the emittance of fluorescence. The black box in the top left depicts an example output of the histone peptide microarray assay. Figure obtained from the Strahl Lab at the University of North Carolina.

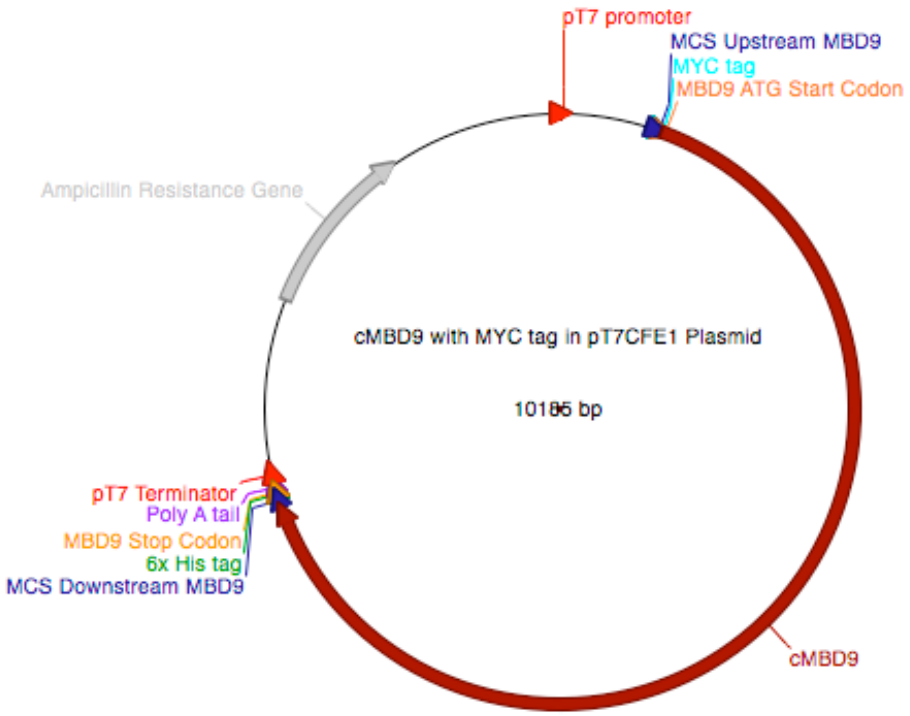

Figure 2. *Plasmid map of cMBD9 with a MYC tag inserted into pT7CFE1 plasmid*. The plasmid map shows cMBD9 inserted into the multiple cloning sequence of pT7CFE1. The plasmid has an ampicillin resistance gene for selection purposes as well as a 6x His tag for protein purification. The MYC tag is located on the N-terminal end of cMBD9. The entire construct is just over 10 thousand bases. Figure created in ApE software.

***att*B1:** 5'-GGGG ACA AGT TTG TAC **AAA AAA** GCA GGC TNN-(template specific sequence)-3'

***att*B2:** 5'-GGGG AC CAC **TTT GTA** CAA GAA AGC TGG GTN- (template specific sequence)-3

Figure 3. *The attB sequences required for BP Recombination*. These sequences were added on to both sides of MBD9 so that BP Clonase II can swap MBD9 into the pDONR221 vector which contains the *att*P sequences. To keep MBD9 in-frame, the "A" nucleotide from the ATG start codon was removed. Figure obtained from Dr. Nikos Pinotsis of the University of Vienna.

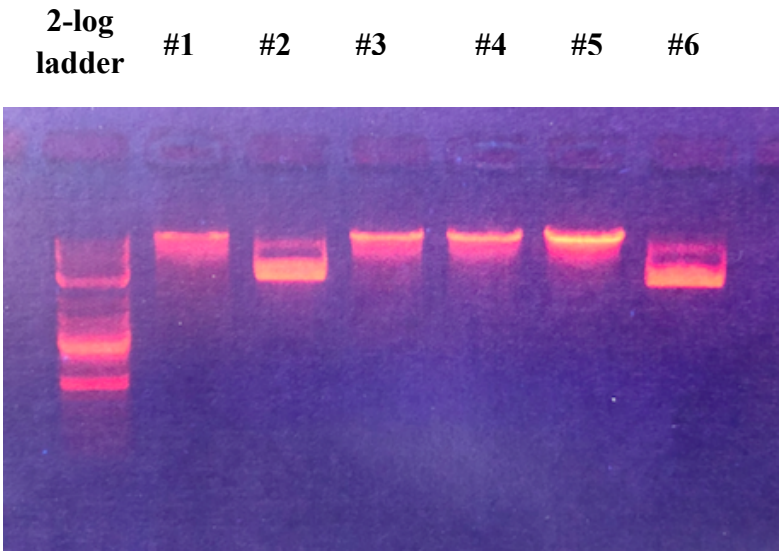

Figure 4. *ClaI digestion to test for cMBD-MYC as the Gibson Assembly insert.* Gibson Assembly was chosen to incorporate cMBD9-MYC into the plasmid pT7CFE1. Following Gibson Assembly and bacterial transformation on Ampicillin-containing plates, six colonies were selected, inoculated, and miniprepped. Presence of cMBD9-MYC was tested with ClaI digestion, a restriction enzyme that cuts in the insert but not in the plasmid. Samples #1, #3, #4, and #5 showed linearized DNA with gel electrophoresis, confirming the presence of cMBD9-MYC. Samples #2 and #6 displayed two bands, characteristic of circular plasmids, indicating that cMBD9-MYC is not present as the enzyme was unable to linearize the plasmid.

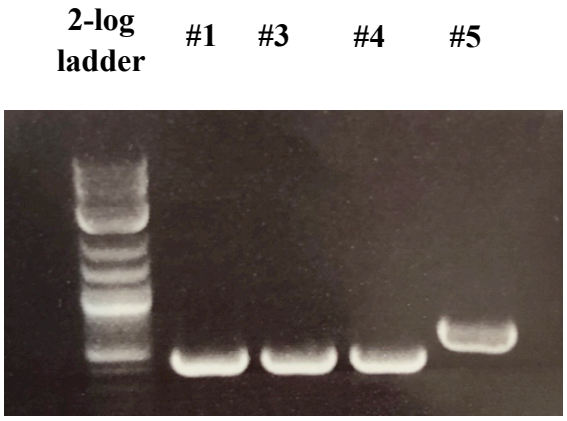

Figure 5. *Gel electrophoresis of a segment of MBD9 to test for the presence of intron 5*. Following Gibson Assembly of cMBD9-MYC into pT7CFE1, primers were designed on either side of a potential intron that was failed to be removed. The presence of the intron causes the PCR product to be roughly 700 base pairs while the lack of intron leads to a PCR product around 500 base pairs. Samples #1, #3, and #4 do not contain an intron. Sample #5 does contain MBD9's intron 5. Lane 1 features a DNA ladder for length estimates.

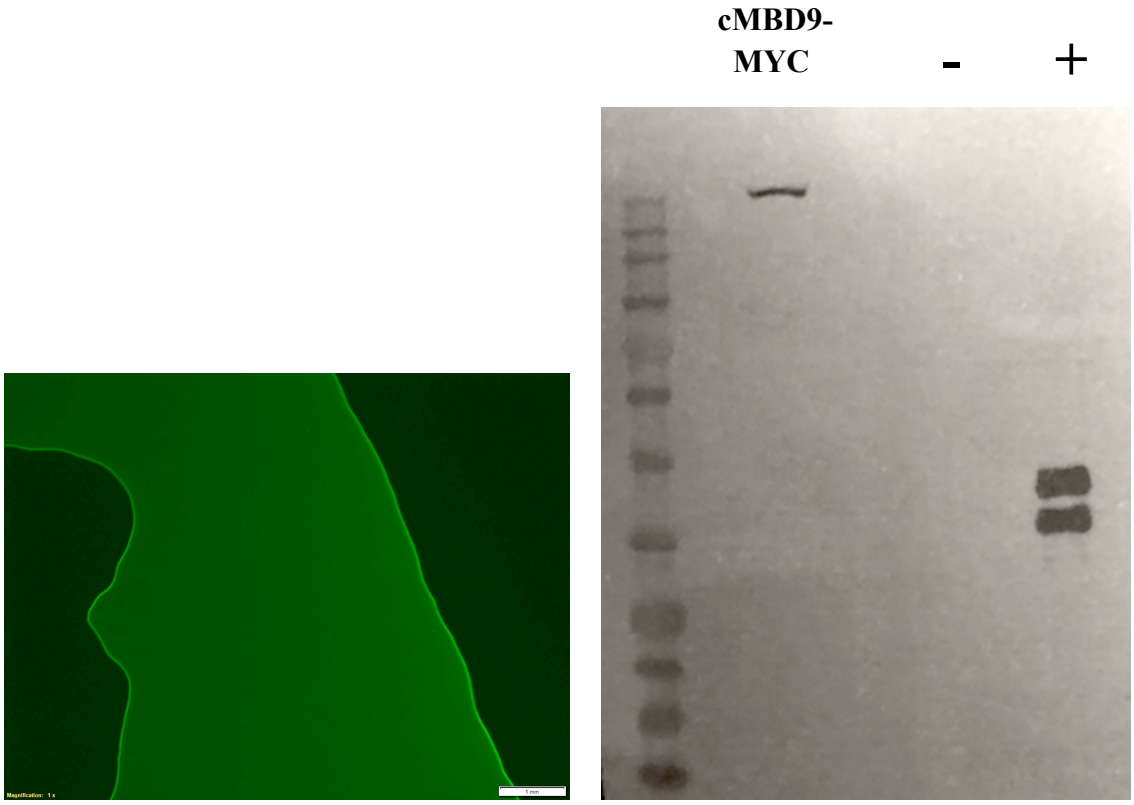

Figure 6. *GFP confirmation of translation and Western Blot of cMBD9-MYC*. Translation of cMBD9-MYC was done using the 1-Step Human Coupled IVT Kit and protocol. A GFP positive control for translation was included and confirmed with a fluorescent microscope. 2 µL of GFP was visualized. For the cMBD9-MYC protein Western Blot using a MYC antibody, *in vitro* translated *E. coli* MYC protein as a positive control (designated by +) and GFP as a negative control (designated by -) were included. Western Blot was performed by Dr. Paja Sijacic. cMBD9-MYC is shown at around 240 kDa, the correct weight.

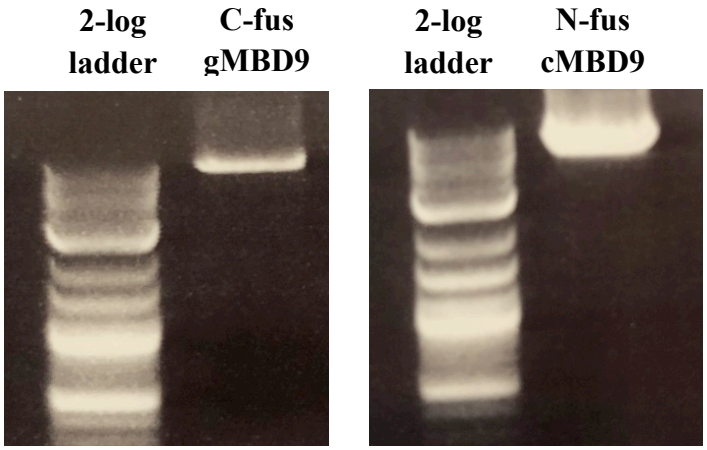

Figure 7. *Gel electrophoresis of gMBD9 and cMBD9 with attB flanking sequences*. PCR products of MBD9 with *att*B flanking sequences on both sides of MBD9 were made using cDNA and gDNA plasmid templates. The *att*B flanking sequences are required for for Gateway Recombination Cloning assays. Both gel electrophoresis images include a DNA ladder to confirm the success of PCR by viewing the length of the PCR product.

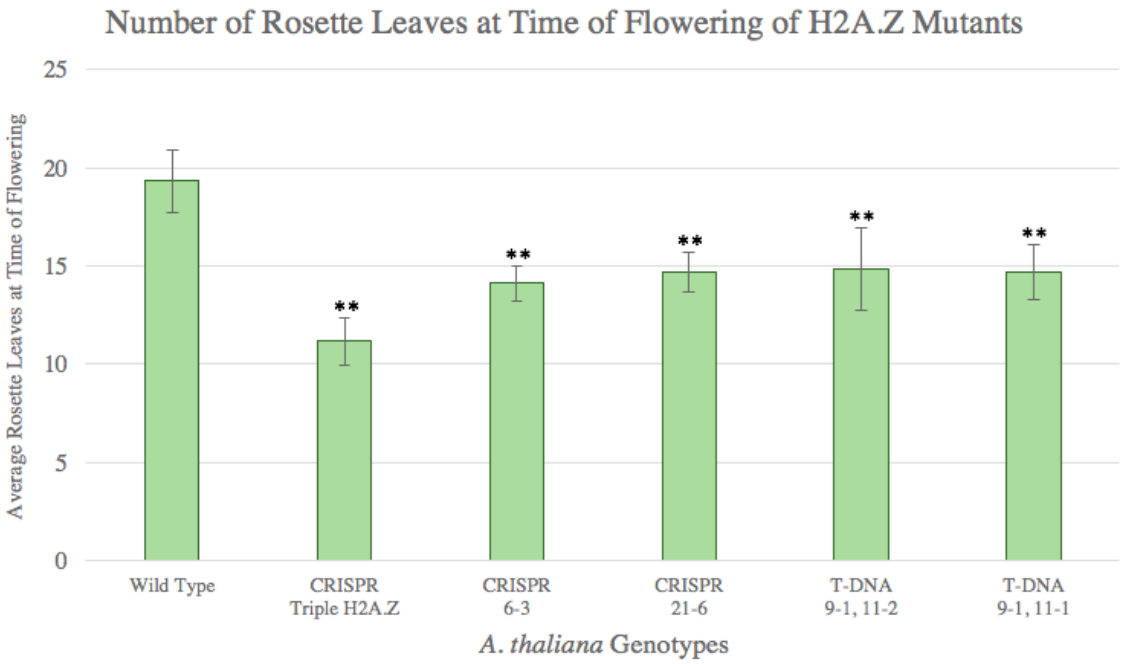

Figure 8. *Number of rosette leaves at time of flowering of A. thaliana H2A.Z mutants*. Plants were grown in at 20°C with long-day cycles (16 hours of light, 8 hours of dark). Genotypes grown were wild type Col-0 ecotype and H2A.Z mutants: *hta9-1*; *hta11-2* T-DNA mutant, *hta9-1*; *hta11-1* T-DNA mutant, CRISPR *6-3* mutant, CRISPR *21-6* mutant, and a triple H2A.Z mutant. Either 6 or 7 plants were grown for each genotype. Counting of rosette leaves was done at the emergence of the shoot when plants switched from leaf production to flower production (time of flowering). Rosette leaves are located at the base of the plant, and cotyledons were counted as rosette leaves. Two-tailed, independent means t-tests were calculated between wild type averages and mutant plant averages of rosette leaves. A single star (*) indicates a p-value <0.05. A double star (**) indicates a p-value <0.005. All mutant genotypes had significantly less rosette leaves at time of flowering compared to the wild type, including the CRISPR lines (p-value <0.005).

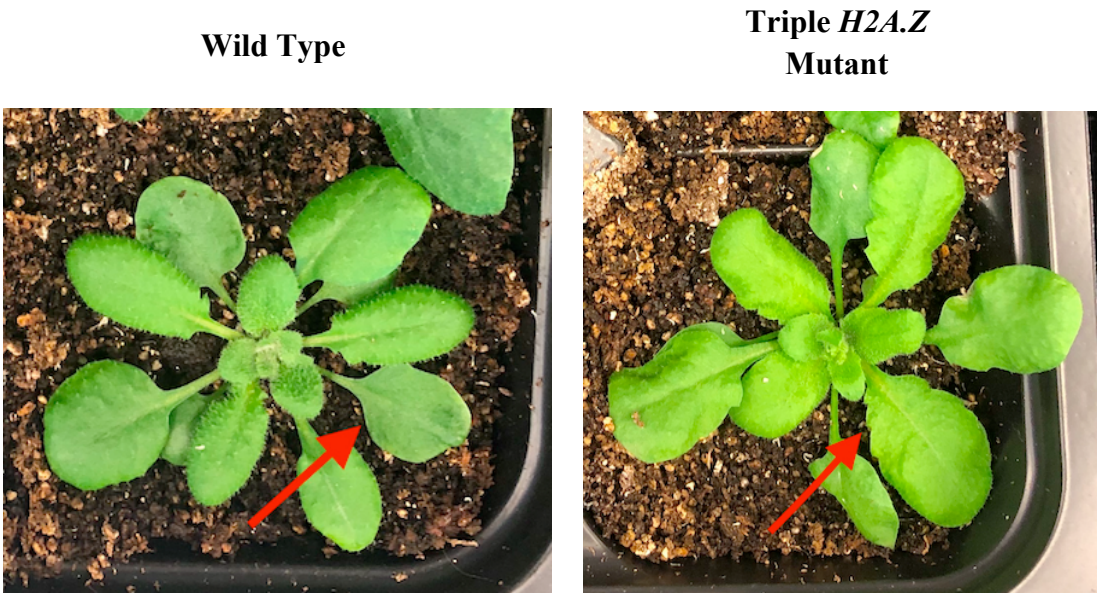

Figure 9. *Serrated rosette leaf phenotype of A. thaliana H2A.Z mutants*. *A. thaliana* wild type Col-0 ecotype plants display a rounded rosette leaf phenotype. In comparison, *H2A.Z* mutant plants (triple *H2A.Z* mutant shown in Figure) display a serrated rosette leaf phenotype with jagged edges visible on leaves. Plants were grown in at 20°C with long-day cycles (16 hours of light, 8 hours of dark).

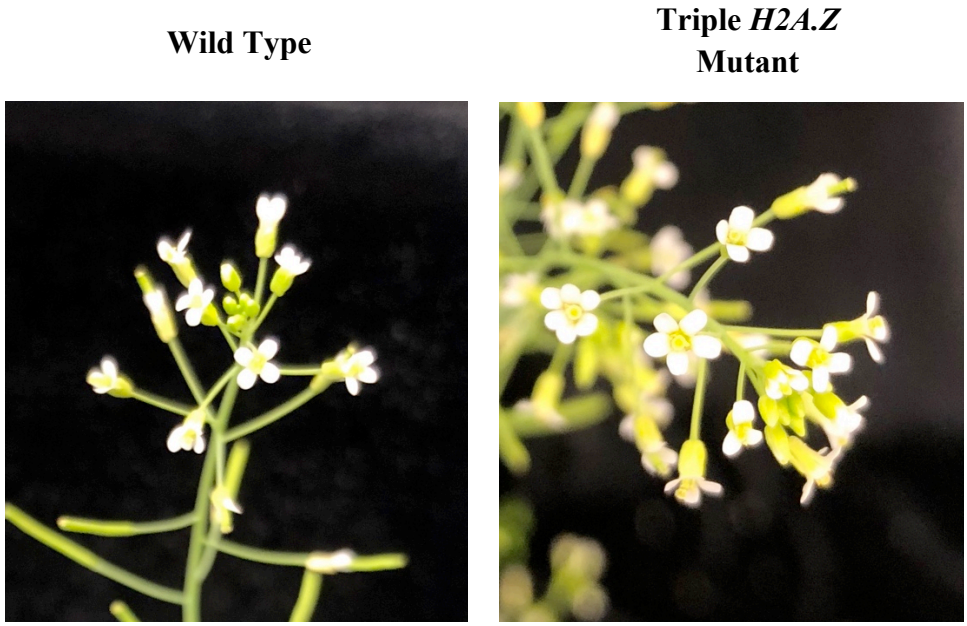

Figure 10. *Abnormal petal number phenotype of A. thaliana H2A.Z mutants*. Wild type Col-0 ecotype *A. thaliana* plants show the expected petal number on flowers (4). *H2A.Z* mutant plants (triple *H2A.Z* mutant genotype shown here) display a mix of normal petal number flowers and flowers with an abnormal number (greater than 4 petals). Plants were grown in at 20°C with long-day cycles (16 hours of light, 8 hours of dark).

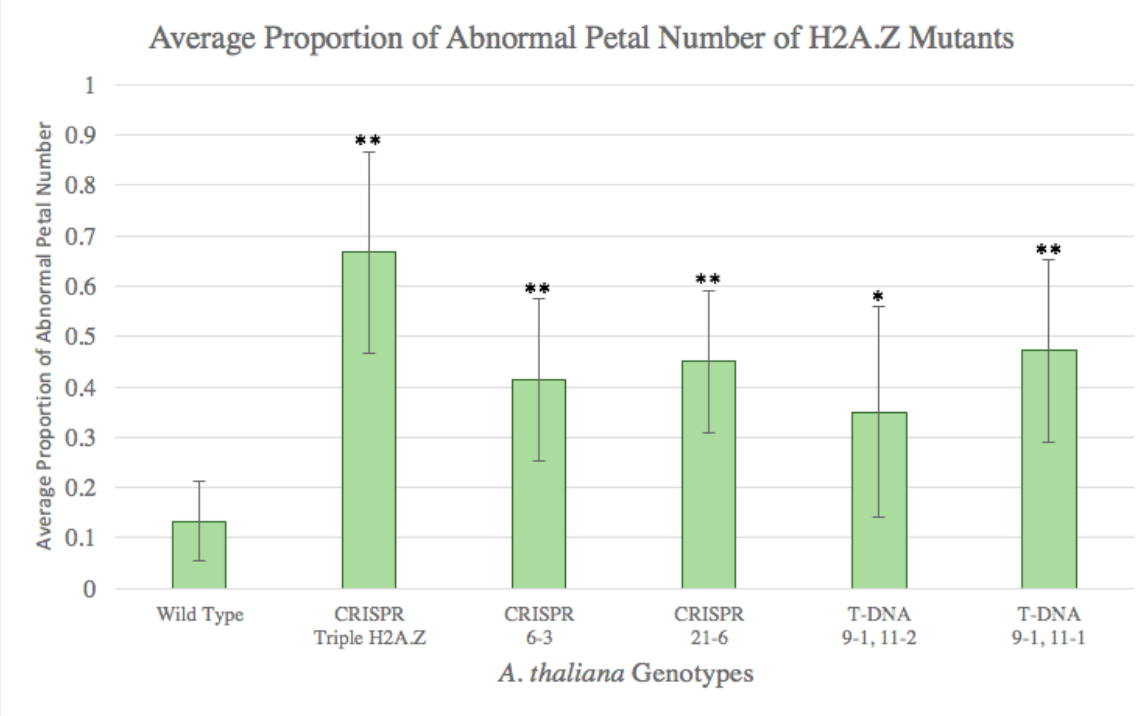

Figure 11. *Average proportion of abnormal petal number of A. thaliana H2A.Z mutants.* Plants were grown in at 20°C with long-day cycles (16 hours of light, 8 hours of dark). Genotypes grown were wild type Col-0 ecotype and H2A.Z mutants: *hta9-1*; *hta11-2* T-DNA mutant, *hta9-1*; *hta11-1* T-DNA mutant, CRISPR *6-3* mutant, CRISPR *21-6* mutant, and a triple H2A.Z mutant. Either 6 or 7 plants were grown for each genotype. Flower petal counting was done at the emergence of distinguishable flower petals. Flowers with more than 4 petals were counted as abnormal. 10 flowers were counted per plant and the proportion of flowers with abnormal petal numbers was calculated. Two-tailed, independent means t-tests were calculated between wild type averages and mutant plant averages of abnormal petal number proportions. A single star (*) indicates a p-value <0.05. A double star (**) indicates a p-value <0.005. All mutant genotypes had significantly less rosette leaves at time of flowering compared to the wild type (p-value <0.005, except for T-DNA *hta9-1*; *hta11-2* which had a p-value <0.05).

References

Adam M, Robert F, Larochelle M, Gaudreau L. (2001) H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. *Mol Cell Biol* 21: 6270-6279

Alonso J.M., Stepanova A.N. (2003) T-DNA Mutagenesis in *Arabidopsis*. *Plant Functional Genomics*. 236:177-88.

Ausio, J & Abbott DW. (2002). The Many Tales of a Tail: Carboxyl-Terminal Tail Heterogeneity Specializes Histone H2A Variants for Defined Chromatin Function. *Biochemistry*, *41*(19), 5945-5949. doi: 10.1021/bi020059d.

Cai Y, Jin J, Florens L, Swanson S, Kusch T, Li B, Workman J, Washburn M, Conaway R,

Conaway J. (2005) The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes. *J Biol Chem* 280: 13665-13670

Choi K, Park C, Lee J, Oh M, Noh B, & Lee I. *Arabidopsis* homologs of components of the SWR1 complex regulate flowering and plant development.*Development*. 2007; **134**: 1931-1941

Coleman-Derr D & Zilberman D. (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet.* 8(10):e1002988.

Cong L, Ann Ran F, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*. Vol. 339, Issue 6121, pp. 819-823. DOI: 10.1126/science.1231143

Creyghton MP, Markoulaki S, & Levine SS. H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. *Cell*. 2008;135(4):649-61.

Cutter AR, & Hayes JJ. (2015). A brief review of nucleosome structure. *FEBS letters*, *589*(20 Pt A), 2914-22.

Deal R, Kandasamy M, McKinney E, & Meagher R. (2005) The nuclear actin-related protein ARP6 is a pleiotropic developmental regulator required for the maintenance of FLOWERING LOCUS C expression and repression of flowering in Arabidopsis. Plant Cell 17: 2633- 2646

Drynurst D, McMullen B, Fazli L, Rennie PS, & Ausio J. (2011). Histone H2A.Z prepares the prostate specific antigen (PSA) gene for androgen receptor-mediated transcription and is upregulated in a model of prostate cancer progression. *Cancer Letters*. 2011;315:1:38-47 doi: 10.1016/j.canlet.2011.10.003

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, & Pikaard CS. (2006) Gateway- compatible vectors for plant functional genomics and proteomics. *The Plant Journal*. 45: 616-629. doi:10.1111/j.1365-313X.2005.02617.

Fry CJ, & Peterson CL. (2001) Chromatin remodeling enzymes: who's on first? *Curr Biol*. 11: R185–R197.

Gelvin SB. (2003) Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev*. 2003;67(1):16-37

Hasegawa Y, Taylor D, Ovchinnikov DA, Wolvetang EJ, de Torrenté L, & Mar JC. (2015) Variability of Gene Expression Identifies Transcriptional Regulators of Early Human

Embryonic Development. *PLoS genetics*, *11*(8), e1005428. doi:10.1371/journal.pgen.1005428.

Krogan NJ, Michael-Christopher K, Datta N, Hughes TR, Buratowski S, & Greenblatt JF. (2003) A snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Molecular Cell*, 2003;12:6:1565-1576. doi:10.1016/S1097-2765(03)00497-0

Luger K, Dechassa ML, & Tremethick DJ. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?. *Nature reviews. Molecular cell biology*, *13*(7), 436-47. doi:10.1038/nrm3382.

Luger K, Mader AW, Richmond RK, Sargent DF, & Richmond TJ. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature.* 389:251–60.

March-Diaz R & Reyes J. (2009) The beauty of being a variant:H2A.Z and the SWR1 complex in plants. *Mol Plant* 2:565-577

Palmer DK, O'Day K, Trong HL, Charbonneau H, & Margolis RL. (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci 88: 3734–3738.

Peng M, Cu Y, Bi Y, & Rothstein S. (2006) AtMBD9: a protein with a methyl-CpG-binding domain regulates flowering time and shoot branching in Arabidopsis. *Plant J* 46:282-296

Rangasamy, D. (2010) Histone Variant H2A.Z Can Serve as a New Target for Breast Cancer Therapy. *Current Medicinal Chemistry.* 17: 3155. doi: 10.2174/092986710792231941

Rothbart SB, Krajewski K, Strahl BD, & Fuchs SM. (2012). Peptide microarrays to interrogate the "histone code". *Methods in Enzymology*. 512, 107-35.

Sijacic P, Holder D, Bajic M, & Deal R. (2019) Methyl-CpG-binding domain 9 (MBD9) is required for H2A.Z incorporation into chromatin at a subset of H2A.Z-enriched regions in the Arabidopsis genome. Manuscript submitted for publication.

Soboleva TA, Nekrasov M, Ryan DP, & Tremethick DJ. (2014) Histone variants at the transcription start-site. *Trends in Genetics.* 30(5):199-209.

Strahl BD & Allis CD. (2000) The language of covalent histone modifications. *Nature.* 403: 41–5.

Sutcliffe EL, Parish IA, He YQ, Juelich T, Tierney ML, Rangasamy D, Milburn PJ, Parish CR, Tremethick DJ, & Rao S. Dynamic histone variant exchange accompanies gene induction in T cells. (2009) *Mol Cell Biol.* 29(7):1972-86.

Thatcher TH & Gorovsky MA. (1994) Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Res.* ;22(2):174-9.

Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E, Vercruysse L, Dedecker M, Verkest A, Vandepoele K, Martens L, Witters E, Gevaert K, & De Jaeger G (2015) An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. *Nat Protoc.* 10:169-187

Yang, XJ. (2004) Lysine acetylation and the bromodomain: a new partnership for signaling. *Bioessays*. 26, 1076-1087.