

EXAMINATION OF EPIGENETIC CONTROL MECHANISMS AND THE GENETIC
CHARACTERIZATION OF VEGETATIVE CLASS PROFILINS IN *ARABIDOPSIS*

THALIANA

by

KRISTOFER JOSEPH MUSSAR

(Under the Direction of Richard B. Meagher)

ABSTRACT

The dynamic nature of the actin cytoskeleton is maintained through interactions with a suite of different Actin Binding Proteins (ABPs) that remodel actin filaments. One such ABP, profilin, is believed to promote both actin filament polymerization and depolymerization via the binding and sequestering of globular actin (G-actin) monomers. In *Arabidopsis thaliana*, profilin is encoded by a five-member gene family that contains two distinct subclasses, vegetative and reproductive. PRF1, PRF2, and PRF3 are expressed in all vegetative tissues, while PRF4 and PRF5 are specifically expressed only in reproductive tissues. The goal of this study was to characterize the three vegetative members in terms of their roles in plant cell and organ development. Using a collection of T-DNA insertion mutants and RNAi knockdowns targeting individual and combinations of PRF1, PRF2, and PRF3, I found that each of these three variants gave rise to specific developmental deficiencies. Plants lacking profilins had defects in rosette leaf morphology, inflorescence stature, petiole elongation, and lateral root initiation and growth. Microscopic examination of these dwarfed plants lacking in profilin variants

indicated that they have smaller cells defective in cell elongation. Evidence is presented that mixtures of independent function, quantitative genetic effects, and functional redundancy have preserved the three vegetative profilin genes.

I also explore the possibility of DNA sequence guiding various epigenetic control mechanisms. My efforts focus on understanding how sequence facilitates the epigenomic landscape of histone post-translational modifications (PTMs). Through interpretation of PTM deposition data at the gene and gene family level, I discovered that recently duplicated gene sequences exhibit varying levels of conservation across their histone modification enrichment profiles. These data suggest that epigenetic controls aid “evolution by gene duplication” by silencing some recent gene duplicates, but not others, until beneficial mutations and subfunctionalization can occur. By searching for correlations among these enrichment profiles I was able to detect combinatorial patterns of histone modification marks within each gene family. Distinct patterns containing known activation marks that are cooperatively interacting were found in gene families where sequence was more conserved, suggesting that sequence may be playing some role in facilitating PTM deposition throughout the genome.

INDEX WORDS: *Arabidopsis thaliana*, actin, profilin, cytoskeletal dynamics, epigenetics, chromatin, histone posttranslational modification, sequence conservation, gene duplication, combinatorial patterns, molecular evolution, subfunctionalization

EXAMINATION OF EPIGENETIC CONTROL MECHANISMS AND THE GENETIC
CHARACTERIZATION OF VEGETATIVE CLASS PROFILINS IN *ARABIDOPSIS*
THALIANA

by

KRISTOFER JOSEPH MUSSAR

BS, University of Illinois Urbana-Champaign, 2006

MS, George Washington University, 2007

MBA, University of Georgia, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

© 2013

Kristofer Joseph Mussar

All Rights Reserved

EXAMINATION OF EPIGENETIC CONTROL MECHANISMS AND THE GENETIC
CHARACTERIZATION OF VEGETATIVE CLASS PROFILINS IN *ARABIDOPSIS*
THALIANA

by

KRISTOFER JOSEPH MUSSAR

Major Professor: Richard B. Meagher

Committee: Sidney Kushner
Jonathan Arnold
Xiaoyu Zhang
Walter Schmidt

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

DEDICATION

To Elizabeth Ann Mussar, her resiliency, work ethic, and dedication were truly inspirational.

ACKNOWLEDGEMENTS

First, I would like to extend my thanks to my adviser, Dr. Rich Meagher, for helping me develop into a better scientist. I also would like to thank my advisory committee, Dr. Sidney Kushner, Dr. Jonathan Arnold, Dr. Xiaoyu Zhang, and Dr. Walter Schmidt for their guidance and critical analysis of my research. Their feedback challenged me to approach the scientific process with more awareness. Overwhelming thanks must be offered to Libby McKinney and Dr. Muthugapatti Kandasamy, for their day to day training and advice were crucial to my success. To Eileen Roy, who provided me with support both in and out of the lab and was always there to lend an ear. A special thanks to the backbone of our lab and often overlooked Yolanda Lay; her assistance will not go without appreciation.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	ix
CHAPTER	
1 Introduction and Literature Review	1
2 Analysis of <i>Arabidopsis thaliana</i> Plants Deficient in Vegetative Class Profilins Identifies Independent and Quantitative Genetic Effects.....	42
3 Phylogenetic Identification of Inherited Patterns of Nucleosomal Histone Modification.....	96
4 Summary and Conclusions	161
APPENDICES	
A The Influence of DNA Sequence on Epigenome-Induced Pathologies.....	177

LIST OF TABLES

	Page
Table 2.S1: Summary of all phenotypic measurements for the various mutant and RNAi silenced plant lines examined	95
Table 3.1: Summary of the 14 different histone modifications analyzed in this study ...	139
Table 3.2: Gene family information for actins, profilins, ADFs, and ARPs	140
Table 3.3: Summary of the enrichment of histone modification marks among gene pairs.....	141
Table 3.4: Summary of the combinatorial patterns of histone modification marks detected within and among gene families	142
Table 3.S1: Summary of ChIP-Chip data for the actin gene family.....	147-148
Table 3.S2: Summary of ChIP-Chip data for the profilin gene family.....	149-150
Table 3.S3: Summary of ChIP-Chip data for the Actin Depolymerizing Factor (ADF) gene family.....	151-152
Table 3.S4: Summary of ChIP-Chip data for the Actin Related Protein (ARP) gene family	153-154
Table 3.S5: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among the actin gene family	155
Table 3.S6: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among the profilin gene family.....	156

Table 3.S7: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among the ADF gene family.....	157
Table 3.S8: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among the ARP gene family	158
Table 3.S9: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among actin, profilin, ADF, and ARP gene families.....	159
Table 3.S10: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among actin, profilin, and ADF gene families	160
Table A1.1: Examples of genes and specific sequences that support or reject the hypothesis that genotype predisposes transgenerationally inherited epitype and phenotype.....	229

LIST OF FIGURES

	Page
Figure 1.1: Defining an epitope	39
Figure 1.2: Remodeling of the actin cytoskeleton by Actin Binding Proteins (ABPs)	40
Figure 1.3: Tissue-specific expression profiles for <i>Arabidopsis</i> profilins (PRF1-PRF5) retrieved from the <i>Arabidopsis</i> e-GFP browser (Winter et al., 2007)	41
Figure 2.1: Analysis of mutants defective in individual vegetative profilins	78
Figure 2.2: Analysis of profilin RNA and protein expression for vegetative PRF single T- DNA mutants and complement lines	80
Figure 2.3: Morphological analysis of vegetative profilin double mutants	81
Figure 2.4: qRT-PCR data and western blot analysis for double and triple mutant/ RNAi line	83
Figure 2.5: Morphological analysis of PRF double and triple RNAi lines.....	85
Figure 2.6: PRF3 knockout displays elongated petioles, while PRF3 overexpression results in no phenotypic effects.....	87
Figure 2.7: Vegetative profilin double and triple RNAi lines show defects in lateral root formation and growth.....	89
Figure 2.8: Cell size and F-actin filament organization in <i>PRF1PRF2</i> -RNAi plants	90
Figure 2.S1: Morphology of vegetative profilin single mutants.....	92
Figure 2.S2: Morphology and qRT-PCR analysis of transcript levels for vegetative PRF single RNAi lines.....	93

Figure 2.S3: Vegetative PRF double and triple RNAi lines that are only weakly silenced for profilin RNA expression show slight defects in lateral root development	94
Figure 3.1: Examples of enrichment profiles generated for this analysis.....	131
Figure 3.2: H3K36me3 and H4K8ac enrichment profiles for ACT1 and ACT3 gene pairs	133
Figure 3.3: H3K4me1 enrichment profiles for the PRF4 and PRF5 gene pair.....	134
Figure 3.4: H3K4me3 enrichment profiles for ADF1 and ADF4 and H3K4me2 enrichment profiles for ADF7 and ADF10 gene pairs.....	135
Figure 3.5: H3K36me3, H3K4me3, and H4K8ac enrichment profiles for ACT7 (actin combinatorial pattern).....	136
Figure 3.6: H3K36me3, H3K4me3, H3K9ac, H3K14ac, and H4K8ac enrichment profiles for PRF2 (profilin combinatorial pattern).....	137
Figure 3.7: H3K36me3 and H4K8ac enrichment profiles for ADF4 (ADF combinatorial pattern)	138
Figure 3.S1: Bayesian phylogenetic analysis of the <i>Arabidopsis</i> actin gene family	143
Figure 3.S2: Bayesian phylogenetic analysis of the <i>Arabidopsis</i> profilin gene family ...	144
Figure 3.S3: Bayesian phylogenetic analysis of the <i>Arabidopsis</i> Actin Depolymerizing Factor (ADF) gene family.....	145
Figure 3.S4: Bayesian phylogenetic analysis of the <i>Arabidopsis</i> Actin Related Protein (ARP) gene family	146
Figure A1.1: Summary of relationship between epitype and DNA sequence	225
Figure A1.2: The relationships among genotype, epitype, and phenotype.....	227

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Overview of Epigenetics

Epigenetics is classically known as the study of inherited gene expression or phenotypic changes that are caused by mechanisms that do not involve changes in DNA sequence. These types of changes are a result of non-genetic factors that cause an organism to differentially express their genes, which can impact phenotype (Hunter, 2008). The term 'epigenetics' was originally defined by C.H. Waddington in 1942 and it was a blend of the two words genetics and epigenesis (Waddington, 1942). Epigenesis is defined as the differentiation of cells from their initial embryonic stage of development. Some years later David Nanney took this one step further by introducing the term 'epigenetic controls' to describe the diverse inherited differences between genetically identical daughter cells (Nanney, 1958). Currently, epigenetics refers to the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by DNA sequence (Russo *et al.*, 1996).

Epigenetics plays an important role in cellular differentiation allowing distinct cell types to have specific characteristics despite sharing the same genomic sequence. For example, during morphogenesis, totipotent stem cells differentiate into the many different cell types by epigenetically activating and inhibiting certain genes (Wolf, 2007).

This makes epigenetic regulation a target for medical therapies. Understanding the myriad of ways epigenetic mechanisms control gene expression can have a profound impact on genetic therapies.

In order to understand the complexities behind epigenetic regulation, one must first understand the importance of chromatin. Chromatin refers to the DNA and proteins that together make up chromosomes. A major component of chromatin is histones. DNA is wrapped around histone proteins in order to compact and strengthen DNA, as well as regulate gene expression. Chromatin can be separated into two forms: a condensed heterochromatic form that is associated with gene repression (since proteins, like transcription factors, are unable to access these DNA regions), and euchromatin, which is a more relaxed and extended form that allows proteins access to the DNA sequence and is therefore normally associated with active gene expression. Chromatin is extremely dynamic. Epigenetic control mechanisms, like DNA methylation and histone modifications, are constantly relaxing and condensing DNA regions in order to facilitate temporal and spatial gene expression.

The two most studied pathways of epigenetic control are DNA methylation and histone modification. In addition to these pathways, we will also examine the roles that histone variants, nucleosome positioning/ phasing, RNA interference (RNAi), and non-coding RNAs (ncRNAs) play in directing epigenetic control in plants, fungi, and animals. In addition, there will be a section dedicated toward some of the more species specific epigenetic pathways. These will include mechanisms like dosage compensation and genomic imprinting. By taking a broad look at the different types of epigenetic pathways across different kingdoms, we are hoping to gain a better understanding of the

evolutionary history of epigenetic control, with hopes to determine which mechanisms are the most evolutionarily conserved.

DNA Methylation

DNA methylation involves the addition of a methyl group (CH_3) to the number 5' carbon of the cytosine pyrimidine ring (DNA methylation can also occur on the number 6 nitrogen of the adenine purine ring, but for epigenetic purposes we will concentrate on 5mCytosine methylation). This type of DNA modification is able to be inherited through cell division, making DNA methylation crucial for cellular and organismal development. Cytosine methylation allows for the stable alteration of gene expression patterns, leading to cellular differentiation in higher organisms. DNA methylation has also been shown to play a major role in the development of many different types of cancers (Jaenisch, Bird, 2003). In terms of epigenetics, DNA methylation is known to play a major role in facilitating chromatin structure.

5mC DNA methylation tends to occur in areas of the genome that are not expressed. Examples of these areas include repeat sequences and transposable elements (Slotkin and Martienssen, 2007). This is because DNA methylation has the effect of reducing gene expression. However, more recent studies have shown methylation to be associated with transcribed genes, demonstrating that transcription and methylation are interwoven processes (Zilberman and Henikoff, 2007). DNA methylation patterns have been shown to be affected by various environmental factors. In order to maintain appropriate methylation profiles, there are families of enzymes called DNA methyltransferases (MTs) that catalyze the transfer of a methyl group to DNA. This

allows the integrity of DNA to be preserved as it is inherited from generation to generation.

Plants and mammals are said to have very similar epigenetic regulation landscapes. When it comes to DNA methylation, however, plants use a different recognition pattern, use different molecular machinery (DMTs), and have the ability to reverse the methylation process in non-dividing cells. In fact, plant genomes are much more highly methylated than mammalian genomes (up to ~30% in plants, while only ~2-7% in mammals) (Antequera *et al.*, 1984). There are two types of methylation: *de novo* and maintenance. *De novo* methylation occurs on unmethylated DNA and can affect CpG, CpNpG, and CpNpN (N= A, T, or C) sequences in plants, whereas in mammals you tend to only detect CpG dinucleotide methylation. Maintenance methylation refers to the process of maintaining methylation patterns on hemimethylated DNA after replication. Their substrates are typically CpG and CpNpG.

Plants have members from all three conserved DNA methyltransferase (DMT) families. The MET1 family of DMTs are mainly considered CpG maintenance methyltransferases. These family members are similar to mammalian DNMT1. *De novo* methyltransferases in plants are referred to as Domains Rearranged Methyltransferases (DRMs). DNMT3 is the mammalian homolog with a different arrangement of protein domains. In plants, DRMs methylate DNA sequences in all three contexts and are thought to play a prominent role in RNA directed DNA methylation (RdDm) (Allis *et al.*, 2007). The plant specific chromomethylase, CMT3, is thought to have both *de novo* and maintenance methylation abilities and acts on CpNpG substrates. Loss of CMT3 results in the reactivation of certain silent transposons. In addition, while it is known that

mammalian *dnmt1* and *dmnt3* mutants die during embryonic development, plant *met1*, *drm*, and *cmt3* mutants are usually viable and fertile (Chan *et al.*, 2005).

DNA methylation can be lost through both passive (failure to maintain methylation through DNA replication) and active (enzymatic activity) mechanisms. This makes DNA methylation a reversible process. In mammals, it was thought that active reversal could be achieved through DNA glycosylases, which are involved with base excision repair (Kress *et al.*, 2001). In plants, Repressor of silencing 1 (ROS1) is a protein with a glycosylase domain that has a nicking activity that only acts on methylated DNA, making it an ideal candidate for removal of methylated bases (Kapoor *et al.*, 2005).

Through restriction and nearest-neighbor analysis, it was determined that most fungal species have very low levels of DNA methylation in their genomes (0.1-0.5%) (Antequera *et al.*, 1984). The exception in the fungal kingdom appears to be *Neurospora crassa*, which has a well characterized DNA methylation system, making this organism the most relevant to understanding epigenetic regulation in other higher eukaryotes. The enzymes involved act mostly on CpG dinucleotides and methylated sites appear to cluster together away from unmethylated regions. The main DNA methyltransferase in *N. crassa* is DIM-2. This protein is unique in that it differs from the four major families of DMTs (Groll, Bestor, 2005). However, when all DNA methylation mechanisms are knocked out in *Neurospora*, there is no effect on sexual reproduction or growth. More experiments will need to be performed in order to determine whether fungi evolved to not require this regulating pathway, or if this type of regulation can be achieved through alternative fungal pathways.

While DNA methylation does not occur in all animals (i.e. *Drosophila melanogaster* and *Caenorhabditis elegans*), it has been shown to be vital for the normal development of most animal species and is also associated with processes like genomic imprinting and X chromosome inactivation. Just as in plants, mammals have both *de novo* and maintenance methylation. DNMT1 is an example of a mammalian maintenance methyltransferase, while DNMT3A and DNMT3B are examples of *de novo* methyltransferases. Each one of these three proteins was found to be essential for the normal development of a mouse embryo. Knocking out any one of these proteins results in lethality (Li *et al.*, 2002). The most common substrate for DNA methylation in animals is CpG dinucleotides. It is estimated that between 60-90% of all CpG dinucleotides in mammals are 5'-cytosine methylated (Tucker, 2001). In the 5' regulatory regions of many animal genes there are groups of unmethylated CpGs that are referred to as CpG islands. CpG islands are thought to play a role in cancer, whereby these regions become methylated causing transcriptional silencing. We must note that although these methylation patterns are transferred from cell to cell, environmental and pathological factors can lead to effects and changes in methylation.

Determining the location of *de novo* DNA methylation remains unclear. Current research has begun to focus on the involvement of small non-coding RNAs in directing *de novo* methylation (RdDm) (discussed below). Overall, it appears as if there are multiple ways that DNA methylation can affect transcription. Methylated DNA could physically block access of transcription activating proteins to the DNA sequence. Alternatively, proteins bound to methylated DNA can recruit other proteins to alter the

chromatin structure thereby rendering it silent. Either way, there is clearly an important relationship between DNA methylation and chromatin structure.

Histone Modifications

Histone modifications represent another major process of epigenetic control. Histones are proteins found in the nucleus of eukaryotic cells, which are responsible for packaging the DNA into structural units called nucleosomes (Cox *et al.*, 2005). Histones are the main component of chromatin, and as a result play a key role in gene regulation. Histones undergo a process known as posttranslational modification (PTM), which alters their interactions with DNA and other nuclear proteins. In particular, histones H3 and H4 have long protruding tails that can be covalently modified via a multitude of mechanisms. Histone tails can be methylated (mono-, di-, or tri-), acetylated, phosphorylated, SUMOylated, ubiquitinated, glycosylated, citrullinated, and ADP-ribosylated (Strahl, Allis, 2000). These covalent modifications influence DNA/protein interactions thereby affecting such processes as gene expression, DNA repair, and chromosome condensation. Collectively, these modifications are thought to guide DNA/ chromatin interactions to different phenotypic outcomes by recruiting specific transcriptional regulators. These interactions are thought to make up what is known as the “Histone code” (Jenuwein, Allis, 2001). However, the exact mechanism behind the influence of these histone modification interactions remains unclear.

The two most common types of histone modifications are methylation and acetylation. Histone methyltransferases (HMTs) are a family of proteins that are able to attach a methyl group(s) to lysine and arginine residues protruding from H3 and H4 histone tails. These HMTs contain a common SET domain. In plant species like

Arabidopsis, there are approximately 30 SET domain proteins. This is much higher than what is found in other organisms like fly (14) and yeast (only 4), but still less than mammals (50 have been found in mice) (Allis *et al.*, 2007). These SET domains appear to be the only conserved portion of these proteins across different species suggesting that plant and animal modifying complexes could be very different. Histone tails can be mono-, di- or tri-methylated at multiple amino acid positions. This creates a complex network of combinatorial modifications that is proving extremely hard to decipher. It is important to note that a certain modification in plants could play a different role than it does in animals and fungi. For example, H3K9me1 and H3K9me2 are associated with heterochromatic formation in plants, while in fungi and animals they are associated with active euchromatin. The opposite is true for H3K9me3, which is associated with active chromatin in plants and silent heterochromatin in fungi and animals (Vanyushin, 2006). This lack of functional conservation among kingdoms makes the understanding of the combinatorial nature of these modifications even more complex.

A common theme that appears when analyzing epigenetic control pathways is reversibility. Currently there are 12 putative histone acetyltransferases (HATs) and 18 histone deacetylases (HDACs) found in plants (Pandey *et al.*, 2002). These numbers are comparable to those found in mammals, but most other animal species have less. Acetyl groups are attached to the lysine residues on H3 and H4 histone tails, and tend to be associated with gene activation. Currently, there are three families of HDACs: the sirtuins (see fungi section below); the classical HDACs; and a family of HD2-like enzymes that are only found in plants (Loidl, 2004). Unlike DMTs, some plant HDACs, when mutated, cause pleiotropic effects on cellular development.

Just as in plants and animals, fungi have a myriad of mechanisms for post-translationally modifying amino acids protruding off of histone tails. Like plants, the two most well studied modifications in fungi are methylation and acetylation. Coincidentally, these are both reversible modifications. An example of this regulation in fungi can be seen with H3K9me₂ and H3K9me₃. These modifications are associated with the outer repeats of the heterochromatic centromeres. The presence of these modifications are needed to attract chromatin remodelers (i.e. Swi6 and Chp2) and to ensure that the chromatin in these regions remain silent. For this to occur, H3 must first be deacetylated by a variety of HDACs (i.e. Sir2, Clr3, and Clr6), which then allows for Clr4 (a HMT) to come and methylate H3K9 (Nakayama *et al.*, 2001). In *Neurospora crassa*, lysines 4, 27, 36, and 79 of histone H3 are all able to undergo methylation *in vivo* (Adhvaryu *et al.*, 2005).

In *Saccharomyces cerevisiae*, Sir2, a member of the sirtuins, is a NAD-dependent histone deacetylase known to deacetylate lysine 16 on histone H4, which is essential for heterochromatin formation. In fact, the initial Sir2 deacetylation serves to provide binding sites for other Sir proteins in the SIR complex, allowing for the spread of heterochromatin to the appropriate regions. While the exact mechanism for this process remains elusive, perhaps combinations of these modifications generate specific sites that can serve as docking stations for regulatory proteins. In fungi, HATs and HDACs can act on a variety of nuclear and cytoplasmic proteins besides histones. Scientists could be overlooking key insights into these regulatory mechanisms by not paying enough attention to the acetylation of these non-histone proteins (Brosch *et al.*, 2007).

Saccharomyces cerevisiae is also known to have H2A, H2B, and H4 sumoylation sites

that have been linked to transcriptional repression and seem to have an interacting role with other modifications (Nathan *et al.*, 2006). One must note that data gathered from species like *Saccharomyces cerevisiae* are not applicable across filamentous fungal species due to significant differences in their metabolic complexities.

As in plants and fungi, mammalian histone tails are subject to a series of covalent modifications. In animals, these covalent modifications have been shown to modulate gene expression throughout development (Goll and Bestor, 2002). There are other processes that histone modifications affect, such as DNA repair, X chromosome inactivation, chromosome inheritance dynamics, maintaining genome stability, and genomic imprinting. The impact of these events and the role they play in epigenetic regulation will be briefly discussed later. Animals possess homologs of all of the histone modifying enzymes discussed above with the exception of the HD2-like family of HDAC enzymes that are only found in plants. In fact, due to organismal complexity, animals have an even more diverse collection of HATs, HDACs, and HKMTs than found in plants and fungi. While loss of DMTs have been shown to lead to lethality in mice, loss of certain HKMTs, like Suv39h (which methylates histone tails as opposed to cytosines in the DNA sequence), showed no global effects on transcription. The same was also seen for mouse embryos lacking HDAC1 (Lagger *et al.*, 2002). In either case, no large-scale transcription defects were detected. This may indicate that DNA cytosine methylation is a more essential epigenetic regulatory pathway in mammals and higher eukaryotes.

Histone Variants

Histone variants usually only differ from a core (canonical) histone by a small number of amino acid changes. It is believed that the incorporation of histone variants can designate regions of the chromatin for specialized functions and can also reverse the effects of histone methyltransferases. Variants can be classified into two classes based upon how divergent their amino acid sequence is from the core histone: homomorphous and heteromorphous. Homomorphous variants have only a few amino acid changes (i.e. H2A.1, H2A.2, H3.1, H3.2, and H3.3), while heteromorphous variants involve changes that affect larger portions of the histone protein (i.e. H2A.X, H2A.Z, and centromeric protein A) (Ausio, 2006). As such, histone variants can create a unique nucleosomal architecture that can regulate a variety of nuclear processes, indicating that these variants do indeed possess distinct functions (Brown, 2001).

There are several known histone variants in plants including H2A, H2B, H2A.X, H2A.Z, H3.1, H3.2 (H3.3 in animals), and CENP-A. Unlike animal H2A and H2B variants, which co-migrate on SDS polyacrylamide gels, plant H2A and H2B have differential electrophoretic motilities indicating that they differ in molecular weight (Spiker, 1982). In addition to this molecular weight discrepancy, further experiments determined that plant and animal variants have different amino acid point substitutions (Patthy *et al.*, 1973). This could mean that these variants are playing slightly different roles in regulating gene expression. Plant specific variant H3.2 has been shown to have large amounts of acetylation. The presence of this H3.2 variant is inversely related to genome size, and associated with active genic regions (Waterborg, 1992). This makes sense since smaller genomes have more actively transcribed genes. CENP-A is a histone

variant that replaces H3 in centromeric chromatin, thereby causing condensed heterochromatin.

Histone variants also play a pivotal role in regulating the chromatin dynamics in fungal species. For example, H2A.Z has been shown to help limit the spread of silent heterochromatin by being placed at the outer regions of active euchromatin. Nearly all euchromatic genes are associated with H2A.Z deposition in the promoter regions (Raisner *et al.*, 2005). For efficient H2A.Z deposition to occur, H3 and H4 histone tails must be acetylated. This demonstrates that there is interdependency between different epigenetic pathways. CENP-A is also prevalent in fungal species associating with the heterochromatic centromeres of the chromosome. This is another example of histone variants coordinating the structure of chromatin. Not all histone variants appear to be essential in fungi. Certain types of H2B variants were shown to be dispensable during the yeast cell cycle (Spiker, 1982). This has not been seen in higher eukaryotes.

Histone variants have been studied extensively in animals. All eukaryotic organisms have a CENH3 histone variant, which associates with nucleosomes in the centromeric regions. Another animal variant is H3.3. This variant has only 4 amino acid changes, and appears to associate with the cell cycle (the canonical H3 protein associates with replication forks). Just as with yeast, animal H2A.Z is associated with active chromatin. It is important to note that many of these variants (i.e. H2A.Z) are less stable than the canonical histones (Henikoff *et al.*, 2004). Histone variants also play a role in mammalian X chromosome inactivation, providing yet another example of interrelated epigenetic pathways. Clearly, these variants play a crucial role in nucleosome assembly pathways and are likely responsible for the inheritance of specific chromatin states.

Altogether, these data show that the distribution of histone variants can have a significant effect on gene expression. Mechanistically there are a few different ways this can be accomplished. Histone variants may affect histone-histone interactions, resulting in changes in nucleosome stability and folding. Another hypothesis is that these variants specifically interact with certain proteins, which upon recruitment can alter chromatin structure. While the mechanisms remain unclear, histone variants clearly affect the makeup of nucleosomes and therefore can change chromatin structure, which directly influences gene regulation.

Nucleosome Positioning

Like histone modifications, nucleosome positioning plays a vital role in epigenetic regulation, since the distribution of nucleosomes directly facilitates the structure of chromatin. Nucleosome positioning can have inhibitory roles (i.e. polymerase occlusion), facilitative roles (i.e. histone protein domains), and roles in transcription factor binding. How these nucleosomes are distributed appears to have a large impact on genetic control. In order for histones to facilitate the wrapping of different DNA sequences into highly organized nucleosomes, they must be highly dependent on DNA sequence (Sekinger, 2005). This sequence preference could facilitate nucleosome distribution and therefore control binding site accessibility. Currently, new experimental and computational approaches are underway to construct and validate a DNA-nucleosome interaction model to predict the genome-wide organization of nucleosomes (Segal *et al.* 2006).

In *Arabidopsis thaliana*, efforts to establish genome-wide nucleosome positioning have been completed. By overlapping this data with genome-wide DNA methylation

data, one can detect higher levels of methylation on nucleosome-bound DNA compared to flanking DNA (Chodavaparu *et al.*, 2010). This means that DMTs prefer to target nucleosome-bound DNA, and that nucleosome positioning influences DNA methylation patterns in plants. This would suggest that nucleosome positioning is dependent on other epigenetic regulatory pathways like DNA methylation and histone variants.

In fungi, genome-wide nucleosome position was determined for *Saccharomyces cerevisiae* and *Aspergillus fumigatus*. In *S. cerevisiae*, over 70,000 nucleosome positions were mapped covering approximately 80% of the genome. Both nucleosome occupancy signatures and overall occupancy correlated with transcript abundance and transcription rate (Lee *et al.*, 2007). From these results, one can cluster gene families based upon their nucleosome occupancy patterns observed at their promoters. In the filamentous fungi, *Aspergillus fumigatus*, scientists have identified both mono- and di-nucleosome positions across the genome (Nishida *et al.*, 2009). Mono-nucleosomal DNA fragments were then compared among active and inactive genes. It was shown that DNA fragments in active gene promoters were not protected by nucleosomes, indicating that in order for a gene to be actively transcribed it needs to allow proteins and transcription factors to access its DNA sequence. If the DNA is tightly wrapped around a nucleosome (or multiple nucleosomes) it would make it less likely to have high gene expression.

As in plant and fungal genomes, animal genomes are packaged into nucleosomes that exclude DNA from interacting with DNA-binding proteins (i.e. chromatin remodeling proteins). However, it appears that certain DNA sequences have a higher tendency to be nucleosome-bound. These results suggest that DNA sequence could facilitate the positioning of the nucleosome core. Higher eukaryotes tend to contain

genes with large introns, and it is hypothesized that these introns guide nucleosomes to their appropriate position. Perhaps introns allow for flexibility when wrapping around nucleosomes. It is possible that higher eukaryotes facilitate nucleosome positioning through different mechanisms than plants and fungi based upon the structure of their genes.

RNA Interference and non-coding RNAs

RNA interference (RNAi) is a natural defense mechanism where small RNA molecules bind to sequence specific messenger RNA transcripts and induce degradation of the transcript. RNAi was first discovered in plants. This type of gene silencing is used to protect against invasive nucleic acids like retrotransposons and viruses, as well as maintain chromosomal stability (Akashi and Taira, 2007). Transcriptional gene silencing is achieved through RNA-directed DNA methylation (RdDm). Currently, researchers are focused on understanding the mechanistic link between RNA interference and DNA methylation pathways.

Many researchers believe that RNAi machinery plays an integral part in epigenetic regulation. Currently, the effects of RNAi on epigenetic mechanisms are being heavily studied in plants, focusing on RNA-directed-DNA-methylation (RdDM). RNAi mediated gene silencing is more prevalent in plants than in any other type of organism. RdDM was first observed as a viroid defense mechanism in tobacco plants (Wassengger *et al.*, 1994). While it is unclear how the cell determines *de novo* DNA methylation sites, many believe that RdDM is involved. RdDM begins with the formation of double-stranded RNA (dsRNA) molecules from genomic DNA. These dsRNA molecules have extensive secondary structure and act either through the

microRNA (miRNA) or small interfering RNA (siRNA) pathway to direct *de novo* DNA methylation (Aufsatz *et al.*, 2002). Since DNA methylation is associated with inactive gene expression in plants, this mechanism could prove to protect against transposable elements and viruses. A few years ago, researchers discovered genes that appear to be inducers of RdDM and transcriptional gene silencing (TGS): the phosphoribosyl anthranilate isomerase (PAI) gene family (4 members); and the *SUPERMAN* gene (Chan *et al.*, 2005). In plants, RdDM requires a plant-specific RNA polymerase (polIV) that has been shown to generate siRNAs. According to the literature, the RdDM machinery in plants is very different than that found in mammals. This indicates that plants and animals (mammals in particular) have different RdDM mechanisms.

Not all fungi have an RNAi system. For example, *S. cerevisiae* does not contain RNAi machinery while *S. pombe* (fission yeast) does. Since some fungi (*S. cerevisiae*) do not have DNA methylation in their genome, there is no need for an RNAi directed RdDM pathway. *Neurospora crassa* does have DNA methylation, rendering RdDM applicable. Despite clear evidence that post transcriptional gene silencing (PTGS) and RNAi does exist in *N. crassa*, there has yet to be evidence that they use RNAi machinery to guide epigenetic modifications (Matzke *et al.*, 2005). Instead, *N. crassa* use an unusual process called repeat-induced point mutation (RIP) to generate signals for *de novo* DNA methylation. In RIP, T:A and C:G transition mutations are created in duplicated DNA regions via an unknown mechanism. These sequences are targeted for DNA methylation, perhaps based on lack of GC content (Tamaru and Selker, 2003).

While an endogenous RNAi pathway does not exist in *S. cerevisiae*, a subset of ncRNAs called Cryptic Unstable Transcripts (CUTs) may be serving comparable

functions. These CUTs are 200-800 bp transcripts that have a 5' cap, a polyadenylated tail, and are rapidly degraded by the combined activity of poly-adenylating polymerases and exosome complexes (Davis and Ares, 2006; Thompson and Parker, 2007; Berretta and Morillion, 2009). CUTs are produced from inter and intra-genic regions. Despite their rapid degradation and lack of in-depth characterization, they have been implicated in numerous gene regulation and silencing pathways (Martens et al., 2004, Berretta et al., 2008). CUTs have been shown to effect histone eviction and chromatin remodeling processes through interactions with various histone side chain modifications (Camblong et al., 2007, Uhler et al., 2007), thus linking CUTs to epigenetic control. CUTs have since been found to also exist in higher eukaryotes including humans, but their roles have yet to be extensively dissected.

There are many RNAi mechanisms that operate in higher eukaryotes. They utilize a plethora of ncRNAs like microRNAs, siRNAs, piwi-interactingRNAs, snoRNAs, scaRNAs, CUTs, and many more. The hallmark for RdDM in plants is non-CG methylation, which has also been detected in mammalian embryonic stem cells and in human L1 retrotransposons. However, it is unknown whether this methylation is directed by RNA (Matkze, Birchler, 2005). One might wonder if mammals have homologs of the protein machinery needed for RdDM in plants. While some methyltransferases like DRM2 and MET1 do have mammalian homologs, other components like CHROMOMETHYLASE 2 (CMT2) and the SNF2-like protein, DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), do not. Some believe that the mammalian protein ATRX can substitute for DRD1 in the RdDM pathway, but it is still unknown if these proteins are essential for RdDM. Since RdDM takes place in the

nucleus and mRNA degradation typically takes place in the cytoplasm, one must ensure that mammalian cells, as in plant cells, can translocate these small RNAs from the cytoplasm to the nucleus. This was experimentally verified by comparing nuclear and cytoplasmic cellular fractions and seeing that mature miRNAs were found in both. Accordingly, it appears that mammals have the ability to operate RdDM, but how frequent this process is occurring remains elusive.

Other Epigenetic Control Pathways

There are other types of epigenetic control that occur only in certain subsets of species: dosage compensation, genomic imprinting, and heterochromatin formation. In organisms like *D. melanogaster*, *H. sapiens*, and *C. elegans*, dosage compensation represents a regulatory mechanism that equilibrates the expression of genes on the sex chromosome (X) so that equal amounts of certain genes that affect specific phenotypes are expressed in both males and females. This can be accomplished through partial repression or complete inactivation of one of the sex chromosomes. In *Drosophila*, dosage compensation operates by hyperactivating the single male X chromosome. In mammals, dosage compensation is achieved through maintaining only one active X chromosome in each cell (X-inactivation). It is important to note that in plants that lack dimorphic sex chromosomes, dosage compensation can still occur but is less understood. This event usually leads to outcomes like polyploidy or aneuploidy (Buzek, 1998). However, fungi do not appear to have any sort of dosage compensation activity.

Genomic imprinting is an epigenetic phenomenon where certain genes are not inherited in a classical Mendelian fashion. Instead these imprinted genes appear to be expressed in a parent of origin manner. Genomic imprinting has been detected in

mammals, insects, and flowering plants and utilizes DNA methylation and histone modification mechanisms to facilitate monoallelic gene expression without any need to change the DNA sequence (Wilkinson, 2007). These marks appear in germline cells and are inherited to all somatic cells. In genomic imprinting, cells inherit one copy of their genes from their father and one from their mother. Genomic imprinting refers to the procedure of epigenetically silencing one of these copies so that only one copy of the imprinted gene is expressed. These epigenetic marks are stable for the life of the organism and are reset during egg and sperm formation (Jaenisch, 1997). Imprinting utilizes other epigenetic mechanisms like DNA methylation and histone modifications to position the epigenetic marks required for gene inactivation and has been shown to be essential for normal development.

The formation of heterochromatin is an epigenetic process that functions to protect chromosome integrity and facilitate gene regulation (Grewal, Jia, 2007). Heterochromatin distribution is epigenetically inherited through cell divisions so that the daughter cells will have an almost identical heterochromatic landscape across the chromosomes. Recent studies have suggested that RNAi machinery plays a role in heterochromatin formation, with RNAi mediated chromatin modifications being shown to effect epigenetic transcriptional gene silencing (Wassenegger, 2005).

As previously discussed, epigenetic mechanisms facilitate the formation of chromatin structure thereby impacting gene regulation. One way that epigenetic inactivation of gene expression can occur is through the formation of heterochromatin. While mechanistically this is still unclear, there are examples of this process occurring in plants, fungi, insects, and mammals. In *S. cerevisiae*, heterochromatin formation is

triggered by the expression of the Sir3 silencing protein leading to the rapid loss of histone acetylation, whereas histone methylations are removed gradually over several generations (Katan-Khaykovich, Strhul, 2005). Thus, the transition between euchromatin and heterochromatin does not occur immediately, but rather takes place over multiple cell divisions. This suggests that certain types of histone modifications can inhibit the formation of heterochromatin. Recent experiments have indicated that RNAi components are involved with heterochromatic formation in *D. melanogaster* (Wassenegger, 2005). Once again, this is another example of the interdependency between these epigenetic pathways.

Chromosome inheritance theory states that inheritance patterns may be generally explained by assuming that Mendelian genes are located in specific sites on chromosomes, and it is the chromosomes that undergo segregation and independent assortment. Recently, it has been discovered that epigenetic mechanisms regulate many crucial functions necessary for genome stability and chromosome inheritance. Some of these functions include but are not limited to: DNA repair and recombination, the initiation of DNA replication, chromosome end protection (telomeres), chromosome movement (centromere), and the segregation of homologous chromosomes during meiosis (Allis *et al.*, 2007). Abnormal chromosome inheritance can lead to birth defects and diseases like cancer. By understanding an organism's epigenetic control mechanisms, we can hope to eventually explain the subtle nuances of chromosome architecture and inheritance. Altogether, we can clearly see that these various pathways have created an interwoven network of genetic control, giving rise to this theme of interdependency among epigenetic control mechanisms.

Defining a Gene's Epitype

The aforementioned sections show that there are a variety of different mechanisms that facilitate epigenetic control. Nucleosome positioning, histone sequence variants, DNA bases from the standard GATC (e.g., cytosine methylation), and histone side chain modifications all represent mechanisms that affect gene expression by changing chromatin structure. These processes facilitate what are known as *cis*-linked changes to chromatin structure because they are occurring in the chromosomal vicinity of genic regions. Individual changes or the sum of all of these *cis*-linked chromatin structures that distinguish a gene sequence from naked DNA may be defined as an “epitype” (Meagher, 2010) (**Figure 1.1**). In order to fully understand epigenetic control, we must first determine how these mechanisms are interacting with one another to constitute a gene's epitype.

The Actin Cytoskeleton

The actin cytoskeleton is vital to a multitude of different cellular processes. It plays a crucial role in stress response, transcription, cytokinesis, cell locomotion, intracellular trafficking, maintenance of cell shape and polarity, and development (Williamson, 1993, Meagher and Williamson, 1994, Jockusch et al., 2007, Perrin and Ervasti, 2010, Miralles and Visa, 2006). The actin cytoskeleton controls spatial development in eukaryotic cells and organs through the construction of polymeric filaments and filament bundles (Kudryashov and Reisler, 2012). Actins are more than 80% conserved in most organisms across kingdoms, the exception being primarily among

the more divergent protists (Kandasamy et al., 2012). The actin cytoskeleton consists of unpolymerized, free globular actin monomers (G-actin) that join together to create polymerized actin filaments (F-actin). The process of constructing F-actin filaments requires energy derived from ATP hydrolysis. To extend F-actin filaments, G-actin monomers bind ATP, thus allowing them to remain in an open-fold state to stably associate with the “barbed ends” (+ end of actin polarity) of filaments. Hydrolysis of ATP to ADP drives a conformational change in the actin protein allowing for the simultaneous disassembly of F-actin-ADP monomers at the “pointed end” (- end of actin polarity) of filaments (Pollard et al., 2000). G-actin monomers cannot be added to the pointed ends of actin filaments. The rapid polymerization and depolymerization of actin filaments is necessary for preserving cell shape, normal cell motility, and development (Yarmola and Bubb, 2009).

In order to control the rapid reorganization of the actin cytoskeleton, actins interact with a suite of different proteins called Actin Binding Proteins (ABPs). In fact, actin is thought to participate in more protein-protein interactions than any known protein (Dominguez and Holmes, 2011). Through these interactions, ABPs facilitate rapid remodeling of the cytoskeleton by regulating the unpolymerized (G-actin monomers) and polymerized (F-actin filaments) actin equilibrium (Yarar et al., 2007). Some examples of different ABPs include: Actin Depolymerizing Factor/Cofilin (ADF/CFL), fimbrin, vilin, CAP1, Arp2/3, tropomyosin, and profilin. **Figure 1.2** depicts some of these roles facilitated by ABPs (Pollard et al., 2000). While they all play some role in facilitating the restructuring of the actin cytoskeleton, we will be focusing on a specific ABP gene family, profilin.

Profilins

Profilins are small (12-15 kDa), ubiquitously expressed, monomeric ABPs that have been identified in numerous organisms ranging from amoebae to higher plants and mammals (Ramachandran et al., 2001). Profilin is folded into a central beta-sheet, flanked on both the N- and C-terminus by alpha-helices. The N-terminal portion on the C-terminal alpha-helix along with the C-terminal portion of the central beta-sheet is where actin binding occurs (Schutt et al., 1993). Initially, profilin was found to distinctively bind globular actin (G-actin) and to facilitate G-actin sequestering in cells (Carlsson et al., 1976). Since then, more recent research has shown that in addition to binding and sequestering G-actin, profilin also plays a specific role in the formation of filamentous actin (F-actin).

Like actin, profilin is known to bind many different proteins. In addition to actin, profilin binds phosphatidylinositol 4,5-bisphosphate (PIP₂) (Sohn et al., 1995), poly-L-proline (Bjorkegren et al., 1993), many proline-rich proteins such as vasodilator-stimulated phosphoprotein (VASP) (Haffner et al., 1995), formin homology domain-containing proteins (Frazier and Field, 1997), the Arp2/3 complex (Mullins et al., 1998), and the annexins (Alvarez-Martinez et al., 1996). It is now thought that PIP₂ binds profilin competitively with actin, with its binding site on the other (N-terminal) alpha-helix (Sohn et al., 1995). Unlike with PIP₂, profilin can bind poly-L-proline (proline-rich sequences) and actin simultaneously, since poly-L-proline interacts with the hydrophobic surface of the protein which is positioned opposite the actin binding site (Bjorkegren et al., 1993). In fact, this occurs with higher affinity than binding either one separately (Ferron et al., 2007). Since proline-rich sequences are abundant among cytoskeletal

proteins, perhaps they serve to help guide profilin-actin complexes from the cellular pool onto actin filaments (Dominguez and Holmes, 2011).

Additionally, there is evidence that profilins are regulating actin networks by connecting signaling to the actin cytoskeleton through PIP2, Cdc42, and Actin Related Protein (ARP) complexes (Stradal et al., 2004 and Witke et al., 1998). Through these interactions, profilins have been shown to participate in cell elongation, cell shape maintenance, and other actin dynamics (Zheng et al., 2012). However, their mechanistic links, their role in signaling, as well as their effects on tissue and organ development remains relatively uncharacterized.

Profilins are approximately 130 amino acids long. Sequence alignments show very little sequence homology in the N- and C-terminal regions when comparisons are made across various eukaryotic kingdoms. Only 25 to 40 residues (~20-30%) are universally conserved across all kingdoms, which is much lower than the 80% seen among actins (Meagher, 1995, Kandasamy et al., 2012). Profilins tend to exist in small gene families, especially in higher eukaryotes. Members of the profilin gene family have distinct tissue and organ-specific expression patterns throughout development (Honore et al., 1993, Kandasamy et al., 2002). In humans (Hs), HsProfilin1 is ubiquitously expressed at high levels in all organs and tissues except muscle. HsProfilin2 is expressed high in brain and muscle, while HsProfilin3 and HsProfilin4 are expressed almost exclusively in sperm (Honore et al., 1993). In higher plants, there are two evolutionarily distinct classes of profilins: vegetative (constitutive) and reproductive (pollen-specific). Through phylogenetic analysis, it has been determined that the two distinct classes of profilin have independently evolved in both plants and animals.

Profilins Role in Actin Polymerization and Depolymerization

Profilin's role in actin cytoskeletal dynamics has been widely debated over the past 20 years. As previously stated, profilin was originally found to act in the sequestration of G-actin monomers, but now it has also been shown to be involved in filamentous actin (F-actin) formation. Profilins are thought to inhibit the spontaneous polymerization of actin filaments by forming a 1:1 complex with actin, thereby lowering the steady-state concentration of ATP-G-actin, and once all barbed ends become blocked by capping proteins, profilin begins to sequester G-actin from pointed-end polymerization (Dominguez, 2009). This would suggest that profilin acts in the depolymerization of actin filaments.

However, extensive research has shown that they play a complex role in F-actin formation through the replenishment of the ATP-actin monomer pool via catalyzing the exchange of ADP for ATP on actin (Dominguez and Holmes, 2011). While profilin does not bind F-actin directly, profilin-ATP-G-actin complexes are essential for rapid filament assembly (Schulter et al., 1997). Although profilin-bound actin monomers cannot add to pointed ends of actin filaments, they have been shown to elongate filament barbed ends at approximately the same rate as free actin monomers (Pollard and Borisy, 2003, dos Remedios et al., 2003). This would lead us to believe that profilin might be facilitating rather than inhibiting polymerization (Yarmola and Bubb, 2006). This idea is further strengthened by results indicating that profilin could lower the critical concentrations of actin needed to drive polymerization (Pantaloni and Carrier, 1993). Altogether, it appears as if profilin may be playing a role in both the polymerization and depolymerization of actin filaments.

Profilin in the Nucleus

In some plant and animal cells, profilin derivatives have been shown to be enriched in the nucleus (Mayboroda et al., 1997; Schluter et al., 1998). *Arabidopsis* PRF1 (and PRF2, unpublished) was detected in the nucleus using indirect immunofluorescence (Wittenmayer et al., 2000; Kandasamy et al., 2002, 2010). Interestingly, there is also evidence that the reproductive class profilin, PRF4, is not in either the vegetative or sperm cell nuclei of pollen (Kandasamy et al., 2002). These data suggest that some profilins are performing nuclear functions while others may not.

In isolated nuclei from human HeLa cells, profilin has been shown to be essential for actin export (along with Exportin-6, which recognizes actin/profilin complexes), with defects in profilin resulting in increased actin accumulation in the nucleus (Stuven et al., 2003). In *Drosophila*, profilin mutants result in impaired nuclear export causing defects in eye development, suggesting that profilin and the organization of the actin cytoskeleton play an important role in nuclear trafficking (Minakhina et al., 2005). Some hypothesize that profilins are involved with the intracellular trafficking of vesicles from the plasma membrane to the nucleus while others believe that profilins and actins are interacting within the nucleus in an effort to regulate gene expression through interactions with chromatin remodeling and transcriptional complexes.

***Arabidopsis thaliana* Profilins**

In *Arabidopsis thaliana*, there are five profilin genes (PRF1-PRF5). PRF1, PRF2, and PRF3 are constitutively expressed throughout all vegetative tissues and in ovules, while PRF4 and PRF5 are classified as reproductive profilins, being predominately expressed in mature pollen. Phylogenetic analysis of this gene family has determined

that PRF1 and PRF2, as well as PRF4 and PRF5 exist as duplicated gene pairs, most likely stemming from the recent genome-wide duplication event occurring in *Arabidopsis* approximately 30-35 million years ago (MYA) (Meagher et al., 1999, Vision et al., 2000, Simillion et al., 2002, Bowers et al., 2003, Blanc et al., 2003). While the expression profiles of PRF4 and PRF5 are very similar (**Figure 1.3**), the vegetative profilins exhibit a wide range in their expression levels. PRF2 is the most highly expressed, followed by PRF1 (~40% of PRF2 levels), and PRF3 (~12% of PRF2 levels). Despite having varying expression levels, PRF1, PRF2, and PRF3 have similar spatial expression patterns (**Figure 1.3**). The *Arabidopsis* vegetative and reproductive profilin proteins share 90% sequence identity within their class while the classes themselves share ~70-75% sequence homology (Huang et al., 1996). In fact, if we look at the sequence conservation among all profilins compared among monocots and dicots, we see that vegetative profilins are more similar to each other than they are to their reproductive counterparts in the same species.

All five *Arabidopsis* profilin genes have conserved intron-exon architecture with two introns at identical positions in each gene (Christensen et al., 1996). The first introns of vegetative PRF1, PRF2, and PRF3 are 250 bp longer than reproductive PRF4 and PRF5, which prompted an investigation into the functional and regulatory roles of these regions. Through transient expression studies using *GUS*-transgenes, it was shown that the first intron of the profilin variants has remarkably different functions and plays a regulatory role in profilin expression (Jeong et al., 2006). By swapping the first introns between PRF2 and PRF5, it was found that the PRF2 intron was required for proper ubiquitous PRF2 expression, while PRF5's pollen specific expression was not affected

with PRF2's intron. These data suggest that the larger vegetative first intron is required for strong, constitutive expression of the vegetative profilins, while in reproductive profilins intron1 appears to have no effect on gene expression. The presence of these regulatory elements in vegetative profilins infers the possibility of additional protein function.

Profilins and Cancer

Profilins have been shown to be involved in various signaling pathways to regulate actin cytoskeleton reorganization, and to interact with many cytoplasmic and nuclear ligands. Recent studies have implicated that profilins play a major role in cancer cells as well. Initially it was found that human cancer cells, expressing low profilin1 levels, adopt a non-tumorigenic phenotype upon raising their profilin1 level (Wittenmayer et al., 2004). These results suggest that profilin1 is functioning as a tumor suppressor. This was further validated when it was discovered that expression of profilin1 was down regulated in breast cancer cells, indicating that loss of profilin1 has a significant effect on breast cancer progression (Zou et al., 2007 and Zou et al., 2010). Profilin2 has also been dissected for its effects on cancer progression. An interaction with the ENA/VASP protein, EVL, aids profilin2 in reducing membrane protrusions and cell migration through an actomyosin contractility mechanism (Mouneimne et al., 2012). Reduction in profilin2 expression appears to correlate with invasive disease and poor prognosis in patients with breast cancer (Mouneimne et al., 2012). These data indicate that profilins could serve as a potential biomarker for cancer diagnosis (Zoidakis et al., 2012). Altogether, there is accumulating evidence that the improper expression and

regulation of these profilin proteins are directly linked to the progression of various cancers.

REFERENCES

- Adhvaryu K, Morris S, Strahl B, Selker E. 2005. Methylation of histone H3 lysine 36 is required for normal development in *Neurospora crassa*. *Euk Cell* **4**:1455–1464.
- Akashi H, Taira K. 2007. RNAi and epigenetics: Pol IV is a matchmaker of small RNAs meeting with chromatin. *Heredity* **98**:125–127.
- Allis CD, Jenuwein T, Reinberg D, and Caparros ML. 2007. Epigenetics. *Cold Spring Harbor Laboratory Press* Plainview, NY.
- Alvarez-Martinez MT, Mani JC, Porte F, Faivre-Sarrailh C, Liautard JP, Sri Widada J. 1996. Characterization of the interaction between annexin I and profilin. *Eur J Biochem* **238**:777–784.
- Antequera F, Tamame M, Villanueva JR, Santos T. 1984. DNA methylation in the fungi. *J Biol Chem* **259**:8033–8036.
- Aufsatz W, Mette MF, van der Winden J, Matzke AJM, and Matzke M. 2002. RNA-directed DNA methylation in Arabidopsis. *PNAS* **99**:16499–16506.
- Ausio J. 2006. Histone variants-the structure behind the function. *Functional Genomics* **5**:228-243.
- Berretta J, Morillon A. 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Reports* **12**:973–982.
- Berretta J, Pinskaya M, Morillon A. 2008. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev* **22**:615–626.
- Bjorkegren C, Rozycki M, Schutt CE, Lindberg U, Karlsson R. 1993. Mutagenesis of human profilin locates its poly (L-proline)-binding site to a hydrophobic patch of aromatic amino acids. *FEBS Letts* **333**:123–126.
- Blanc G, Hokamp K, Wolfe KH. 2003. A Recent Polyploidy Superimposed on Older Large-Scale Duplications in the *Arabidopsis* Genome. *Genome Res* **13**:137–144.
- Brosch G, Loidl P, Graessle S. 2007. Histone modifications and chromatin dynamics: a focus on filamentous fungi. *FEMS Microbiol Rev* **32**:409-439.
- Brown D. 2001. Histone variants: are they functionally heterogeneous? *Genome Biology* **2**: reviews0006.
- Buzek J. 1998. Gene dosage compensation in plant cells. *Biologické listy* **63**:39-60.

- Camblong J et al. 2007. Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in *S. cerevisiae*. *Cell* **131**:706–717.
- Carlsson L, Nystrom L, Sundkvist I, Markey F, Lindberg U. 1976. Profilin, a low-molecular weight protein controlling actin polymerisability. In *Contractile Systems in Non Muscle Tissues*. Elsevier Amsterdam, 39-49.
- Chan SW, Henderson IR, and Jacobsen SE. 2005. Gardenening the genome: DNA methylation in *Arabidopsis thaliana*. *Na. Rev Genet* **6**:351-360.
- Chodavarapu R, Feng S, Bernatavichute Y, et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature Lett* **466**:388-392.
- Christensen HEM, Ramachandran S, Tan CT, Surana U, Dong CH, Chua NH. 1996. Arabidopsis profilins are functionally similar to yeast profilins: identification of a vascular bundle-specific profilin and a pollen-specific profilin. *Plant J* **10**:269–279.
- Davis CA, Ares M. 2006. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *PNAS* **103**:3262–3267.
- Dominguez R. 2009. Actin filament nucleation and elongation factors – structure function Relationships. *Crit. Rev. Biochem. Mol Bio* **44**:351-366.
- Dominguez R, Holmes KC. 2011. Actin Structure and Function. *Annu Rev Biophys* **40**:169-186.
- dos Remedios CG, Chhabra D, Kekic M, Dedova IV, Tsubakihara M, et al. 2003. Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol Rev* **83**:433–73.
- Ferron F, Rebowski G, Lee SH, Dominguez R. 2007. Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP. *EMBO J* **26**:4597–4606.
- Goll MG, Bestor TH. 2002. Histone modification and replacement in chromatin activation. *Genes Dev* **16**:1739-1742.
- Grewal S, Jia S. 2007. Heterochromatin revisited. *Nature Reviews Genetics* **8**: 35-46.
- Groll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry* **74**:481-514.

- Haffner C, Jarchau T, Reinhard M, Hoppe J, Lohmann SM, Walter U. 1995. Molecular cloning, structural analysis and functional expression of the proline-rich focal adhesion and microfilament-associated protein VASP. *EMBO J* **14**:19–27.
- Henifoff S, Furuyama T, and Ahmad K. 2004. Histone variants, nucleosome assembly and epigenetic inheritance. *Trends in Genetics* **20**:7.
- Honore B, Madsen P, Andersen AH, Leffers H. 1993. Cloning and expression of a novel human profilin variant, profilin II. *FEBS Lett* **330**:151–155.
- Huang S, McDowell JM, Weise MJ, Meagher RB. 1996. The Arabidopsis Profilin Gene Family: Evidence for an Ancient Split between Constitutive and Pollen-Specific Profilin Genes. *Plant Physiol* **111**:115-126.
- Hunter P. 2008. What genes remember. *Prospect Magazine* **146**.
- Jaenisch, R. 1997. DNA methylation and imprinting: why bother? *Trends in Genetics* **13**:323-329.
- Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* **33**:245–254.
- Jenuwein T, Allis C. 2001. Translating the histone code. *Science* **293**:1074–80.
- Jeong YM, Mun JH, Lee I, Woo JC, Hong CB, Kim SG. 2006. Distinct Roles of the First Introns on the Expression of *Arabidopsis* Profilin Gene Family Members. *Plant Physiology* **140**:196-209.
- Jockusch BM, Murk K, Rothkegel M. 2007. The profile of profilins. *Rev Physiol Biochem Pharmacol* **159**:131-149.
- Kandasamy MK, McKinney EC, Meagher RB. 2002. Plant Profilin Isovariants Are Distinctly Regulated in Vegetative and Reproductive Tissues. *Cell Motil Cytoskeleton* **52**:22-32.
- Kandasamy MK, McKinney EC, Meagher RB. 2010. Differential sublocalization of Actin variants within the nucleus. *Cytoskeleton* **67**:729-743.
- Kandasamy MK, McKinney EC, Roy EM, Meagher RB. 2012. Plant vegetative and animal cytoplasmic actins share functional competence for spatial development with protists. *Plant Cell* **24**:2041-2057.
- Kapoor A, Agius F, Zhu JK. 2005. Preventing transcriptional gene silencing by active DNA methylation. *FEBS Lett* **579**:5889-5898.

- Katan-Khaykovich Y, Struhl K. 2005. Heterochromatin formation involves changes in histone modifications over multiple cell generations. *EMBO* **24**:2138-2149.
- Kress C, Thomassin H, and Grange T. 2001. Local DNA demethylation in vertebrates: How could it be performed and targeted? *FEBS Lett* **494**:135-140.
- Kudryashov DS and Reisler E. 2012. ATP and ADP Actin States. *Biopolymers* **99**:245-256.
- Lagger G, O'Carroll D, Rembold M, Khier H, Tischler J, Weitzer G, Schuettengruber B, Hauser C, Brunmeir R, Jenuwein T, *et al.* 2002. Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J* **21**:2672–2681.
- Lee W, Tillo D, Bray N, Morse R, *et al.* 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics* **39**:1235–1244.
- Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69** (6):915–926.
- Loidl P. 2004. A plant dialect of the histone language. *Trends Plant Sci* **9**:84–90.
- Martens J, Laprade L, Winston F. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**:571–574.
- Matzke M, Birchler J. 2005. RNA directed pathways in the nucleus. *Nat Rev Gen* **6**:24-35.
- Matzke M, Kanno T, Huettel B, *et al.* 2005. RNA-directed DNA methylation. Meyer book: *Plant Epigenetics* (Chapter 3).
- Mayboroda O, Schlueter K, Jockusch BM. 1997. Differential colocalization of profilin with microfilaments in PtK2 cells. *Cell Motil Cytoskeleton* **37**:166-177.
- Meagher RB. 1995. The impact of historical contingency on gene phylogeny: Plant actin diversity. *Evolutionary Biology* **28**:195-215.
- Meagher RB. 2010. The evolution of epitype. *The Plant Cell* **22**:1658-1666.
- Meagher RB, McKinney EC, Vitale AV. 1999. The evolution of new structures: Clues from plant cytoskeletal genes. *Trends Genet* **15**:278–284.
- Meagher RB, Williamson RE. 1994. The Plant Cytoskeleton. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY, 1049–1084.

- Michael N, David R, Lehninger A. 2005. *Lehninger Principles of Biochemistry*. San Francisco: W.H. Freeman. Press.
- Minakhina S, Myers R, Druzhinina M, and Steward R. 2005. Crosstalk between the actin cytoskeleton and Ran-mediated nuclear transport. *BMC Cell Biology* **6**:32-42.
- Miralles F, Visa N. 2006. Actin in transcription and transcription regulation. *Curr Opin Cell Biol* **18**:261–66.
- Mouneimne, G. et al. 2012. Differential remodeling of actin cytoskeleton architecture by profilin isoforms leads to distinct effects on cell migration and invasion. *Cancer Cell* **22**:615–630.
- Mullins RD, Kelleher JF, Xu J, Pollard TD. 1998. Arp2/3 complex from *Acanthamoeba* binds profilin and crosslinks actin filaments. *Mol Biol Cell* **9**:841–852.
- Nakayama J, Rice JC, Strahl BD, Allis CD, and Grewal SI. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**:110-113.
- Nanney DL. 1958. Epigenetic control systems. *PNAS* **44**:712-717.
- Nathan D, Ingvarsdottir K, Sterner DE, Bylebyl GR, Dokmanovic M, Dorsey JA, Whelan KA, Krsmanovic M, Lane WS, Meluh PB, Johnson ES, Berger SL. 2006. Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Genes Dev* **20**:966–976.
- Nishida H, Motoyama T, Yamamoto S, et al. 2009. Genome-wide maps of mono- and dinucleosomes of *Aspergillus fumigatus*. *Bioinformatics* **25**:2295-2297.
- Pandey R, Muller A, Napoli CA, Selinger DA, et al. 2002. Analysis of histone acetyltransferase and histone deacetylase families in *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucl Acids Res* **30**:5036-5055.
- Pantaloni D, Carlier MF. 1993. How profilin promotes actin filament assembly in the presence of thymosin b4. *Cell* **75**:1007–1014.
- Patthy L, Smith, E, and Johnson J. 1973. Histone 3.V. The amino acid sequence of pea embryo histone 3. *J Biol Chem* **248**:6834-6840.
- Perrin BJ, Ervasti JM. 2010. The Actin Gene Family: Function Follows Isoform. *Cytoskeleton* **67**:630–634.

- Pollard TD, Borisy GG. 2003. Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112**:453–65.
- Pollard TD, Blanchoin L, Mullins RD. 2000. Molecular mechanism controlling actin filament dynamics in nonmuscle cells. *Annu Rev Biophys Biomol Struct* **29**:545–576.
- Raisner R, Hartley P, Meneghini M, *et al.* 2005. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**:233–48.
- Ramachandran S, Christensen H, Ishimaru Y, Dong CH, Chao-Ming W, Cleary AL, Chua NH. 2001. Profilin Plays a Role in Cell Elongation, Cell Shape Maintenance, and Flowering in Arabidopsis. *Plant Physiology* **124**:1637–1647.
- Reik, W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**:425–432.
- Russo V, Martienssen R, Riggs D. 1996. Epigenetic mechanisms of gene regulation. *Cold Spring Harbor Laboratory Press* Plainview, NY.
- Schlüter K, Jockusch BM, Rothkegel M. 1997. Profilins as regulators of actin dynamics. *Biochim Biophys Acta* **1359**:97–109.
- Schlüter K, Schleicher M, Jockusch BM. 1998. Effects of single amino acid substitutions in the actin-binding site on the biological activity of bovine profilin I. *J. Cell Sci* **111**:3261–3273.
- Schutt CE, Myslik JC, Rozycki MD, Goonesekere NC, Lindberg U. 1993. The structure of crystalline profilin-beta-actin. *Nature* **365**:810–816.
- Segal E, Fondufe-Mittendorf Y, Chen L, *et al.* 2006. A genomic code for nucleosome positioning. *Nature* **442**:772–778.
- Sekinger E, Moqtaderi Z, Struhl K. 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol Cell* **18**:735–748.
- Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of Arabidopsis thaliana. *PNAS* **99**:13627–13632.
- Slotkin, R, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**:272–85.
- Sohn RH, Chen J, Koblan KS, Bray PF, Goldschmidt-Clermont PJ. 1995. Localization of a binding site for phosphatidylinositol 4, 5-bisphosphate on human profilin. *J Biol Chem* **270**:21114–21120.

- Spiker S. 1982. Histone variants in plants. *Jour Biol Chem* **257**:14250-14255.
- Stradal TE, Rottner K, Disanza A, Confalonieri S, Innocenti M, Scita G. 2004. Regulation of actin dynamics by WASP and WAVE family proteins. *Trends Cell Biol* **14**:303–311.
- Strahl B, Allis C. 2000. The language of covalent histone modifications. *Nature* **403**:41–45.
- Stüven T, Hartmann E, Görlich D. 2003. Exportin 6: a novel nuclear export receptor that is specific for profilin.actin complexes. *EMBO J* **22**:5928–5940.
- Tamaru H, Selker EU. 2003. Synthesis of signals for de novo DNA methylation in *Neurospora crassa*. *Molecular and Cellular Biology* **23**:2379–2394.
- Thompson D, Parker R. 2007. Cytoplasmic Decay of Intergenic Transcripts in *Saccharomyces cerevisiae*. *Molecular and Cell Biology* **27**:92–101.
- Tucker KL. 2001. Methylated cytosine and the brain: a new base for neuroscience. *Neuron* **30**:649–652.
- Uhler J, Hertel C, Svejstrup J. 2007. A role for noncoding transcription in activation of the yeast PHO5 gene. *PNAS* **104**:8011–8016.
- Vanyushin BF. 2006. DNA methylation in plants. *Curr Top Microbiol Immunol* **301**:67-122.
- Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* **290**:2114–2117.
- Waddington CH. 1942. The epigenotype. *Endeavour* **1**:18–20.
- Wassenegger M, Heimes S, Riedel L, and Sanger HL. 1994. RNA directed de novo methylation of genomic sequences in plants. *Cell* **76**:567-576.
- Wassenegger M. 2005. The Role of the RNAi Machinery in Heterochromatin Formation. *Cell* **122**:13-16.
- Waterborg JH. 1992. Existence of two histone H3 variants in dicotyledonous plants and correlation between their acetylation and plant genome size. *Plant Mol Biol* **18**:181-187.
- Wilkinson L, Davies W, and Isles AR. 2007. Genomic imprinting effects on brain development and function. *Nature Reviews Neuroscience* **8**:832–843.

- Williamson RE. 1993. Organelle movements. *Annu Rev Plant Physiol Plant Mol Biol* **44**:181-202.
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An electronic fluorescent pictograph browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* **2**:e718.
- Witke W, Podtelejnikov AV, Di Nardo A, Sutherland JD, Gurniak CB, et al. 1998. In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly. *EMBO J* **17**:967–976.
- Wittenmayer N, Rothkegel M, Jockusch BM, Schluter K. 2000. Functional characterization of green fluorescent protein-profilin fusion proteins. *Eur J Biochem* **267**:5247–5256.
- Wittenmayer N, et al. 2004. Tumor Suppressor Activity of Profilin Requires a Functional Actin Binding Site. *Mol Biol Cell* **15**:1600-1608.
- Xu Z et al. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**:1033–1037.
- Yarar D, Waterman-Storer CM, Schmid SL. 2007. SNX9 couples actin assembly to phosphoinositide signals and is required for membrane remodeling during endocytosis. *Dev Cell* **13**:43–56.
- Yarmola EG, Bubb MR. 2006. Profilin: emerging concepts and lingering misconceptions. *Trends in Biochem Sciences* **31**:197-205.
- Yarmola EG, Bubb MR. 2009. How depolymerization can promote polymerization: the case of actin and profilin. *BioEssays* **31**:1150–1160.
- Zheng Y, Xin H, Lin J, Liu CM, Huang S. 2012. An Arabidopsis class II formin, AtFH19, nucleates actin assembly, binds to the barbed end of actin filaments, and antagonizes the effect of AtFH1 on actin dynamics. *J Integr Plant Biol* **54**:800-13.
- Zilberman D, Henikoff S. 2007. Genome-wide analysis of DNA methylation patterns. *Development* **134**:3959-3965.
- Zoidakis et al. 2012. Profilin 1 is a potential biomarker for bladder cancer aggressiveness. *Mol Cell Proteomics* **11**: M111.009449.
- Zou L, Ding Z, Roy P. 2010. Profilin-1 overexpression inhibits proliferation of MDA-MB-231 breast cancer cells partly through p27kip1 upregulation. *J Cell Physiology* **223**:623-629.

Zou L, Jaramillo M, Whaley D, Wells A, Panchapaksea V, Das T, Roy P. 2007. Profilin-1 is a negative regulator of mammary carcinoma aggressiveness *British Journal of Cancer* **97**:1361–1371.

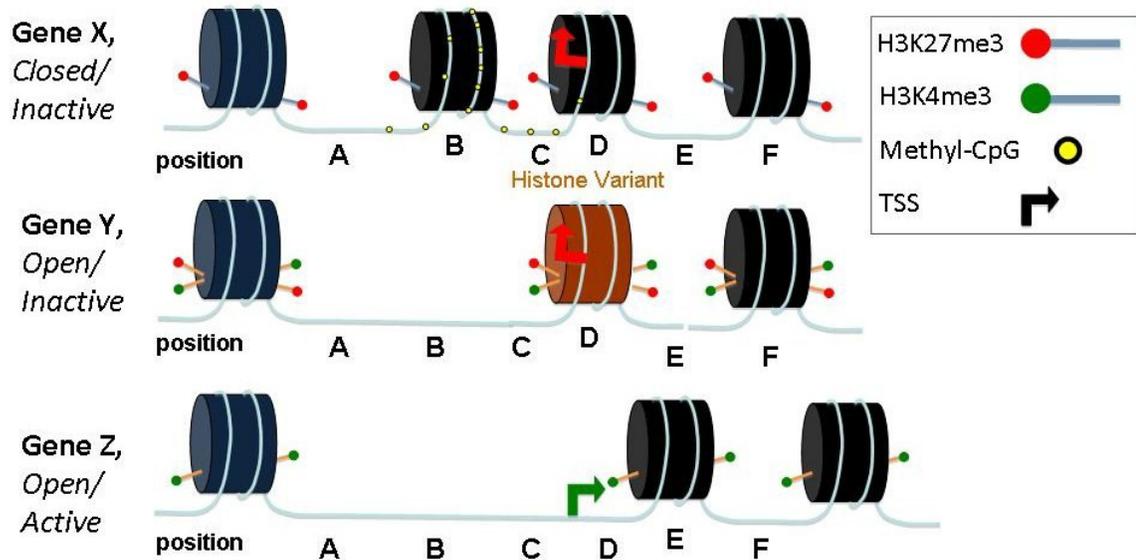


Figure 1.1: Defining an epitype. The epitype of a gene or genome is defined as the sum of all cis-linked chromatin structures that distinguish it from naked DNA, including but not limited to nucleosome phasing, base methylation (e.g., 5-methylcytosine, 5MeC), various histone side chain modifications, and deposition of histone variants (indicated by brown color change) within nucleosomes. Genes X/Y/Z all represent different epigenetic states, and hence have different epitypes. TSS: transcriptional start site.

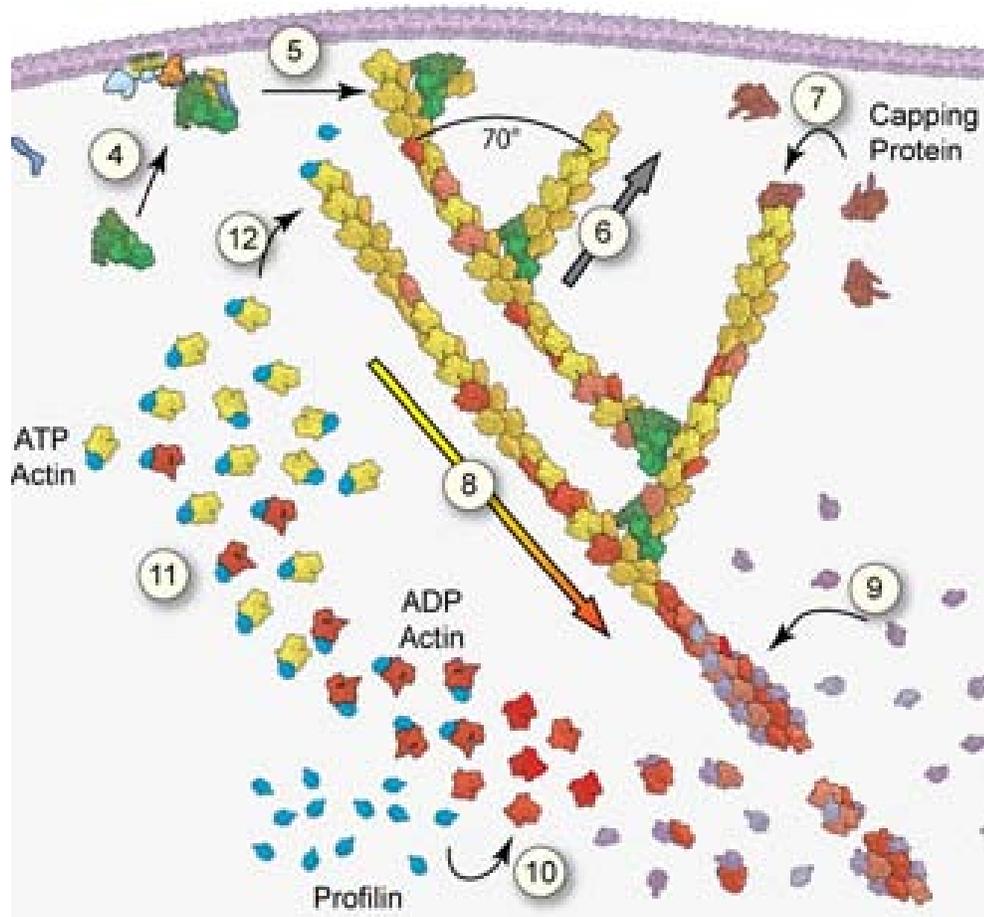


Figure 1.2: Remodeling of the actin cytoskeleton by Actin Binding Proteins (ABPs) (Adapted from Pollard et al., 2000). The rapid remodeling of the actin cytoskeleton is facilitated through protein-protein interactions with a suite of ABPs. The roles of capping proteins, actin depolymerizing factors (ADFs), and profilins are depicted.

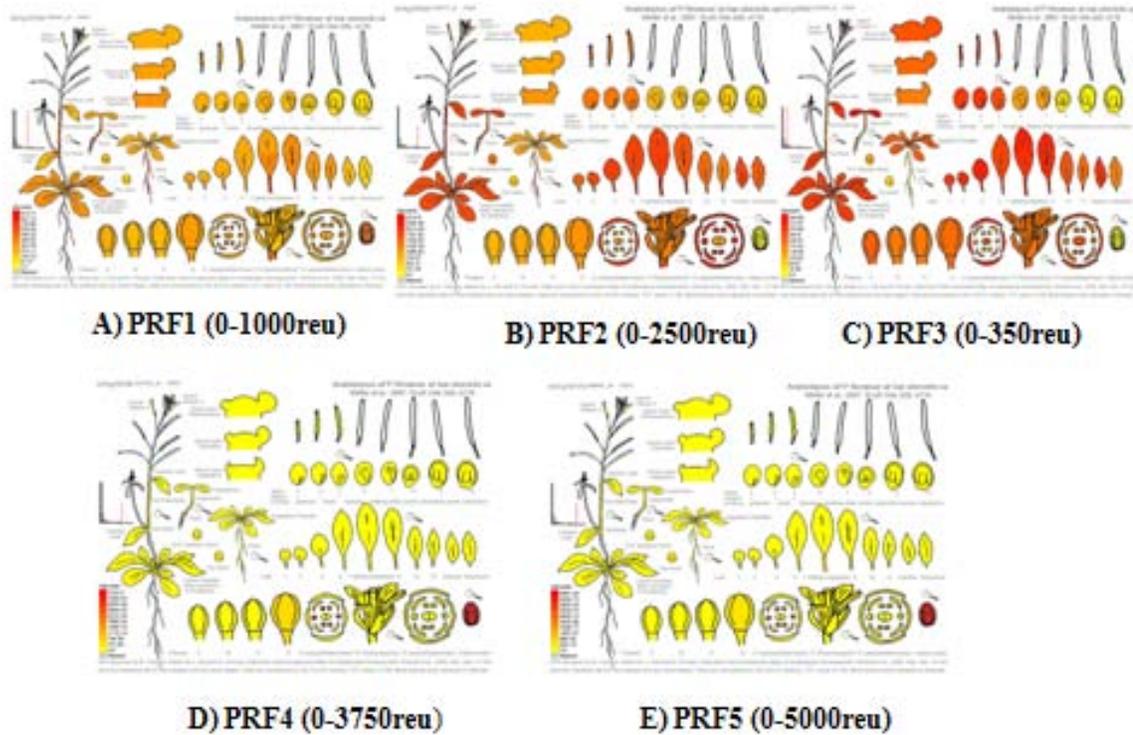


Figure 1.3: Tissue-specific expression profiles for *Arabidopsis* profilins (PRF1-PRF5) retrieved from the *Arabidopsis* e-GFP browser (Winter et al., 2007). Scale of expression is in relative expression units (reu), with red representing high expression levels, orange moderate expression levels, and yellow low expression levels. A) PRF1. B) PRF2. C) PRF3. D) PRF4. E) PRF5.

CHAPTER 2

ANALYSIS OF ARABIDOPSIS THALIANA PLANTS DEFICIENT IN VEGETATIVE
CLASS PROFILINS IDENTIFIES INDEPENDENT AND QUANTITATIVE GENETIC
EFFECTS¹

¹ Kristofer J. Müssar, Muthugapatti Kandasamy, Elizabeth McKinney, and Richard B. Meagher. To be submitted to the Journal of Plant Physiology.

Abstract

The actin cytoskeleton is involved in an array of integral structural and developmental processes throughout the cell. One of actin's best studied binding partners is the small ubiquitously expressed protein, profilin. *Arabidopsis thaliana* is known to encode a family of five profilin sequence variants: three vegetative profilins that are expressed in all vegetative tissues and ovules, and two that are specifically expressed in pollen. This paper analyzes the roles of three vegetative profilin members, PRF1, PRF2, and PRF3, in plant cell and organ development. Using a collection of knockout or severe knockdown T-DNA single mutants, we found that defects in each of the three variants gave rise to specific developmental deficiencies. Plants lacking PRF1 or PRF2 had defects in rosette leaf morphology and inflorescence stature, while those lacking PRF3 led to plants with slightly elongated petioles. To further examine these effects, double mutants and multiple gene silenced RNAi epialleles were created. These plants displayed significantly compounded developmental defects, as well as novel lateral root growth morphological phenotypes. Microscopic examination of dwarfed plants lacking in profilin variants indicated that they have smaller cells defective in cell elongation. Evidence is presented that mixtures of independent function, quantitative genetic effects, and functional redundancy have preserved the three vegetative profilin genes. We conclude with discussing a model for profilin's role in cell elongation based upon overall profilin concentrations in the cell.

Introduction

Actin is a highly conserved, abundant, and multifunctional cytoskeletal protein that makes polymeric filaments and filament bundles to control spatial development in eukaryotic cells and organs (Kudryashov and Reisler, 2012). Actin participates in more protein-protein interactions than any known protein (Dominguez and Holmes, 2011). Through these interactions, Actin Binding Proteins (ABPs) facilitate rapid remodeling of the cytoskeleton by regulating the unpolymerized (G-actin monomers) and polymerized (F-actin filaments) actin equilibrium (Yarar et al., 2007). Actin is critical player in such processes as stress response, transcription, cytokineses, cell locomotion, maintenance of cell shape and polarity, and development (Williamson, 1993; Meagher and Williamson, 1994; Jockusch et al., 2007; Perrin and Ervasti, 2010; Miralles and Visa, 2006). The rapid polymerization and depolymerization of actin filaments is necessary for preserving cell shape, appropriate cell division, and development (Yarmola and Bubb, 2009).

Profilins are small (12-15 kDa), ubiquitously expressed, monomeric ABPs that have been identified in organisms ranging from protists and fungi through higher plants and animals (Ramachandran et al., 2001). Originally, profilin was shown to specifically bind G-actin (globular actin) and was thought primarily responsible for G-actin sequestering in cells (Carlsson et al., 1976). Recently, profilin has also been found to inhibit the spontaneous polymerization of actin filaments by forming a 1:1 complex with G-actin, thereby lowering ATP-G-actin steady-state concentrations, and once all barbed ends (+ end of actin polarity) become blocked by capping proteins, profilin begins to sequester G-actin from pointed-end polymerization (Dominguez, 2009).

However, extensive research has shown that they also play a complex role in the formation of F-actin (filamentous actin) through the replenishment of the ATP-actin monomer pool via catalyzing the exchange of ADP for ATP on Actin (Dominguez and Holmes, 2011). While profilin does not bind F-actin directly, profilin-ATP-G-actin complexes are essential for rapid filament assembly (Schulter et al., 1997). Although profilin-bound actin monomers cannot add to pointed ends of actin filaments, they have been shown to elongate filament barbed ends at approximately the same rate as free actin monomers (Pollard and Borisy, 2003 and dos Remedios et al., 2003). This would lead us to believe that profilin might be facilitating rather than inhibiting polymerization (Yarmola and Bubb, 2006). This idea is further strengthened by results indicating that profilin could lower the critical concentration of actin needed to drive polymerization (Pantaloni and Carlier, 1993).

Besides binding actin, profilin also binds phosphatidylinositol 4,5-bisphosphate (PIP₂) (Sohn et al., 1995), poly-L-proline (Bjorkegren et al., 1993), many proline-rich proteins such as vasodilator-stimulated phosphoprotein (VASP) (Haffner et al., 1995), formin homology domain-containing proteins (Frazier and Field, 1997), the *Arp2/3* complex (Mullins et al., 1998), and the annexins (Alvarez-Martinez et al., 1996). Profilin can bind proline-rich sequences and actin simultaneously. In fact, it binds to both with higher affinity as a ternary complex than to either one separately (Ferron et al., 2007). The knowledge that proline-rich sequences are abundant among cytoskeletal proteins suggests that these proteins may serve to guide profilin-actin complexes from the cellular pool onto actin filaments (Dominguez and Holmes, 2011). Through these complex interactions, profilins have been shown to participate in cell elongation, cell shape

maintenance, and other processes requiring actin dynamics (Zheng et al., 2012). In addition, there is evidence showing that profilins are regulating actin networks by signaling to the actin cytoskeleton through PIP2, Cdc42, and Actin Related Protein (ARP) complexes (Stradal et al., 2004 and Witke et al., 1998). However, this link, as well as profilins effects on tissue and organ development remains relatively unexplored.

Profilin sequence alignments show small amounts of sequence homology in the N- and C-terminal regions, even when comparisons are made across various eukaryotic kingdoms. Only 25 to 40 residues (~20-30%) are universally conserved across all kingdoms (Meagher, 1995). This is unlike what is seen with actin, which is more than 80% conserved in most organisms across kingdoms, the exception being primarily among the more divergent protists (Kandasamy et al., 2012).

In plants, profilins are encoded by small gene families, which exhibit distinct tissue and organ-specific expression patterns throughout development (Kandasamy et al., 2002). There are two separate ancient classes of profilins in higher plants, vegetative (constitutive) and reproductive (pollen-specific). These classes show significant amounts (~25%) of amino acid sequence divergence. By looking at conservations among all profilins compared among monocots and dicots, we see that vegetative profilins in monocots and dicots are more similar to each other than they are to their own rep. profilin counterparts. Based on this and other divergence data it has been estimated that reproductive and vegetative class profilins arose ~200 to 400 million years ago in early land plant evolution before the split between monocot and dicot angiosperms (Blanc and Wolfe, 2004; Krom and Ramakrishna, 2008). This tissue specific partitioning can also be seen in humans (Hs), where HsProfilin1 is ubiquitously expressed at high levels in all

organs and tissues except muscle. HsProfilin2 is expressed high in brain and muscle, and HsProfilin3 and HsProfilin4 are expressed almost exclusively in sperm (Honore et al., 1993). Phylogenetic analysis demonstrates that distinct classes of profilin have independently evolved in both plants and animals.

In *Arabidopsis thaliana*, there are five profilin genes (PRF1-PRF5). PRF1, PRF2, and PRF3 are vegetative, being constitutively expressed throughout all vegetative tissues and in ovules, and were originally classified as “constitutive profilins”. PRF4 and PRF5 are classified as reproductive profilins and are predominately expressed in mature pollen. Phylogenetic relationships among this gene family can be seen in **Figure 2.1A** (Meagher et al., 1999). The three veg. proteins share 90% sequence identity, while the vegetative and reproductive classes share ~70-75% sequence homology (Huang et al., 1996). While the expression levels of PRF4 and PRF5 are almost identical, the vegetative profilins exhibit a varying range of expression. PRF2 is the most highly expressed, PRF1 is only expressed at moderate levels (~40% of PRF2 levels), and PRF3 is weakly expressed (~12% of PRF2 levels). Despite varying in their amounts of expression, PRF1, PRF2, and PRF3 are expressed spatially similar (Winter et al., 2007).

While there has been some research depicting the function of these *Arabidopsis* profilins, their effects on overall plant development still remain a mystery. Previous analysis has shown that a partial knockdown (RNA and protein levels 50% of WT) of the vegetative profilin, PRF1, results in altered seedling development, elongated hypocotyls, loss of light regulation, as well as defects in root hair development, flowering time, cell elongation, and overall cell shape maintenance (McKinney et al., 2001, Ramachandran et al., 2001). However, due to the leaky nature of the mutants being examined these

phenotypes were not overwhelming, suggesting that complete knockouts as well as double and triple knockouts will need to be established and dissected in detail.

Biochemical analysis and localization observations have shown that PRF1 has a higher affinity for binding poly-L-proline and G-actin than PRF2, and that while PRF1 is more likely associated with filamentous actin, PRF2 localizes to polygonal meshes resembling the endoplasmic reticulum (Wang et al., 2009). A detailed functional analysis of PRF3 has yet to be reported.

We describe here, using various knockout T-DNA insertion mutants and RNAi knockdown plants in multiple combinations, the roles of the three *Arabidopsis* vegetative protein variants in cell, tissue, and organ development. After morphological analysis we saw that plants deficient in PRF1 or PRF2 lead to similar defects in rosette leaf morphology and inflorescence stature, while the loss of PRF3 led to less evident phenotypes like elongated petioles. The creation of double mutants showed combinations of the single mutant phenotypes, while knocking down all three profilins showed the most drastic dwarfed phenotypes as well as problems with lateral root initiation and growth, indicating the presence of quantitative genetic effects. Through microscopic analysis of dwarfed plants lacking PRF1 and PRF2, we saw what appear to be defects in cell elongation that result in undersized cells along the leaf epidermis. These results help further elucidate profilin's role in actin dynamics, which has been widely debated over the past 20 years.

Materials and Methods

Plant Materials and Growth Conditions

All *Arabidopsis thaliana* seeds were of the Columbia (Col) ecotype. Wild-type, mutant, and transgenic seeds were grown in conditions and media described previously (McKinney et al., 1995 and 2001, Gilliland et al., 2002). T-DNA insertion lines were obtained from the *Arabidopsis* Biological Resource Center (ARBC, Ohio St. University). *prf1-4* (GK_614F01) and *prf3-2* (GK_055A02) were from the Gabi Kat mutant collection, while *prf2-1* (SALK_129071) was generously provided to us from Dr. Brad Day (Michigan St. University), and is derived from the SALK mutant collection. T-DNA mutant lines were cleaned up by backcrossing to WT-Col, allowing heterozygotes for the insertion to self-pollinate, and then repeating the process for a second and third time to ensure that these lines are free of other T-DNA insertions. These plants were screened each generation for the presence of their respective mutant alleles by PCR using methods previously described (Kandasamy et al., 2005) and the following sets of mutant Left Border (LB) and WT primers: *prf1-4*, PRF1_WT_S (5'-TAGACCATTAGTCTATTGTGAGAT-3'), Prf1-4_GK_LB (5'-CGTCGGAGAATTCAGTACTCG-3'), and PRF1_WT_AS (5'-TTCGCCACCGAGAAATAGTCCGGTT-3'), *prf2-1*, PRF2_WT_S (5'-ATCGACTTTCACACAAAACAT-3'), Prf2-1_SALK_LB (5'-GCAATTAGCTTCAACCGACTG-3'), and PRF2_WT_AS (5'-TTGCCTTCGACCTCGCACATGAGAT-3'), *prf3-2*, PRF3_WT_S (5'-AGATGAGGGCCTTATAATGGA-3'), Prf3-2_GK_LB_S (5'-ATCATCGATCGGCTCATATTG-3'), and PRF3_WT_AS (5'-

GTAGTCGGTATAGAAATA-3'). DNA for PCR was extracted using the REDExtract N-Amp Plant PCR Kit (Sigma-Aldrich). Following confirmation via PCR, clean mutant lines were sent off for DNA sequencing to confirm the exact location of the insertions. *prf1-4* had an insertion 74 bp upstream of the second exon in the first intron, *prf2-1* had an insertion 113 bp upstream of the translational start site in the promoter, and *prf3-2* had an insertion 15 bp from the end of the first exon (**Figure 2.1B**). All plants were grown at 22°C with 16-h days/ 8-h nights.

Double Mutant Generation

Double mutants were then generated through the following plant crosses between the individual T-DNA mutants: *prf1-4/prf1-4* pollen crossed with emasculated *prf2-1/prf2-1* (*prf1-4 prf2-1*), *prf1-4/prf1-4* pollen crossed with emasculated *prf3-2/prf3-2* (*prf1-4 prf3-2*), *prf2-1/prf2-1* pollen crossed with emasculated *prf3-2/prf3-2* (*prf2-1 prf3-2*). F1 progeny were screened by PCR for the presence of both alleles (using primers above), and then allowed to self-pollinate. PCR was used to check F2 progeny displaying the dwarfed leaves phenotypes for the presence of both mutant alleles and the absence of both wild-type alleles.

Simplified Construction of RNAi Transgenes

Single, double, and triple RNAi constructs were designed based on previously described methods (Pawloski et al., 2006 and Tian et al., 2009) with an important simplification. Previous constructs used a large 1,400 bp petunia intron to separate the forward and reverse facing sequences and RNAi gene constructions required going through multiple rounds of overlapping PCR or a multistep cloning process to make the assembly. Instead we used a 79 bp Actin2 intron flanked by two “A” residues on either

side and had it synthesized by GenScript (Piscataway, NJ). This design allowed for a much smaller gene construct to be assembled. The constructs consisted of 100 (*PRF1*-RNAi), 200 (*PRF2L*-RNAi and *PRF1 PRF2*-RNAi), or 300 (*PRF1 PRF2 PRF3*-RNAi) bp inverted repeats (depending on how many genes being targeted) separated by the “A” residues and the 79 bp Actin2 intron, all under the control of the Actin2 promoter terminator (A2pt) (Kandasamy et al., 2002). Once the intron was removed, we were left with a stable “AAAA” loop connected to the RNA stem consisting of the inverted repeats that hybridize to the first 100 bp (200 bp for *prf2*-RNAi) of the 3’-UTRs of their corresponding profilin target genes. *PRF2*-RNAi required a longer inverted stem of 200 bp in order to achieve sufficient silencing of *PRF2*.

Complementation Constructs were made by cloning full-length *PRF1*, *PRF2*, and *PRF3* cDNAs under the control of the A2pt construct as described in Kandasamy et al., 2002. This ensured the proper expression in the appropriate tissues. Fimbrin-GFP reporter constructs (*35S:GFP-FABD2*); previously described in (Wang et al., 2004) were transformed into our WT and *PRF1 PRF2*-RNAi plants to allow for visualization of actin filaments. The *35S:GFP-FABD2* construct consists of GFP fused to the C-terminal half of *Arabidopsis* Fimbrin1. For our constructs we exchanged the hygromycin resistance marker for a Basta resistance marker. All transformations were performed with *Agrobacterium tumefaciens* strain C58C1 using the floral dip method (Clough, 2005).

Leaf, Root, and Plant Measurements

All leaf measurements were taken on day 28 of development (i.e., 28 days after seed germination on soil). Measurements were collected using a standard metric ruler by the same individual to ensure consistency. For each measurement a total of 52 rosette

leaves (largest two leaves per plant on 26 plants) were analyzed for each WT, mutant, complement, and RNAi line. Plant height measurements were taken on day 40 of development using a standard metric ruler after laying plants flat on the bench and measuring to the top of the inflorescence. For each measurement a total of 30 plants were analyzed for each WT, mutant, complement, and RNAi line. All root quantifications were taken on day 15 of development using a standard metric ruler or by counting (lateral root initiation). For each measurement a total of 30 roots were analyzed for WT, *PRF1 PRF2*-RNAi, *PRF1 PRF2 PRF3*-RNAi, *prf3-2*, and *A2p:PRF3* overexpression lines. To measure the hypocotyls, seeds were grown vertically in dark growth conditions with measurements taken on day 10. All measurements were taken to the nearest 0.1 mm. Graphs of resulting data were constructed in Excel (Microsoft).

qRT-PCR RNA Analysis

RNA was isolated, treated, and cDNA was made from leaf tissues of wild-type and various transgenic plants as previously described (Kandasamy et al., 2007). cDNA populations were analyzed using the following qRT-PCR primers: Ubiquitin10 was the endogenous control, Ubiquitin10_Sense (5'-AGAAGTTCAATGTTTCGTTTCATGTAA-3') and Ubiquitin10_Antisense (5'-GAACGGAAACATAGTAGAACACTTATT-3'), PRF1, PRF1_3utr_Sense (5'-TCTCCTTCGTTACCGAGTTTGAG-3') and PRF1_3utr_Antisense (5'-ACTCAATACATATGGAGAAAAAAGAT-3'), PRF2, PRF2_3utr_Sense (5'-CTGCCATGTATTGTGATTTGATTG-3') and PRF2_3utr_Antisense (5'-GAGAGGATCAAAACCATAACAAATAT-3'), PRF3, PRF3_3utr_Sense (5'-GTGTCGTGAGAGAAAAACTATTCGAT-3') and PRF3_3utr_Antisense (5'-CCCCAAGATCCATCACAAGGT-3'). All primer sets were

designed to detect the 3'-UTR of their respective genes, thus ensuring distinct specificity and that primers were downstream of all T-DNA insertions. Reactions were performed on an Applied Biosystems 7500 real-time PCR system using SYBR Green detection chemistry (Applied Biosystems) as described previously (Deal et al., 2007). In all experiments, the $2^{-\text{ddCT}}$ method (Livak and Schmittgen, 2001) was used to detect the relative quantification of gene expression.

Western Blot Analysis

Arabidopsis protein samples were prepared by grinding 50 mg of frozen leaf tissue in liquid nitrogen and then extracted in 125 μL of extraction buffer containing 25 mM Tris-HCl, pH 7.5, 10 mM NaCl, 10 mM MgCl_2 , 5 mM EDTA, and a protease inhibitor cocktail (Roche Diagnostics; one tablet/10 mL). After 10 min centrifugation, the supernatant was mixed 1:1 with 2 \times Sodium Dodecyl Sulfate (SDS) sample buffer (O'Farrell, 1975) and boiled for 5 min. $\sim 15\text{-}20$ μL were loaded per well (i.e., ~ 25 μg protein). Protein samples were then separated on 12% SDS-PAGE gels and transferred to a Immobilon transfer membrane (Millipore, MA) by semi-dry blotting (Hofer, CA). Membranes were blocked for 30 min in TBST (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween 20) containing 20% goat serum and 5% dry milk, and then probed with the primary antibody (MAbPRF1a or MAbPRF12a, see Kandasamy et al., 2002) at 0.5 g/ml concentration for 1 h, and then washed thoroughly with TBST. Then membranes were probed with IgG-antimouse horseradish peroxidase-conjugated secondary antibody at a 1:2000 dilution in blocking solution for 30 min. The blots were washed again in TBST (3 \times 5 min), treated with ECL detection solution (Amersham, NJ) for about 2 min and then exposed to the Hyperfilm ECL (Amersham). Western blot analysis was repeated at least

twice for each experiment. Coomassie Brilliant Blue staining of duplicate gels was used to monitor the equal loading of proteins and to adjust loading if necessary.

Quantification of bands was calculated using ImageJ (NIH), a Java-based image processing program.

Microscopy Analysis

Day 28 leaf samples from WT and transgenic *PRF1 PRF2*-RNAi plants containing the *35S:GFP-FABD2* construct (Wang et al., 2004) were suspending in a single drop of distilled water on a 25 mm x 75 mm microscope slide. The actin microfilaments in the labeled cells were visualized with a Leica confocal laser scanning microscope (TCS-SP2). The images were further processed using Adobe Photoshop software (Adobe Systems, Mountain View, CA).

Actin-Inhibiting Drug Treatments

Both WT-*35S:GFP-FABD2* and *PRF1 PRF2*-RNAi - *35S:GFP-FABD2* seedlings were grown up in liquid germination medium (Murashige and Skoog salts [Life Technologies, Rockville, MD]) with 1% sucrose on a 96 well MICROTTEST flat bottom Tissue Culture Plate (Becton Dickinson), and grown at 22°C with 16-h days/ 8-h nights. Latrunculin-B, from Red Sea Sponge (Sigma-Aldrich) and Cytochalasin D, from *Zygosporium mansonii* (Sigma-Aldrich) were dissolved in dimethyl sulfoxide (DMSO) to varying concentrations (25-100 mM). Five day old seedlings were then treated with 25, 50, 75, or 100 mM of either Latrunculin-B or Cytochalasin D, and allowed to incubate for one to five hours. The treated seedlings were then examined using a Leica confocal laser scanning microscope (TCS-SP2).

Results

Vegetative Profilin Single Mutants Show Defects in Leaf and Inflorescence

Development

To determine the *in vivo* functional roles these vegetative profilins play in plant growth, actin dynamics, and *Arabidopsis* development, we characterized single T-DNA insertion mutants for PRF1, PRF2, and PRF3. *prf1-4* has an insertion in the first intron 74 bp upstream of the second exon, *prf2-1* has an insertion 113 bp upstream of the translational start site in the promoter, and *prf3-2* has an insertion at the end of the first exon (**Figure 2.1B**). To ensure that resulting mutant phenotypes were indeed caused by these specific insertions, we constructed complemented lines by overexpressing endogenous complementary PRF1, PRF2, or PRF3 cDNAs, respectively, under the control of the constitutive Actin2 promoter (A2pt). Multiple (2-3) independent transgenic complementation lines were analyzed.

At day 28 after germination there are significant visible defects in leaf and inflorescence development for *prf1-4* and *prf2-1* plants, while *prf3-2* plants appeared relatively normal (with slightly elongated petioles) as shown in **Figure 2.1C**. *prf1-4* and *prf2-1* plants developed leaves that are significantly shorter in total length, width, and blade length (**Figure 2.1D-E**). Individual leaf morphology pictures can be seen in **Figure 2.1G**. These mutant plants were also shorter in overall plant height, with inflorescences appearing less stable than that of WT (**Figure 2.1F**). This lack of structure suggested that the actin cytoskeletons of these cells in the inflorescences' was impaired. Pictures of these mutant plants at other developmental time points can be seen in the supplemental data (**Figure 2.S1**).

To ensure that these mutants and RNAi knockdowns were not expressing any RNA or protein, qRT-PCR and western blot analysis were performed. qRT-PCR primers were designed to target the distinct 3'-UTR sequence of each gene respectively, and all monoclonal antibodies (mAb) used in this study recognized a 13- to 14-kD profilin band (Kandasamy et al., 2002). MAbPRF1a reacts strongly and specifically with PRF1, while MAbPRF12a reacts strongly with PRF1 and PRF2 and only modestly with PRF3 (Kandasamy et al., 2002). qRT-PCR and western blot analysis revealed that these mutants had very little to no detectable RNA or protein expression (**Figure 2.2A-B**). We also demonstrated that the complement lines did contain much higher levels of RNA and protein than WT (**Figure 2.2A-B**). However, no visible phenotypes were observed in these complemented plant lines overexpressing any of the three vegetative profilins.

To further confirm our analysis of phenotypes, we also created single RNAi silencing epialleles for PRF1 and PRF2 (*PRF1*-RNAi and *PRF2*-RNAi). We developed a new efficient method for constructing RNAi genes that expressed simple stem-loop structures that will silence RNA expression (see Materials and Methods). These stem-loop structures were designed to target and silence the 3'-UTR of each gene. We saw similar morphological phenotypes compared to our T-DNA insertion mutants and the resulting measurements for leaf length, width, and plant height were confirmed (**Figure 2.S2**). qRT-PCR of these epialleles exposed that RNA levels were less than 10% of WT (**Figure 2.S2**).

Vegetative Profilin Double Mutants and Double/ Triple RNAi lines Show More Severe Effects on Development

In order to assess both functional redundancy and quantitative effects of the vegetative profilins in *Arabidopsis* development, three double T-DNA mutants were generated: *prf1-4 prf2-1*, *prf1-4 prf3-2*, and *prf2-1 prf3-2*. The phenotypes shown in **Figure 2.3** make it clear that all three double homozygous null mutants exhibit even stronger and more distinct developmental phenotypes than any single PRF1, PRF2, or PRF3 defective plant. The double mutants had noticeably smaller leaves than the single mutants that are shorter in total length, blade length, and width. These defects are seen throughout development. Interestingly, double mutants containing the *prf3-2* allele show remarkably longer petioles (**Figure 2.3B-C**, see the next section). The double mutants are also shorter in overall plant height (**Figure 2.3E**).

qRT-PCR and western blot analysis again revealed that these mutants had very little to no detectable RNA and even more greatly reduced protein expression (**Figure 2.4A, D**). Notice how the *prf1-4 prf3-2* double mutant had the strongest profilin expression, which is in agreement with PRF2 being the most highly expressed member of the gene family. *prf1-4 prf2-1* had the lowest profilin expression, consistent with known expression levels for these two genes, and therefore produced the most drastic developmental phenotypes. These data suggested that unlike *Arabidopsis* vegetative actin ACT7, which is up-regulated in response to deficiencies in ACT2 and ACT8, none of the vegetative profilins were significantly up-regulated in response to profilin deficiency.

Based on previous studies that silenced four late pollen actins and four Actin Depolymerizing Factors (ADFs) by stacking four different 100 bp 3'-UTR sequences in

the stem of a stem-loop RNA (Pawloski et al., 2006, Tian et al., 2009), we created a simplified design for RNA interference constructs, and silenced PRF1 and PRF2 (*PRF1 PRF2*-RNAi), as well as PRF1, PRF2, and PRF3 (*PRF1 PRF2 PRF3*-RNAi), simultaneously (see Materials and Methods). Two strongly silenced independent transgenic lines (#23 and #6 for *PRF1 PRF2*-RNAi, #26 and #19 for *PRF1 PRF2 PRF3*-RNAi) and one intermediately silenced line (#11 for *PRF1 PRF2*-RNAi, #6 for *PRF1 PRF2 PRF3*-RNAi) were used in these analyses. The strongly silenced lines were severely dwarfed throughout development, with the triple RNAi epiallele showing much more drastic phenotypes than any of the double mutants (**Figure 2.5**). qRT-PCR and western blot analysis of these lines show minute levels of RNA and no detectable protein expression (**Figure 2.4B,C,E**). *PRF1 PRF2 PRF3*-RNAi plants produced much less seeds, had fewer siliques, and were significantly shorter than any of the other single or double mutant lines (**Figure 2.5D**). These results indicated that when *Arabidopsis* plants were deficient in all three vegetative profilins there appears to be a quantitative genetic effect leading to severely dwarfed plants and a wide range of tissues not being fully developed. Interestingly, the double *PRF1 PRF2*-RNAi line exhibited slightly more radical phenotypes than the *prf1-4 prf2-1* double mutant. However, by looking at the western blot data (**Figure 2.4D-E**) we saw that the *PRF1 PRF2*-RNAi line had even less protein than the *prf1-4 prf2-1* double mutant, which explains the more severe phenotypes.

PRF3 Deficient Plants Exhibit Slightly Elongated Petioles

While PRF3 deficient plants did not seem to display the strong dwarfed leaf phenotype similar to plants deficient in PRF1 or PRF2, they did exhibit elongated petioles compared to WT (**Figure 2.6A**). The elongated petiole phenotype can be seen in

all plant lines that have the PRF3 gene knocked down (*prf3-2*, *prf1-4 prf3-2*, and *prf2-1 prf3-2*), with the exception of *PRF1 PRF2 PRF3*-RNAi plant lines whose leaves and petioles were so dwarfed that this difference was not statistically significant (**Figures 2.6A, 2.3B, and 2.5B**). These data suggest that PRF3 was performing some sort of independent function during development. Previous experiments using *PRF3* promoter-*GUS* fusion constructs confirmed that PRF3 was being expressed in petioles (Fan et al., 2013), suggesting that PRF3 may play a distinct role in petiole development, a role that cannot be played by either PRF1 or PRF2.

Multiple independent PRF3 overexpression lines were analyzed by growing them vertically on plates containing 0.5 MS salts and 1% sucrose germination media, yielding no phenotypic result (**Figure 2.6B**). Despite previously published data indicating that the overexpression of PRF3 causes stunted roots with defects in cell elongation and F-actin organization (Fan et al., 2013), we saw no phenotype in our PRF3 overexpression lines. In addition, no effects were seen on hypocotyl development in PRF3 overexpression plants (**Figure 2.6C**). These contrasting phenotypic results could be due to differences in how PRF3 was overexpressed. The loss of PRF3 RNA expression in these lines was demonstrated using qRT-PCR (**Figure 2.6D**). Western blot analysis was unable to clearly determine the extent of PRF3 protein present, since PRF3 represents such a small part of total profilin expression, and because MAbPRF12a reacts stronger with the more highly expressed PRF1 and PRF2 proteins than PRF3 (**Figure 2.6E**). A complete list of phenotypic measurements for all plant lines is presented in **Table 2.S1**.

Vegetative Profilins are Essential to Lateral Root Initiation

We grew all mutant and epiallele plant lines vertically on plates containing 0.5 MS salts and 1% sucrose germination media to look for defects in root growth. Most lines showed no significant root growth and formation phenotypes (not shown). For example, lines lacking PRF1 and/or PRF2, the two most highly expressed profilins, had nearly normal primary and lateral root development as shown in **Figure 2.7A**. However, multiple *PRF1 PRF2 PRF3*-RNAi plant lines showed drastic differences in its root architecture (**Figure 2.7A**). These plants appeared to form normal length primary roots (**Figure 2.7B**), but have severe deficiencies in lateral root growth and formation. They not only produce a lower number of lateral roots (**Figure 2.7C**), but these lateral roots were also much shorter (**Figure 2.7D**), indicating that there could likely be a problem in cell elongation among these lateral roots. The *PRF1 PRF2*-RNAi plants (or any of the T-DNA double mutants, not pictured) do not show a significant loss in the number of lateral roots initiated, but do show shorter lateral roots than WT. This suggested that the total amount of profilin needs to be at some minimal level in order to properly initiate lateral root formation. This would indicate that there was functional redundancy among the vegetative profilin gene family, and that having any of these three profilins is sufficient for proper lateral root initiation and growth. Root architecture of the intermediately silenced *PRF1 PRF2*-RNAi and *PRF1 PRF2 PRF3*-RNAi epialleles revealed intermediate phenotypes that appear proportional to PRF expression levels (**Figure 2.S3**).

PRF1PRF2-RNAi Plants Show Effects on Cell Size and Actin Filament Organization

In order to determine how the defects described above affected the F-actin cytoskeleton, we cloned the *GFP-FABD2* fusion (a green fluorescent reporter with one binding site for F-actin) into *PRF1 PRF2*-RNAi lines as well as WT control plants. The doubly silenced plants were chosen over the triply silenced lines because *PRF1 PRF2 PRF3*-RNAi lines were too sick to manipulate genetically. The GFP fusion allowed us to visualize actin cellular dynamics by binding and revealing dynamic arrays of F-actin filaments *in vivo* (Wang et al., 2004 and Voigt et al., 2005). These lines were then observed using confocal microscopy. We first looked at pavement cells on the *Arabidopsis* leaf epidermis (**Figure 2.8A**). These cells are known for being shaped like the interlocking pieces of a jigsaw puzzle (Glover, 1999). As shown in **Figure 2.8A**, *PRF1 PRF2*-RNAi plants have dramatically smaller cells than those seen in WT, suggesting that vegetative profilins were playing an active role in cell elongation. This result likely explains why the resulting leaves were much smaller in size throughout development. To further examine actin filament organization, we looked at vein patterning in leaf cells (**Figure 2.8B**). It appeared as if plants deficient in PRF1 and PRF2 exhibited slightly disorganized filamentous structures, with an increased aggregation of F-actin filaments. These plants were also treated with the actin inhibitors, latrunculin B and cytochalasin D, but we failed to see a significant change in their F-actin networks compared to WT (not pictured). Further studies will need to be performed to mechanistically deduce the effects that lowering the concentration of the profilin pool has on actin filament dynamics.

Discussion

The actin binding protein, profilin, is abundantly found throughout complex multicellular organisms. The *Arabidopsis thaliana* genome encodes a five member profilin gene family, three vegetative and two reproductive. We have focused on dissecting the functional consequences of knocking out the three vegetative gene members both individually and in combinations. Previous studies on a PRF1 mutant that partially reduced PRF1 expression led to the discovery of minor defects in *Arabidopsis* development (McKinney et al., 2001, Ramachandran et al., 2001). This paper serves to extend these analyses to all three vegetative gene members by utilizing newly obtained null T-DNA insertion mutations, double mutants, and by taking advantage of a slightly modified RNAi approach to silencing multiple profilin genes simultaneously. With these techniques we have identified additional abnormal developmental phenotypes that are much more severe, and hope to use these results to develop a better understanding of profilin's functions as an actin binding protein.

Vegetative Profilins Effect Normal Leaf and Inflorescence Development

To complete this analysis we began by acquiring and establishing T-DNA insertion mutants to each of the three vegetative profilins found in *Arabidopsis*. These lines were then confirmed by qRT-PCR and western blot analysis (**Figure 2.2**). For *prf1-4* and *prf2-1*, we saw extremely similar phenotypic effects, with plants showing defects in normal rosette leaf morphology as well as inflorescence development, leading to shorter overall plant height for these mutants (**Figure 2.1**). The inflorescences of these mutants were thinner and weaker than WT and not stable enough to stand up on their own. This suggests that there are structural deficiencies in these mutant tissues. Since profilin is

thought to be responsible for shuttling monomeric actin to promote filament formation, perhaps the absence of these highly expressed profilins is inhibiting or slowing the formation of actin-filaments at the leading edge of cells resulting in a lack of appropriate cell elongation in these tissues. The *prf3-2* mutant showed no significant inflorescence phenotypes (**Figure 2.1**).

GUS-constructs have been made to show that all three of these profilins are expressed in the same tissues all throughout the plant (Jeong et al., 2006 and Fan et al., 2013), so it is slightly surprising that similar phenotypes are not seen in the *prf3-2* mutant. One must remember that expression profiles have indicated that PRF1 and PRF2 are more highly expressed than PRF3 (Winter et al., 2007). This can explain why the lower expressed PRF3 does not have such obvious developmental defects. This suggests the possibility that there is partial functional redundancy among the vegetative members in this gene family.

While the rosette leaf and inflorescence phenotypes were absent from *prf3-2*, this mutant did appear to show effects on petiole development. In particular, plants lacking PRF3 showed slightly elongated petioles compared to WT (**Figure 2.6**), indicating that PRF3 may be specifically required for proper petiole formation. Perhaps PRF3 is responsible for controlling the proper spatial sequestering of actin monomers in petioles, thereby guiding normal petiole development. This result is the first documented instance where a loss of PRF3 has shown any type of developmental phenotype. Further studies into this particular mutant will be necessary in order to fully understand all of the specific functions of PRF3 in petiole and overall plant development. PRF3 overexpression lines were recently analyzed to show defects in seedling development, in particular, stunted

primary root and hypocotyl length (Fan et al., 2013). However, after constructing PRF3 overexpression lines using the strong constitutive ACTIN2 promoter, and performing qRT-PCR and western analysis to confirm the overexpression of PRF3 RNA and protein, these phenotypes were not detected. In particular, we did not see any significant deviation from WT primary root and hypocotyl length (**Figure 2.6B**). We are uncertain as to what to conclude from these two conflicting results. Regardless, single mutant analysis has revealed that when each of these genes are knocked out, abnormalities arise in plant development.

Knocking Out Multiple Vegetative Profilins Leads to Compounded Phenotypic Defects

After seeing that individual single gene mutants gave rise to noticeable phenotypes, we developed novel plant lines where multiple PRF genes were knocked out. This was the first time *Arabidopsis* plants with multiple profilin mutants have ever been constructed. As we were hoping to expect, we saw the same developmental phenotypes as the single mutants only they were more extreme. While *prf1-4* and *prf2-1* had similar leaf and inflorescence defects, the *prf1-4 prf2-1* double mutant gave rise to plants with even smaller leaves than the respective single mutants (**Figure 2.3**). However, the *prf1-4 prf2-1* double mutant did not lead to plants that were significantly shorter in overall plant height than the single mutants. This was unexpected, but after establishing the *PRF1 PRF2*-RNAi lines we did in fact see a significant drop in plant height as well as leaf size (**Figure 2.4**). While qRT-PCR data has both *prf1-4 prf2-1* and *PRF1 PRF2*-RNAi plant lines expressing very little PRF1 and PRF2 RNA, after performing western blot analysis we can clearly see that the RNAi lines have absolutely no protein expression while double mutants show a faint band indicating that there may still be small amounts

of protein present (**Figure 2.4D-E**). This could be attributed to the fact that the *prf2-1* insertion is in the promoter region upstream of the transcriptional start site leading to some leakiness when crossed with another mutant. Regardless, there is an agreement between the two approaches that by knocking out the two most highly expressed vegetative profilins; you see the most dramatic leaf and inflorescence phenotypes.

While the *prf1-4 prf2-1* mutant showed the most compounded developmental defects, we saw a combination of single mutant phenotypes in the *prf1-4 prf3-2* and the *prf2-1 prf3-2* double mutants. These plants both displayed dwarfed leaves, shorter, less stable inflorescences, and elongated petioles (**Figure 2.3**). This was fortunately exactly what we expected. This would indicate that while PRF1 and PRF2 are playing major roles in rosette and inflorescence development, PRF3 must be involved in the proper development of petioles. It makes sense that the much lower expressed PRF3 seems to have evolved to function specifically in the assistance of petiole development, while PRF1 and PRF2 serve to function in multiple tissues. In addition, since the PRF1 and PRF2 deficient plants exhibit very similar phenotypic effects, this suggests the possibility of there being partial functional redundancy. However, since the single mutants each have strong phenotypes on their own, we believe this may be more of a quantitative genetic effect. Further experiments will be needed to determine the extent, if any, of this possible functional redundancy.

PRF1 PRF2 PRF3-RNAi Plants Show the Most Drastic Dwarfed Phenotypes and Exhibit Defects in Lateral Root Growth and Formation

To further dissect the role vegetative profilins are playing in *Arabidopsis* development and examine possible quantitative genetic effects, we created an RNAi

construct using a modification of a published method (Pawloski et al., 2006 and Tian et al., 2009), that silences all three profilin genes simultaneously, *PRF1 PRF2 PRF3*-RNAi. Molecular characterization has shown that these plants are not expressing any vegetative profilin protein (**Figure 2.4E**). These plants are completely stunted in almost all aspects of development. They produce much smaller rosette and cauline leaves, have less siliques that are shorter leading to lower seed production, a lower degree of branching, and are completely dwarfed in overall plant height (**Figure 2.5**). When all three vegetative profilins are knocked down, the plants exhibit severe developmental defects. It appears that when the vegetative profilin pool is completely depleted, plants are unable to fully form many of its above ground tissues and organs. This stunted phenotype is indicative of defects in cell elongation. Surprisingly, these plants are not sterile; they produce on average 40-50 seeds per plant, but due to their sickening health they were very difficult to genetically manipulate.

The overwhelming amount of visible above ground phenotypes in these plants prompted a more detailed analysis into their root development. It was recently found that at root tips, actin polymerization is facilitated by the Arp2/3 complex and profilin through interactions with phosphatidylinositol 4,5-bisphosphate (Pei et al., 2012), thereby implicating profilin with proper root elongation. All single PRF mutants and most double mutants yielded no effects on root development. When PRF1 and PRF2 are knocked down, we begin to see slight root defects. Once all three PRFs are knocked down we begin to observe major deficiencies in the formation of the overall root architecture (**Figure 2.7A**). In particular, *PRF1 PRF2 PRF3*-RNAi plants lack the ability to initiate lateral root growth. Not only can we detect an absence of lateral root growth, there are

also much less lateral roots being formed. While primary root growth appears unaffected in these plant lines, defects in lateral root growth and formation are abundantly clear. These results would lead one to believe that perhaps it is PRF3 that controls lateral root development. However, these phenotypes are not seen in the *prf3-2* single mutant alone (**Figure 2.6B**), indicating that this interaction must be more complicated. One explanation would involve the overall pool of profilin in the cell. Perhaps once overall vegetative profilin protein levels reach a certain threshold, the cells conserve what is present and only initiates cell elongation in certain tissues and organs (possibly those more essential to development or survival). Another explanation is that the reproductive, PRF4 and PRF5, gene members are functionally filling in for the vegetative PRFs, but are less efficient or unable to promote growth in all tissues leading to visible developmental deficiencies in roots. However, since we do see a partial phenotypic rescue of root architecture and lateral root growth in our intermediately silenced *PRF1 PRF2 PRF3*-RNAi line (**Figure 2.S3**); we believe that this is a result of a quantitative genetic effect.

Lowering PRF Concentrations Lead to Altered Cytoskeletal Dynamics and Decreased Cell Size

Complex data from a variety of studies suggests that profilin may be functioning in either actin polymerization or depolymerization, but perhaps the most accurate description is that profilin is doing both. This is true for another class of small ABPs, the Actin depolymerizing factors (ADFs), which are known to stabilize and/or sever F-actin filaments in a concentration dependent manner (Bamburg, 1999 and Andrianantoandro and Pollard, 2006). In the presence of profilin, filament elongation occurs exclusively at the barbed ends, while elongation at the pointed ends appears inhibited. However, barbed

end elongation also occurs at the same rate in the absence of profilin suggesting that the dissociation of profilin could be a prerequisite for further filament elongation (Yarmola and Bubb, 2006). Furthermore, X-ray structure analysis has shown that profilin is required for the open-nucleotide-pocket of actin to remain open and stable (Minehardt et al, 2006). This conformation is a crucial intermediate in the actin depolymerization/polymerization cycle, and facilitates ADP to ATP nucleotide exchange in actin monomers, thereby linking profilin to both actin polymerization and depolymerization. This fluctuation-based process of nucleotide exchange diffusion has been shown to regulate actin polymerization and depolymerization (Yarmola and Bubb, 2009). Recent studies have found that the slow release of inorganic phosphate (Pi) from the barbed end of actin filaments is linked to an increase in the rate of filament disassembly, and is further accelerated by profilin (Jegou et al, 2011). This is evidence that profilin is facilitating disassembly of actin filaments. Other studies have shown that the overexpression of profilin by microinjection inhibited pollen tube elongation (McKenna et al. 2004 and Vidali et al. 2001). Yet, in order for the cell to extend protrusions and promote elongation there must be a rapid treadmilling (turnover) of actin, which is facilitated through ABPs like profilin and ADF (Clainche and Carlier, 2008). Altogether, these data suggest that profilin is involved in both the polymerization and depolymerization of actin filaments.

In order to determine possible alterations of the F-actin cytoskeleton in our dwarfed profilin deficient plants, we transformed a *35S::GFP-FABD2* fusion construct into WT and *PRF1 PRF2-RNAi* plants. This allowed for the visualization of F-actin filaments. Using this construct, we examined the consequences of having a decreased

profilin pool and its impact on actin cytoskeletal dynamics. Actin filament organization was examined in greatest detail in leaf epidermal cells. We saw that plants lacking PRF1 and PRF2 were associated with more disorganized filamentous structures (**Figure 2.8B**). There appears to be elevated levels of aggregated F-actin filaments in profilin deficient plants. We believe that by lowering profilin concentrations, there will be less profilin-actin complexes leading to an increase in unbound G-actin monomers. With little to no profilin present, actin monomers will not be properly sequestered within the cell resulting in the enhancement of actin polymerization and aggregation of actin filaments.

Further examination of the pavement cells on the leaf's epidermis indicated that the sizes of these cells in *PRF1 PRF2*-RNAi plants are significantly smaller than WT (**Figure 2.8A**). While the cells maintain the same jigsaw puzzle shape, they appear to lack the ability to expand and form protrusions to reach the size of normal puzzle shaped pavement cells. This is indicative of what we would expect if there were defects in cell expansion.

Our original hypothesis was that decreasing profilin levels would act to free up more actin monomers to form F-actin filaments, which would in turn lead to more rapid cell elongation. This is based off the idea that profilin is known to bind and sequester G-actin monomers (Carlsson et al., 1976). We previously reported longer hypocotyl cell lengths in mutant plants with lowered PRF1 concentrations (McKinney et al., 2001). However, this is not what we see in our plants that are severely deficient in PRF1 and PRF2. These initial hypotheses were based off of having knocked down a very small amount of just PRF1, thus only decreasing the overall profilin pool by approximately 15%. In the plant lines being studied now, we know that we have dropped overall

profilin concentrations more dramatically (~90%), by simultaneously silencing PRF1 and PRF2. Based on these previous findings and the results presented here, we suggest that decreasing the profilin pool by small amounts might lead to a sensory signal that tells the cell to start elongating quickly, whereas major reductions of profilin will lead to a type of arrest in cell elongation. This quantitative genetic effect of lowering profilin pool concentrations suggest that if actin monomers are unable to be bound by profilin, there will not be enough profilin-actin complexes being properly sequestered to the cell periphery to promote appropriate cell elongation. Furthermore, the lack of profilin would lead to defects in actin treadmilling, which is required for cell elongation. This would inevitably cause arrest in actin filament protrusion leading to plants with smaller leaves, as was seen in our profilin mutants. This model agrees with our findings that there is a direct correlation between the number of profilin genes that were knocked out, and the severity of the dwarfed plant phenotype.

In conclusion, we have demonstrated that vegetative profilins play an essential role in *Arabidopsis* plant development and the regulation of the actin cytoskeleton. We show that as you continue to decrease vegetative profilin gene expression you produce more compounded phenotypes, suggesting that there is a direct correlation between profilin concentrations and defects in development. While our model discussed above serves to provide explanations for these phenotypic effects, the exact mechanisms still need to be clarified in future studies. This paper analyzed profilin's role in promoting proper cell elongation, but additional research is needed to examine their roles in signal transduction, intracellular transport, and communication. We believe that a systems biology approach needs to be taken in order to explain how all of these processes are

interacting with each other through a profilin intermediate. Further experiments are also needed to determine the extent of functional redundancy between the highly similar PRF1 and PRF2.

REFERENCES

- Alvarez-Martinez MT, Mani JC, Porte F, Faivre-Sarrailh C, Liautard JP, Sri Widada J. (1996). Characterization of the interaction between annexin I and profilin. *Eur J Biochem* 238, 777–784.
- Andrianantoandro E, Pollard TD. (2006). Mechanism of actin filament turnover by severing and nucleation at different concentrations of ADF/Cofilin. *Mol Cell* 24, 13–23.
- Bamburg JR. (1999). Proteins of the ADF/Cofilin Family: Essential Regulators of Actin Dynamics. *Annual Review of Cellular Developmental Biology* 15, 185-230.
- Bjorkegren C, Rozycki M, Schutt CE, Lindberg U, Karlsson R. (1993). Mutagenesis of human profilin locates its poly (L-proline)-binding site to a hydrophobic patch of aromatic amino acids. *FEBS Letts* 333, 123–126.
- Blanc G and Wolfe KH. (2004). Functional Divergence of Duplicated Genes formed by Polyploidy during *Arabidopsis* Evolution. *Plant Cell* 16, 1679-1691.
- Carlsson L, Nystrom L, Sundkvist I, Markey F, Lindberg U. (1976). Profilin, a low-molecular weight protein controlling actin polymerisability. In *Contractile Systems in Non Muscle Tissues* (Perry, S.V., Margreth, A. & Adelstein, R.S., eds), Elsevier, Amsterdam, 39-49.
- Clainche CL and Carlier MF. (2008). Regulation of Actin Assembly Associated with Protrusion and Adhesion in Cell Migration. *Physiol Rev* 88, 489-513.
- Clough SJ and Bent AF. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* 16, 735-743.
- Clough SJ. (2005). Floral dip: agrobacterium-mediated germ line transformation. *Methods Mol Biol* 286, 91–102.
- Deal RB, Topp CN, McKinney EC, Meagher RB. (2007). Repression of flowering in *Arabidopsis* requires activation of FLOWERING LOCUS C expression by the histone variant H2A.Z. *Plant Cell* 19, 74–83.
- Dominguez R. (2009). Actin filament nucleation and elongation factors – structure function Relationships. *Crit Rev Biochem Mol Bio* 44, 351-366.
- Dominguez R and Holmes KC. (2011). Actin Structure and Function. *Annu Rev Biophys* 40, 169-186.

- dos Remedios CG, Chhabra D, Kekic M, Dedova IV, Tsubakihara M, et al. (2003). Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol Rev* 83, 433–473.
- Fan T, Zhai H, Shi W, Wang J, Jia H, Xiang Y, An L. (2013). Overexpression of profilin 3 affects cell elongation and F-actin organization in *Arabidopsis thaliana*. *Plant Cell Rep* 32, 149–60.
- Ferron F, Rebowski G, Lee SH, Dominguez R. (2007). Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP. *EMBO J* 26, 4597–606.
- Frazier JA and Field CM. (1997). Actin cytoskeleton: are FH proteins local organizers? *Curr Biol* 7, R414–417.
- Gilliland LU, Kandasamy MK, Pawloski LC, Meagher RB. (2002). Both vegetative and reproductive actin isoforms complement the stunted root hair phenotype of the *Arabidopsis act2-1* mutation. *Plant Physiol* 130, 2199–2209.
- Glover BJ. (2000). Differentiation in plant epidermal cells. *J Exp Bot* 51, 497–505.
- Haffner C, Jarchau T, Reinhard M, Hoppe J, Lohmann SM, Walter U. (1995). Molecular cloning, structural analysis and functional expression of the proline-rich focal adhesion and microfilament-associated protein VASP. *EMBO J* 14, 19–27.
- Harbeck B, Hüttelmaier S, Schluter K, Jockusch BM, Illenberger S. (2000). Phosphorylation of the vasodilator-stimulated phosphoprotein regulates its interaction with actin. *J Biol Chem* 275, 30817–25.
- Honore B, Madsen P, Andersen AH, Leffers H. (1993). Cloning and expression of a novel human profilin variant, profilin II. *FEBS Lett* 330, 151–155.
- Huang S, McDowell JM, Weise MJ, Meagher RB. (1996). The *Arabidopsis* Profilin Gene Family: Evidence for an Ancient Split between Constitutive and Pollen-Specific Profilin Genes. *Plant Physiology* 111, 115–126.
- Jegou A, Niedermayer T, Orban J, Didry D, Lipowsky R, Carlier MF, Romet-Lemonne G. (2011). Individual actin filaments in a microfluidic flow reveal the mechanism of ATP hydrolysis and give insight into the properties of profilin. *PLoS Biol* 9, e1001161.
- Jeong YM, Mun JH, Lee I, Woo JC, Hong CB, Kim SG. (2006). Distinct Roles of the First Introns on the Expression of *Arabidopsis* Profilin Gene Family Members. *Plant Physiology* 140, 196–209.

- Jockusch BM, Murk K, Rothkegel M. (2007). The profile of profilins. *Rev Physiol Biochem Pharmacol* 159, 131-149.
- Kandasamy MK, Deal RB, McKinney EC, Meagher RB. (2005). Silencing the nuclear actin-related protein AtARP4 in *Arabidopsis* has multiple effects on plant development, including early flowering and delayed floral senescence. *Plant Journal* 41, 845-58.
- Kandasamy MK, McKinney EC, Meagher RB. (1999). The late pollen specific actins in angiosperms. *Plant Journal* 18, 681–691.
- Kandasamy MK, McKinney EC, Meagher RB. (2002). Plant Profilin Isovariants Are Distinctly Regulated in Vegetative and Reproductive Tissues. *Cell Motility and the Cytoskeleton* 52, 22-32.
- Kandasamy MK, Burgos-Rivera B, McKinney EC, Ruzicka DR, Meagher RB. (2007). Class-Specific Interaction of Profilin and ADF Isovariants with Actin in the Regulation of Plant Development. *The Plant Cell* 19, 3111–3126.
- Kandasamy MK, McKinney EC, Meagher RB. (2010). Differential sublocalization of Actin variants within the nucleus. *Cytoskeleton* 67, 729-743.
- Kandasamy MK, McKinney EC, Roy EM, Meagher RB. (2012). Plant vegetative and animal cytoplasmic actins share functional competence for spatial development with protists. *Plant Cell* 24, 2041-2057.
- Krom N and Ramakrishna W. (2008). Comparative Analysis of Divergent and Convergent Gene Pairs and Their Expression Patterns in Rice, *Arabidopsis*, and *Populus*. *Plant Physiology* 147, 1763-1773.
- Krysan PJ, Young JC, Sussman MR. (1999). T-DNA as an Insertional Mutagen in *Arabidopsis*. *Plant Cell* 11, 2283-2290.
- Kudryashov DS and Reisler E. (2012). ATP and ADP Actin States. *Biopolymers* 99, 245-256.
- Livak KJ and Schmittgen TD. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* 25, 402–408.
- McKenna ST, Vidali L, Hepler PK. (2004). Profilin inhibits pollen tube growth through actin-binding, but not poly-L-proline binding. *Planta* 218, 906–915.
- McKinney EC, Ali N, Traut A, Feldmann KA, Belostotsky DA, McDowell JM, Meagher RB. (1995). Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants act2-1 and act4-1. *Plant J* 8, 613–622.

- McKinney EC, Kandasamy MK, Meagher RB. (2001). Small Changes in the Regulation of One Arabidopsis Profilin Isovariant, PRF1, Alter Seedling Development. *Plant Cell* 13, 117-1191.
- Meagher RB. (1995). The impact of historical contingency on gene phylogeny: Plant actin diversity. *Evolutionary Biology* 28, 195-215.
- Meagher RB, McKinney EC, Vitale AV. (1999). The evolution of new structures: Clues from plant cytoskeletal genes. *Trends Genet* 15, 278-284.
- Meagher RB and Williamson RE. (1994). *The Plant Cytoskeleton*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 1049-1084.
- Minehardt TJ, Kollman PA, Cooke R, Pate E. (2006). The open nucleotide pocket of the profilin/actin x-ray structure is unstable and closes in the absence of profilin. *Biophys J* 90, 2445-2449.
- Miralles F and Visa N. (2006). Actin in transcription and transcription regulation. *Curr Opin Cell Biol* 18, 261-66.
- Mullins RD, Kelleher JF, Xu J, Pollard TD. (1998). Arp2/3 complex from *Acanthamoeba* binds profilin and crosslinks actin filaments. *Mol Biol Cell* 9, 841-852.
- Obermann H, Raabe I, Balvers M, Brunswig B, Schulze W, Kirchhoff C. (2004). Novel testis-expressed profilin IV associated with acrosome biogenesis and spermatid elongation. *Molecular Human Reproduction* 11, 53-64.
- O'Farrell PH. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem.* 250, 4007-4021.
- Pantaloni D and Carlier MF. (1993). How profilin promotes actin filament assembly in the presence of thymosin b4. *Cell* 75, 1007-1014.
- Pawloski LC, Kandasamy MK, Meagher RB. (2006). The late pollen actins are essential for normal male and female development in *Arabidopsis*. *Plant Mol Biol* DOI 10.1007/s11103-006-9063-5.
- Pei W, Du F, Zhang Y, He T, Ren H. (2012). Control of the actin cytoskeleton in root hair development. *Plant Sci* 187, 10-8.
- Perrin BJ and Ervasti JM. (2010). The Actin Gene Family: Function Follows Isoform. *Cytoskeleton* 67, 630-634.
- Pollard TD and Borisy GG. (2003). Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112, 453-65.

- Ramachandran S, Christensen H, Ishimaru Y, Dong CH, Chao-Ming W, Cleary AL, ChuaNH. (2001). Profilin Plays a Role in Cell Elongation, Cell Shape Maintenance, and Flowering in *Arabidopsis*. *Plant Physiology* 124, 1637–1647.
- Schlüter K, Jockusch BM, Rothkegel M. (1997). Profilins as regulators of actin dynamics. *Biochem Biophys Acta* 1359, 97-109.
- Schutt CE, Myslik JC, Rozycki MD, Goonesekere NC, Lindberg U. (1993). The structure of crystalline profilin-beta-actin. *Nature* 365, 810-816.
- Sohn RH and Goldschmidt-Clermont PJ. (1994). Profilin: at the crossroads of signal transduction and the actin cytoskeleton. *Bioessays* 16, 465-472.
- Sohn RH, Chen J, Koblan KS, Bray PF, Goldschmidt-Clermont PJ. (1995). Localization of a binding site for phosphatidylinositol 4, 5-bisphosphate on human profilin. *J Biol Chem* 270, 21114–21120.
- Staiger CJ, Yuan M, Valenta R, Shaw PJ, Warn RM, Lloyd CW. (1994). Microinjected profilin affects cytoplasmic streaming in plant cells by rapidly depolymerizing actin microfilaments. *Curr Biol* 4, 215–219.
- Stradal TE, Rottner K, Disanza A, Confalonieri S, Innocenti M, Scita G. (2004). Regulation of actin dynamics by WASP and WAVE family proteins. *Trends Cell Biol* 14, 303–311.
- Taka T, Pechan T, Richter H, Muller J, Cek C, Bohm N, Obert B, Ren H, Miehaus K, Amaj JS. (2011). Proteomics on Brefeldin A-Treated *Arabidopsis* Roots Reveals Profilin 2 as a New Protein Involved in the Cross-Talk between Vesicular Trafficking and the Actin Cytoskeleton. *Journal of Proteome Research* 10, 488-501.
- Tian M, Chaudhry F, Ruzicka D, Meagher R, Staiger C, Day B. (2009). *Arabidopsis* Actin-Depolymerizing Factor AtADF4 Mediates Defense Signal Transduction Triggered by the *Pseudomonas syringae* Effector AvrPphB1. *Plant Physiology* 150, 815-824.
- Vidali L, McKenna ST, Hepler PK. (2001). Actin polymerization is essential for pollen tube growth. *Mol Biol Cell* 12, 2534–2545.
- Voigt B, Timmers AC, Samaj J, Muller J, Baluska F, Menzel D. (2005). GFP-FABD2 fusion construct allows *in vivo* visualization of the dynamic actin cytoskeleton in all cells of *Arabidopsis* seedlings. *European Journal of Cell Biology* 6, 595-608.
- Wang F, Jing Y, Wang Z, Mao T, Samaj J, Yuan M, Ren H. (2009). *Arabidopsis* profilin

- isoforms, PRF1 and PRF2 show distinctive binding activities and subcellular distributions. *J Integr Plant Biol* 51, 113-21.
- Wang YS, Motes CM, Mohamalawari DR, Blancaflor EB. (2004). Green fluorescent protein fusions to *Arabidopsis* fimbrin 1 for spatio-temporal imaging of F-actin dynamics in roots. *Cell Motil Cytoskeleton* 59, 79–93.
- Williamson RE. (1993). Organelle movements. *Annu RevPlant Physiol Plant Mol Biol* 44, 181–202.
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. (2007). An electronic fluorescent pictograph browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* 2, e718.
- Witke W, Podtelejnikov AV, Di Nardo A, Sutherland JD, Gurniak CB, et al. (1998). In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly. *EMBO J* 17, 967–976.
- Yarar D, Waterman-Storer CM, Schmid SL. (2007). SNX9 couples actin assembly to phosphoinositide signals and is required for membrane remodeling during endocytosis. *Dev Cell* 13, 43–56.
- Yarmola EG and Bubb MR. (2006). Profilin: emerging concepts and lingering misconceptions. *TRENDS in Biochem. Sciences* 31, 197-205.
- Yarmola EG and Bubb MR. (2009). How depolymerization can promote polymerization: the case of actin and profilin. *BioEssays* 31, 1150–1160.
- Zheng Y, Xin H, Lin J, Liu CM, Huang S. (2012). An *Arabidopsis* class II formin, AtFH19, nucleates actin assembly, binds to the barbed end of actin filaments, and antagonizes the effect of AtFH1 on actin dynamics. *J Integr Plant Biol* 54, 800-13.

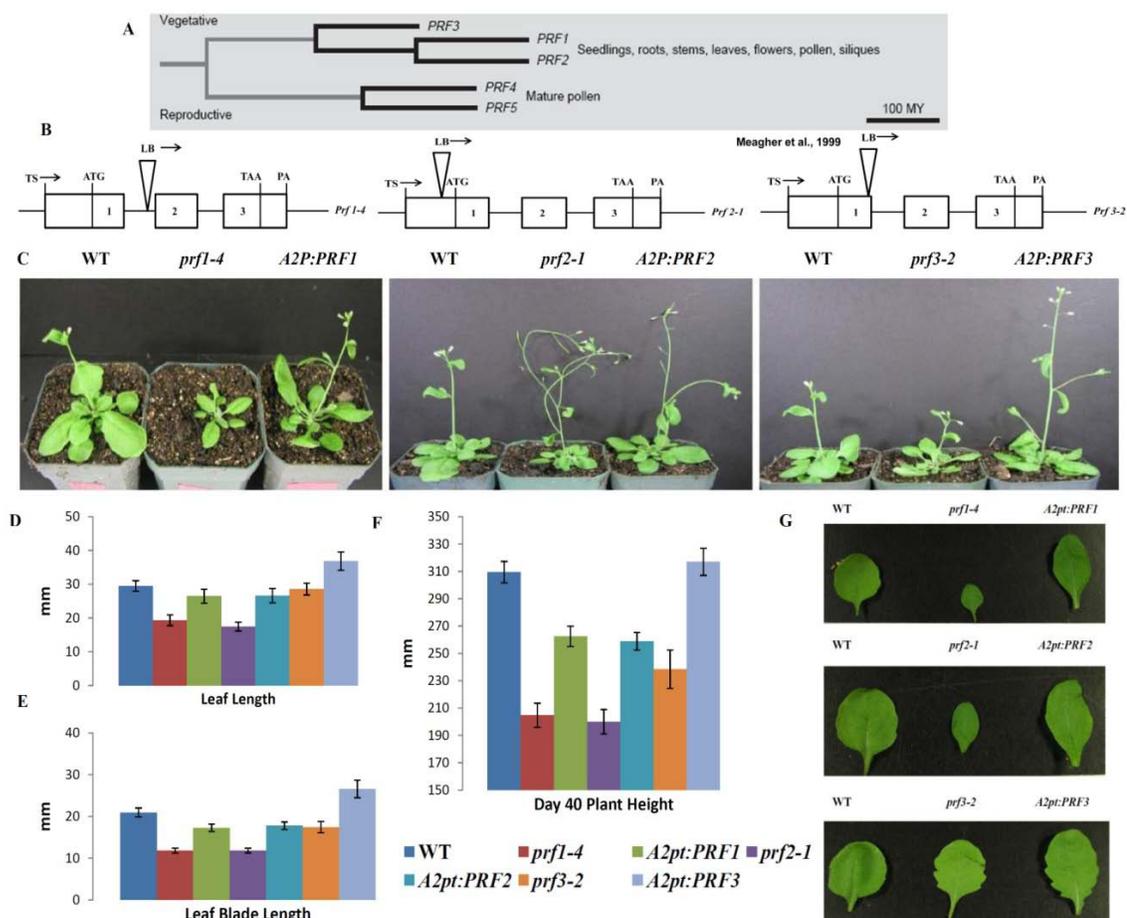


Figure 2.1: Analysis of mutants defective in individual vegetative profilins. **A)** Phylogeny of the profilin gene family in *Arabidopsis thaliana* based on amino acid sequence divergence (adapted from Meagher et al., 1999). **B)** Schematic drawings indicating the location of each T-DNA insertion in mutant plants *prf1-4*, *prf2-1*, and *prf3-2*. **C)** Visualization of adult plant morphological phenotypes of profilin T-DNA mutants, wild type (WT), and each mutant (*prf1-4*, *prf2-1*, *prf3-2*) complemented with the appropriate transgene (*A2P:PRF1*, *A2:PRF2*, and *A2:PRF3*, respectively). Pictures were taken 4 weeks (4w) after seed germination. **D)** Leaf length for single vegetative PRF T-DNA mutants and their complemented lines. **E)** Leaf blade length for single vegetative

PRF T-DNA mutants and their complemented lines. **F)** Mature plant height for single vegetative PRF T-DNA mutants and their complemented lines. **G)** 4 week (4w) leaf morphology pictures of individual leaves from single PRF T-DNA mutants, WT, and complement lines. Leaf measurements (D and E) were taken on day 28 (4w) (n=52) following seed germination, while plant height measurements were taken on day 40 (~5 ½ w) (n=30). All measurements are in mm. Error bars represent +/- 1 SD.

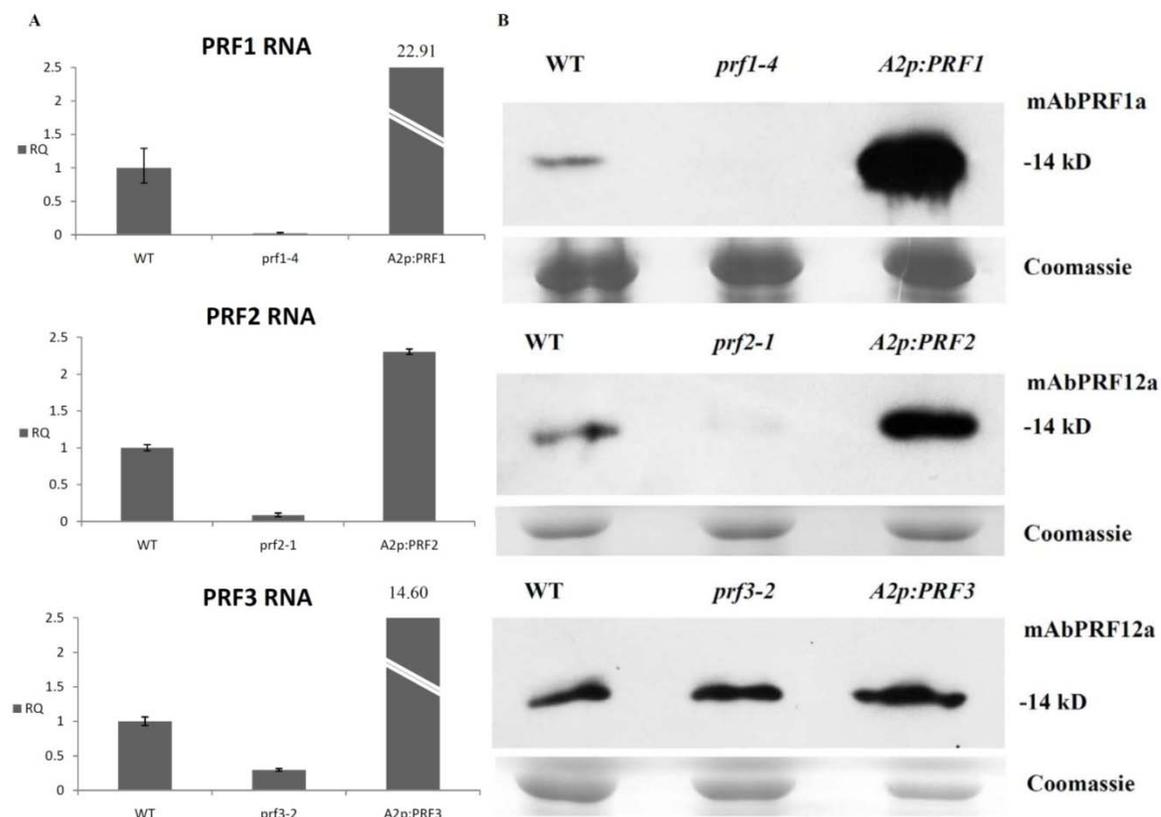


Figure 2.2: Analysis of profilin RNA and protein expression for vegetative PRF single T-DNA mutants and complement lines. **A)** The Relative Quantities (RQ) of *PRF1* RNA for WT, *prf1-4*, and *A2p:PRF1* plants were determined by quantitative Real Time PCR (qRT-PCR). The RQ of *PRF2* RNA for WT, *prf2-1*, and *A2p:PRF2* plants and the RQ of *PRF3* RNA for WT, *prf3-2*, and *A2p:PRF3* plants are also shown. Error bars represent +/- 1 SD. **B)** Profilin protein expression was examined by Western blot analysis: WT, *prf1-4*, and *A2p:PRF1* plants using the PRF1 specific monoclonal antibody mAbPRF1a; WT, *prf2-1*, and *A2p:PRF2* plants using the PRF1 and PRF2 specific monoclonal antibody mAbPRF12a; WT, *prf3-2*, and *A2p:PRF3* plants using the PRF1 and PRF2 specific monoclonal antibody mAbPRF12a, which also has weak affinity for PRF3. Coomassie stained gels showing rubisco protein expression are located beneath each blot to show equal loading across lanes. All samples were taken from 4w old leaf tissue.

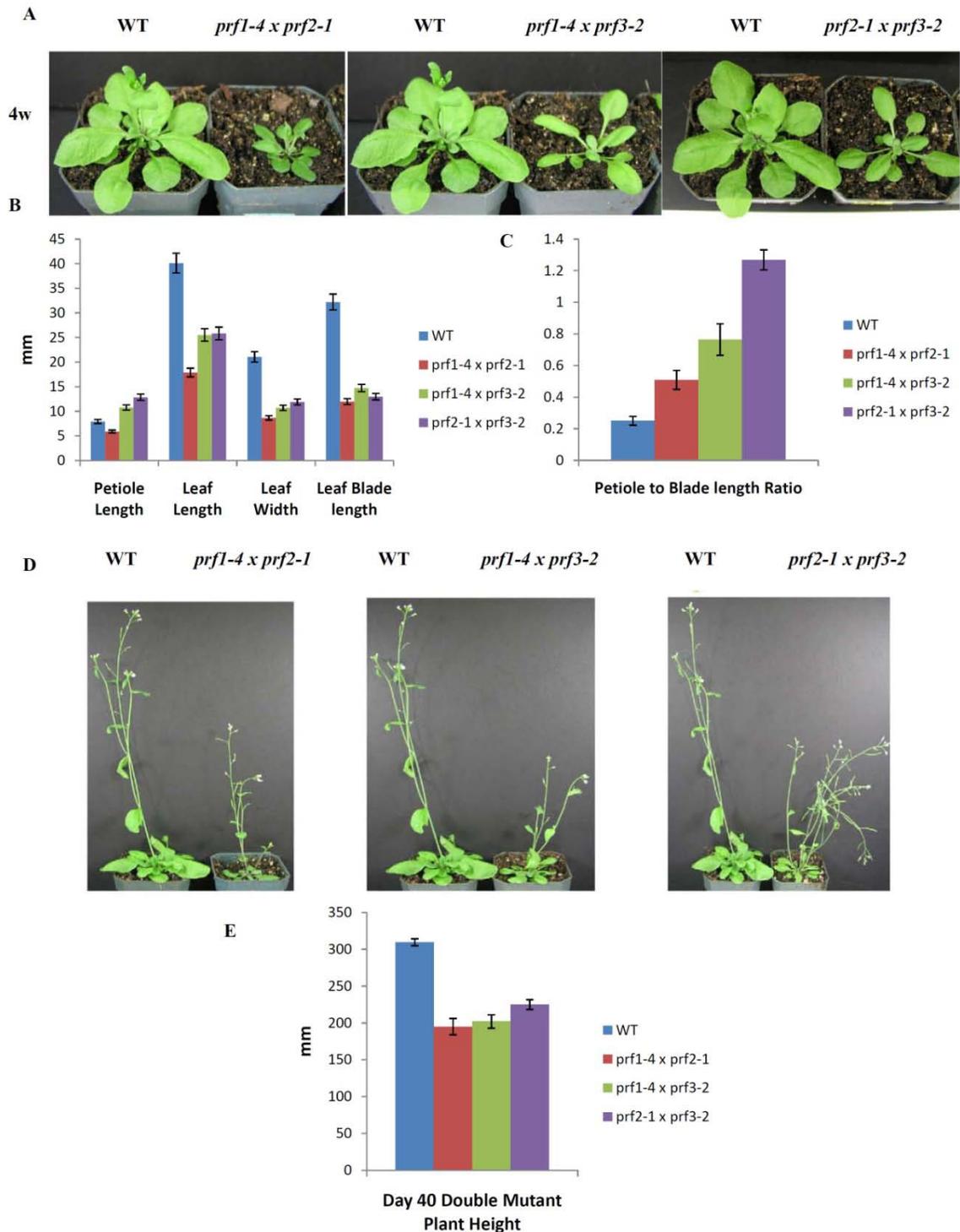


Figure 2.3: Morphological analysis of vegetative profilin double mutants. A)

Visualization of morphological phenotypes observed for profilin double T-DNA mutants.

Pictures were taken at 4 weeks (4w) after seed germination. **B)** Petiole length, leaf

length, leaf width, and leaf blade length for double vegetative PRF T-DNA mutants. **C)** Petiole to leaf blade length ratio for double PRF T-DNA mutants. **D)** Pictures of double mutant plants showing morphological phenotypes at 5 weeks (5w). **E)** Mature plant height for double mutants. Leaf measurements were taken on day 28 (4w, n=52 for each measurement), while plant height measurements were taken on day 40 (~5 ½ w, n=30). All measurements are in mm. Error bars represent +/- 1 SD.

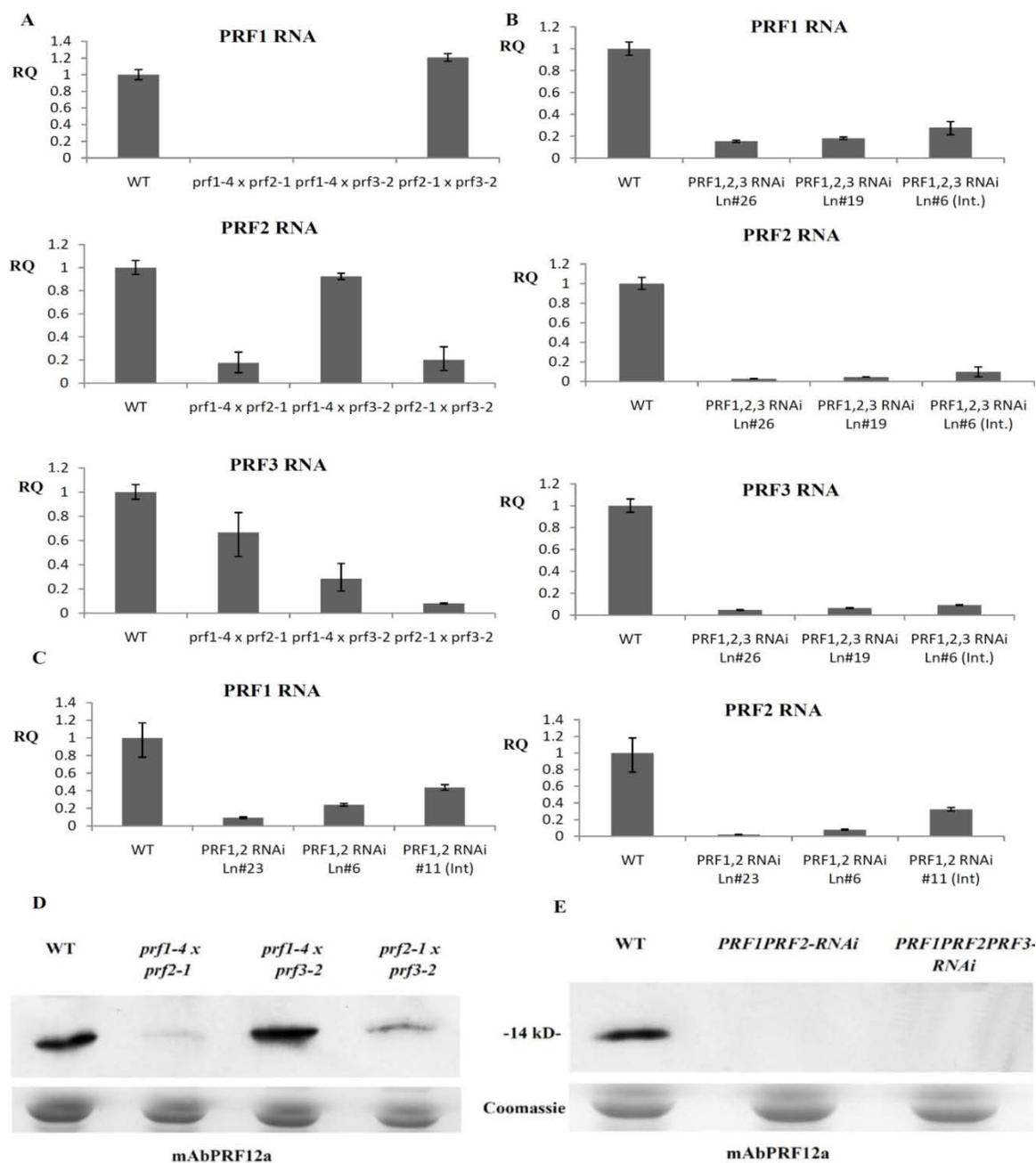


Figure 2.4: qRT-PCR data and western blot analysis for double and triple mutant/RNAi lines. *Transcript expression (A, B, & C)* **A)** qRT-PCR data for T-DNA double mutants. Each graph shows the RQ of *PRF1*, *PRF2*, or *PRF3* expression levels for each of the T-DNA double mutants. **B)** qRT-PCR data for *PRF1PRF2PRF3*-RNAi lines

(three lines shown). Each graph shows the RQ of *PRF1*, *PRF2*, or *PRF3* expression levels for each of the *PRF1PRF2PRF3*-RNAi lines. **C)** qRT-PCR data for *PRF1PRF2*-RNAi lines (three lines shown). Each graph shows the RQ of either *PRF1* or *PRF2* expression levels for each of the *PRF1PRF2*-RNAi lines. Error bars represent +/- 1 SD.

Protein expression (D & E) **D)** Western analysis of protein levels in profilin double mutants (all three combinations) using the PRF1 and PRF2 specific monoclonal antibody mAbPRF12a. **E)** Western analysis of *PRF1PRF2*-RNAi and *PRF1PRF2PRF3*-RNAi lines using the PRF1 and PRF2 specific monoclonal antibody mAbPRF12a. Coomassie stained gels showing rubisco protein expression are located beneath each blot to show equal loading across lanes. All samples were taken from 4w old leaf tissue.

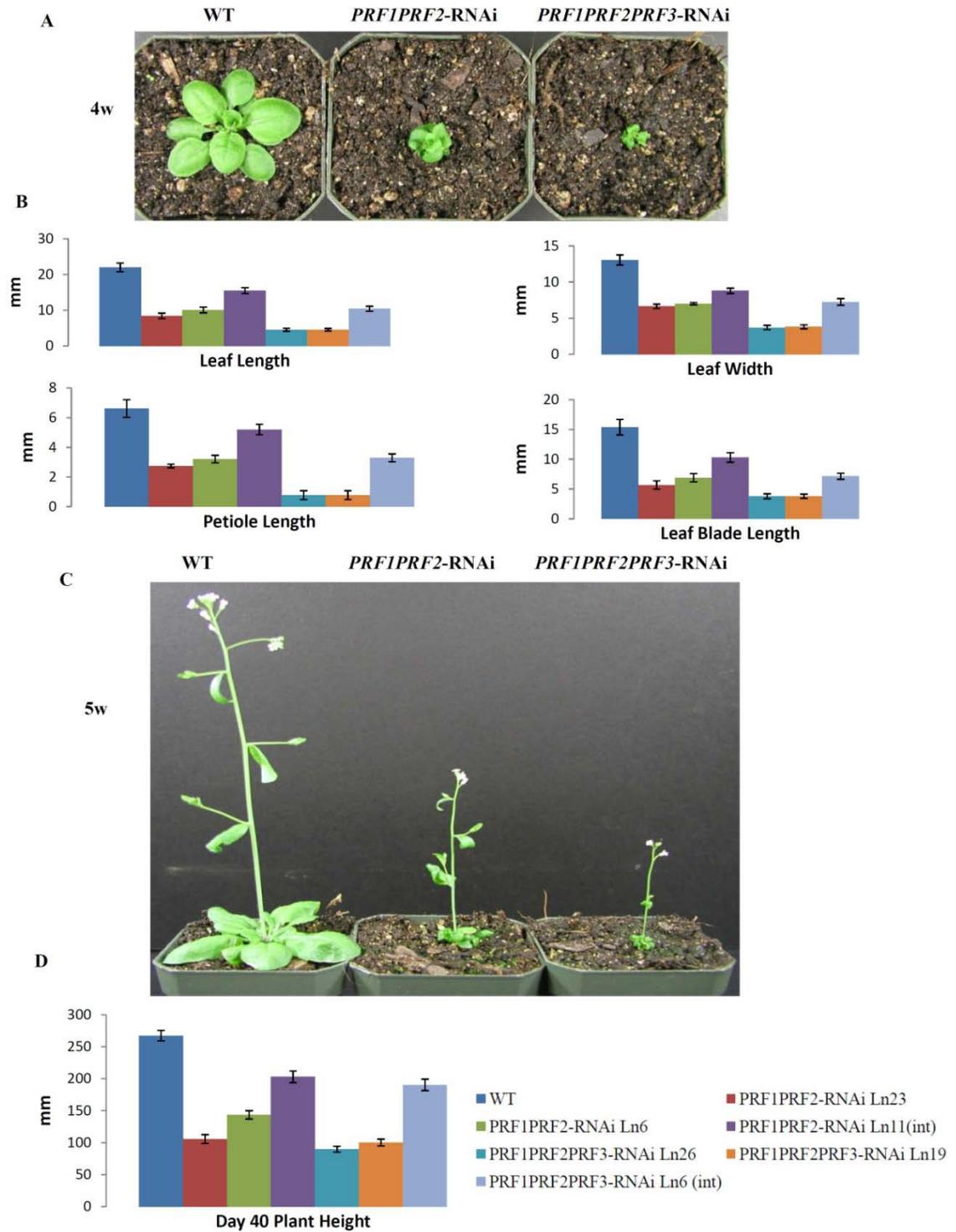


Figure 2.5: Morphological analysis of PRF double and triple RNAi lines. A) The morphology of PRF double and triple RNAi lines (lines silenced for *PRF1* and *PRF2* and

for *PRF1-3*, respectively) were examined 4 weeks (4w) post-germination. **B)** Leaf length, leaf width, petiole length, and leaf blade length for PRF double and triple RNAi lines. **C)** PRF double and triple RNAi plants show severe morphological phenotypes at 5 weeks (5w). **D)** Mature plant height for double and triple RNAi lines. Leaf measurements were taken on day 28 (4w) during development (n=52), while plant height measurements were taken on day 40 (~5 ½ w, n=30). All measurements are in mm. Error bars represent +/- 1 SD.

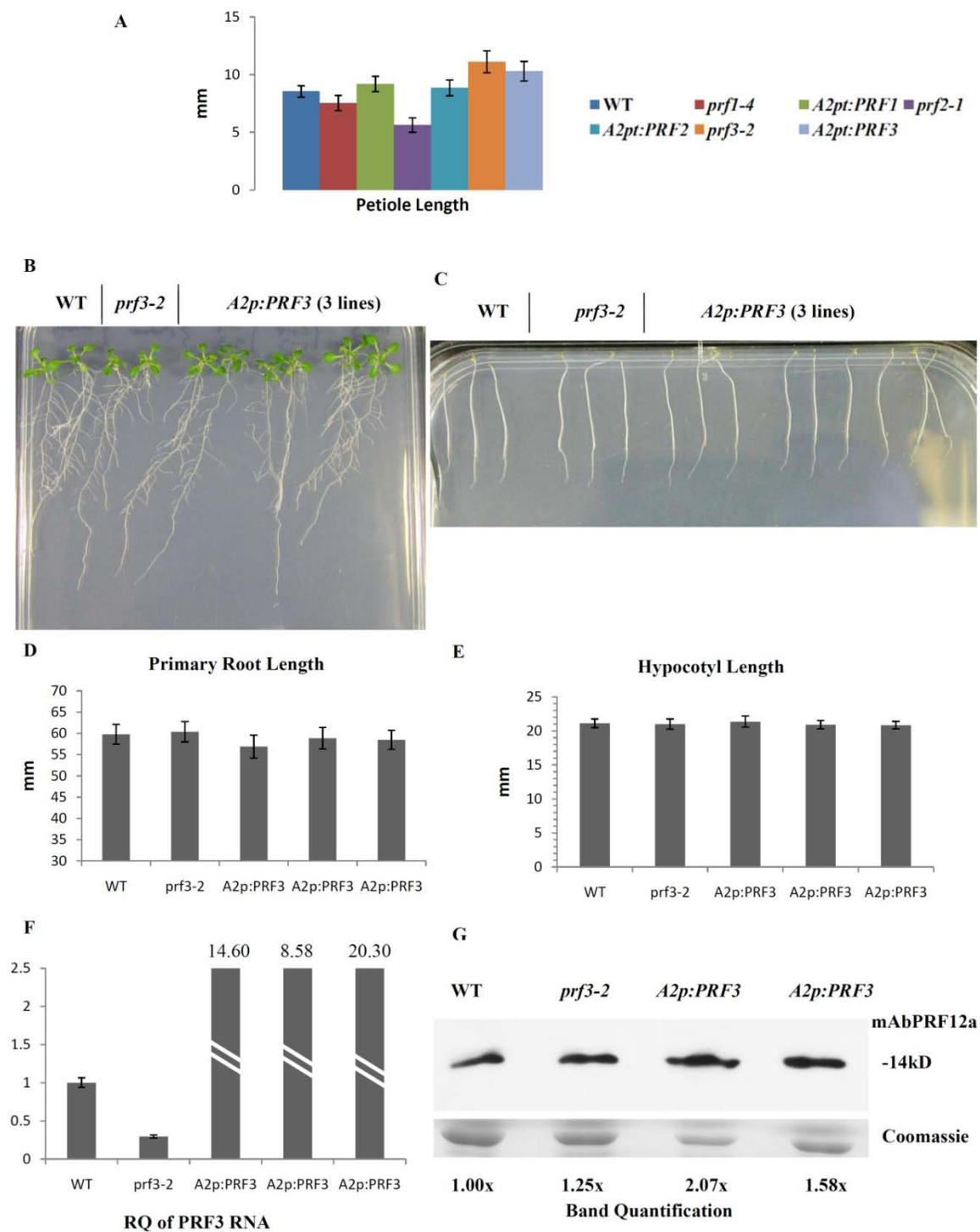


Figure 2.6: PRF3 knockout displays elongated petioles, while PRF3 overexpression results in no phenotypic effects. A) Petiole length for vegetative PRF T-DNA mutant

prf3-2 and its complement lines. **B)** Day 14 primary root length comparison between WT, *prf3-2*, and three independent *A2p:PRF3* overexpression lines. **C)** Day 10 hypocotyl length comparison between WT, *prf3-2*, and three independent *A2p:PRF3* overexpression lines. **D)** Quantification of primary root length measurements from lines pictured in **B**. **E)** Quantification of hypocotyl length measurements from lines pictured in **C**. **F)** qRT-PCR data on the relative quantity of PRF3 RNA expression for WT, *prf3-2*, and three independent *A2p:PRF3* overexpression lines. **G)** Western blot analysis for WT, *prf3-2*, and two independent *A2p:PRF3* overexpression lines. Western blot bands were quantified using the ImageJ software. All samples were taken from 4w old leaf tissue. Leaf measurements (A) were taken on day 28 (4w) during development, primary root length measurements (D) were taken on day 15, and hypocotyl length measurements (E) were taken on day 10. All measurements' are in mm. Leaf measurements were generated with a sample of n=52, while root and hypocotyls measurements have an n=30. Error bars represent +/- 1 SD.

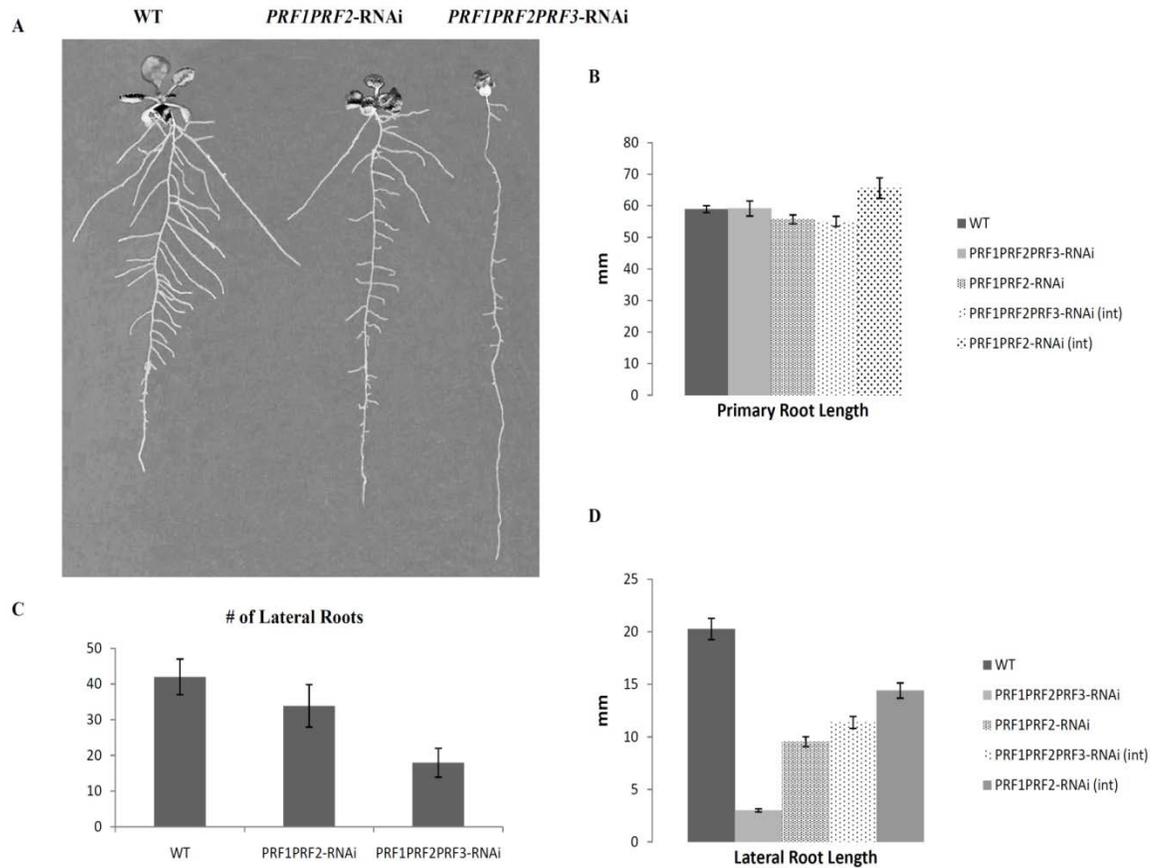
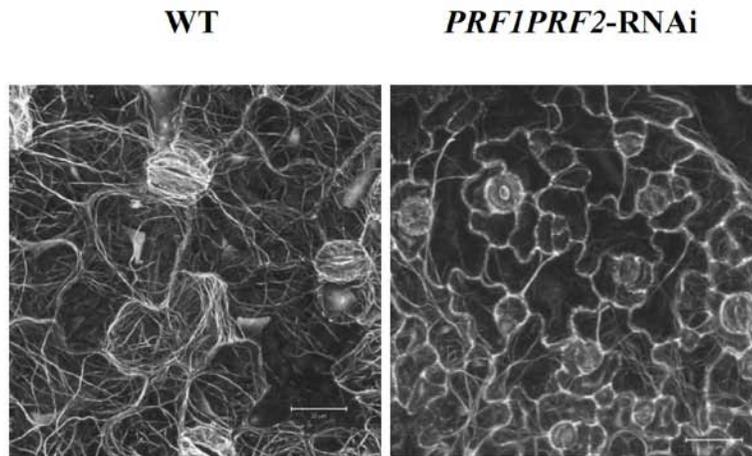


Figure 2.7: Vegetative profilin double and triple RNAi lines show defects in lateral root formation and growth. **A)** Visualization of defects in root development for PRF double and triple RNAi lines. **B)** Quantification of primary root length. **C)** Quantification of the number of lateral roots formed/ initiated. **D)** Quantification of lateral root length. Pictures and measurements were taken on day 15 of development. Sample size was 30 (n=30) and error bars represent +/- 1 SD.

A



B

Organization of actin filaments in the leaf cells of PRF 1/2 RNAi plants

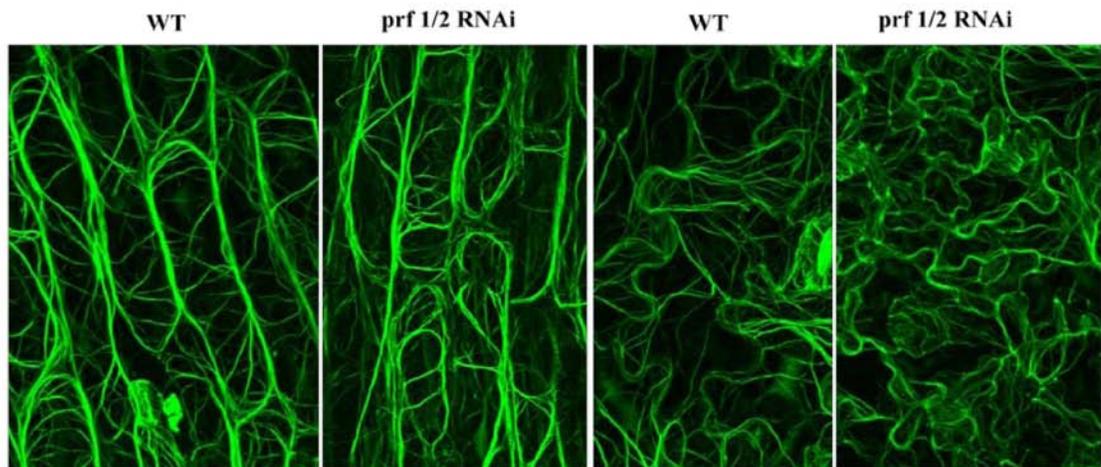


Figure 2.8: Cell size and F-actin filament organization in *PRF1PRF2*-RNAi plants.

A) Confocal images showing the organization of actin filaments (*35S:GFP-FABD2*) and cell size in wild-type (WT) and *PRF1PRF2*-RNAi leaf epidermal cells at 4 weeks (4w).

The pavement cells and stomatal guard cells of the upper leaf side are shown. Size bar

represents 20 μm . **B)** Confocal images revealing the organization of actin filaments in wild-type (WT) and *PRF1PRF2*-RNAi upper leaf epidermal cells at 4 weeks (4w). The first two images are cells at the vein region and the last two are pavement cells.

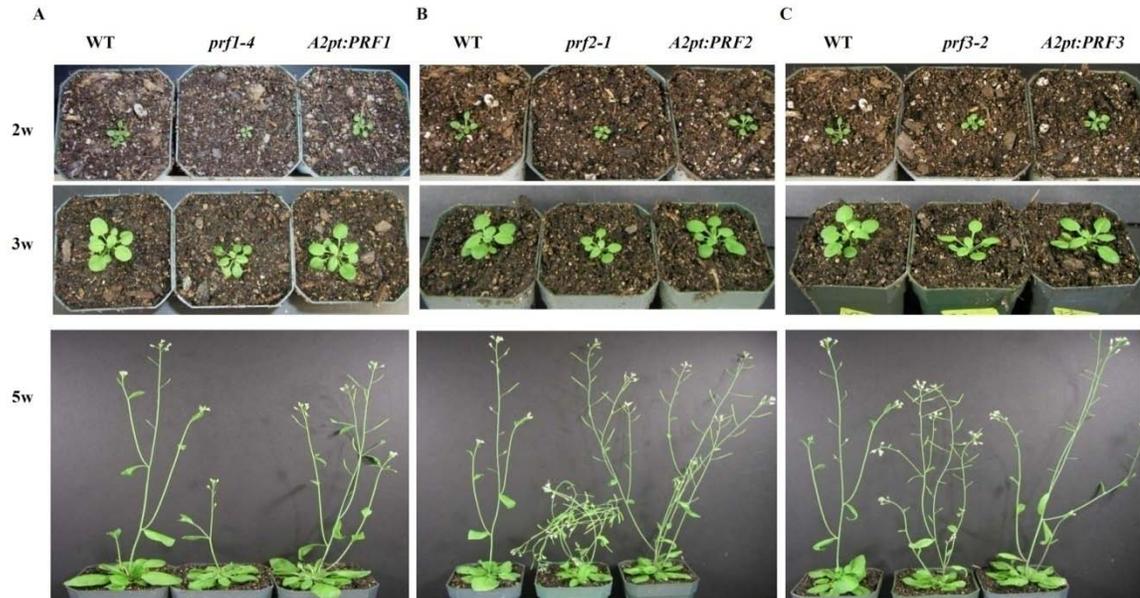


Figure 2.S1: Morphology of vegetative profilin single mutants. Visualization of the morphological phenotypes seen in profilin single T-DNA mutants at 2 weeks (2w), 3 weeks (3w), and 5 weeks (5w) post germination. **A)** WT, *prf1-4*, and *A2p:PRF1* complemented plants across development. **B)** WT, *prf2-1*, and *A2p:PRF2* complemented plants across development. **C)** WT, *prf3-2*, and *A2p:PR3* complemented plant across development.

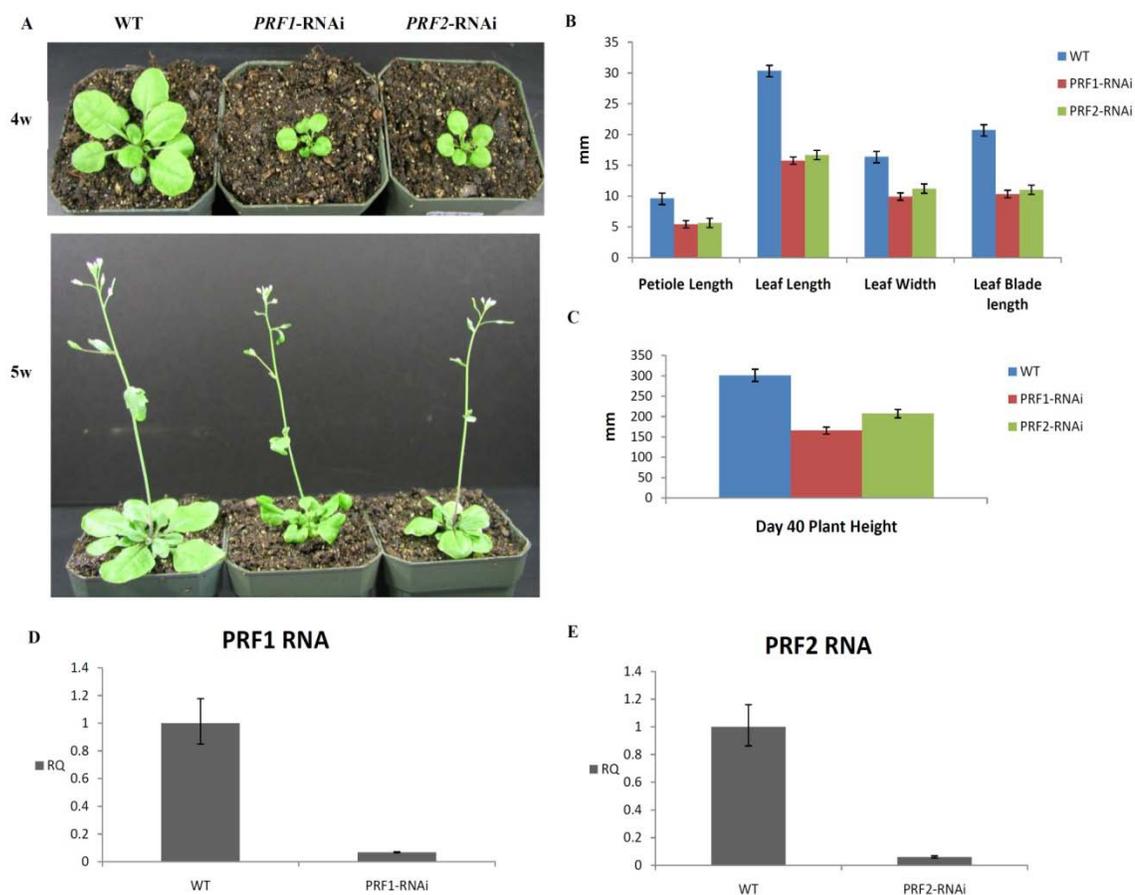


Figure 2.S2: Morphology and qRT-PCR analysis of transcript levels for vegetative PRF single RNAi lines. **A)** Morphological phenotypes of *PRF1*-RNAi and *PRF2*-RNAi lines across development at 4 weeks (4w) and 5 weeks (5w) post germination. **B)** Quantification of petiole length, leaf length, leaf width, and leaf blade length for *PRF1*-RNAi and *PRF2*-RNAi lines. **C)** Quantification of mature plant height for *PRF1*-RNAi and *PRF2*-RNAi lines. Leaf measurements were taken on day 28 (4w) during development (n=52), while plant height measurements were taken on day 40 (~5 ½ w, n=30). All measurements are in mm. **D)** qRT-PCR data representing the RQ of *PRF1* RNA for WT and *PRF1*-RNAi. **E)** qRT-PCR data representing the RQ of *PRF2* RNA for WT and *PRF2*-RNAi. Error bars represent +/- 1 SD.

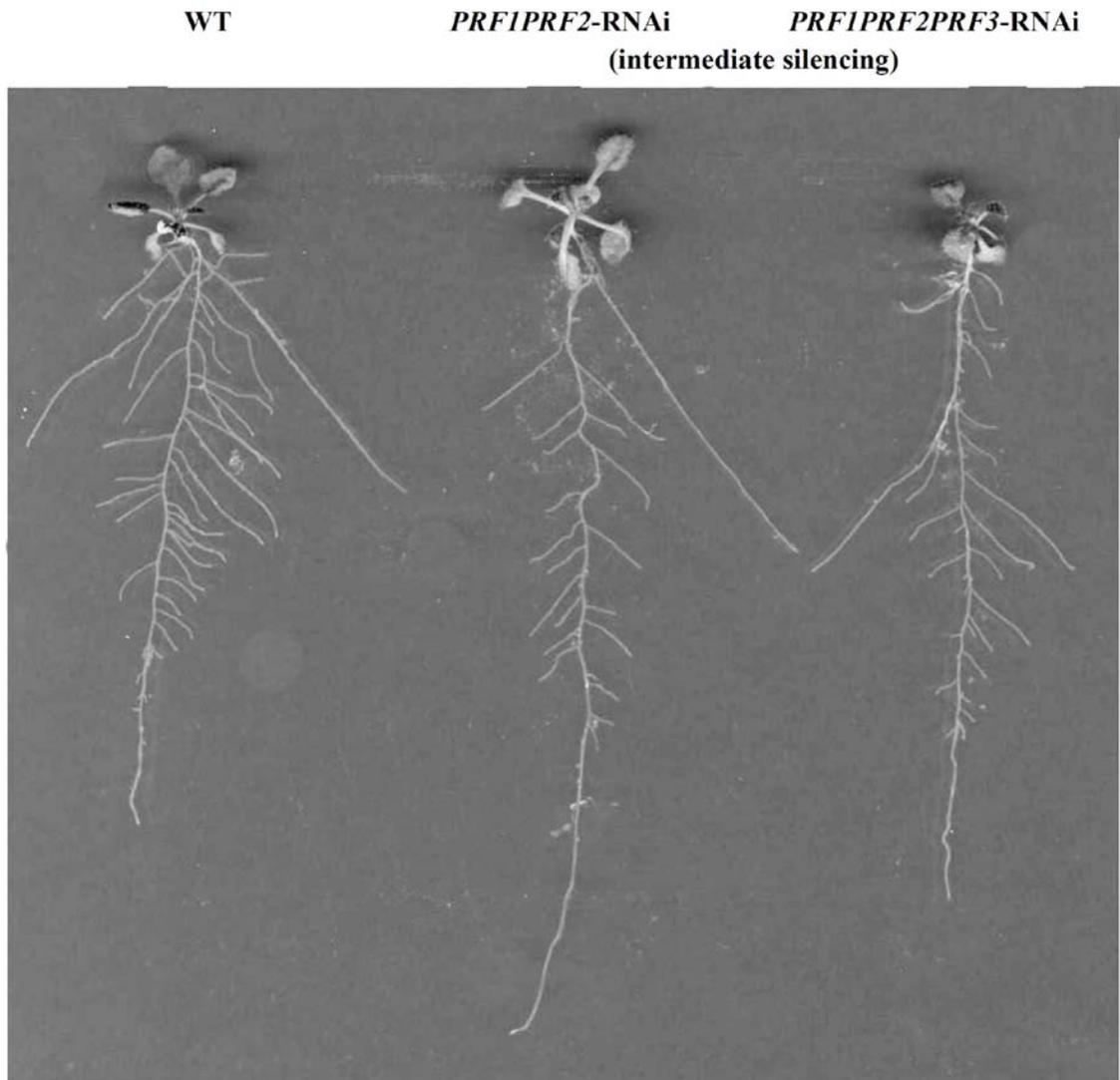


Figure 2.S3: Vegetative PRF double and triple RNAi lines that are only weakly silenced for profilin RNA expression show slight defects in lateral root development. Visualization of slight defects in root development for PRF double and triple RNAi lines with intermediate silencing (~40% of WT levels). Pictures were taken 15 days after seed germination. Measurements can be seen in **Figure 7B** and **7D**.

	Day 28 Petiole Length (mm)	Day 28 Leaf Length (mm)	Day 28 Leaf Width (mm)	Day 28 Leaf Blade Length (mm)	Petiole to Blade Length Ratio	Day 40 Plant Height (mm)
WT (single T-DNA mutants)	8.97	35.65	17.91	26.67	0.417	309.42
<i>prf1-4</i>	7.54	19.34	9.93	11.80	0.664	204.77
<i>A2p:PRF1</i>	9.95	30.69	12.16	20.74	0.569	262.46
<i>prf2-1</i>	5.63	17.45	10.36	11.82	0.485	200
<i>A2p:PRF2</i>	10.85	34.57	13.52	23.71	0.526	258.85
<i>prf3-2</i>	11.11	28.53	13.35	17.41	0.668	238.35
<i>A2p:PRF3</i>	10.12	31.49	14.83	21.37	0.404	316.92
WT (RNAi lines)	9.63	30.38	16.40	20.75	0.472	301.08
<i>PRF1 RNAi</i>	5.42	15.75	9.91	10.33	0.540	165.44
<i>PRF2 RNAi</i>	5.64	16.66	11.20	11.01	0.520	207.08
WT (double T-DNA mutants)	7.92	40.12	21.07	32.21	0.250	309.53
<i>prf1-4 prf2-1</i>	5.88	17.87	8.66	11.98	0.509	195.00
<i>prf1-4 prf3-2</i>	10.77	25.50	10.70	14.72	0.764	202.05
<i>prf2-1 prf3-2</i>	12.84	25.82	11.91	12.98	1.269	225.05
<i>PRF1PRF2- RNAi Ln#23</i>	2.74	8.40	6.64	5.66	0.509	105.52
<i>PRF1PRF2- RNAi Ln#6</i>	3.21	10.08	6.99	6.87	0.477	143.28
<i>PRF1PRF2- RNAi Ln#11 (int)</i>	5.20	15.48	8.79	10.28	0.516	202.84
<i>PRF1PRF2PRF3- RNAi Ln#26</i>	0.78	4.55	3.70	3.78	0.211	89.76
<i>PRF1PRF2PRF3- RNAi Ln#19</i>	0.78	4.56	3.81	3.78	0.209	100.12
<i>PRF1PRF2PRF3- RNAi Ln#6 (int)</i>	3.29	10.43	7.25	7.14	0.465	190.28

Table 2.S1: Summary of all phenotypic measurements for the various mutant and RNAi silenced plant lines examined.

CHAPTER 3

PHYLOGENETIC IDENTIFICATION OF INHERITED PATTERNS OF
NUCLEOSOMAL HISTONE MODIFICATION¹

¹Kristofer J. Müssar, Xiaoyu Zhang, and Richard B. Meagher. To be submitted to the Journal of Epigenetics and Chromatin.

Abstract

This paper explores the hypothesis that nucleosomal histone modifications are conserved following gene duplication by focusing on the sequence specific localization of histone marks within defined gene family members. We selected four cytoskeletal gene families that comprise a range of sequence divergence. Three of them (actin, profilin, actin-depolymerizing factor (ADF)) have paired family members representing a well-established, recent genome-wide duplication event ~25-40 MYA within brassicaceae, while the origin of most of the members in the fourth family, Actin-Related Proteins (ARPs), date back to common ancestry with protists. Using chromatin immunoprecipitation and high-resolution whole-genome tiling microarray (ChIP-chip) data, we examined 14 different histone post translational modification (PTM) marks in *Arabidopsis thaliana* for specific enrichment patterns. We found subsets of data supporting and rejecting this hypothesis, with recently duplicated gene pairs exhibiting varying levels of conservation across their histone modification enrichment profiles. We discuss the idea that epigenetic controls aid “evolution by gene duplication” by silencing some recent gene duplicates, but not others, until beneficial mutations and subfunctionalization can occur. By looking for correlations among enrichment profiles for each gene family, we also detected distinct combinatorial patterns of histone modification marks. These patterns include, H3K36me3; H3K4me3; H4K8ac for actin gene members, H3K36me3; H3K4me3; H4K8ac; H3K9ac; H3K14ac for profilin gene members, and H3K36me3; H4K8ac for ADF gene members. Interestingly, the ancient and highly sequence divergent ARP gene family did not contain any of these

combinatorial patterns, suggesting that sequence may be playing some role in facilitating the epigenomic landscape of histone post-translational modifications.

Introduction

Epigenetics has been defined in a multitude of ways over the past 50 years (Haig, 2004). For the sake of clarity, we will be discussing epigenetics as “somatic cell or transgenerationally inherited changes in gene function that are unable to be explained by the classical central dogma of molecular genetics” (Meagher and Müssar, 2012, Nanney, 1958). We have previously explored the hypothesis that DNA sequence is facilitating epigenetic control for transgenerationally inherited epigenome changes controlling phenotypes (Meagher & Mussar, 2012). While we found evidence that sequence may be guiding epigenetic processes like nucleosome positioning and cytosine methylation (5meC), existing genome-wide data on histone PTMs generally did not support this hypothesis. However, it is believed that histone PTMs may be inherited through somatic cell duplication leading to strong effects on gene expression and phenotype. Therefore, in this study we continue to search for a link between DNA sequence and histone PTM deposition by analyzing PTM enrichment profiles among recently duplicated genes and exploring sequence’s role in guiding combinatorial PTM interactions.

Histones undergo a process known as posttranslational modification (PTM), which alters their interactions with DNA and other nuclear proteins. The N- and C-terminal amino acid residues (e.g., Lys, Arg, Ser, Thr, Tyr) of most histone polypeptides may undergo modifying processes like methylation (mono-, di-, or tri-), acetylation (mono-, di-, or tri-), phosphorylation, SUMOylation, ubiquitination, glycosylation, citrullination, and ADP-ribosylation (Strahl and Allis, 2000). These covalent

modifications influence DNA/protein interactions thereby affecting gene expression, DNA repair, and chromosome condensation. Collectively, these histone PTMs are thought to help facilitate changes to chromatin structure and subsequently recruit specific transcriptional regulators or factors involved in recombination and repair. The idea that histone PTMs are controlling transcriptional regulation constitutes what is known as the “histone code hypothesis” (Jenuwein and Allis, 2001). However, due to the several dozen different types of modifications and the multitude of combinatorial interactions, deciphering the exact mechanism behind the influence of histone PTMs on gene expression is often obscure (Zhang, 2010, Bannister and Kouzarides, 2011).

There have been many studies trying to depict how specific histone modifications interact with gene expression and protein recruitment. For example, H3K27me3 has been found to be associated with many tissue-specific genes and is a hallmark of transcriptional repression (Cao et al., 2002, Zhang et al., 2007). CHIP-chip analysis has revealed that there is a correlation between H3K4me3 and transcriptional activation, which occurs through the recruitment of histone acetyltransferases and the transcriptional pre-initiation complex (Taverna et al., 2006, Vermeulen et al., 2007, Hung et al., 2009). H3K4me2 is also involved in gene activation, but in this case it has been linked to the recruitment of the Set3 histone deacetylase complex (Lippman et al., 2004, Kim and Buratowski, 2009). So it is obvious that different modifications are operating through various mechanisms in order to achieve the same end goal, transcriptional regulation. A summary of the 14 different types of histone modifications used in this analysis, their associated effects, and estimated locations can be found in **Table 3.1**.

Currently most studies investigating histone modification interactions are being performed on a genome-wide scale. This is ideal for identifying significant genomic level interactions, but can lead to copious amounts of noise/bias that can mask or cause one to overlook specific unapparent or atypical trends (Pearson, 2008). In order to further dissect how various PTMs correlate with each other, more focused studies need to be initiated. We believe that histone modification profiles from individual genes and gene families need to be compared in order to gain additional insight into these overarching genomic patterns. Through the understanding of these interactions on the individual gene level, we can begin to specifically define the steps by which these mechanisms operate. Some gene families consist of gene duplicates that arise during evolution to gain additional functions or specificities. Their sequences tend to be fairly conserved. By comparing PTM profiles among recently duplicated gene pairs, we can examine if sequence is facilitating the deposition of these various combinatorial modification patterns, perhaps in an attempt to aid evolutionary mechanisms in controlling spatial and temporal gene expression. Comprehending how these PTMs operate within a gene family can then be extrapolated to explain what is seen throughout other regions of the genome.

Zheng, 2008 uses a variation of this gene-specific approach by analyzing original and derived loci in human segmental duplications (SDs). The study looked at 20 different histone modifications across 1,646 non-redundant pairs of genomic regions (SD pairs). The analysis showed that there were obvious asymmetries among histone modification profiles between original and derived (duplicated) loci (Zheng, 2008). The study discovered that the parental loci of these SDs had much higher levels of histone

modifications than in derivative loci. Interestingly, they also found that these asymmetries tended to increase with the age of the segmental duplication event, but they failed to continue their analyses down to an examination of the individual gene sequences within these SDs.

In contrast, recent studies in yeast have shown that duplicate gene pairs, on average, share more common histone modification patterns than random singleton pairs (Zou et al., 2012). Like Zheng, 2008, they found that histone modification profiles between duplicated genes begin to demonstrate even more deviation as sequence divergence increases (Zou et al., 2012). Since these studies were performed in yeast, which lack other epigenetic control mechanisms like DNA methylation, it is thought that histone modifications play a more major role. These analyses were performed on all gene duplicates found throughout the yeast genome, and failed to identify specific individual examples. In this paper, we have performed a more focused examination of gene families in *Arabidopsis* and more particularly an analysis of well characterized gene duplicates within these families.

It is well documented that there were three separate whole-genome duplication events over the past ~350 million years in the ancestry of the *Arabidopsis thaliana* genome, with the most recent occurring ~25-40 million years ago (MYA) (McDowell et al., 1996, Vision et al., 2000, Simillion et al., 2002, Bowers et al., 2003, Blanc et al., 2003, Mun et al., 2009). These events directly gave rise to a greater than 90% increase in transcription factors, signal transducers, and developmental genes in higher plants (Maere et al., 2005). To determine how histone modifications are inherited following gene duplication, we focused most of our analysis on gene families that contained

representatives of this most recent duplication event and are well characterized in terms of their functions, spatial and temporal expression patterns, and evolutionary divergence. The actins, actin depolymerizing factors (ADFs), and profilins are cytoskeletal gene families that fit these criteria and exhibit varying levels of sequence divergence among family members (McDowell et al., 1996, Huang et al., 1996, McKinney et al., 2002, Ruzika et al., 2007). We will be using the actin related protein (ARP) gene family as a potential negative control. The origin of most of its members date back to protists, there are no documented gene pairs, they have high levels of nucleotide sequence divergence (e.g., >70%), and they seemed the least likely to have evolutionarily conserved patterns of histone modifications that would correlate with DNA sequence (**Table 3.2**).

Using ChIP-chip analysis for the 14 modifications summarized in **Table 3.1** (7 histone methylations and 7 histone acetylations), we constructed PTM enrichment profiles for each modification type for every gene family member and found that recently duplicated gene pairs are mostly distinct in their epigenomic properties. They exhibit varying levels of conservation across their histone modification enrichment profiles, suggesting that epigenetic controls aid “evolution by gene duplication” by silencing some recent gene duplicates, but not others, until beneficial mutations and subfunctionalization can occur (Rodin and Riggs, 2003). By looking for correlations among these PTM enrichment profiles for each gene family, we detected distinct combinatorial patterns of histone modification marks. These patterns include, H3K36me₃; H3K4me₃; H4K8ac for actin gene members, H3K36me₃; H3K4me₃; H4K8ac; H3K9ac; H3K14ac for profilin gene members, and H3K36me₃; H4K8ac for ADF gene members. Interestingly, the ancient and highly sequence divergent ARP gene family did not contain any of these

combinatorial interactions, suggesting that sequence may be playing some role in facilitating the epigenomic landscape of histone post-translational modifications.

Results

The genome wide distribution of 14 histone PTMs previously reported for two week old *Arabidopsis* above ground leaf tissue were examined at the gene family level. We constructed 14 histone PTM enrichment profiles for all of the members in the *Arabidopsis* actin, profilin, ADF, and ARP gene families. Each profile encompassed the genes entire coding region, and was extended 500 bp upstream of the transcriptional start site (TSS) and 500 bp downstream of the transcriptional termination site (TTS) in order to ensure the inclusion of proximal promoter and enhancer information. These profiles indicate the presence or absence of each of the 14 modifications shown in **Table 3.1**. **Figure 3.1** presents examples of three enrichment profiles and how we have deemed individual genes to be enriched (yes) or not enriched (no) for any one particular histone PTM. Using these profiles we searched for potentially conserved patterns of nucleosomal histone modifications within gene families and among recently duplicated gene pairs. A detailed summary of these results are given in **Tables 3.S1 to 3.S4**.

Gene Pair Analysis

Actin gene family (low DNA sequence divergence)

Actins are highly conserved, abundant, and multifunctional cytoskeletal proteins that construct polymeric filaments and filament bundles to control spatial development in eukaryotic cells and organs (Kudryashov and Reisler, 2012). In *Arabidopsis thaliana*, there are 9 actin genes, with only one (ACT9) being a non-expressed and highly diverged pseudogene (McDowell et al., 1996). Evolutionary relationships among these 9 actin

variants based on nucleotide sequence divergence can be seen in **Figure 3.S1**.

Phylogenetic analysis indicates that there are three duplicated gene pairs in this family:

ACT1 and ACT3, ACT4 and ACT12, ACT2 and ACT8 remaining from the most recent genome-wide duplication event in Brassicaceae ~25 to 40 MYA (McDowell et al., 1996,

Mun et al., 2009). The first two pairs are mostly pollen specific and not expressed in the young leaf tissue examined, while the third pair is constitutively expressed in all

vegetative tissues including young leaves. From **Table 3.S1**, we see that ACT1 and

ACT3 are enriched for similar histone modifications. They share the same outcome (presence or absence of a certain modification type) for 11 out of the 14 modifications

examined (**Table 3.3**). Examples of how their enrichments compare for H3K36me3 and

H4K8ac can be seen in **Figure 3.2**. The only differences are that ACT1 is enriched for

H3K9ac (activation mark), while ACT3 is enriched for H3K27me3 and H4K16ac (both repressive marks). These two pollen specific actins are both thought to be equally and

very poorly expressed in these tissues during this time of development (**Table 3.S1**), but this data would suggest that ACT1 is more likely being expressed, while ACT3

repressed. Nevertheless, we still see much agreement between the histone modification

profiles among this gene pair. The two other pollen specific actins, ACT4 and ACT12,

share the same enrichment outcomes for 9 out of 14 PTMs. ACT4 appears to be enriched for more modifications than ACT12 (5 vs. 2), which could have something to do with it

being expressed ~10 times higher at this developmental time point, despite some of these modifications being linked to transcriptional repression. Unlike the previous gene pairs,

the two vegetative constitutive actins (ACT2 and ACT8) that are strongly expressed in

these tissues are not enriched for any of the same histone PTMs. The only similarities

amongst this pair are that they both lack H3K36me1, H3K27me3, H3K5ac, and H4K12ac enrichment. For the other 10 histone PTMs, these two genes have opposite patterns of enrichment (**Table 3.3**). These 10 modifications are found to be enriched in ACT2 and absent in ACT8. There are no histone PTMs present in ACT8 that are absent in ACT2. ACT2 is a very highly expressed gene and is several fold more highly expressed than ACT8, but ACT8 is still a well expressed gene in these tissues. Hence, the reasons for this large difference are not at all clear. So for the three gene pairs in the actin gene family, we see varying results ranging from near complete agreement among modification enrichment to almost complete disagreement.

Profilin gene family (moderate DNA sequence divergence)

Profilins are small, ubiquitously expressed, monomeric actin binding proteins (ABPs) that sequester monomeric actin and promote actin filament dynamics (Jockusch et al., 2007). In *Arabidopsis thaliana*, there are 5 profilin genes. Evolutionary relationships among these 5 profilin variants can be seen in **Figure 3.S2**. Based on the low percent of silent nucleotide substitutions (43% to 45% synonymous substitutions) we were able to detect two duplicated gene pairs in this family: PRF1 and PRF2, PRF4 and PRF5, with their divergent times consistent with their duplication during the most recent genome wide duplication of ~25 to 40 MYA (Huang et al, 1996, Mun et al., Vision et al., 2000, Simillion et al., 2002, Bowers et al., 2003, Blanc et al., 2003). PRF1 and PRF2 are vegetative constitutively expressed genes with PRF2 being expressed at ~2.5-fold higher levels than PRF1. Using the data from **Table 3.S2**, we see that PRF1 and PRF2 show many discrepancies between their enrichment profiles (**Table 3.3**). For all 7 modifications enriched in PRF2, they are absent in PRF1. As with ACT2 and ACT8, we

notice an inverse relationship in modification deposition among this gene pair. However, the two pollen specific profilins (PRF4 and PRF5) that are very poorly expressed in the tissues being examined contain the same enrichment outcomes for nearly all 14 PTMs, with the exception of H4K16ac, which is present in PRF4 and absent in PRF5 (**Table 3.3**). An example of how PRF4 and PRF5 compare in their H3K4me1 enrichment can be seen in **Figure 3.3**. In summary, we see an extremely wide range of results in the conservation of the histone modification landscape for these recently duplicated gene pairs.

ADF gene family (moderate to high DNA sequence divergence)

Actin depolymerizing factors (ADFs) are small ABPs that stabilize and/or sever F-actin filaments in an ADF concentration dependent manner (Bamburg, 1999, Andrianantoandro and Pollard, 2006). There are 11 known ADF genes in *Arabidopsis thaliana*. **Figure 3.S3** shows the evolutionary relationships among these 11 ADF genes. Phylogenetic analysis indicates that there are four gene pairs remaining from the recent genome wide duplication in *Arabidopsis* ancestry (Ruzicka et al., 2007): ADF1 and ADF4, ADF7 and ADF10, ADF8 and ADF11, ADF5 and ADF9. Using the data in **Table 3.S3**, we can again see that the vegetative constitutively expressed ADF1 and ADF4 gene pair is exhibiting inverse enrichments for their histone modifications (**Table 3.3**). They differ in the enrichment of 9 of these 14 modifications. H3K4me3 is the only modification type for which both genes are enriched (**Figure 3.4A** and **3.4C**). This is also the case for the root and trichome specific ADF8 and ADF11 gene pair, which differ in the presence of 7 histone PTMs and are both enriched for only two of the same modifications. However, the other two ADF gene pairs exhibit strikingly similar histone

modification enrichment patterns. Pollen specific ADF7 and ADF10 share 11/14 outcomes for enrichment, while ADF5 and ADF9, which are weakly expressed in the meristem and other fast growing tissues, share 12/14 (**Table 3.3**). An example showing H3K4me2 enrichment in ADF7 and ADF10 can be seen in **Figure 3.4B** and **3.4D**. Once more we see this range for gene pairs that have varying degrees of agreement among their histone modification enrichments.

Detection of Combinatorial Patterns within and among Cytoskeletal Gene Families

We explored the possibility that the presence of any one of the 14 histone modification types influenced the presence or absence of another. We calculated the Pearson product-moment correlation coefficient (Pearson's r) by converting our summary data tables (**Tables 3.S1-3.S4**) to binary variables (see Materials and Methods).

Pearson's r will quantitatively score a histone modification type that influences the presence (positive correlations, 0.0 to +1.0) or absence (negative correlations, 0.0 to -1.0) of another histone PTM. This allows us to examine how different modifications interact or cooperate with each other in order to modulate transcription. We used two-tailed significance tests to determine p -values, and only report interactions between PTMs (correlation sets) that have corresponding p -values that are less than 0.05. To interpret the size of the correlation coefficients, we used the criteria and cut-offs discussed in Cohen, 1988 and Choudhury, 2009 (see Materials and Methods).

Pearson's r calculations for the actin gene family only revealed a total of three, strong, statistically significant PTM correlation sets out of a possible 91 (see Materials and Methods) (**Table 3.S5**). The size of these correlation coefficients ranged from +0.791-1.000. However, these three correlation sets contained interactions between the

same three histone modifications, H3K36me3, H3K4me3, and H4K8ac (**Table 3.4**). All three of these modifications are thought to help arrange chromatin for transcriptional activation. Interestingly, there were no negative correlation sets detected in this analysis. Meaning that for this family, the enrichment of certain modifications only acted cooperatively. Thus, among members of the *Arabidopsis* actin gene family, the presence of one of these three modifications correlates with the presence of the other two PTMs (and vice versa), in an effort to help prepare chromatin for activating gene expression. Examples of enrichment profiles for ACT7 showing the presence and location of the three PTMs in this combinatorial pattern can be seen in **Figure 3.5**. The complete correlation table for the actin gene family is represented in **Table 3.S5**.

Extending this analysis to the profilin gene family revealed a total of 10 statistically significant correlation sets out of a possible 91 (**Table 3.S6**). Notably, all of these correlations were a perfect positive 1.00. Just as with the actin gene family, the profilin analysis exposed five specific modifications all acting cooperatively with each other. This combinatorial pattern consisted of the following modifications: H3K36me3, H3K4me3, H4K8ac (the same three seen in the actin pattern), H3K9ac and H3K14ac (**Table 3.4**). All five modifications are associated with transcriptional activation (**Table 3.1**), with the presence of one being linked to causing a greater likelihood of the other four being enriched for a given gene in this family. Examples of enrichment profiles for PRF2 showing the presence and location of the five PTMs in this combinatorial pattern can be seen in **Figure 3.6**. The complete correlation table for the profilin gene family is represented in **Table 3.S6**.

Moving onto the ADF gene family, Pearson's r analysis indicated that there are 20 total statistically significant correlation sets out of a possible 91 (**Table 3.S7**). This is by far the highest number of correlation sets seen in any of the gene families examined. Perhaps this is not surprising since the ADF gene family contains the most members (11). All 20 of these correlations were deemed strong, positive correlations with coefficients ranging from +0.624-1.000. Unlike the iterative interactions seen between the histone modifications that correlated in the actin and profilin gene families, we only found two modifications that showed this relationship in the ADFs, H3K36me3 and H4K8ac (**Table 3.4**). Coincidentally, these interactions are also identified within the actin and profilin patterns. This suggests that H3K36me3 and H4K8ac are clearly exhibiting a positive influence on whether or not the other will be present throughout these three cytoskeletal gene families (likely in an attempt to activate transcription). **Figure 3.7** shows H3K36me3 and H4K8ac enrichment profiles for ADF4. The complete correlation table for the ADF gene family can be seen in **Table 3.S7**.

We also performed this analysis on the anciently and highly divergent Actin Related Protein (ARP) gene family. There are 8 members in the *Arabidopsis thaliana* ARP gene family. A phylogenetic tree showing their relationships and levels of sequence divergence is presented in **Figure 3.S4**. The ancestry of ARP2, 3, 4, 5 and 6 all predate the divergence of algae and plants from protist ancestors. There no recently duplicated gene pairs among the eight *Arabidopsis* ARPs, with all genes being constitutively expressed in nearly all tissues and cell types (McKinney et al., 2002, Mathur et al., 2003, Meagher et al., 2009). Due to the lack of sequence conservation, we will be using the ARPs as a type of negative control for this analysis.

Pearson's r analysis for the ARPs found a total of five, strong, statistically significant correlation sets out of a possible 91 (**Table 3.S8**), with coefficients ranging from ± 0.745 to ± 1.000 . Interestingly, three of the five interactions detected were negative. These are the first and only negative interactions noticed in our study. Unlike what we saw for the actin, profilin, and ADF gene families, there were no combinatorial interactions identified among the ARPs (**Table 3.4**). All five correlation sets involved modifications that were completely independent of each other. In fact, out of the 10 possible interacting PTMs (5 correlation sets \times 2 PTMs per set = 10 interacting PTMs), 9 are different, with only H3K4me2 being involved in multiple interactions. These results may suggest that these combinatorial patterns only exist in gene families where sequence is more highly conserved, implying that sequence is facilitating these interactions. The complete correlation table for the ARP gene family can be seen in **Table 3.S8**.

After performing these analyses on each gene family individually, we combined all 33 genes (thus increasing our sample size) and searched for the same types of interacting patterns. The results of this analysis can be seen in **Table 3.S9**. We detected a total of 23 statistically significant correlation sets, with all interactions being positive correlations indicating that they are all cooperative in nature. 9 of the 23 correlation sets were deemed strong (coefficients ranging from $+0.527$ to 0.782), while the other 14 moderate (coefficients ranging from $+0.383$ to 0.487). However, we did not detect the exact same combinatorial patterns as in the individual gene family analyses. Instead we noticed a pattern containing H3K9ac, H3K14ac, H3K23ac, H4K8ac, H4K16ac, and H3K36me3 (**Table 3.4**). Interestingly, five of the six modifications detected in these interactions are acetylations (H3K36me3 was the only methylation), including the

repressive mark H4K16ac, which recruits deacetylases to silence transcription. The other five PTMs in this pattern are associated with activating transcription.

In an attempt to understand the inclusion of H4K16ac amongst these combinatorial patterns, further analysis was performed by removing the ARPs from the overall dataset and generating new correlation coefficients for the actins, profilins, and ADFs combined. Nonetheless, we detected the same 23 interactions as previously stated, but improved the strength of these correlations to 15 strong and 8 moderate (**Table 3.S10**). In addition to the 14 histone modifications analyzed in this study, we also included data for 5-methyl cytosine (5meC) DNA methylation (Zhang et al., 2006). However, we did not find any of the 14 PTMs to correlate/ interact with the presence or absence of DNA methylation in any of our analyses (**Table 3.S5-3.S10**).

Discussion

Genome-wide histone modification deposition studies have been ongoing for some time now, with one of its main purposes being to decipher the role of the complex “histone code” on gene expression (Jenuwein and Allis, 2001, Bannister and Kouzarides, 2011). It is popular thought that we can decipher how these various modification patterns are facilitating gene regulation through chromatin preparation by understanding how these different histone PTMs are interacting with one another (Rando, 2012, Linghu et al., 2013). We believe that taking a genome-wide approach could be masking or distorting some of the findings that are occurring on the individual gene level. For this reason we felt that a more focused, smaller scale approach to analyzing this problem could shed additional light on the overall assumptions that have previously been made. Our analysis takes advantage of genome-wide ChIP-chip data on 14 different histone

post-translational modifications to perform a study on PTM deposition among individual gene families in Arabidopsis (Zhang et al., 2006, 2007, and 2009).

There were two separate parts to our analysis. First we examined histone modification enrichment profiles between duplicated gene pairs in the actin, profilin, and ADF gene families. Second, we performed statistical analysis on the presence or absence of each modification type within an among these gene families to search for combinatorial patterns that may exist to help facilitate the preparation of chromatin in an effort to modulate transcription. These patterns could serve as data points in the attempt to decode the overall histone code.

Analysis of Duplicated Gene Pairs

By examining how recently duplicated gene pairs compare in the overall deposition of various histone modification types, we can try to understand if and how evolutionary forces act on epigenetic control mechanisms to assist in regulating different duplicated genes for spatial and temporal development. Detailed analyses of the molecular processes that guide the evolution and regulation of duplicated sequences have yet to be extensively explored. Previous studies have systematically looked at 20 histone methylation modifications in whole segmental duplications (SDs) in mammals (Zheng, 2008). This study was not focused on specific expressed genes but on recent and essentially random SDs. They found no definitive evidence for conservation of histone PTMs following SD. In fact, they discovered that most derivative (or duplicated) loci differed significantly from the original loci with respect to many histone PTMs. They also noticed that these asymmetries increased with the age of the duplication event. By looking at PTM patterns for four specific segmental duplication pairs, they saw a clear

variable range of agreement among histone modification enrichments between different SD pairs. Two of the pairs exhibited inverse enrichment patterns, one showed conserved enrichment, and the last contained an intermediate level of PTM conservation. These results suggest that histone modifications are impacting the local chromatin environments to discriminate against duplicated loci. We take this analysis one step further by looking at specific duplicated gene pairs, and found a similar varying range of results as was seen in Zheng, 2008.

Arabidopsis' recent, well-documented genome wide duplication event 25 to 40 MYA (Vision et al., 2000, Simillion et al., 2002, Bowers et al., 2003, Blanc et al., 2003), as well as their extensive and widely accessible epigenomics databases make them a great model for this study. As previously stated, examination of PTM enrichment profiles for three well characterized cytoskeletal gene families with known sets of duplicated gene pairs (actins, profilins, and ADFs) revealed a varying range of results, but with more support for conserved PTM profiles after duplication. Out of the nine recently duplicated gene pairs analyzed (Table 3.3); we see four pairs [ACT1/ACT3, PRF4/PRF5, ADF7/ADF10, and ADF5/ADF9] that exhibit near exact enrichment landscapes for the same histone modification types. Three of the nine gene pairs [ACT2/ACT8, PRF1/PRF2, and ADF1/ADF4] show an almost inverse relationship in their enrichment profiles, meaning that when one gene has a certain histone modification present, the other gene in that pair does not. Interestingly, examples of these types of pairs are found in all three gene families analyzed. In fact, these three pairs all represent vegetative constitutively expressed genes, suggesting that these more highly expressed gene pairs need to have vastly different PTM compositions in order to guide proper spatial and

temporal expression. Finally, the other two gene pairs [ACT4/ACT12 and ADF8/ADF11] show an intermediate of the aforementioned two results, with certain modifications matching up whereas others don't. A summary of these results can be seen in **Table 3.3**. While it is known that duplicate gene pairs in yeast, on average, share more common histone modification patterns than random singleton pairs (Zou et al., 2012); further analysis will be needed to determine if this is also the case within the more complex *Arabidopsis* genome.

Since it appears as if some recently duplicated gene pairs have matching histone modification profiles and others do not, we have reason to believe that this epigenetic control mechanism is helping aid evolution when it comes to gene duplication, by utilizing varying levels in PTM conservation to control spatial and temporal gene expression. Our data, as well as published data on larger segmental duplications (Zheng, 2008), support the hypothesis that epigenetic controls aid “evolution by gene duplication” by mechanistically silencing some recent gene duplicates, but not others, until beneficial mutations and subfunctionalization can occur (Rodin and Riggs, 2003, Meagher, 2010). The next step is to determine the role, if any, sequence is playing in facilitating the deposition of various histone modifications.

Since recently duplicated gene pairs have very similar sequences, we have reason to believe that sequence may be driving the control of histone modification deposition (Meagher and Müssar, 2012). In yeast, histone modification profiles between gene duplicates begin to exhibit even more deviation as sequence divergence increases, suggesting that there is a link between evolutionary divergence and conservation among histone modifications (Zou et al., 2012). This evidence supports the notion that there

may be a co-evolution between both genetic and epigenetic elements following gene duplication that is contributing to the expression divergence seen among recently duplicated genes. By selecting gene families that contain varying levels of sequence divergence (**Table 3.2**), our analysis was trying to further explore these histone modification interactions with respect to sequence degeneration. Unfortunately, we see very similar results among the gene pairs in these different families that vary in their sequence conservation. Actins and ADFs massively differ in the amount of sequence conservation between their gene family members, but in both cases we saw duplicate gene pairs that run the gambit in their overall PTM profile divergences (**Table 3.3**). Since sequence is more involved in facilitating nucleosome position (Meagher and Müssar, 2012), perhaps it is this positioning that is facilitating PTM deposition, thus making it difficult to establish a direct link between sequence and PTM enrichment.

By examining the locations of these PTMs within duplicated gene pairs, we can clearly see that PTM enrichments are occurring at different locations within each gene (**Figures 3.2-3.4**). In fact, in most cases the PTM enrichments seem to be located on the neighboring nucleosome in the respective gene duplicate (**Figures 3.2-3.4**). **Figure 3.4** shows two very clear examples of this result. H3K4me3 enrichment for ADF1 occurs directly at the TSS, while in ADF4 it occurs right after the TSS. This on/off enrichment pattern on neighboring nucleosomes is also present in the ADF7 and ADF10 gene pair for H3K4me3 enrichment. These results would suggest that sequence is not playing a role in facilitating PTM deposition amongst gene duplicates; instead PTMs may be acting to promote the subfunctionalization of the gene pair. Perhaps, the specific locations of these

PTMs are determining which roles these variants will be playing throughout spatial and temporal development, thus leading to subfunctionalization.

We believe that in order to fully understand this association, this study needs to be expanded to encompass more gene families in *Arabidopsis*, keeping with the theme of selecting families that vastly range in their sequence conservation. Finding well documented examples of extremely recent duplicated gene pairs (~5-10 MYA) could also help elucidate this connection. Nevertheless, continued analysis of gene duplication, in the context of histone modifications, could help explain how PTM deposition diverges over time to regulate chromatin structure and therefore gene expression.

Combinatorial Patterns of Histone Modifications within Gene Families

Previous genome-wide analysis of 39 different histone modifications in human CD4+ T cells revealed that a large number of combinatorial patterns are associated with genic, promoter, and enhancer regions (Wang et al., 2008). In particular, they found a pattern consisting of 17 different modifications (containing both methylations and acetylations) that colocalize and correlate with each other on an individual nucleosome level. This pattern associated with genes that tend to be more highly expressed, suggesting that these modifications are acting cooperatively to prepare chromatin for active transcription (Wang et al., 2008). This is not surprising since the addition of more modifications tends to lead to further activation of gene expression. Using a statistical hybrid clustering algorithm, this dataset was further analyzed to find that 15 modifications (H2BK120ac, H4K91ac, H2BK20ac, etc.), three histone acetylations (H2AK9ac, H4K16ac, and H4K12ac) and five histone methylations (H3K79me1, H3K79me2, H3K79me3, H4K20me1, and H2BK5me1) were most likely prone to coexist

respectively in these patterns (Linghu et al., 2013). Our analysis tries to detect such patterns among and between individual gene families in Arabidopsis.

By performing Pearson product-moment correlation coefficient (Pearson's r) analyses using the data gathered from **Tables 3.S1-3.S4**, we were able to detect if any of the 14 histone modification types used in this analysis are interacting with one another. In particular, we are determining if the presence of one PTM encourages the presence or absence of another PTM. We found that cooperative histone modification patterns do exist on the individual gene family level. By first performing correlation analysis on our four individual gene families (actins, profilins, ADFs, and ARPs), we found that for the actin gene family, a combinatorial pattern of H3K36me3, H3K4me3, and H4K8ac exists (**Table 3.4**). The presence of one of these three modifications correlates with the enrichment of the other two PTMs (and vice versa). This type of combinatorial PTM enrichment was even more clearly seen in the profilin gene family. Here, a pattern of five modification types was detected: H3K36me3, H3K4me3, H4K8ac, H3K9ac, and H3K14ac (**Table 3.4**). Notice how the same three modifications that were seen in the actin family are also included amongst the profilin pattern. Analysis of the ADF gene family only revealed an interaction containing two histone PTM types, H3K36me3 and H4K8ac (**Table 3.4**). Still, these two PTMs were also contained within the actin and profilin patterns, providing additional evidence that these findings are not just artifacts or a result of small sample size. We performed the same analysis on the ancient and highly divergent ARP gene family. The ARPs were primarily being used as a negative control family because its members exhibit extremely high amounts of sequence divergence. ARP4, 5, 6 and perhaps the other ARPs each appear to be independently evolved from

actin more than 1.5 billion years ago (Meagher, 2010). We detected no combinatorial patterns or interactions among PTMs enriched in the ARP gene family (**Table 3.4**). This suggests that sequence divergence could be facilitating these combinatorial interactions.

Next, we wanted to determine if these same combinatorial PTM interactions could be detected across multiple gene families, or if they were gene family specific. This was achieved by performing Pearson's r analysis on all members of the actin, profilin, ADF, and ARP gene families. Through this method we detected a pattern consisting of six PTMs: H3K36me3, H3K23ac, H4K8ac, H3K9ac, H3K14ac, and H4K16ac (**Table 3.4**). While the majority of these modifications are contained within the individual gene family patterns (i.e. H3K36me3, H4K8ac, H3K9ac, and H3K14ac), we believe that by extending out this analysis we are beginning to lose valuable gene family-specific information.

Notice that all of the PTM combinatorial patterns mentioned for the individual gene families consist of known activation marks. When we extended our analysis to include all four gene families, we saw the inclusion of H4K16ac, which is thought to be associated with transcriptional repression. This is somewhat confusing, and additional experiments will need to be performed to understand why this modification is interacting (cooperatively) with other known activating marks. However, it does appear that H4K16ac enrichment is not occurring on the same nucleosome with the other activating PTMs in this combinatorial pattern. In fact, upon closer analysis it seems as if H4K16ac is enriched on the neighboring nucleosome to the rest of the activating PTMs in this pattern. Perhaps H4K16ac enrichment on the neighboring nucleosome is acting to help facilitate the rest of the PTMs in this pattern to activate transcription. This would explain why some highly expressed genes are showing enrichment for this known repressive

modification. It is important to stress that all of these interactions detected in the actin, profilin, and ADF gene families are of a cooperative nature (yielding positive correlation coefficients). The only statistically significant negative correlations were detected in the ARP gene family analysis, perhaps further authenticating them as a proper negative control.

Since we did not detect any interacting modification patterns in the highly sequence divergent ARP gene family, whereas the more conserved actin, profilin, and ADF gene families did contain such interactions, perhaps sequence conservation is facilitating the deposition of these combinatorial patterns. By examining enrichment profiles for the PTMs contained in each of the actin, profilin, and ADF combinatorial patterns (**Table 3.4**); we can see that all of these enrichments appear to occur on the same nucleosomes within a single gene (**Figures 3.5-3.7**). This would suggest that there may be patterns imbedded in the nucleotide sequence that are recruiting specific histone modifications and excluding others. Perhaps these sequence patterns are conserved in an effort to guide nucleosome position, which might then recruit specific enzymes to deposit these PTMs giving rise to these combinatorial patterns. Most likely, these patterns are acting to recruit proteins like histone acetyltransferases (HATs), histone deacetylases (HDACs), and histone methyltransferases (HMTs) to the nucleosomes to facilitate post-translational modifications, just as promoter regions recruit proteins for transcription (Taverna et al., 2006, Vermeulen et al., 2007, Hung et al., 2009). These results agree with what has been found in *Saccharomyces cerevisiae*, in that as sequence divergence increases the conservation of histone modification deposition begins to deviate (Zou et al., 2012). This would explain why no patterns were detected within the highly sequence

divergent ARPs. Since we do not detect any interactions between these 14 histone modifications and enrichment for 5meC, DNA methylation (which is absent in *Saccharomyces cerevisiae*), we can predict that this phenomenon noticed in yeast may also exist in the more complex *Arabidopsis thaliana*.

While it is unrealistic to expect there to be a single specific PTM activation pattern that can apply to all genes throughout the genome, perhaps these combinatorial patterns are associated with relative spatial locations on certain chromosomes, or are coupled with a gene or gene family's relative expression levels. These patterns are likely altered throughout development to correspond with the changing need for temporal specific functions. Performing this analysis at different time points during development could provide insight into these inquiries.

When performing this analysis, we must be aware of the technological flaws that exist with such methodologies. In order to perform ChIP-chip or ChIP-seq experiments, we are relying heavily on antibody recognition to bind the appropriate protein associated with each histone modification type. It has recently been discovered that modification-specific antibodies are more promiscuous in their PTM recognition than initially expected. Furthermore, these antibodies appear to be highly influenced by neighboring modifications (Fuchs et al., 2011). This makes trusting our data in the context of trying to detect specific neighboring combinatorial patterns quite ambiguous. Additional investigations into the extent of this antibody-specific promiscuity will be needed to further validate these types of studies.

To better understand the biological significance of our results, we must expand these studies to include more gene families, thus giving us more gene pairs and family-

specific combinatorial patterns to analyze. We can use previously designed algorithms (Gu et al., 2002) to uncover novel recently duplicated gene pairs in *Arabidopsis*, and then classify the extent of their divergence within their respective gene families through phylogenetic methods. We can also extend our analysis by incorporating additional types of histone modifications. Here we look at seven different histone acetylations and methylations, but there are a multitude of other PTMs that can be included to try and help deduce these interactions. By performing our analysis on these newly discovered families with these other histone modification types, we will have additional data points to try and determine the role sequence is playing in facilitating histone modification deposition. Using this data, we can then extract the specific sequence portions of each gene where these combinatorial interactions are occurring. By using pattern-based discovery algorithms, like the TEIRESIAS algorithm (Rigoutsos and Floratos, 1998), we can search for statistically significant patterns among these regions that may be attracting specific PTMs. By searching the genome for regions containing these sequence patterns and determining if specific PTMs are enriched at these locations as well, we can further confirm the link between nucleotide sequence and PTM deposition.

Conclusion. Using ChIP-chip data and correlation coefficient analysis we dissected histone PTM deposition at the gene family level. Investigation of the *Arabidopsis* actin, profilin, ADF, and ARP cytoskeletal gene families revealed that recently duplicated gene pairs exhibit varying levels of conservation across their histone modification enrichment profiles. This suggests that histone modifications might aid “evolution by gene duplication” by silencing some recent gene duplicates, but not others, until beneficial mutations and subfunctionalization can occur. *Pearson’s r* analysis

detected distinct combinatorial patterns for some of these histone modification marks within and amongst the individual gene families, with the exception of the most highly sequence divergent ARPs. These results provide evidence that sequence may be helping facilitate the epigenomic landscape of histone post-translational modifications.

Materials and Methods

ChIP-chip analysis

Arabidopsis thaliana plants of Col-ecotype were grown on soil for two weeks, and then the above ground tissues were harvested for analysis. ChIP-chip analysis was performed as previously described in Zhang et al., 2006, 2007, and 2009, using an Affymetrix platform (Santa Clara, CA). The antibodies used in this study were: anti-H3K4me1 (ab8895); anti-H3K4me2 (ab7766); anti-H3K4me3 (ab8580); anti-H3K36me1 (ab8895); anti-H3K36me2 (ab9049); anti-H3K36me3 (ab9050); anti-H3K27me3 (ab6002); anti-H3K9ac (ab4441); anti-H3K14ac (07-353); anti-H3K23ac (07-355); anti-H4K5ac (06-759 MN); anti-H4K8ac (07-328); anti-H4K12ac (07-595); anti-H4K16ac (07-762); anti-H3 (ab1791); anti-5-methyl Cytidine (ab10805), and were purchased from Abcam [ab] (Cambridge, MA, USA) and Millipore (Billerica, MA, USA). H3 ChIP analysis was performed to isolate nucleosomal control DNA.

Data analysis and enrichment profile construction

ChIP-chip data was processed, normalized, and analyzed as previously described in Zhang et al., 2006 and Zhang et al., 2007, using TileMap with the Hidden Markov model option (Ji and Wong, 2005). Posterior probabilities of enrichment (ranging from 0-1) were ascribed to probes across the entire *Arabidopsis* genome. Chromosomal coordinates for the 33 members that comprise the actin, profilin, ADF, and ARP gene families in *Arabidopsis thaliana* were retrieved through the NCBI and TAIR online

databases. Coordinates were extended 500 bp upstream and downstream of the gene coding regions to ensure inclusion of promoter and nearby enhancer information. The full genome dataset from each ChIP-chip analysis was uploaded into SPSS Statistics Version 21 (IBM). The data corresponding to each gene's coordinates were then extracted and exported into Excel (Microsoft). Enrichment profiles were constructed from this data by creating XY scatter plots for each modification type in each gene. The x-axis of these graphs denotes chromosomal location, while the y-axis corresponded to the posterior probabilities of enrichment (**Figure 3.1**). Graphs were generated using either Excel (Microsoft) or RLPlot version 1.5 (University of Innsbruck). As previously described (Zhang et al., 2007), profiles with neighboring probes that yielded posterior probabilities of 0.50 or higher were deemed to be enriched for that modification type ("yes"), while profiles with values below 0.50 were said to be free of that modification type ("no"). To be deemed "yes", neighboring probe regions required a minimal run of 100 bp, allowing for a maximal gap of 200 bp (Zhang et al., 2007). This data was then used to create the summary tables and subsequent correlations for each of the four gene families examined.

Statistical analysis, generation of Pearson's r correlation coefficients

The summary data from the enrichment profiles were converted into binary variables for statistical analysis. The profiles deemed "yes" for enrichment of each modification type were converted to "1", and "no" was converted into "0". The binary data was then uploaded into the program SPSS Statistics Version 21 (IBM). We performed correlation analysis by calculating the Pearson product-moment correlation coefficient (Pearson's r) to determine if there are any modification types that influenced

the presence (positive correlations, 0.0 to +1.0) or absence (negative correlations, -1.0 to 0.0) of another modification type. Two-tailed significance tests were used to calculate the corresponding p -values between each possible correlation. Only p -values less than or equal to 0.05 were signified as being statistically significant. Using the equation $[(n*(n-1))/2]$, meant that for 14 PTMs there were a total of 91 possible correlation sets for each analysis run ($14 \times 13 / 2 = 91$). Interpretation of the size of the correlation coefficients was done using the criteria and cut-offs discussed in Cohen, 1988 and Choudhury, 2009. Coefficients ranging from 0.0 - +/-0.3 were deemed as weak correlations, +/-0.3 - +/-0.5 were moderate correlations, and +/-0.5 - +/-1.0 were considered strong correlations.

Phylogenetic Analysis

Arabidopsis nucleotide coding sequences (CDS) for the actin, profilin, ADF, and ARP gene families were retrieved through the NCBI and TAIR online databases. Sequences were aligned using the ClustalW program (Larkin et al., 2007) within MEGA5.05 (Tamura et al., 2011). A BLOSUM protein weight matrix was used to align the sequences. Once aligned, Bayesian inference phylogenetic trees were built for each of the four gene families using the program Mr.Bayes version 3.1 (Ronquist and Huelsenbeck, 2003). We ran each of our analyses for 5,000,000 generations with a tree sample frequency of every 100 generations. The final tree was compiled after a burnin of 12,500 trees (25%). Trees were then converted to newick format using the ape library (Paradis et al., 2004), and then visualized in MEGA5.05. Trees were rooted after the analysis. To ensure congruence, we also performed maximum-likelihood and neighbor-joining phylogenetic analyses using Mega 5.05, and verified that the relationships seen in our trees are consistent.

REFERENCES

- Andrianantoandro E, Pollard TD: **Mechanism of actin filament turnover by severing and nucleation at different concentrations of ADF/Cofilin.** *Mol Cell* 2006, **24**:13–23.
- Bamburg JR: **Proteins of the ADF/Cofilin Family: Essential Regulators of Actin Dynamics.** *Annual Review of Cellular Developmental Biology* 1999, **15**:185-230.
- Bannister AJ, Kouzarides K: **Regulation of chromatin by histone modifications.** *Cell Research* 2011, **21**:381-395.
- Benevolenskaya EV: **Histone H3K4 demethylases are essential in development and differentiation.** *Biochem Cell Biol* 2007, **85**:435–443.
- Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T, Schreiber SL: **Methylation of histone H3 Lys 4 in coding regions of active genes.** *PNAS* 2002, **99**:8695-8700.
- Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, Gingeras TR, Schreiber SL, Lander ES: **Genomic maps and comparative analysis of histone modifications in human and mouse.** *Cell* 2005, **120**:169–81.
- Blanc G, Hokamp K, Wolfe KH: **A Recent Polyploidy Superimposed on Older Large-Scale Duplications in the *Arabidopsis* Genome.** *Genome Res* 2003, **13**:137–144.
- Bowers JE, Chapman BA, Rong J, Paterson AH: **Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events.** *Nature* 2003, **422**:433–438.
- Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P et al: **Role of histone H3 lysine 27 methylation in Polycomb-group silencing.** *Science* 2002, **298**:1039–1043.
- Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP, Workman JL: **Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription.** *Cell* 2005, **123**:581–92.
- Choudhury A: **Statistical Correlation.** *Explorable.com* 2009: <http://explorable.com/statistical-correlation>.
- Cohen J: **Statistical power analysis for the behavioral sciences.** 2nd ed. 1988.

- Fnu S, Williamson EA, De Haro LP, Brennehan M, Wray J, Shaheen M, Radhakrishnan K, Lee SH, Nickoloff JA, Hromas R: **Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining.** *PNAS* 2010, **108**:540-545.
- Fuchs SM, Krajewski K, Baker RW, Miller VL, Shultz BD: **Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition.** *Current Biology* 2011, **21**:53–58.
- Gu Z, Cavalcanti A, Chen F-C, Bouman P, Li W-H: **Extent of Gene Duplication in the Genomes of Drosophila, Nematode, and Yeast.** *Mol Biol Evol* 2002, **19**:256-262.
- Haig D: **The (dual) origin of epigenetics.** *Cold Spring Harb Symp Quant Biol* 2004, **69**:67-70.
- Huang S, McDowell JM, Weise MJ, Meagher RB: **The Arabidopsis Profilin Gene Family: Evidence for an Ancient Split between Constitutive and Pollen-Specific Profilin Genes.** *Plant Physiology* 1996, **111**:115-126.
- Hung T, Binda O, Champagne KS, Kuo AJ, Johnson K, Chang HY, Simon MD, Kutateladze TG, Gozani O: **ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation.** *Mol Cell* 2009, **33**:248–256.
- Jenuwein T, Allis C: Translating the histone code. *Science* 2001, **293**:1074–80.
- Jeppesen P, Turner BM: **The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression.** *Cell* 1993, **74**:281-289.
- Ji H and Wong WH: **TileMap: create chromosomal map of tiling array hybridizations.** *Bioinformatics* 2005, **21**:3629–3636.
- Jockusch BM, Murk K, Rothkegel M: **The profile of profilins.** *Rev Physiol Biochem Pharmacol* 2007, **159**:131-149.
- Joshi AA, Struhl K: **Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation.** *Mol Cell* 2005, **20**:971–978.
- Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ, Boone C, Emili A, Weissman JS, Hughes TR, Strahl BD, Grunstein M, Greenblatt JF, Buratowski S, Krogan NJ: **Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex.** *Cell* 2005, **123**:593–605.

- Kim T, Buratowski S: **Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 50 transcribed regions.** *Cell* 2009, **137**:259–272.
- Koch CM, Andrews RM, Flicek P, Dillon SC, Karaöz U, Clelland GK, Wilcox S, Beare DM, Fowler JC, Couttet P, James KD, Lefebvre GC, Bruce AW, Dovey OM, Ellis PD, Dhimi P, Langford CF, Weng Z, Birney E, Carter NP, Vetrici D, Dunham I: **The landscape of histone modifications across 1% of the human genome in five human cell lines.** *Genome Res* 2007, **17**:691–707.
- Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M, Greenblatt JF, Shilatifard A: **The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation.** *Mol Cell* 2003, **11**:721–729.
- Kudryashov DS, Reisler E: **ATP and ADP Actin States.** *Biopolymers* 2012, **99**:245-256.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **ClustalW and ClustalX version 2.** *Bioinformatics* 2007, **23**:2947-2948.
- Li J, Moazed D, Gygi SP: **Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation.** *J Biol Chem* 2002, **277**:49383–49388.
- Linghu C, Zheng H, Zhang L, Zhang J: **Discovering common combinatorial histone modification patterns in the human genome.** *Gene* 2013, **518**:171-178.
- Lippman Z, et al.: **Role of transposable elements in heterochromatin and epigenetic control.** *Nature* 2004, **430**:471-476.
- Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: **Modeling gene and genome duplications in eukaryotes.** *PNAS* 2005, **102**:5454–5459.
- Mathur J, Mathur N, Kerneback B, Hülskamp M: **Mutations in actin-related proteins 2 and 3 affect cell shape development in Arabidopsis.** *Plant Cell* 2003, **15**:1632-1645.
- McDowell JM, Huang S, McKinney EC, An Y-Q, Meagher RB: **Structure and evolution of the actin gene family in Arabidopsis thaliana.** *Genetics* 1996, **142**:587–602.
- McKinney EC, Kandasamy MK, and Meagher RB: **Arabidopsis Contains Ancient Classes of Differentially Expressed Actin-Related Protein Genes.** *Plant Physiology* 2002, **128**:997-1007.
- Meagher RB: **The evolution of epitype.** *The Plant Cell* 2010, **22**:1658-1666.

- Meagher RB, Müssar KJ: **The Influence of DNA Sequence on Epigenome-induced Pathologies.** *Epigenetics & Chromatin* 2012, **5**:11.
- Meagher RB, Kandasamy MK, McKinney EC, Roy E: **Chapter 5- Nuclear actin-related proteins in epigenetic control.** *Int Rev Cell Mol Biol* 2009, **277**:157-215.
- Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, et al.: **Genome-wide comparative analysis of the Brassica rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication.** *Genome Biol* 2009, **10**:R111.
- Nanney DL: **Epigenetic Control Systems.** *Proc Natl Acad Sci U S A* 1958, **44**:712-717.
- Ng HH, Robert F, Young RA, Struhl K: **Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity.** *Mol Cell* 2003, **11**:709–719.
- Paradis E, Claude J, Strimmer K: **APE: analyses of phylogenetics and evolution in R language.** *Bioinformatics* 2004, **20**:289–290.
- Pearson TA: **How to Interpret a Genome-wide Association Study.** *JAMA* 2008, **299**:1335-1344.
- Rando OJ: **Combinatorial complexity in chromatin structure and function: revisiting the histone code.** *Curr Opin Genet Dev* 2012, **22**:148-155.
- Rigoutsos I, Floratos A: **Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm.** *Bioinformatics* 1998, **14**:55-67.
- Robyr D, Kurdistani SK, Grunstein M, Kurdistani SK, Grunstein M, Grunstein M: **Analysis of genome-wide histone acetylation state and enzyme binding using DNA microarrays.** *Methods Enzymol* 2004, **376**:289–304.
- Rodin SN, Riggs AD: **Epigenetic Silencing May Aid Evolution by Gene Duplication.** *Molecular Evolution* 2003, **56**:718-729.
- Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
- Ruzicka DR, Kandasamy MK, McKinney EC, Burgos-Rivera B, Meagher RB: **The ancient subclasses of Arabidopsis ACTIN DEPOLYMERIZING FACTOR genes exhibit novel and differential expression.** *The Plant Journal* 2007, **52**:460-472.
- Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y: **The hidden duplication past of Arabidopsis thaliana.** *PNAS* 2002, **99**:13627–13632.

- Strahl B, Allis C: **The language of covalent histone modifications**. *Nature* 2000, **403**:41–45.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar, S: **MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods**. *Mol Biol Evol* 2011, **28**:2731–2739.
- Tanaka Y, Katagiri Z, Kawahashi K, Kioussis D, Kitajima S: **Trithorax-group protein ASH1 methylates histone H3 lysine 36**. *Gene* 2007, **397**:161-168.
- Taverna SD, Ilin S, Rogers RS, Tanny JC, Lavender H, Li H, Baker L, Boyle J, Blair LP, Chait BT, et al: **Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs**. *Mol Cell* 2006, **24**:785–796.
- The modENCODE Consortium- Roy S et al: **Identification of Functional Elements and Regulatory Circuits by *Drosophila* modENCODE**. *Science* 2010, **330**:1787-1797.
- Tsai WW, Wang Z, Yiu TT, Akdemir KC, Xia W, Winter S, Tsai CY, Shi X, Schwarzer D, Plunkett W, Aronow B, Gozani O, Fischle W, Hung MC, Patel DJ, Barton MC: **TRIM24 links a non-canonical histone signature to breast cancer**. *Nature* 2010, **468**:927-932.
- Vaquero A, Scher M, Lee D, Erdjument-Bromage H, Tempst P, Reinberg D: **Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin**. *Mol Cell* 2004, **16**:93-105.
- Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT: **Selective anchoring of TFIIID to nucleosomes by trimethylation of histone H3 lysine 4**. *Cell* 2007, **131**:58–69.
- Vision TJ, Brown DG, Tanksley SD: **The origins of genomic duplications in *Arabidopsis***. *Science* 2000, **290**:2114–2117.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K: **Combinatorial patterns of histone acetylation and methylation in the human genome**. *Nat Genet* 2008, **40**:897-903.
- Yuan W, Xu M, Huang C, Lin N, Chen S, Zhu B: **H3K36 methylation antagonizes PRC2-mediated H3K27 methylation**. *J Biol Chem* 2011, **286**:7983-7989.

- Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu JK: **Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana***. *PLoS ONE* 2007, **2**:e1210.
- Zhang X: **The Epigenetic Landscape of Plants**. *Science* 2008, **320**:489-492.
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, and Jacobsen SE: **Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana***. *Genome Biology* 2009, **10**:R62.
- Zhang X, Clarenz O, Cokus S, Pelligrini M, Jacobsen SE: **Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis***. *PLoS Biol* 2007, **5**:e129.
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, et al.: **Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis***. *Cell* 2006, **126**:1189–1201.
- Zheng D: **Asymmetric histone modifications between the original and derived loci of human segmental duplications**. *Genome Biology* 2008, **9**:R105.
- Zou Y, Su Z, Huang W, Gu X: **Histone modification pattern evolution after yeast gene duplication**. *BMC Evolutionary Biology* 2012, **12**:111.

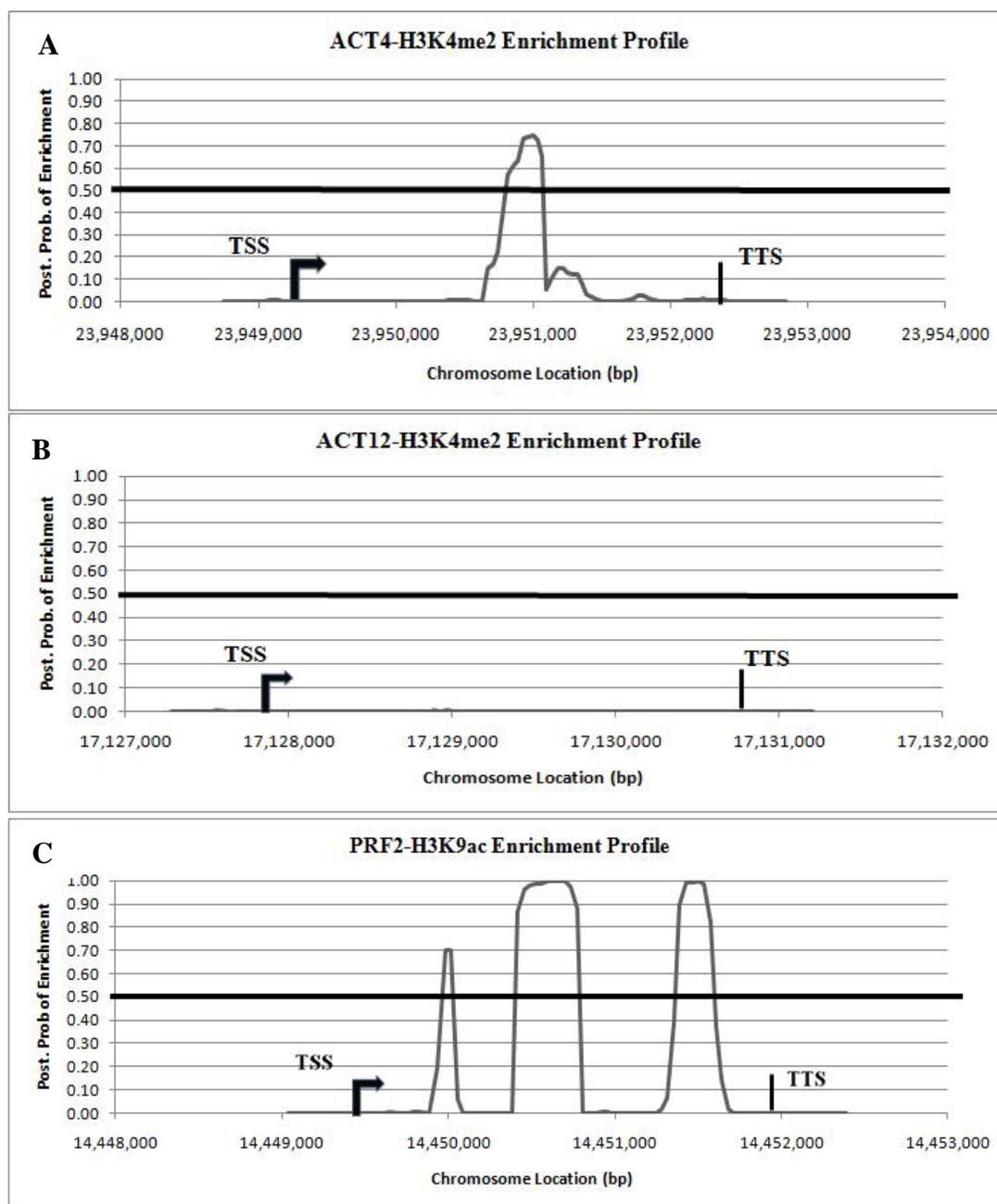


Figure 3.1: Examples of enrichment profiles generated for this analysis.

A) H3K4me2 enrichment profile for ACT4. This profile shows an enrichment peak above our 0.50 threshold, so H3K4me2 is deemed present in ACT4 chromatin (“yes”). This is an example of seeing enrichment [in the gene body]. **B)** H3K4me2 enrichment profile for ACT12. This is an example of “no” enrichment for this modification since it does not reach our 0.50 threshold. **C)** H3K9ac enrichment profile for PRF2. This profile shows that H3K9ac is present in PRF2 chromatin (“yes”). This is an example of seeing enrichment [throughout profile], since multiple peaks were detected. TSS= Transcriptional Start Site; TTS=Transcriptional Termination Site. X-axis= Chromosome Location (bp); Y-axis= Posterior Probability of Enrichment (0.00-1.00).

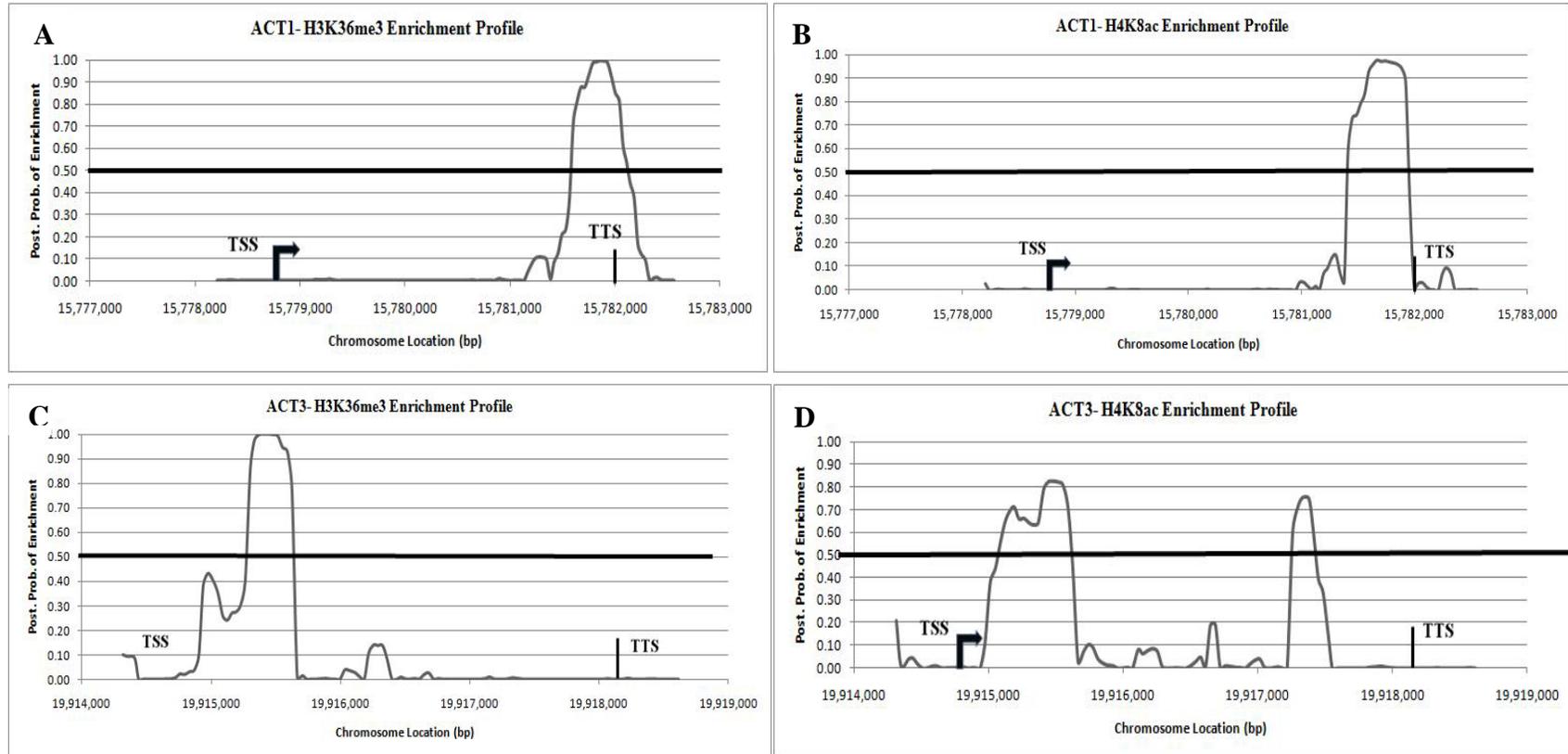
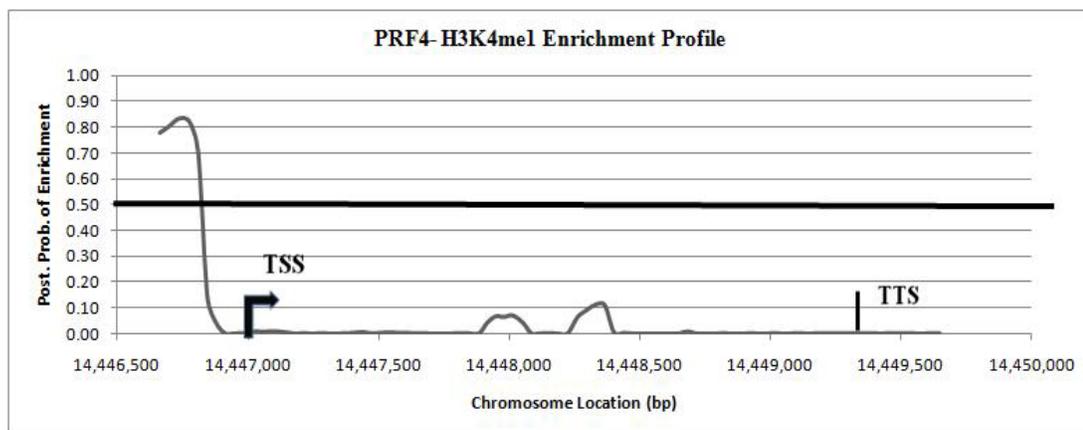


Figure 3.2: H3K36me3 and H4K8ac enrichment profiles for ACT1 and ACT3 gene pairs. A) H3K36me3 enrichment profile for ACT1. **B)** H3K36me3 enrichment profile for ACT3. **C)** H4K8ac enrichment profile for ACT1. **D)** H4K8ac enrichment profile for ACT3. TSS= Transcriptional Start Site; TTS=Transcriptional Termination Site.

A



B

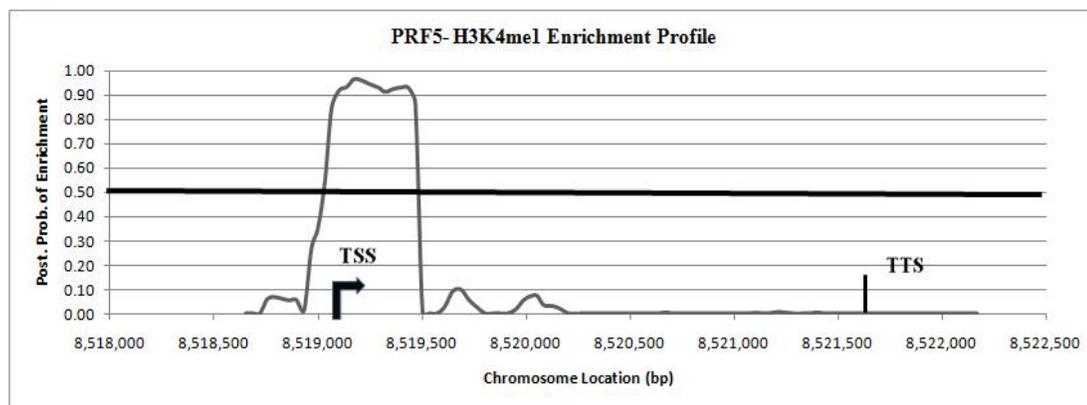


Figure 3.3: H3K4me1 enrichment profiles for the PRF4 and PRF5 gene pair. A)

H3K4me1 enrichment profile for PRF4. **B)** H3K4me1 enrichment profile for PRF5.

TSS= Transcriptional Start Site; TTS=Transcriptional Termination Site.

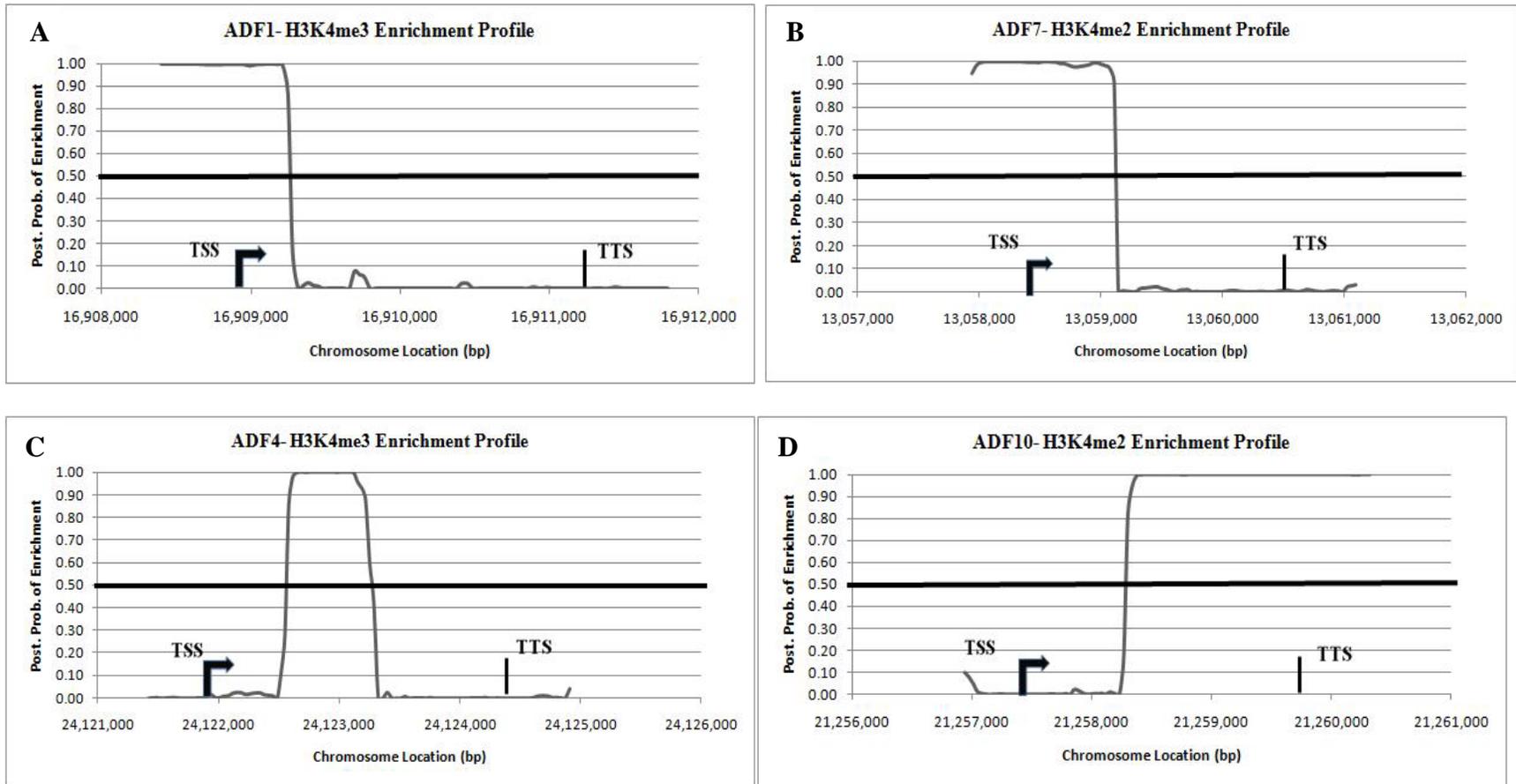


Figure 3.4: H3K4me3 enrichment profiles for ADF1 and ADF4 and H3K4me2 enrichment profiles for ADF7 and ADF10 gene pairs. A) H3K4me3 enrichment profile for ADF1. B) H3K4me3 enrichment profile for ADF4. C) H3K4me2 enrichment profile for ADF7. D) H3K4me2 enrichment profile for ADF10. TSS= Transcriptional Start Site; TTS=Transcriptional Termination Site.

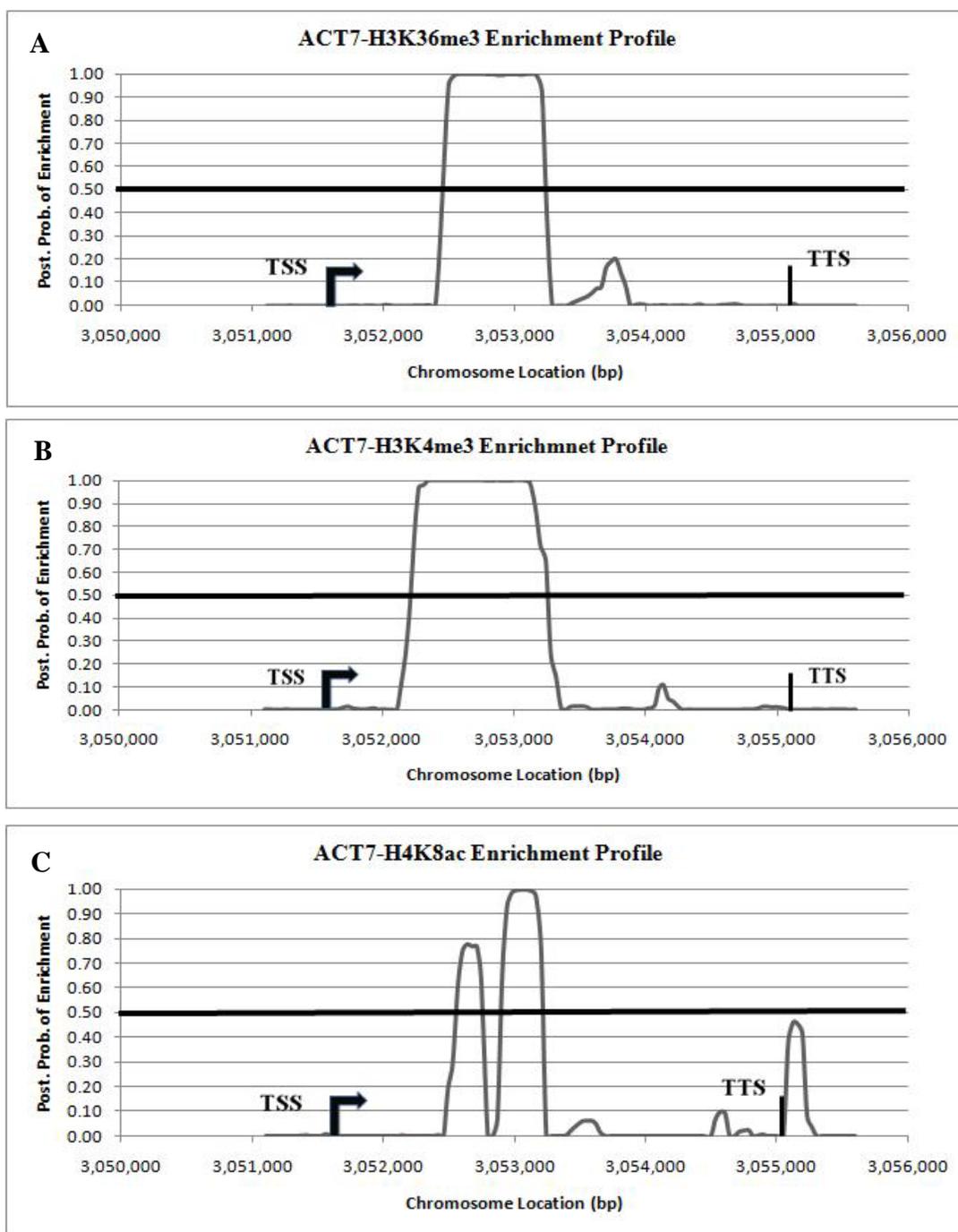


Figure 3.5: H3K36me3, H3K4me3, and H4K8ac enrichment profiles for ACT7 (actin combinatorial pattern). A) H3K36me3 enrichment profile for ACT7. B) H3K4me3 enrichment profile for ACT7. C) H4K8ac enrichment profile for ACT7.

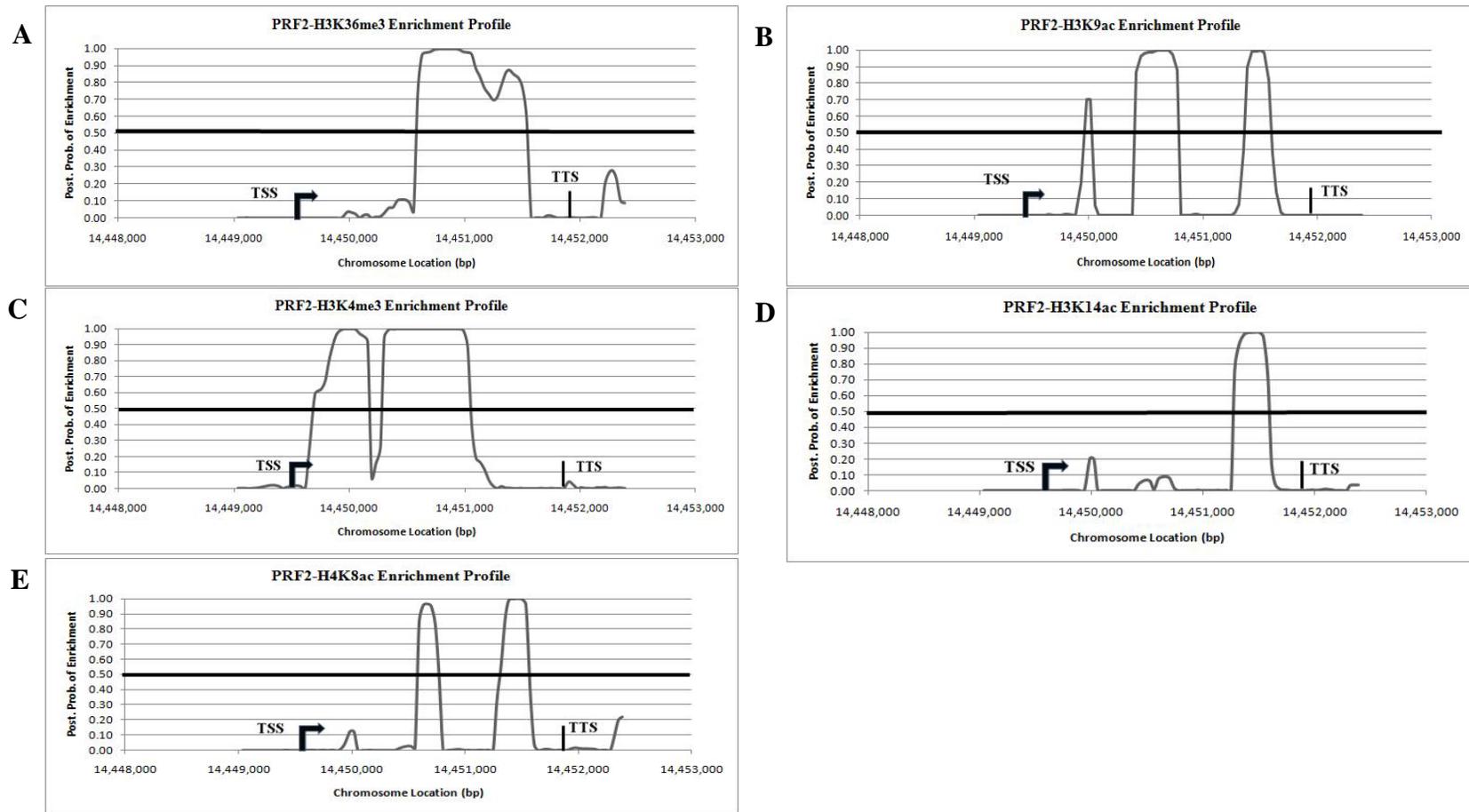


Figure 3.6: H3K36me3, H3K4me3, H3K9ac, H3K14ac, and H4K8ac enrichment profiles for PRF2 (profilin combinatorial pattern). A) H3K36me3 enrichment profile for PRF2. B) H3K9ac enrichment profile for PRF2. C) H3K4me3 enrichment profile for PRF2. D) H3K14ac enrichment profile for PRF2. E) H4K8ac enrichment profile for PRF2.

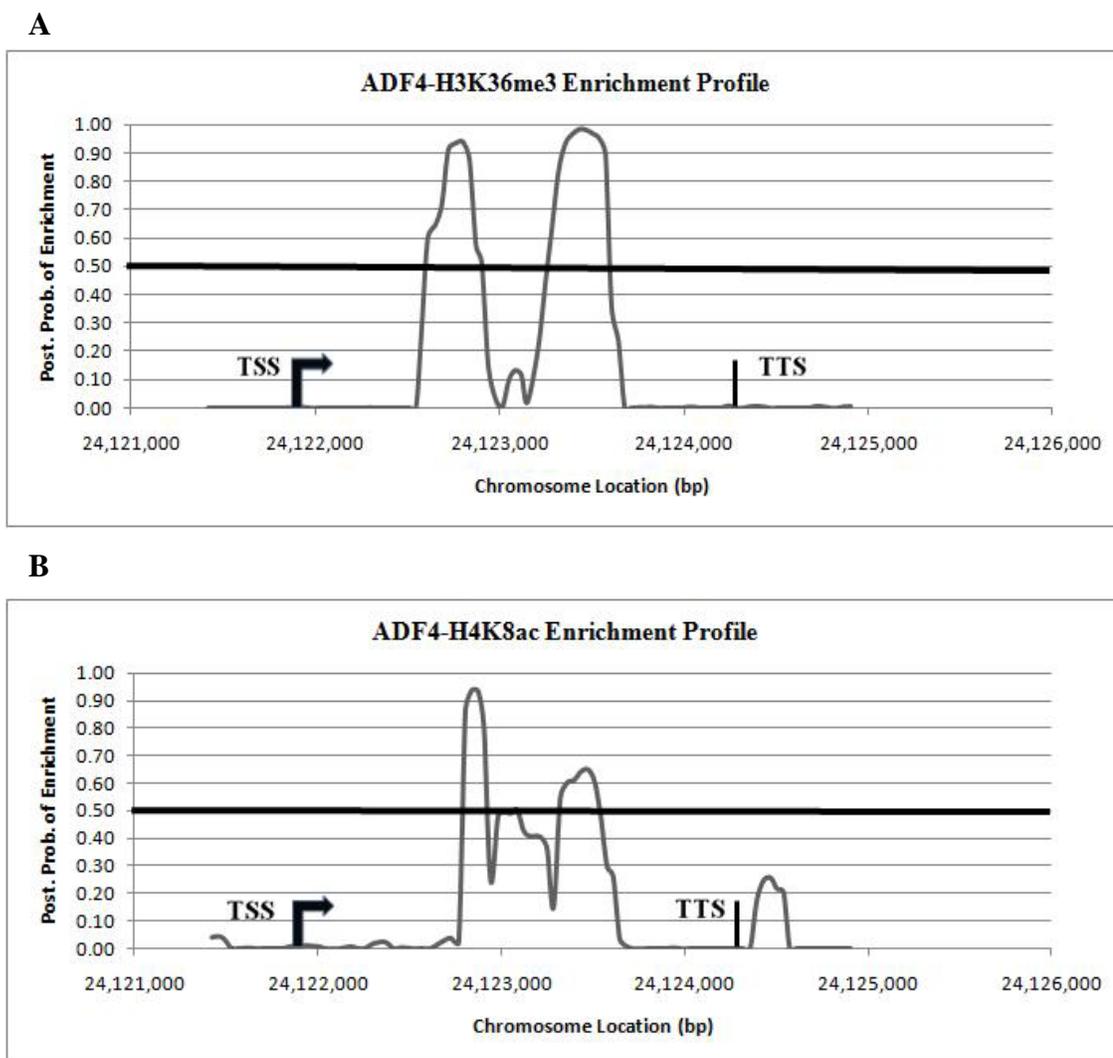


Figure 3.7: H3K36me3 and H4K8ac enrichment profiles for ADF4 (ADF combinatorial pattern). A) H3K36me3 enrichment profile for ADF4. B) H4K8ac enrichment profile for ADF4.

Type of Histone Modification	Known Associated Effects (Activation, Repression)	Location (if known) of Resulting Histone Mark	References
H3K36me1	Active Gene Transcription	Intergenic domains located outside of annotated genes	Tanaka et al., 2007 Roy et al., 2010
H3K36me2	Active Gene Transcription, prevents H3K27me3 spreading	DNA double strand breaks	Roy et al., 2010 Fnu et al., 2010 Yuan et al., 2011
H3K36me3	Active Gene Transcription	Gene body	Li et al., 2002 Carrozza et al., 2005 Keogh et al., 2005 Joshi and Struhl 2005
H3K27me3	Repression of transcription	Upstream/ Promoter	Cao et al., 2002
H3K4me1	Maintaining active transcription	Gene body	Benevolenskaya 2007 Bernstein et al., 2002
H3K4me2	Active Gene Transcription	Transcriptional Start Site (TSS)	Bernstein et al., 2002
H3K4me3	Active Gene Transcription	Promoter	Krogan et al., 2003 Ng et al., 2003 Bernstein et al., 2005
H3K9ac	Active Gene Transcription	Gene body & Upstream of TSS	Robyr et al, 2004 Koch et al., 2007
H3K14ac	Active Gene Transcription	Gene body & Upstream of TSS	Robyr et al, 2004 Koch et al., 2007
H3K23ac	Active Gene Transcription	Undetermined	Tsai et al., 2010
H4K5ac	Active Gene Transcription	TSS & along Gene Body	Jeppesen and Turner 1993 Wang et al., 2008
H4K8ac	Active Gene Transcription	TSS & along Gene Body	Wang et al., 2008
H4K12ac	Active Gene Transcription	TSS & along Gene Body	Wang et al., 2008
H4K16ac	Recruits Deacetylases to Silence Transcription	TSS & along Gene Body	Vaquero et al., 2004 Wang et al., 2008

Table 3.1: Summary of the 14 different histone modifications analyzed in this study.

Maximum Nucleotide Sequence Divergence Among Gene Family	Example Gene Family	# of Genes in Family	# of Gene Pairs in Family
~18.5% {Low}	Actins	9	3
~31% {Moderate}	Profilins	5	2
~52% {Mod.-High}	ADFs	11	4
~73% {High}	ARPs	8	0

Table 3.2: Gene family information for actins, profilins, ADFs, and ARPs.

	ACT2 / ACT8	ACT1 / ACT3	ACT4 / ACT12	PRF1 / PRF2	PRF4 / PRF5	ADF1 / ADF4	ADF7 / ADF10	ADF8 / ADF11	ADF5 / ADF9
H3K36me1	- / -	- / -	- / -	- / -	- / -	- / -	- / +	+ / +	- / -
H3K36me2	+ / -	- / -	- / -	- / -	- / -	- / +	- / -	- / +	- / -
H3K36me3	+ / -	+ / +	- / -	- / +	- / -	- / +	- / -	- / +	- / -
H3K27me3	- / -	- / +	+ / -	- / -	- / -	- / -	- / -	+ / -	- / -
H3K4me1	+ / -	- / -	- / -	- / -	+ / +	- / +	- / -	- / -	- / -
H3K4me2	+ / -	- / -	+ / -	- / -	- / -	- / +	+ / +	+ / +	- / +
H3K4me3	+ / -	+ / +	- / -	- / +	- / -	+ / +	- / +	- / -	- / +
H3K9ac	+ / -	+ / -	+ / -	- / +	- / -	- / +	- / +	- / -	- / -
H3K14ac	+ / -	- / -	- / +	- / +	- / -	- / +	- / -	- / +	- / -
H3K23ac	+ / -	+ / +	+ / +	- / +	- / -	- / +	- / -	- / +	- / -
H3K5ac	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -
H4K8ac	+ / -	+ / +	- / -	- / +	- / -	- / +	- / -	- / +	- / -
H4K12ac	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -
H4K16ac	+ / -	- / +	+ / -	- / +	- / +	- / +	- / -	- / -	- / -
Overall marks in common (%)	4/14 (28.6%)	11/14 (78.6%)	9/14 (64.3%)	7/14 (50.0%)	13/14 (92.9%)	5/14 (35.7%)	11/14 (78.6%)	8/14 (57.1%)	12/14 (85.7%)

+ = modification present - = modification absent

Table 3.3: Summary of the enrichment of histone modification marks among gene pairs.

Gene Family	Maximum Nucleotide Sequence Divergence Among Gene Family	Histone Modification Marks that Positively Correlate with each other (PTM Combinatorial Patterns)
Actins	~18.5% {Low}	<ul style="list-style-type: none"> • H3K36me3 • H3K4me3 • H4K8ac
Profilins	~31% {Moderate}	<ul style="list-style-type: none"> • H3K36me3 • H3K4me3 • H4K8ac • H3K9ac • H3K14ac
ADFs	~52% {Mod.-High}	<ul style="list-style-type: none"> • H3K36me3 • H4K8ac
ARPs	~73% {High}	None
All 4 Families Combined	N/A	<ul style="list-style-type: none"> • H3K9ac • H3K14ac • H3K23ac • H4K8ac • H4K16ac • H3K36me3

Table 3.4: Summary of the combinatorial patterns of histone modification marks detected within and among gene families.

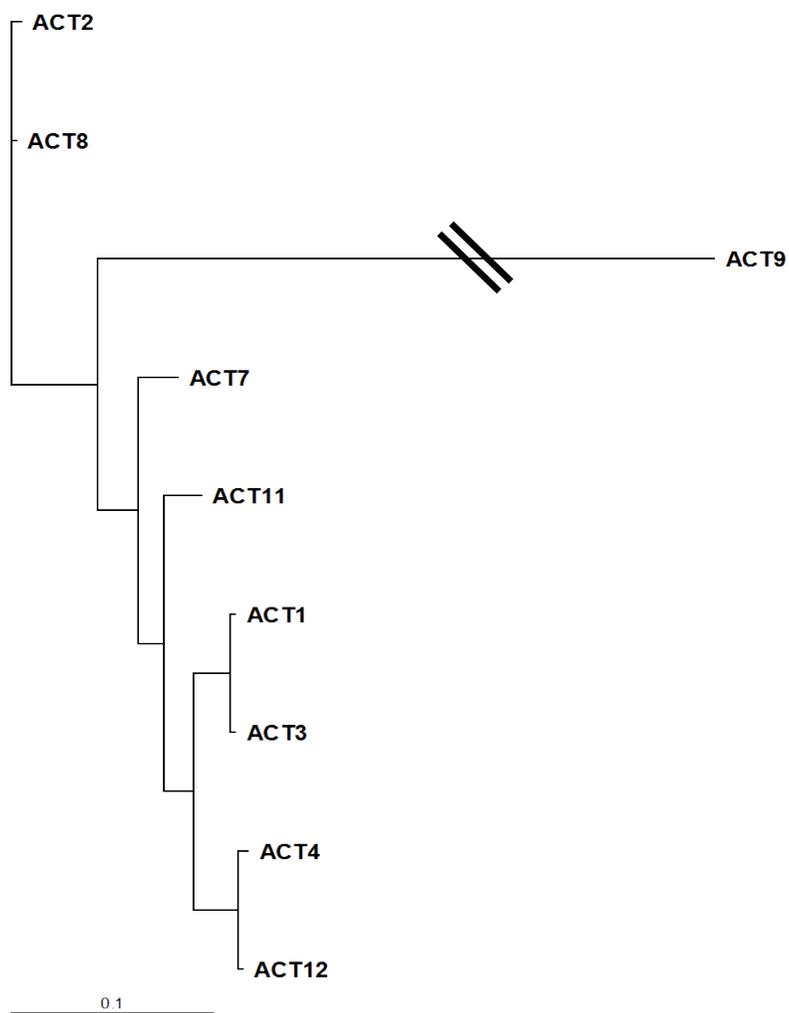


Figure 3.S1: Bayesian phylogenetic analysis of the *Arabidopsis* actin gene family.

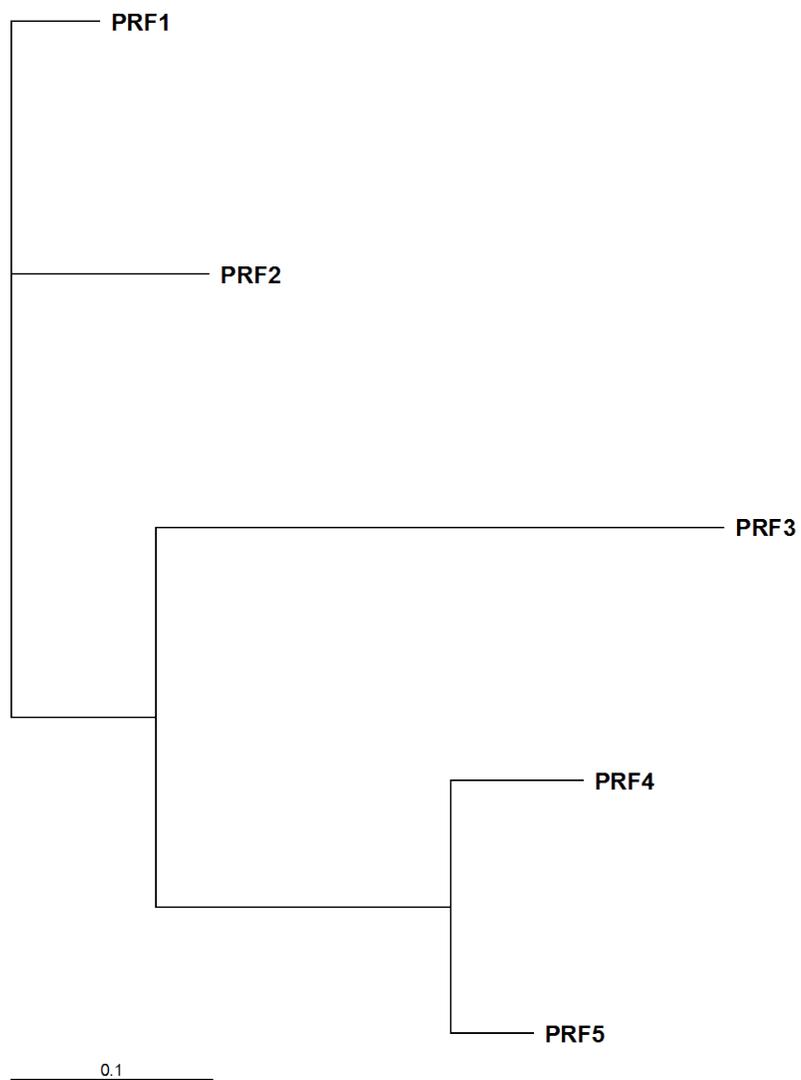


Figure 3.S2: Bayesian phylogenetic analysis of the *Arabidopsis* profilin gene family.

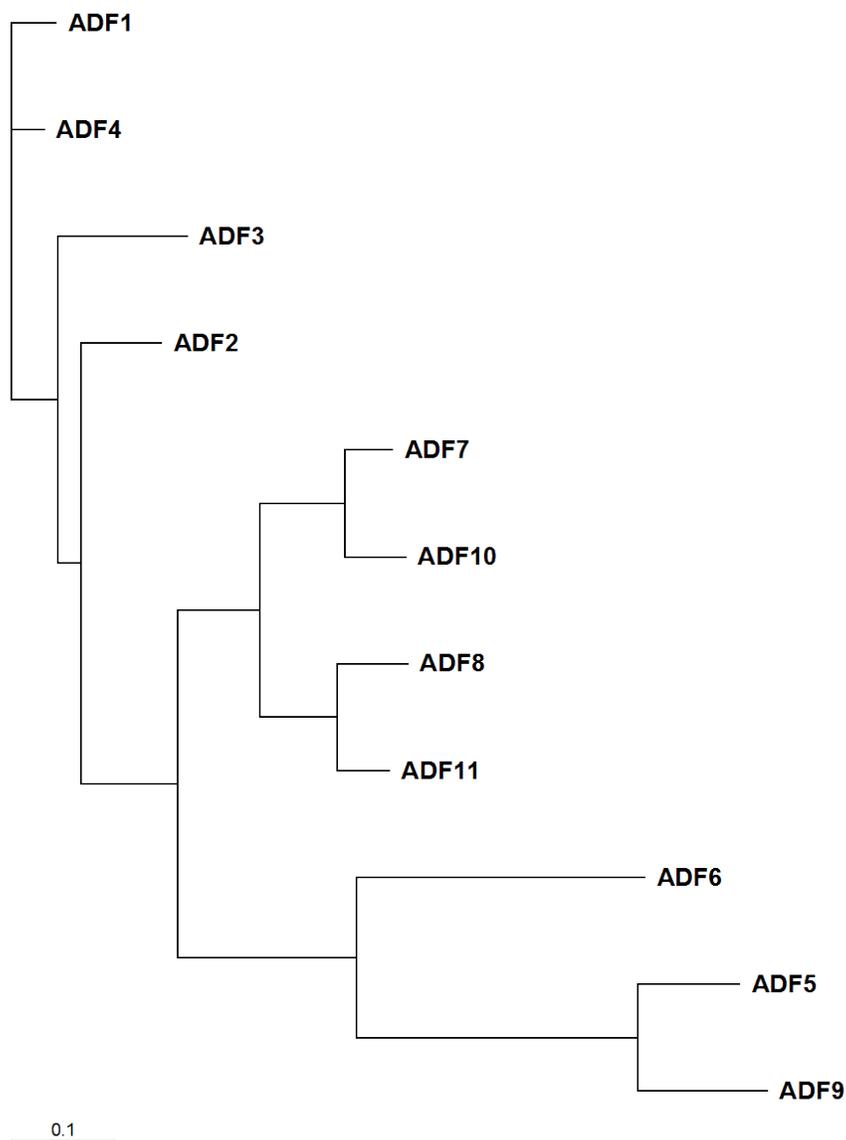


Figure 3.S3: Bayesian phylogenetic analysis of the *Arabidopsis* Actin Depolymerizing Factor (ADF) gene family.

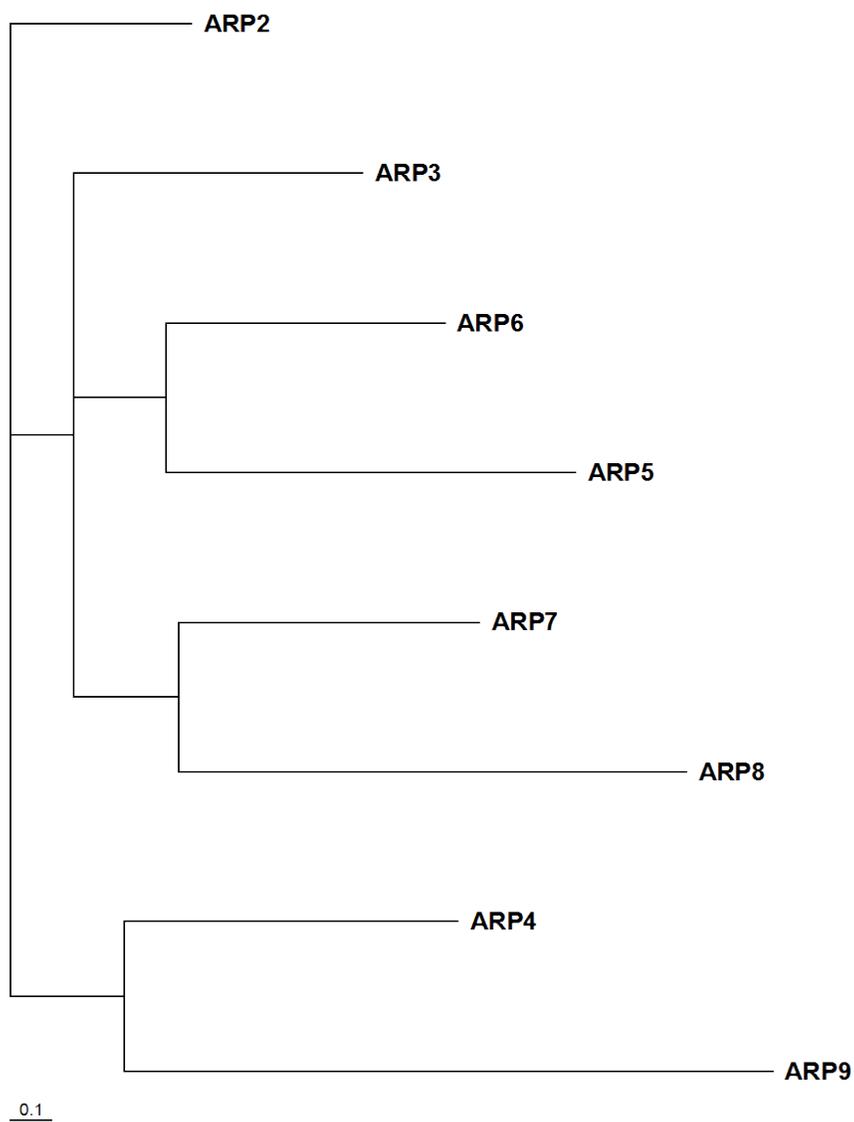


Figure 3.S4: Bayesian phylogenetic analysis of the *Arabidopsis* Actin Related Protein (ARP) gene family.

Gene Name	Expression Level (RFU)	H3K36me1 [Loc.]	H3K36me2 [Loc.]	H3K36me3 [Loc.]	H3K27me3 [Loc.]	H3K4me1 [Loc.]	H3K4me2 [Loc.]	H3K4me3 [Loc.]
ACT2*	1713.81	no	yes [gene body]	yes [gene body]	no	yes [gene body]	yes [gene body]	yes [throughout profile]
ACT8*	595.98	no	no	no	no	no	no	no
ACT9	3.58	no	no	yes [gene body]	no	yes [upstream]	no	yes [gene body]
ACT7	1486.31	yes [gene body]	no	yes [gene body]	no	no	no	yes [gene body]
ACT11	72.73	yes [gene body]	no	yes [gene body]	no	yes [gene body]	yes [throughout profile]	no
ACT1*	131.15	no	no	yes [downstream]	no	no	no	yes [downstream]
ACT3*	131.15	no	no	yes [before & start of gene]	yes [upstream & start of gene]	no	no	yes [throughout profile]
ACT4*	9.18	no	no	no	yes [throughout profile]	no	yes [gene body]	no
ACT12*	0.95	no	no	no	no	no	no	no

*Gene Pairs

Table 3.S1: Summary of ChIP-Chip data for the actin gene family.

Gene Name	Expression Level (RFU)	H3K 9ac [Loc.]	H3K14ac [Loc.]	H3K23ac [Loc.]	H4K5ac [Loc.]	H4K8ac [Loc.]	H4K12ac [Loc.]	H4K16ac [Loc.]
ACT2*	1713.81	yes [throughout profile]	yes [gene body]	yes [throughout profile]	no	yes [gene body]	no	yes [gene body]
ACT8*	595.98	no	no	no	no	no	no	no
ACT9	3.58	yes [gene body]	yes [upstream]	yes [gene body]	no	yes [throughout profile]	no	no
ACT7	1486.31	yes [gene body]	yes [gene body]	yes [gene body]	no	yes [gene body]	no	yes [gene body]
ACT11	72.73	yes [throughout profile]	yes [gene body]	yes [gene body]	no	yes [gene body & downstream]	no	yes [gene body]
ACT1*	131.15	yes [gene body]	no	yes [gene body & downstream]	no	yes [gene body]	no	no
ACT3*	131.15	no	no	yes [upstream & start of gene]	no	yes [gene body]	no	yes [gene body]
ACT4*	9.18	yes [gene body]	no	yes [gene body]	no	no	no	yes [gene body]
ACT12*	0.95	no	yes [downstream]	yes [downstream]	no	no	no	no

*Gene Pairs

Table 3.S1 (cont.): Summary of ChIP-Chip data for the actin gene family.

Gene Name	Expression Level (RFU)	H3K36me1 [Loc.]	H3K36me2 [Loc.]	H3K36me3 [Loc.]	H3K27me3 [Loc.]	H3K4me1 [Loc.]	H3K4me2 [Loc.]	H3K4me3 [Loc.]
PRF1*	24.01	no	no	no	no	no	no	no
PRF2*	94.87	no	no	yes [gene body]	no	no	no	yes [gene body]
PRF3	22.56	no	no	no	no	no	no	no
PRF4*	0.91	no	no	no	no	yes [upstream]	no	no
PRF5*	2.39	no	no	no	no	yes [start of gene]	no	no

*Gene Pairs

Table 3.S2: Summary of ChIP-Chip data for the profilin gene family.

Gene Name	Expression Level (RFU)	H3K 9ac [Loc.]	H3K14ac [Loc.]	H3K23ac [Loc.]	H4K5ac [Loc.]	H4K8ac [Loc.]	H4K12ac [Loc.]	H4K16ac [Loc.]
PRF1*	24.01	no	no	no	no	no	no	no
PRF2*	94.87	yes [throughout profile]	yes [gene body]	yes [throughout profile]	no	yes [gene body & downstream]	no	yes [gene body]
PRF3	22.56	no	no	yes [upstream]	no	no	no	no
PRF4*	0.91	no	no	no	no	no	no	no
PRF5*	2.39	no	no	no	no	no	no	yes [downstream]

*Gene Pairs

Table 3.S2 (cont.): Summary of ChIP-Chip data for the profilin gene family.

Gene Name	Expression Level (RFU)	H3K36me1 [Loc.]	H3K36me2 [Loc.]	H3K36me3 [Loc.]	H3K27me3 [Loc.]	H3K4me1 [Loc.]	H3K4me2 [Loc.]	H3K4me3 [Loc.]
ADF1*	1044.28	no	no	no	no	no	no	yes [upstream]
ADF4*	598.9	no	yes [gene body]	yes [gene body]	no	yes [gene body]	yes [gene body]	yes [gene body]
ADF3	1467.81	no	no	yes [downstream]	yes [gene body]	no	yes [gene body]	yes [gene body]
ADF2	502.33	no	no	no	no	no	no	yes [throughout profile]
ADF7*	0.43	no	no	no	no	no	yes [upstream & start of gene]	no
ADF10*	0.91	yes [gene body]	no	no	no	no	yes [gene body]	yes [gene body]
ADF8*	1.63	yes [downstream]	no	no	yes [gene body]	no	yes [gene body]	no
ADF11*	5.88	yes [gene body]	yes [gene body]	yes [downstream]	no	no	yes [upstream]	no
ADF6	391.58	no	no	no	no	no	no	no
ADF5*	190.93	no	no	no	no	no	no	no
ADF9*	29.86	no	no	no	no	no	yes [gene body]	yes [gene body]

*Gene Pairs

Table 3.S3: Summary of ChIP-Chip data for the Actin Depolymerizing Factor (ADF) gene family.

Gene Name	Expression Level (RFU)	H3K9ac [Loc.]	H3K14ac [Loc.]	H3K23ac [Loc.]	H4K5ac [Loc.]	H4K8ac [Loc.]	H4K12ac [Loc.]	H4K16ac [Loc.]
ADF1*	1044.28	no	no	no	no	no	no	no
ADF4*	598.9	yes [gene body]	yes [gene body]	yes [throughout profile]	no	yes [gene body]	no	yes [gene body & downstream]
ADF3	1467.81	yes [throughout profile]	yes [gene body]	yes [throughout profile]	no	yes [start of gene]	no	yes [gene body]
ADF2	502.33	yes [throughout profile]	no	yes [gene body]	no	yes [gene body]	no	no
ADF7*	0.43	no	no	no	no	no	no	no
ADF10*	0.91	yes [upstream]	no	no	no	no	no	no
ADF8*	1.63	no	no	no	no	no	no	no
ADF11*	5.88	no	yes [start of gene]	yes [start of gene]	no	yes [start of gene]	no	no
ADF6	391.58	no	yes [gene body]	no	no	no	no	no
ADF5*	190.93	no	no	no	no	no	no	no
ADF9*	29.86	no	no	no	no	no	no	no

*Gene Pairs

Table 3.S3 (cont.): Summary of ChIP-Chip data for the Actin Depolymerizing Factor (ADF) gene family.

Gene Name	Expression Level (RFU)	H3K36me1 [Loc.]	H3K36me2 [Loc.]	H3K36me3 [Loc.]	H3K27me3 [Loc.]	H3K4me1 [Loc.]	H3K4me2 [Loc.]	H3K4me3 [Loc.]
ARP2	33.01	yes [gene body]	yes [downstream]	yes [throughout profile]	no	yes [gene body]	no	yes [start of gene]
ARP3	53.03	no	no	yes [gene body]	no	no	yes [gene body]	yes [start of gene]
ARP6	99.91	no	no	yes [upstream]	no	no	no	yes [upstream]
ARP5	44.4	no	no	yes [gene body]	no	yes [gene body]	yes [gene body]	yes [throughout profile]
ARP7	52.65	no	no	no	yes [throughout profile]	no	no	no
ARP8	100.41	yes [gene body]	no	yes [upstream & start of gene]	no	yes [start of gene]	yes [start of gene]	yes [upstream]
ARP4	138.68	no	yes [gene body]	yes [throughout profile]	no	yes [gene body]	yes [start of gene]	yes [upstream & start of gene]
ARP9	42.3	no	no	no	no	no	yes [start & end of gene]	yes [throughout profile]

Table 3.S4: Summary of ChIP-Chip data for the Actin Related Protein (ARP) gene family.

Gene Name	Expression Level (RFU)	H3K 9 ac [Loc.]	H3K14 ac [Loc.]	H3K23 ac [Loc.]	H4K5 ac [Loc.]	H4K8 ac [Loc.]	H4K12 ac [Loc.]	H4K16 ac [Loc.]
ARP2	33.01	yes [gene body]	yes [throughout profile]	yes [upstream & gene body]	no	yes [throughout profile]	no	yes [gene body]
ARP3	53.03	no	no	yes [gene body]	no	no	no	no
ARP6	99.91	yes [upstream]	yes [upstream]	yes [upstream]	no	yes [upstream]	no	yes [upstream]
ARP5	44.4	no	no	yes [throughout profile]	no	yes [throughout profile]	no	yes [gene body]
ARP7	52.65	yes [upstream]	no	no	no	yes [upstream]	no	no
ARP8	100.41	no	no	yes [upstream]	no	yes [upstream]	no	no
ARP4	138.68	yes [downstream]	no	yes [gene body]	no	no	no	no
ARP9	42.3	no	no	no	no	no	no	no

Table 3.S4 (cont.): Summary of ChIP-Chip data for the Actin Related Protein (ARP) gene family.

Correlations from Actin Gene Family Analysis																
		H3K36me1	H3K36me2	H3K36me3	H3K27me3	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K14ac	H3K23ac	H4K5ac	H4K8ac	H4K12ac	H4K16ac	5meC
H3K36me1	Pearson Correlation	1	-.189	.378	-.286	.189	.189	-.060	.378	.478	.189	^a	.378	^a	.478	.189
	Sig. (2-tailed)		.626	.316	.456	.626	.626	.879	.316	.193	.626		.316		.193	.626
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K36me2	Pearson Correlation	-.189	1	.250	-.189	.500	.500	.316	.250	.316	.125	^a	.250	^a	.316	-.250
	Sig. (2-tailed)			.516	.626	.170	.170	.407	.516	.407	.749		.516		.407	.516
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K36me3	Pearson Correlation	.378	.250	1	-.189	.500	.000	.791*	.500	.316	.500	^a	1.000**	^a	.316	-.500
	Sig. (2-tailed)				.626	.170	1.000	.011	.170	.407	.170		0.000		.407	.170
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K27me3	Pearson Correlation	-.286	-.189	-.189	1	-.378	.189	-.060	-.189	-.598	.189	^a	-.189	^a	.478	-.378
	Sig. (2-tailed)					.316	.626	.879	.626	.089	.626		.626		.193	.316
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K4me1	Pearson Correlation	.189	.500	.500	-.378	1	.500	.158	.500	.632	.250	^a	.500	^a	.158	-.500
	Sig. (2-tailed)				.316		.170	.685	.170	.068	.516		.170		.685	.170
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K4me2	Pearson Correlation	.189	.500	.000	.189	.500	1	-.316	.500	.158	.250	^a	.000	^a	.632	-.500
	Sig. (2-tailed)			1.000	.626	.170		.407	.170	.685	.516		1.000		.068	.170
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K4me3	Pearson Correlation	-.060	.316	.791*	-.060	.158	-.316	1	.316	.100	.395	^a	.791*	^a	.100	-.316
	Sig. (2-tailed)			.011	.879	.685	.407		.407	.798	.292		.011		.798	.407
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K9ac	Pearson Correlation	.378	.250	.500	-.189	.500	.500	.316	1	.316	.500	^a	.500	^a	.316	-.500
	Sig. (2-tailed)			.170	.626	.170	.170	.407		.407	.170		.170		.407	.170
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K14ac	Pearson Correlation	.478	.316	.316	-.598	.632	.158	.100	.316	1	.395	^a	.316	^a	.100	.158
	Sig. (2-tailed)				.089	.068	.685	.798	.407		.292		.407		.798	.685
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H3K23ac	Pearson Correlation	.189	.125	.500	.189	.250	.250	.395	.500	.395	1	^a	.500	^a	.395	-.500
	Sig. (2-tailed)				.626	.516	.516	.292	.170	.292			.170		.292	.170
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H4K5ac	Pearson Correlation	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a
	Sig. (2-tailed)															
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H4K8ac	Pearson Correlation	.378	.250	1.000**	-.189	.500	.000	.791*	.500	.316	.500	^a	1	^a	.316	-.500
	Sig. (2-tailed)			0.000	.626	.170	1.000	.011	.170	.407	.170				.407	.170
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H4K12ac	Pearson Correlation	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a	^a
	Sig. (2-tailed)															
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
H4K16ac	Pearson Correlation	.478	.316	.316	.478	.158	.632	.100	.316	.100	.395	^a	.316	^a	1	-.316
	Sig. (2-tailed)				.193	.685	.068	.798	.407	.798	.292		.407			.407
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
5meC	Pearson Correlation	.189	-.250	-.500	-.378	-.500	-.500	-.316	-.500	.158	-.500	^a	-.500	^a	-.316	1
	Sig. (2-tailed)			.170	.316	.170	.170	.407	.170	.685	.170		.170		.407	
	N	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
a. Cannot be computed because at least one of the variables is constant.

Table 3.S5: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among the actin gene family (9 genes total).

Correlations from Profilin Gene Family Analysis																
		H3K36me1	H3K36me2	H3K36me3	H3K27me3	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K14ac	H3K23ac	H4K5ac	H4K8ac	H4K12ac	H4K16ac	SmeC
H3K36me1	Pearson Correlation	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
	Sig. (2-tailed)															
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K36me2	Pearson Correlation	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
	Sig. (2-tailed)															
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K36me3	Pearson Correlation	a	a	1	a	-.408	a	1.000**	1.000**	1.000**	.612	a	1.000**	a	.612	a
	Sig. (2-tailed)					.495		0.000	0.000	0.000	.272		0.000		.272	
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K27me3	Pearson Correlation	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
	Sig. (2-tailed)															
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K4me1	Pearson Correlation	a	a	-.408	a	1	a	-.408	-.408	-.408	-.667	a	-.408	a	.167	a
	Sig. (2-tailed)			.495				.495	.495	.495	.219		.495		.789	
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K4me2	Pearson Correlation	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
	Sig. (2-tailed)															
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K4me3	Pearson Correlation	a	a	1.000**	a	-.408	a	1	1.000**	1.000**	.612	a	1.000**	a	.612	a
	Sig. (2-tailed)			0.000		.495			0.000	0.000	.272		0.000		.272	
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K9ac	Pearson Correlation	a	a	1.000**	a	-.408	a	1.000**	1	1.000**	.612	a	1.000**	a	.612	a
	Sig. (2-tailed)			0.000		.495		0.000		0.000	.272		0.000		.272	
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K14ac	Pearson Correlation	a	a	1.000**	a	-.408	a	1.000**	1.000**	1	.612	a	1.000**	a	.612	a
	Sig. (2-tailed)			0.000		.495		0.000	0.000		.272		0.000		.272	
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H3K23ac	Pearson Correlation	a	a	.612	a	-.667	a	.612	.612	.612	1	a	.612	a	.167	a
	Sig. (2-tailed)			.272		.219		.272	.272	.272			.272		.789	
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H4K5ac	Pearson Correlation	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
	Sig. (2-tailed)															
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H4K8ac	Pearson Correlation	a	a	1.000**	a	-.408	a	1.000**	1.000**	1.000**	.612	a	1	a	.612	a
	Sig. (2-tailed)			0.000		.495		0.000	0.000	0.000	.272				.272	
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H4K12ac	Pearson Correlation	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
	Sig. (2-tailed)															
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
H4K16ac	Pearson Correlation	a	a	.612	a	.167	a	.612	.612	.612	.167	a	.612	a	1	a
	Sig. (2-tailed)			.272		.789		.272	.272	.272	.789		.272			
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
SmeC	Pearson Correlation	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
	Sig. (2-tailed)															
	N	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

** . Correlation is significant at the 0.01 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Table 3.S6: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among the profilin gene family (5 genes total).

Correlations from ADF Gene Family Analysis																
		H3K36me1	H3K36me2	H3K36me3	H3K27me3	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K14ac	H3K23ac	H4K5ac	H4K8ac	H4K12ac	H4K16ac	5meC
H3K36me1	Pearson Correlation	1	.241	.083	.241	-.194	.463	-.261	-.039	-.039	-.039	. ^a	-.039	. ^a	-.289	-.375
	Sig. (2-tailed)		.476	.808	.476	.568	.152	.438	.910	.910	.910		.910		.389	.256
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K36me2	Pearson Correlation	.241	1	.770**	-.222	.671*	.356	-.043	.134	.624*	.624*	. ^a	.624*	. ^a	.389	.241
	Sig. (2-tailed)	.476		.006	.511	.024	.282	.900	.695	.040	.040		.040		.237	.476
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K36me3	Pearson Correlation	.083	.770**	1	.241	.516	.463	.149	.386	.810**	.810**	. ^a	.810**	. ^a	.770**	.083
	Sig. (2-tailed)	.808	.006		.476	.104	.152	.662	.241	.003	.003		.003		.006	.808
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K27me3	Pearson Correlation	.241	-.222	.241	1	-.149	.356	-.043	.134	.134	.134	. ^a	.134	. ^a	.389	-.289
	Sig. (2-tailed)	.476	.511	.476		.662	.282	.900	.695	.695	.695		.695		.237	.389
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K4me1	Pearson Correlation	-.194	.671*	.516	-.149	1	.239	.289	.418	.418	.418	. ^a	.418	. ^a	.671*	.516
	Sig. (2-tailed)	.568	.024	.104	.662		.479	.389	.200	.200	.200		.200		.024	.104
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K4me2	Pearson Correlation	.463	.356	.463	.356	.239	1	.069	.179	.179	.179	. ^a	.179	. ^a	.356	-.386
	Sig. (2-tailed)	.152	.282	.152	.282	.479		.840	.599	.599	.599		.599		.282	.241
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K4me3	Pearson Correlation	-.261	-.043	.149	-.043	.289	.069	1	.690*	-.069	.311	. ^a	.311	. ^a	.430	.149
	Sig. (2-tailed)	.438	.900	.662	.900	.389	.840		.019	.840	.353		.353		.186	.662
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K9ac	Pearson Correlation	-.039	.134	.386	.134	.418	.179	.690*	1	.214	.607*	. ^a	.607*	. ^a	.624*	-.039
	Sig. (2-tailed)	.910	.695	.241	.695	.200	.599	.019		.527	.048		.048		.040	.910
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K14ac	Pearson Correlation	-.039	.624*	.810**	.134	.418	.179	-.069	.214	1	.607*	. ^a	.607*	. ^a	.624*	-.039
	Sig. (2-tailed)	.910	.040	.003	.695	.200	.599	.840	.527		.048		.048		.040	.910
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H3K23ac	Pearson Correlation	-.039	.624*	.810**	.134	.418	.179	.311	.607*	.607*	1	. ^a	1.000**	. ^a	.624*	-.039
	Sig. (2-tailed)	.910	.040	.003	.695	.200	.599	.353	.048	.048			0.000		.040	.910
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H4K5ac	Pearson Correlation	. ^a														
	Sig. (2-tailed)															
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H4K8ac	Pearson Correlation	-.039	.624*	.810**	.134	.418	.179	.311	.607*	.607*	1.000**	. ^a	1	. ^a	.624*	-.039
	Sig. (2-tailed)	.910	.040	.003	.695	.200	.599	.353	.048	.048	0.000				.040	.910
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H4K12ac	Pearson Correlation	. ^a														
	Sig. (2-tailed)															
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
H4K16ac	Pearson Correlation	-.289	.389	.770**	.389	.671*	.356	.430	.624*	.624*	.624*	. ^a	.624*	. ^a	1	.241
	Sig. (2-tailed)	.389	.237	.006	.237	.024	.282	.186	.040	.040	.040		.040			.476
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
5meC	Pearson Correlation	-.375	.241	.083	-.289	.516	-.386	.149	-.039	-.039	-.039	. ^a	-.039	. ^a	.241	1
	Sig. (2-tailed)	.256	.476	.808	.389	.104	.241	.662	.910	.910	.910		.910		.476	
	N	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11

** . Correlation is significant at the 0.01 level (2-tailed).
* . Correlation is significant at the 0.05 level (2-tailed).
a . Cannot be computed because at least one of the variables is constant.

Table 3.S7: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among the ADF gene family (11 genes total).

Correlations from ARP Gene Family Analysis																
		H3K36me1	H3K36me2	H3K36me3	H3K27me3	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K14ac	H3K23ac	H4K5ac	H4K8ac	H4K12ac	H4K16ac	5meC
H3K36me1	Pearson Correlation	1	.333	.333	-.218	.577	-.149	.218	0.000	.333	.333	.a	.447	.a	.149	.333
	Sig. (2-tailed)		.420	.420	.604	.134	.725	.604	1.000	.420	.420		.267		.725	.420
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K36me2	Pearson Correlation	.333	1	.333	-.218	.577	-.149	.218	.577	.333	.333	.a	-.149	.a	.149	.333
	Sig. (2-tailed)	.420		.420	.604	.134	.725	.604	.134	.420	.420		.725		.725	.420
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K36me3	Pearson Correlation	.333	.333	1	-.655	.577	.149	.655	0.000	.333	1.000**	.a	.149	.a	.447	.333
	Sig. (2-tailed)	.420	.420		.078	.134	.725	.078	1.000	.420	0.000		.725		.267	.420
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K27me3	Pearson Correlation	-.218	-.218	-.655	1	-.378	-.488	-1.000**	.378	-.218	-.655	.a	.293	.a	-.293	-.655
	Sig. (2-tailed)	.604	.604	.078		.356	.220	0.000	.356	.604	.078		.482		.482	.078
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K4me1	Pearson Correlation	.577	.577	.577	-.378	1	.258	.378	0.000	0.000	.577	.a	.258	.a	.258	.577
	Sig. (2-tailed)	.134	.134	.134	.356		.537	.356	1.000	1.000	.134		.537		.537	.134
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K4me2	Pearson Correlation	-.149	-.149	.149	-.488	.258	1	.488	-.775*	-.745*	.149	.a	-.600	.a	-.467	.149
	Sig. (2-tailed)	.725	.725	.725	.220	.537		.220	.024	.034	.725		.116		.244	.725
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K4me3	Pearson Correlation	.218	.218	.655	-1.000**	.378	.488	1	-.378	.218	.655	.a	-.293	.a	.293	.655
	Sig. (2-tailed)	.604	.604	.078	0.000	.356	.220		.356	.604	.078		.482		.482	.078
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K9ac	Pearson Correlation	0.000	.577	0.000	.378	0.000	-.775*	-.378	1	.577	0.000	.a	.258	.a	.258	0.000
	Sig. (2-tailed)	1.000	.134	1.000	.356	1.000	.024	.356		.134	1.000		.537		.537	1.000
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K14ac	Pearson Correlation	.333	.333	.333	-.218	0.000	-.745*	.218	.577	1	.333	.a	.447	.a	.745*	.333
	Sig. (2-tailed)	.420	.420	.420	.604	1.000	.034	.604	.134		.420		.267		.034	.420
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H3K23ac	Pearson Correlation	.333	.333	1.000**	-.655	.577	.149	.655	0.000	.333	1	.a	.149	.a	.447	.333
	Sig. (2-tailed)	.420	.420	0.000	.078	.134	.725	.078	1.000	.420			.725		.267	.420
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H4K5ac	Pearson Correlation	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a
	Sig. (2-tailed)															
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H4K8ac	Pearson Correlation	.447	-.149	.149	.293	.258	-.600	-.293	.258	.447	.149	.a	1	.a	.600	.149
	Sig. (2-tailed)	.267	.725	.725	.482	.537	.116	.482	.537	.267	.725				.116	.725
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H4K12ac	Pearson Correlation	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a	.a
	Sig. (2-tailed)															
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
H4K16ac	Pearson Correlation	.149	.149	.447	-.293	.258	-.467	.293	.258	.745*	.447	.a	.600	.a	1	.447
	Sig. (2-tailed)	.725	.725	.267	.482	.537	.244	.482	.537	.034	.267		.116			.267
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
5meC	Pearson Correlation	.333	.333	.333	-.655	.577	.149	.655	0.000	.333	.333	.a	.149	.a	.447	1
	Sig. (2-tailed)	.420	.420	.420	.078	.134	.725	.078	1.000	.420	.420		.725		.267	
	N	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
a. Cannot be computed because at least one of the variables is constant.

Table 3.S8: Correlation analysis for 14 histone modifications and DNA Methylation (5meC) among the ARP gene family (8 genes total).

Correlations from Combined Gene Family Analysis																
		H3K36me1	H3K36me2	H3K36me3	H3K27me3	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K14ac	H3K23ac	H4K5ac	H4K8ac	H4K12ac	H4K16ac	SmeC
H3K36me1	Pearson Correlation	1	.194	.238	-.013	.142	.271	-.005	.122	.224	.115	. ^a	.238	. ^a	.070	.070
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K36me2	Pearson Correlation	.194	1	.436*	-.179	.497**	.293	.192	.293	.383*	.341	. ^a	.266	. ^a	.208	.208
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K36me3	Pearson Correlation	.238	.436*	1	-.072	.416*	.210	.587**	.454**	.527**	.782**	. ^a	.757**	. ^a	.527**	.149
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K27me3	Pearson Correlation	-.013	-.179	-.072	1	-.279	.123	-.150	.123	-.144	-.005	. ^a	.097	. ^a	.208	-.319
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K4me1	Pearson Correlation	.142	.457**	.416*	-.279	1	.193	.166	.193	.187	.262	. ^a	.284	. ^a	.324	.187
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K4me2	Pearson Correlation	.271	.293	.210	.123	.193	1	.168	.022	-.058	.113	. ^a	-.033	. ^a	.069	-.058
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K4me3	Pearson Correlation	-.005	.192	.587**	-.150	.166	.168	1	.414*	.139	.437*	. ^a	.465**	. ^a	.267	.267
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K9ac	Pearson Correlation	.122	.293	.454**	.123	.193	.022	.414*	1	.449**	.487**	. ^a	.576**	. ^a	.449**	-.058
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K14ac	Pearson Correlation	.224	.383*	.527**	-.144	.187	-.058	.139	.449**	1	.481**	. ^a	.527**	. ^a	.476**	.083
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H3K23ac	Pearson Correlation	.115	.341	.782**	-.005	.262	.113	.437*	.487**	.481**	1	. ^a	.658**	. ^a	.481**	.094
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H4K5ac	Pearson Correlation	. ^a														
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H4K8ac	Pearson Correlation	.238	.266	.757**	.097	.284	-.033	.465**	.576**	.527**	.658**	. ^a	1	. ^a	.527**	.023
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H4K12ac	Pearson Correlation	. ^a														
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
H4K16ac	Pearson Correlation	.070	.208	.527**	.208	.324	.069	.267	.449**	.476**	.481**	. ^a	.527**	. ^a	1	.083
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
SmeC	Pearson Correlation	.070	.208	.149	-.319	.187	-.058	.267	-.058	.083	.094	. ^a	.023	. ^a	.083	1
	Sig. (2-tailed)															
	N	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33

*. Correlation is significant at the 0.05 level (2-tailed).

***. Correlation is significant at the 0.01 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Table 3.S9: Correlation analysis for 14 histone modifications and DNA Methylation (SmeC) among actin, profilin, ADF, and ARP gene families (33 genes total).

Correlations from Combined Gene Family Analysis (Minus the ARPs)																
		H3K36me1	H3K36me2	H3K36me3	H3K27me3	H3K4me1	H3K4me2	H3K4me3	H3K9ac	H3K14ac	H3K23ac	H4K5ac	H4K8ac	H4K12ac	H4K16ac	5meC
H3K36me1	Pearson Correlation	1	.123	.204	.055	-.047	.408*	-.080	.161	.204	.040	^b	.161	^b	.042	-.047
	Sig. (2-tailed)		.558	.328	.796	.824	.043	.704	.442	.328	.848		.442		.843	.824
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K36me2	Pearson Correlation	.123	1	.452*	-.161	.369	.452*	.138	.169	.452*	.327	^b	.417*	^b	.236	.081
	Sig. (2-tailed)			.023	.442	.070	.023	.511	.420	.023	.110		.038		.256	.701
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K36me3	Pearson Correlation	.204	.452*	1	.089	.306	.167	.523**	.592**	.667**	.724**	^b	.921**	^b	.578**	-.076
	Sig. (2-tailed)				.672	.137	.426	.007	.002	.000	.000		.000		.002	.716
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K27me3	Pearson Correlation	.055	-.161	.089	1	-.245	.312	.017	.053	-.134	.167	^b	.053	^b	.355	-.245
	Sig. (2-tailed)		.796	.442	.672		.237	.129	.934	.802	.524	.425		.802	.082	.237
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K4me1	Pearson Correlation	-.047	.369	.306	-.245	1	.115	.022	.257	.306	.121	^b	.257	^b	.359	-.096
	Sig. (2-tailed)		.824	.070	.137	.237		.585	.915	.216	.137	.565		.216	.078	.646
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K4me2	Pearson Correlation	.408*	.452*	.167	.312	.115	1	.033	.263	.167	.066	^b	.099	^b	.238	-.268
	Sig. (2-tailed)		.043	.023	.426	.129	.585		.877	.204	.426	.755		.639	.252	.196
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K4me3	Pearson Correlation	-.080	.138	.523**	.017	.022	.033	1	.600**	.196	.368	^b	.600**	^b	.280	.022
	Sig. (2-tailed)		.704	.511	.007	.934	.915	.877		.002	.347	.071		.002	.175	.915
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K9ac	Pearson Correlation	.161	.169	.592**	.053	.257	.263	.600**	1	.428*	.623**	^b	.675**	^b	.510**	-.121
	Sig. (2-tailed)		.442	.420	.002	.802	.216	.204	.002		.033	.001		.000	.009	.565
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K14ac	Pearson Correlation	.204	.452*	.667**	-.134	.306	.167	.196	.428*	1	.559**	^b	.592**	^b	.408*	.115
	Sig. (2-tailed)		.328	.023	.000	.524	.137	.426	.347	.033		.004		.002	.043	.585
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H3K23ac	Pearson Correlation	.040	.327	.724**	.167	.121	.066	.368	.623**	.559**	1	^b	.786**	^b	.497*	-.068
	Sig. (2-tailed)		.848	.110	.000	.425	.565	.755	.071	.001	.004			.000	.012	.747
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H4K5ac	Pearson Correlation	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b
	Sig. (2-tailed)															
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H4K8ac	Pearson Correlation	.161	.417*	.921**	.053	.257	.099	.600**	.675**	.592**	.786**	^b	1	^b	.510**	-.121
	Sig. (2-tailed)		.442	.038	.000	.802	.216	.639	.002	.000	.002	.000			.009	.565
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H4K12ac	Pearson Correlation	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b
	Sig. (2-tailed)															
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
H4K16ac	Pearson Correlation	.042	.236	.578**	.355	.359	.238	.280	.510**	.408*	.497*	^b	.510**	^b	1	-.031
	Sig. (2-tailed)		.843	.256	.002	.082	.078	.252	.175	.009	.043	.012		.009		.882
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
5meC	Pearson Correlation	-.047	.081	-.076	-.245	-.096	-.268	.022	-.121	.115	-.068	^b	-.121	^b	-.031	1
	Sig. (2-tailed)		.824	.701	.716	.237	.646	.196	.915	.565	.585	.747		.565	.882	
	N	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25

*. Correlation is significant at the 0.05 level (2-tailed).

***. Correlation is significant at the 0.01 level (2-tailed).

^b. Cannot be computed because at least one of the variables is constant.

^b. Cannot be computed because at least one of the variables is constant.

Table 3.S10: Correlation analysis for 14 histone modifications and DNA

Methylation (5meC) among actin, profilin, and ADF gene families (25 genes total).

CHAPTER 4

SUMMARY AND CONCLUSIONS

This dissertation has covered two widely different topics, epigenetics and the genetic characterization of a family of actin binding proteins (ABPs) in *Arabidopsis thaliana*. For the sake of clarity, I will conclude each subject matter separately.

Genetic analysis of vegetative profilins in *Arabidopsis thaliana*

The actin cytoskeleton plays fundamental roles in development and cellular function. Its involvement in cellular processes such as organelle movement, chromatin remodeling, and cell division are well documented (Bettinger et al., 2004; Miralles and Visa, 2006; Kandasamy et al., 2012). In order for the actin cytoskeleton to facilitate this wide array of functions, it must interact with a multitude of ABPs. There are various classes of proteins that contribute to the dynamic nature of the actin cytoskeleton; one such group of proteins is encoded by the profilin gene family.

Similar to actin, profilins are highly dynamic and bind to a number of different proteins. Through these interactions they have been implicated in signaling pathways, intracellular transport, communication, organ and tissue development, cell elongation, and recently cancer progression (Stradal et al., 2004; Witke et al., 1998; Zou et al., 2010; Mouneimne et al., 2012). While profilins were originally thought to primarily be

facilitating actin depolymerization through the process of binding and sequestering G-actin monomers (Carlsson et al., 1976), recent evidence has emerged showing that profilin-bound actin monomers can elongate filament barbed ends at approximately the same rate as free actin monomers (Pollard and Borisy, 2003 and dos Remedios et al., 2003), thus implicating profilins in actin polymerization as well. What drew my interest to the profilin gene family was that after decades of research their mechanistic roles in modulating actin dynamics have continued to remain unclear. Additionally, their effects on the overall development of the model organism *Arabidopsis thaliana* are unknown. This prompted my investigation into the genetic characterization of the vegetative subclass of profilins in *Arabidopsis*.

Arabidopsis thaliana encodes a five member multi-gene family of profilin proteins. PRF1, PRF2, and PRF3 belong to the vegetative class and are constitutively expressed throughout all vegetative tissues and in ovules. PRF2 is the most highly expressed member in this family, followed by PRF1 and finally PRF3. PRF4 and PRF5 are classified as reproductive profilins, being predominately expressed in mature pollen. Since I was trying to study the effects of profilins on overall plant development, I decided to only characterize the three vegetative profilin proteins. Previous analyses have shown that a partial knockdown of the vegetative profilin PRF1 results in altered seedling development, elongated hypocotyls, loss of light regulation, defects in root hair development, flowering time, cell elongation, and overall cell shape maintenance (McKinney et al., 2001; Ramachandran et al., 2001). These studies only looked at one of the three members in the vegetative profilin family. Furthermore, these phenotypes were not abundantly noticeable, suggesting that complete knockouts as well as double and

triple knockouts containing all combinations of these three genes were needed. To dissect the functions of these vegetative proteins, I utilized various T-DNA single and double mutants and RNAi knockdown plants that silenced multiple profilins simultaneously. This is the first documented case where null profilin mutants and double mutants were characterized in plants. By knocking down multiple profilins at once, I was able to explore quantitative genetic effects and the likelihood of functional redundancy amongst the vegetative members of this gene family.

Morphological analysis of single T-DNA insertion mutants revealed that plants deficient in PRF1 or PRF2 led to similar defects in rosette leaf morphology and inflorescence stature, while the loss of PRF3 led to less evident phenotypes like elongated petioles. PRF1 and PRF2 mutant plants developed leaves that were significantly shorter in total length, width, and blade length, and were also shorter in overall plant height, with inflorescences appearing less stable than that of WT (Chapter 2, Figure 2.1). This lack of structure suggested that the actin cytoskeleton was impaired in the cells that make up the inflorescences in these mutants. If profilins were responsible for shuttling monomeric actin to promote filamentous formation, perhaps the lack of these highly expressed profilins was inhibiting the formation of actin-filaments at the leading edge of cells resulting in a lack of structure in these tissues. This idea was in agreement with profilins being involved in proper cell elongation. PRF3 mutants did not display a strong, dwarfed leaf or unstable inflorescence phenotype, but did appear to have slightly elongated petioles compared to WT (Chapter 2, Figure 2.6). This result indicated that PRF3 could be required for proper petiole formation. Since this observation was the first documented instance where a PRF3 mutant plant showed any type of developmental phenotype,

further studies will be needed in order to fully understand PRF3's role in petiole and overall plant development.

The generation of double mutants showed combinations of the single mutant phenotypes, with knocking out the two most highly expressed vegetative profilins (PRF1 and PRF2) leading to the most drastic and severe leaf and inflorescence phenotypes (Chapter 2, Figure 2.3). These results confirmed that while PRF1 and PRF2 was playing major roles in rosette and inflorescence development, PRF3 must be involved in the proper development of petioles. The much lower expressed PRF3 seemed to have evolved to function specifically in the assistance of petiole development, while PRF1 and PRF2 serve to function in multiple tissues. The fact that PRF1 and PRF2 deficient plants exhibited very similar phenotypic effects suggested the possibility of there being partial functional redundancy. However, since the single mutants each had strong phenotypes we believe that this is more indicative of a quantitative genetic effect. Additional experiments will be required to determine the extent, if any, of functional redundancy between PRF1 and PRF2.

Further examination of possible quantitative genetic effects among *Arabidopsis* profilins was done by silencing all three vegetative proteins simultaneously. Plants deficient in PRF1, PRF2, and PRF3 showed the most drastic dwarfed phenotypes as well as problems with lateral root initiation and growth (Chapter 2, Figures 2.5 and 2.7). When the vegetative profilin pool was completely depleted, plants were unable to fully form many of its above ground tissues and organs. This stunted phenomenon was indicative of defects in cell elongation. I propose that once vegetative profilin protein levels reach a certain threshold, the cell must conserve the profilin that is present and

only initiate cell elongation in certain tissues and organs that are more essential to development and survival. Another possible explanation is that the reproductive profilins are functionally filling in for the vegetative PRFs, but are less efficient or unable to promote cell elongation throughout all tissues in the plant. Since weakly silenced lines gave rise to intermediate phenotypes (Chapter 2, Figure 2.5 and 2.S3), it appears as if there is a concentration dependent aspect to this relationship.

After completing morphological examination of these various profilin deficient plants, I performed microscopic analysis to search for possible defects in cell elongation. By looking at leaf epidermal cells in WT and plants lacking PRF1 and PRF2, I found that profilin deficient plants had dramatically smaller cells than those seen in WT, suggesting that vegetative profilins were playing an active role in cell elongation (Chapter 2, Figure 2.8). This result explained why the leaves in these mutant plants were dwarfed throughout development. Based on these results I have developed a model for profilin's role in cell elongation based upon overall profilin concentrations in the cell. I propose that significantly decreasing the profilin pool causes an arrest in cell elongation. The lowering of profilin pool concentrations leads to less profilin-actin complexes being properly sequestered to the cell periphery to promote appropriate cell expansion. In addition, since profilin is involved in actin treadmilling, which is required for proper cell elongation, the lack of profilin would inevitably cause arrest in actin filament protrusions leading to plants with smaller leaves, as was seen in my PRF mutants. This model agrees with my findings that there are quantitative genetic effects that provide a direct correlation between the number of profilin genes that were knocked out and the severity of the dwarfed plant phenotype.

While this model serves to explain the phenotypic effects detected in profilin deficient plants, the exact mechanisms still need to be clarified in future experiments. Preliminary results from actin-inhibiting drug treatments were inconclusive, but additional focus on these types of experiments could be useful in deducing profilin's mechanistic interactions in actin cytoskeletal dynamics. This study analyzed profilin's role in plant cell and organ development, but additional research will be needed to examine their roles in signal transduction, intracellular transport, and communication. In order to determine how profilins diverse functions are interacting, a systems biology approach needs to be taken. I suggest beginning with genome-wide transcriptional assays examining these profilin deficient plants. By determining which genes are being differentially expressed, we could help link specific pathways through previously determined interactions. Furthermore, experiments are also needed to establish if there is any functional redundancy between PRF1 and PRF2. By overexpressing one of these profilins in plants lacking the other, we can see if phenotypes still persist or are rescued due to the restoration of overall profilin protein concentrations. Studies dissecting the importance of reproductive PRF4 and PRF5 are also necessary. While initial analysis of plants lacking these profilins did not reveal any noticeable phenotypes, perhaps PRF4/PRF5 double mutants will lead to defects in pollen development and sterility. Creation of this double mutant could also assist in understanding the extent of functional redundancy among reproductive profilins as well.

Phylogenetic analysis has shown that PRF1 and PRF2 as well as PRF4 and PRF5 exist as duplicated gene pairs stemming from a genome wide duplication event occurring in *Arabidopsis thaliana* ~30-35 million years ago. These gene pairs are commonly

noticed across a number of cytoskeletal gene families in *Arabidopsis*. The next section will utilize histone modification deposition data in an effort to determine if this epigenetic control mechanism helps aid evolution through the silencing of certain duplicated gene pair members until subfunctionalization and mutations can occur.

Phylogenetic Identification of Inherited Patterns of Nucleosomal Histone

Modification

Epigenetics can be described as inherited changes in gene function that are unable to be explained by the classical central dogma of molecular genetics (Riggs et al., 1996). There are a variety of different types of epigenetic control mechanisms, including nucleosome positioning/phasing and histone variants (higher order chromatin structure), DNA based alteration (cytosine methylation), and histone side chain modifications. In order to further the understanding of the epigenetic field, I believe that each of these processes needs to be dissected thoroughly to decipher their mechanistic links. Determining how these processes are interacting will help elucidate the wider picture of how epigenetics is controlling global gene regulation through the preparation of chromatin structure. Much of my research has suggested that DNA sequence could be playing a significant role in guiding these control mechanisms. While the review article in the appendix of this dissertation discusses such issues in more detail, Chapter 3 serves to extend the analysis of histone modification deposition in an effort to comprehend how DNA sequence may be facilitating the epigenomic landscape of histone post-translational modifications.

Genome-wide histone modification deposition studies have been ongoing for some time now, with one of their main purposes being to decipher the complicated “histone code” (Jenuwein and Allis, 2001). My analysis has taken advantage of previously generated ChIP-chip data for 14 different modifications across the *Arabidopsis thaliana* genome (Chapter 3, Table 3.1). I took a novel approach to try to understand histone modification deposition by extrapolating this data to examine specific enrichment patterns in individual gene family phylogenies. I selected four cytoskeletal gene families that comprise a wide range of sequence divergence in *Arabidopsis thaliana* (Chapter 3, Table 3.2). Three of them (actin, profilin, actin-depolymerizing factor (ADF)) have paired family members representing a well-established, recent genome-wide duplication event ~30-35 MYA, while the origin of most of the members in the fourth family, Actin-Related Proteins (ARPs), date back to protists and have no discernable gene pairs (Vision et al., 2000; Simillion et al., 2002; Bowers et al., 2003; Blanc et al., 2003). By evaluating how recently duplicated gene pairs compared in their overall deposition of various histone modification types, I attempted to understand if and how evolution was utilizing epigenetic control mechanisms to assist in regulating different genic variants for spatial and temporal development.

Using this data, I constructed enrichment profiles for each modification type for every gene member in these four families and found that recently duplicated gene pairs exhibited varying levels of conservation across their histone modification enrichment profiles (Chapter 3, Table 3.3). These results suggested that histone modifications were aiding “evolution by gene duplication” by mechanistically silencing some recent gene duplicates, but not others, until beneficial mutations and subfunctionalization could occur

(Rodin and Riggs, 2003; Meagher, 2010). These results agreed with previously published findings in humans showing that there was a clear variable range of agreement among histone modification enrichments between different segmental duplication pairs, with parental (original) loci exhibiting higher levels of histone modifications than derivative (duplicated) loci (Zheng, 2008). Some of my results were also supported by studies in yeast showing that, on average, duplicate gene pairs share more common histone modification patterns than random singleton pairs (Zou et al., 2012).

Since recently duplicated gene pairs have very similar sequences, I had reason to believe that sequence might be driving the control of histone modification deposition (Meagher and Müssar, 2012). Evidence in yeast showed that histone modification profiles between gene duplicates begin to exhibit even more deviation as sequence divergence increases (Zou et al., 2012). This notion supports the idea that there may be a co-evolution between both genetic and epigenetic elements following gene duplication that is contributing to the expression divergence seen among recently duplicated genes. To explore these histone modification interactions with respect to sequence degeneration, I selected gene families that contained varying levels of sequence divergence (Chapter 3, Table 3.2), but failed to detect any differences between the gene pairs from the different cytoskeletal gene families. Each family had pairs that agreed, disagreed, and intermediately agreed with respect to their conservations of histone modification enrichments.

By examining the locations of these PTMs within duplicated gene pairs, I was able to clearly see that PTM enrichments were occurring at different locations within each gene. In fact, in most cases the PTM enrichments seem to be located on the

neighboring nucleosome location in the respective gene duplicate (Chapter 3, Figures 3.3-5). We showed clear examples of this on/off enrichment patterning, suggesting that sequence was not playing a role in facilitating PTM deposition amongst gene duplicates. Instead, it appeared that PTMs were acting to promote the subfunctionalization of the gene pair. Perhaps the specific locations of these PTMs were determining which roles these variants were playing throughout spatial and temporal development, thus leading to subfunctionalization.

To determine the extent, if any, of the association between sequence divergence and histone modification deposition, this study could benefit from expanding out to include more gene families in *Arabidopsis*, keeping with the theme of selecting families that vastly range in their sequence conservation. Another option would be to use previously designed algorithms (Gu et al., 2002) to uncover extremely recent duplicated gene pairs (~5-10 Mya) to perform this analysis. Nonetheless, continued analysis of gene duplication, in the context of histone modifications, could help explain how PTM deposition diverges over time to regulate chromatin structure and therefore gene expression.

These gene specific enrichment profiles were then analyzed to look for distinct combinatorial patterns of histone modification marks within each family. Previous genome-wide analysis of 39 different histone modifications in human CD4+ T cells lead to the discovery of a pattern consisting of 17 different modifications (containing both methylations and acetylations) that colocalize in the genome and correlate with each other on an individual nucleosome level. This pattern associated with genes that tended to be more highly expressed, suggesting that these modifications were acting

cooperatively to prepare chromatin for active transcription (Wang et al., 2008). Using our *Arabidopsis* data, I tried to detect if such patterns existed within and among the actin, profilin, ADF, and ARP gene families.

By calculating Pearson product-moment correlation coefficients (Pearson's r) using the data gathered from Tables S1-S4 in Chapter 3, I was able to detect interactions between different histone modification types in three of the four gene families analyzed. I found that for the actin gene family, a pattern of H3K36me3, H3K4me3, and H4K8ac existed (Chapter 3, Table 3.4). This means that each of these modifications were more likely to be present when the other two were enriched on that same gene (and vice versa). This type of interaction was more extensively seen in the profilin gene family, where a pattern of five modification types was detected: H3K36me3, H3K4me3, H4K8ac, H3K9ac, and H3K14ac (Chapter 3, Table 3.4). Analysis of the ADF gene family only revealed an interacting pattern containing two modification types, H3K36me3 and H4K8ac (Chapter 3, Table 4). It should be noted that H3K36me3 and H4K8ac were detected in patterns throughout these three gene families, providing additional evidence that these findings are not just artifacts or a result of small sample size. All modifications in these combinatorial patterns were known activation marks, and correlations were positive indicating that these interactions are all acting cooperatively. The ancient and highly divergent ARPs contained no combinatorial patterns suggesting that sequence divergence could be facilitating these combinatorial interactions. This result was further supported by the fact that these specific interacting PTMs are enriched on the same nucleosome within a single gene (Chapter 3, Figures 3.6-8). Perhaps these patterns are imbedded in the nucleotide sequence and acting to recruit proteins like histone

acetyltransferases (HATs), histone deacetylases (HDACs), and histone methyltransferases (HMTs) to the nucleosomes to facilitate PTM deposition, in a sequence dependent manner.

By extending this analysis to include more gene families, I could then extract the specific sequence portions of each gene where these combinatorial patterns of enrichment are occurring, and use pattern-based discovery algorithms, like the TEIRESIAS algorithm (Rigoutsos and Floratos, 1998), to search for statistically significant patterns among these regions that may be attracting specific PTMs. By searching the genome for regions containing these patterns and determining if the same PTMs are enriched at these locations as well, I could further confirm the link between nucleotide sequence and PTM deposition.

It is improbable to expect that there is a single, specific activation module that can apply to all genes throughout the genome. Maybe these combinatorial patterns are associated with relative spatial locations on certain chromosomes, or are coupled with expression levels at a specific developmental time-point. I believe that these PTM patterns are likely altered throughout development to correspond with the changing need for temporal specific functions. Performing this analysis at different time points during development, thereby incorporating PTM turnover could provide insight into these inquiries. Recently, quantitative proteomics studies have suggested that histone turnover rates are dependent upon site-specific post-translational modifications and sequence variants (Zee et al., 2010), thus providing a link between sequence, PTMs, and histone variants. Determining how these interacting PTM patterns change at various time points and under certain stress conditions could help understand how mechanistically these

modifications are communicating to facilitate gene expression regulation throughout development and in response to specific stressed environments.

Another aspect worth examining is the role that certain classes of non-coding RNAs (ncRNAs) play in facilitating histone modification deposition. ncRNAs have already been implicated in a number of epigenetic control mechanisms (Chapter 1) and many are beginning to believe that these small ncRNAs have much more complex roles than initially perceived. However, since these species of RNAs are known to act over large distances in the genome, it will be difficult to determine which ones are acting on specific gene sequences through bioinformatic approaches alone. Efforts combining *in silico* data mining and biochemical functional validation could eventually deduce the mechanistic links I am searching for among these control processes.

In summation, I believe that taking more focused approaches to analyzing epigenetic control mechanisms could help us appreciate the complexity of these interactions. While this study has attempted to dissect these interactions among certain histone modifications, further analysis is required. I have provided initial evidence throughout this dissertation suggesting that DNA sequence may be playing a much larger role in facilitating the epigenomic landscape than previously thought. As this field continues to gain momentum, it will be interesting to see how important sequence is in guiding epigenetic control.

REFERENCES

- Bettinger BT, Gilbert DM, Amberg DC. 2004. Actin up in the nucleus. *Nature Reviews Molecular Cell Biology* **5**:410-415.
- Blanc G, Hokamp K, Wolfe KH. 2003. A Recent Polyploidy Superimposed on Older Large-Scale Duplications in the *Arabidopsis* Genome. *Genome Res* **13**:137-144.
- Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**:433-438.
- Carlsson L, Nystrom L, Sundkvist I, Markey F, Lindberg U. 1976. Profilin, a low-molecular weight protein controlling actin polymerisability. In *Contractile Systems in Non Muscle Tissues*. Elsevier Amsterdam, 39-49.
- dos Remedios CG, Chhabra D, Kekic M, Dedova IV, Tsubakihara M, et al. 2003. Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol Rev* **83**:433-473.
- Gu Z, Cavalcanti A, Chen F-C, Bouman P, Li W-H. 2002. Extent of Gene Duplication in the Genomes of Drosophila, Nematode, and Yeast. *Mol Biol Evol* **19**:256-262.
- Jenuwein T, Allis C. 2001. Translating the histone code. *Science* **293**:1074-80.
- Kandasamy MK, McKinney EC, Roy EM, Meagher RB. 2012. Plant vegetative and animal cytoplasmic actins share functional competence for spatial development with protists. *Plant Cell* **24**:2041-2057.
- Meagher RB. 2010. The evolution of epitype. *The Plant Cell* **22**:1658-1666.
- Meagher RB, Müssar KJ. 2012. The Influence of DNA Sequence on Epigenome-induced Pathologies. *Epigenetics & Chromatin* **5**:11.
- McKinney EC, Kandasamy MK, Meagher RB. 2001. Small Changes in the Regulation of One Arabidopsis Profilin Isovariant, PRF1, Alter Seedling Development. *Plant Cell* **13**:117-1191.
- Miralles F and Visa N. 2006. Actin in transcription and transcription regulation. *Curr Opin Cell Biol* **18**:261-66.
- Mouneimne, G. et al. 2012. Differential remodeling of actin cytoskeleton architecture by profilin isoforms leads to distinct effects on cell migration and invasion. *Cancer Cell* **22**:615-630.

- Pollard TD and Borisy GG. 2003. Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112**:453–65.
- Ramachandran S, Christensen H, Ishimaru Y, Dong CH, Chao-Ming W, Cleary AL, Chua NH. 2001. Profilin Plays a Role in Cell Elongation, Cell Shape Maintenance, and Flowering in Arabidopsis. *Plant Physiology* **124**:1637–1647.
- Riggs AD, Martienssen RA, Russo VEA. 1996. Introduction. In Epigenetic mechanisms of gene regulation. Edited by Russo VEA. *Cold Spring Harbor, NY Press*.
- Rigoutsos I, Floratos A. 1998. Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm. *Bioinformatics* **14**:55–67.
- Rodin SN, Riggs AD. 2003. Epigenetic Silencing May Aid Evolution by Gene Duplication. *Molecular Evolution* **56**:718–729.
- Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of . *PNAS* **99**:13627–13632.
- Stradal TE, Rottner K, Disanza A, Confalonieri S, Innocenti M, Scita G. 2004. Regulation of actin dynamics by WASP and WAVE family proteins. *Trends Cell Biol* **14**:303–311.
- Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* **290**:2114–2117.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K. 2008. Combinatorial patterns of histone acetylation and methylation in the human genome. *Nat Genet* **40**:897–903.
- Witke W, Podtelejnikov AV, Di Nardo A, Sutherland JD, Gurniak CB, et al. 1998. In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly. *EMBO J* **17**:967–976.
- Zee BM, Levin RS, Dimaggio PA, Garcia BA. 2010. Global turnover of histone post-translational modifications and variants in human cells. *Epigenetics and Chromatin* **3**:22.
- Zheng D. 2008. Asymmetric histone modifications between the original and derived loci of human segmental duplications. *Genome Biology* **9**:R105.
- Zou L, Jaramillo M, Whaley D, Wells A, Panchapaksea V, Das T, Roy P. 2007. Profilin-1 is a negative regulator of mammary carcinoma aggressiveness. *British Journal of Cancer* **97**:1361–1371.

Zou Y, Su Z, Huang W, Gu X. 2012. Histone modification pattern evolution after yeast gene duplication. *BMC Evolutionary Biology* **12**:111.

APPENDIX A

THE INFLUENCE OF DNA SEQUENCE ON EPIGENOME-INDUCED
PATHOLOGIES¹

¹Richard B. Meagher and Kristofer J. Müssar. 2012. *Epigenetics & Chromatin* 5:11.
Reprinted here with permission of publisher. Copyright 2012.

Abstract

Background

Clear cause-and-effect relationships are commonly established between genotype and the inherited risk of acquiring human and plant diseases and aberrant phenotypes. By contrast, few such cause-and-effect relationships are established linking a chromatin structure (i.e., the epitype), with the transgenerational risk of acquiring a disease or abnormal phenotype. It is not entirely clear how epitypes are inherited from parent to offspring as populations evolve, even though epigenetics is proposed to be fundamental to evolution and the likelihood of acquiring many diseases. This article explores the hypothesis that for transgenerationally inherited chromatin structures “*genotype predisposes epitype*”, and that epitype functions as a modifier of gene expression within the classical central dogma of molecular biology.

Results

Evidence for the causal contribution of genotype to inherited epitypes and epigenetic risk comes primarily from two different kinds of studies discussed herein. The first and direct method of research proceeds by the examination of the transgenerational inheritance of epitype and the penetrance of phenotype among genetically related individuals. The second approach identifies epitypes that are duplicated as DNA sequences are duplicated and evolutionarily conserved among repeated patterns in the DNA sequence. The body of this article summarizes particularly robust examples of these studies from humans, mice, Arabidopsis, and other organisms.

Conclusions

The bulk of the data from both areas of research support the hypothesis that genotypes predispose the likelihood of displaying various epitypes, but for only a few classes of epitype. This analysis suggests that renewed efforts are needed in identifying polymorphic DNA sequences that determine variable nucleosome positioning and DNA methylation as the primary cause of inherited epigenome-induced pathologies. By contrast, there is very little evidence that DNA sequence directly determines the inherited positioning of numerous and diverse post-translational modifications of histone side chains within nucleosomes. We discuss the medical and scientific implications of these observations on future research and on the development of solutions to epigenetically induced disorders.

Introduction

Cause-and-effect and epigenetic risk: The inheritance of numerous genetic risk factors for human and plant diseases as well as biotic and abiotic stress susceptibility phenotypes are well established [1-6]. Particular DNA mutations and their mechanistic effect on the timing, level, or quality of gene expression produce the risk of disease. Thus, a clear cause-and-effect relationship is established between the inherited aberrant **genotype** and the risk **phenotype** (i.e., the increased chance or certainty of presenting a disease).

Epigenetics is cited as contributing to the risk of acquiring numerous diseases and aberrant phenotypes in human and plant populations based primarily on correlations between changes in chromatin structure and penetrance of the undesired phenotype [7-10]. There has been a growing suspicion, particularly since the 1980s, that along with

classical genetics, epigenetics was required to explain many complex phenotypes associated with disease [11, 12]. The influences of age and environment (e.g., chemicals, heat, nutrition, daylight) on various pathologies and the seemingly stochastic penetrance of developmental abnormalities are particularly difficult to interpret using purely molecular genetic models and more easily explained by considering epigenetic control mechanisms [13-18]. However, few cause-and-effect relationships have been established proving that particular inherited cis-linked chromatin structures (epitypes) are in fact useful in predicting the inherited risk of acquiring disease phenotypes. Exceptions are the epigenetic silencing of the *skeletal-muscle ryanodine-receptor gene (RYR1)* that causes congenital myopathies and the *MutL Homolog 1 gene (MLH1)* that causes increased risk of colorectal or endometrial tumors, which are discussed in the following section.

Inherited risk epitypes should evolve in populations in ways similar to the evolution of genotypes [19]. The problem is that the transgenerational inheritance of epigenetic controls is not well understood in any multicellular organism and often difficult to prove. This is particularly true in humans or agricultural crops, where the need for understanding epigenetic risk is the greatest [20-27]. Without knowledge about the molecular basis for the transgenerational inheritance or generational reprogramming of defined epigenetic risk factors that contribute to disease, it is difficult to design effective targeted therapeutics for humans or to knowledgeably alter breeding programs for crops that will avoid the onset of a disease phenotype [28-32].

This study explores the cause-and-effect relationships among **genotype**, **epitype** and **phenotype**, where the epitype of a single gene or an entire genome is defined as its various cis-linked chromatin structures (**Figure A1.1**) [19]. Thus, epitype includes, but is

not limited to chromatin domain structures, such as large DNA loops, the position of all nucleosomes and of subclasses of nucleosomes with particular histone variant compositions (e.g., H2A or H2AZ or H2AX), DNA cytosine methylation, and a myriad of histone post-translational modifications (PTMs) [33-35]. By focusing on epitype, we eliminate from consideration several other classes of epigenetic controls such as cell-to-cell communication by morphogens or the inheritance of cell surface patterning [36-40]. Addressing these other epigenetic controls would distract this discussion from a focus on the transgenerational inheritance of chromatin structures.

A working hypothesis that emerged from a preliminary examination of the inheritance and evolution of various epitypes [19] is that “*genotype predisposes epitype*” for most transgenerationally inherited chromatin structures. Only epitypes that are transgenerationally inherited at significant frequencies may contribute to the primary cause of inherited epigenetic risk. Within this hypothesis, epitype and the machinery that alter epitype function as modifiers of the central dogma of molecular biology (DNA → RNA → Protein) influencing the activity of DNA and RNA, as shown in **Figure A1.2A**. In addition, we will discuss how particular DNA and RNA sequences strongly influence the penetrance of some epitypes and resulting phenotypes. By this view transgenerationally inherited epitypes are not acting at a higher level than or independent of DNA sequence in determining phenotype (e.g., RNA and protein expression, disease phenotype).

It will be useful at this point to make the distinction between the transgenerational epigenetic inheritance among parents and offspring and the somatic inheritance between mother and daughter cells within developing tissues and organs [20, 23, 26, 41-45]. The

inheritance of epitypes between dividing somatic cells, such as the transmission of histone PTM [46], is undoubtedly essential to tissue and organ development [47-49] and may be subject to various environmental influences that reveal a phenotype [50], however, inheritance among somatic cells need not contribute causally to epigenetic inheritance across organismal generations. Again, we are interested herein in identifying epitypes that may be the primary cause of transgenerationally inherited epigenetic risk of acquiring a disease phenotype.

To test this hypothesis our discussion is focused on finding evidence for gene-specific epitypes that supports or rejects cause-and-effect relationships between **genotype, epitype** and **phenotype**. Some of the strongest evidence we found, for or against our hypothesis, is summarized in **Table A1.1**, and comes from two different research strategies. The first approach (A) examines the penetrance of transgenerationally inherited epitypes that are known to activate or silence the expression of disease related gene(s), which in turn correlate with onset of the aberrant “disease” phenotypes. This direct approach requires the measurement of the frequency of the transgenerational inheritance of causative epitype(s), the relevant gene expression pattern(s), and the aberrant phenotype(s) among related individuals in a population known to be at risk. This method is powerful, produces convincing data, and in several cases reveals the clear contribution of genotype to epitype. But transgenerational measurements are very expensive and time consuming, particularly in the early stages of establishing cause-and-effect relationships to human or agricultural diseases.

The second and less direct approach (B) searches out epitypes that are duplicated, as DNA sequences are duplicated, and examines multiple copies of DNA sequence and

epitype that have been evolutionarily co-conserved. This approach establishes an unambiguous and in many cases a statistically significant correlation of a particular epitope with highly reiterated DNA sequence motifs, and/or examines the conservation of an epitope among duplicated gene sequences. With this strategy, the evolutionary conservation of epitopes among conserved sequences is used as a filter to identify epitopes that were transgenerationally inherited [19, 51]. In other words, those epitopes that are widely conserved in their sequence position across the genome or may be shown to evolve by gene duplication within a gene family have almost certainly been inherited through past generations. Again, only epitopes that are transgenerationally inherited have the potential to contribute causally to inherited risk. This second approach simplifies analyses, because the initial screening for likely transgenerationally inherited epitopes may be made within a single genome and in one generation. Conversely, an epitope that is not inherited after gene duplication is less likely to be closely and causally related to phenotype even if its presence in an allele correlates well with the disease phenotype. Hence, epitopes not inherited via DNA sequence duplications are likely to be poor predictors of epigenetic risk. The disadvantage of this second genome-centered approach is precisely that it is not focused on finding associated risk phenotypes and during the early stages of analysis we are frequently left with very large data sets describing relationships among epitopes and genotypes without yet knowing correlated pathologies.

A. Direct measurement of transgenerational epigenetic inheritance: Only a handful of studies have succeeded in fully demonstrating that the transgenerational transmission of an epitope produces changes in known target gene expression, which results in a disease or its risk of penetrance (i.e., a causal relationship between genotype,

epitype, and risk). Two of the best examples from humans concern chromatin structure at the *RYR1* and *MLH1* genes, resulting in muscle myopathies and cancer, respectively.

However, the complexity of the data on these two systems highlights the problems that arise when trying to establish such cause-and-effect epigenetic relationships, particularly in humans.

(1) *RYR1*: Genetic mutations causing a loss of expression of *RYR1* function are associated with susceptibility to malignant hyperthermia and congenital myopathies (e.g., central core disease, multiminicore disease) [52-54]. However, many individuals with core myopathy disease are known to be heterozygous for a mutant defective *ryr1* allele [54, 55]. The epigenetic silencing of the otherwise functional *RYR1* allele appears to account for the loss of functional RYR1 protein expression. For example, among a sampling of eleven patients with the disease, six patients showed tissue-specific silencing of the maternally inherited functional *RYR1* allele, which apparently resulted from cytosine hypermethylation of that allele [56]. Treating skeletal-muscle myoblasts cultured from these patients with 5-aza-deoxycytidine, an inhibitor of cytosine methylation in newly replicated DNA, reactivates the transcription of the epigenetically silenced, but otherwise functional allele. These data strongly support the view that hypermethylation is the primary cause of *RYR1* silencing and onset of an epigenetically determined form of the disease (**Figure A1.2**). However, the particular region(s) of DNA in which cytosine residues are methylated to cause gene silencing has not been identified in spite of intense efforts to identify it among three CG islands within the gene. This leaves open the possibility that an epigenetically controlled transacting factor is the causative agent [56]. Thus, for *RYR1* there is not yet a clear causal link between an aberrant genotype, epitype,

and the silenced *RYRI* gene expression producing the disease (**Table A1.1**).

(2) *MLH1*: The human *MLH1* gene encodes a homologue of the bacterial mismatch DNA repair protein MutL, and hence, *MLH1* is classified as a tumor suppressor. Hypermethylation of DNA cytosine residues and silencing of a particular functional *MLH1* alleles (e.g., -93 SNP) [57], when paired with a dysfunctional mutant allele of the same gene, correlates with relatively young individuals developing tumors of the colorectum or endometrium [27, 58]. The tumors and tumor-derived cell lines from individuals with these hypermethylation epimutations fail to express MLH1 protein from this otherwise functional allele [59]. The hypermethylation of the potentially functional *MLH1* allele and its transcriptional silencing is found in most organs and tissues of individuals, who also have hypermethylation of this *MLH1* allele in their tumors. Hence, one might expect that this *heritable epimutation* resulted from the transgenerational inheritance of this epitype. However, studies of the children of these individuals generally show loss of hypermethylation and loss of silencing of this *MLH1* allele in the first generation of transmission. Out of several individuals examined, only in one case was the epitype of hypermethylation and silencing inherited through the male parent to the individual with the disease. The *MHL1* silencing phenotype in females with colorectal cancer was associated with a particular CG island centered at -93 bp from the start of transcription in a particular *MHL1* allele containing a Single Nucleotide Polymorphism, -93 SNP, in this region as illustrated for the more general case in **Figure A1.1 (C and D)** [57]. While 5-aza-2'-deoxycytidine will reactivate silenced alleles in cultured cancer cell lines, demethylation is also correlated with a shift in nucleosome position and increased nucleosome density in the promoter region [60]. In a very recent study, laser capture

microdissection of the ovarian epithelium from ovarian tumors of cancer patients was used to analyze the cell type specific epitype and shows that the hypermethylation of *MHL1* is an early somatic event in the malignant transformation of these cells [61]. Cogent to a theme of this article is the fact that the *MHL1* epitypes of aberrant nucleosome position and cytosine methylation appear to be dependent upon the genotype of the epigenetically silenced *MHL1* allele (**Table A1.1**). Epimutations of other tumor suppressor genes including *MSH2*, *MSH6*, *PMS2*, and *BRCA1* have also been associated with colorectal cancers, but the cause-and-effect relationships with disease are less clear than they are for *MHL1* [62].

There are considerably more robust examples of the transgenerational epigenetic inheritance from model genetic organisms and wild plants, where it is easier to analyze aberrant epitypes and associated phenotypes through multiple generations. A few of the best cases with solid supporting evidence for a relationship between epitype and phenotype will be summarized.

(3) *AGOUTI*: In mice, the secreted *AGOUTI* peptide functions normally as a paracrine regulator of pigmentation. However, the dominant constitutive expression of the *AGOUTI* gene also targets changes in the hypothalamus and adipose tissues and this aberrant expression causes obesity. Hypomethylated, transcriptionally active dominant epialleles of the *agouti* gene may be maternally inherited through meiosis. Variation in the penetrance of different active epialleles generates a distribution of offspring from abnormal yellow (*agouti*) obese mice to darker mice with normal amounts of fat [63-65]. Several of the best-characterized hypomethylated active and dominant alleles of *agouti* (*Agoutiipy*, *Agoutiy*, *Agoutivy*) that are associated with a high penetrance of the yellow

coat color and obesity phenotypes have promoter-containing retrotransposons positioned just up stream of the natural *Agouti* promoter [66, 67]. For the best-studied alleles, these altered promoter structures are correlated with the hypomethylation of *agouti* and constitutive AGOUTI protein expression. However, a recent detailed examination of the DNA methylation profiles of active and silent alleles suggest that hypomethylation alone may not fully account for the complex ectopic expression of Agouti [18]. Nonetheless, the Agouti examples give reasonable support for the hypothesis (**Table A1.1, Figure A1.1, C and D**) that *genotype predisposes epitype* and aberrant phenotype. It would not be surprising to find a shift in promoter nucleosome position resulting from the various retrotransposon insertions contributing to the causative epitype.

(4) *AXINI-FUSED*: Axin1 is an inhibitor of Wnt signaling that regulates embryonic axis formation in deuterostome animals. In mice, Axin1 is the product of the mouse *Fused* locus. Some murine alleles of *Axin1-fused* (*Axin1Fu*) show variable and stochastic expression levels, where high expression of a hypomethylated allele correlates with an abnormal kinked tail. Highly penetrant *Axin1Fu* alleles contain an upstream retrotransposon or retrotransposon-mediated DNA rearrangement that alters chromatin structure and contributes to dominant transcript expression [68, 69]. An active highly penetrant mutant allele may be inherited maternally or paternally for multiple generations. Both cytosine hypomethylation and histone acetylation patterns are reported to correlate with increased *Axin1Fu* expression and risk of abnormal tail development [70-72]. The causal relationships between genotype, the DNA methylation epitype, gene expression, and the kinked tail phenotype are supported by the fact that methyl donor dietary supplementation of the mothers, a treatment known to increase DNA methylation,

reduced *Axin1Fu* expression and halved the incidence of kinked tails. Conversely, treatment of mice with the histone hyperacetylation agent Trichostatin A increased *Axin1Fu* expression and the frequency of a kinked tail phenotype. A recent study examining the chromatin from blastocyst stage heterozygous *Axin1Fu*⁺ embryos shows that dimethylation of lysine-4 on histone H3 (H3K4Me₂) as well as acetylation of lysine-9 on histone H3 (H3K9Ac) correlate with penetrant alleles [72]. By contrast, there was no correlation of blastocyst stage cytosine methylation with penetrant alleles. However, both the drug treatments and studies of development after the blastocyst stage only prove the importance of somatic epigenetic inheritance during tail development. Again, it is reasonable to propose that the presence of retrotransposon mediated changes in DNA sequence, which are present in all the aberrantly expressed *Axin1Fu* alleles, is the primary cause of the transgenerational inheritance of epigenetic risk. A shift in nucleosome position in penetrant alleles could affect downstream cytosine methylation and histone PTM, resulting in higher *Axin1Fu* gene expression and the kinked tail phenotype. By this view, genotype determines the nucleosomal epitype, which produces other aberrant hypomethylation and histone PTM epitypes, leading to increased gene expression and the novel kinked tail phenotype (**Figure A1.1, Figure A1.2A, Table A1.1**).

(5) *CNR*: The tomato Colorless Non-Ripening gene *CNR* encodes a homolog of the animal *SQUAMOSA* promoter binding protein (SPB box protein). *CNR* is essential to normal carotenoid biosynthesis and fruit ripening in tomato and provides one of the best examples of a stable transgenerationally inherited epitype producing an abnormal phenotype. The natural epialleles of *CNR* in the tomato *Lycopersicon esculentum* contain

18 methylated cytosine residues (5MeCG or 5MeCHG, where H is C, A, or T) in a 286 bp contiguous region [73]. Hypermethylation of this region and silencing of the *CNR* gene leads to colorless tomatoes low in carotenoids (**Figure A1.1C**). Because the phenotype is relatively stable, these epialleles were originally mistaken as mutant alleles. The silenced *cnr* epiallele and active wild type *CNR* gene do not have any encoded DNA sequence differences for thousands of base pairs within or flanking this hypermethylated region. Thus, while there is no mutational basis for the change in epitype, the *CNR* gene is potentiated for a stochastic DNA methylation event, because it contains such a large number of strategically positioned cytosine residues in its sequence. While this example supports a link between the *CNR* gene sequence, epitype, and risk phenotype (**Table A1.1**), there does not appear to be a particular genotype that predisposes the cytosine hypermethylation epitype. The significant question becomes, once the aberrant epitype is established, how is this hypermethylation epitype stably inherited through the germ line?

(6) *CYCLOIDEA*: The perennial plant in which *CYCLOIDEA* was first identified, *Linaria vulgaris* (Toadflax, Butter and Eggs), normally produces yellow and orange asymmetric flowers composed of three petals of different morphologies. “Mutant” plants are found in wild populations with aberrant abnormally symmetrical “peloric” flowers that are comprised of five evenly arrayed petals of similar morphology. Plants with these aberrant flowers were first characterized by Carl Linnaeus 260 years ago and collected as herbarium specimens [74]. The peloric floral phenotype is produced by the hypermethylation and transcriptional silencing of the gene encoding a transcription factor *CYCLOIDEA* (*CYC*) [75]. Inheritance of the recessive peloric floral phenotype and silenced *cyc* epiallele is relatively stable, follows Mendelian segregation, and hence,

appeared upon initial investigation to be a normal mutant allele. However, gene silencing always maps to a DNA polymorphic *cyc308G* allele with a single nucleotide polymorphism in an unmethylated region 308 nt downstream of the stop codon and never to the more common wild type *CYC308A* allele. Peloric individuals are homozygous recessive for the *cyc308G* allele with both copies being hypermethylated and completely silenced for RNA expression. Thus, it is reasonable to conclude that *genotype predisposes epitype*, gene silencing, and the peloric phenotype (**Table A1.1, Figure A1.1, C and D**).

(7) *Histone H3K4Me2 Demethylase erases epigenetic memory in each generation:*

A number of histone PTMs such as Histone H3 Dimethyl Lysine4 (H3K4Me2) are acquired during transcription and are associated with active genes [76]. These epigenetic marks are removed at different stages in development by an H3K4Me2 demethylase, known as LSD1 in humans and SPR-5 in *Caenorhabditis elegans* (**Figure A1.1F**). Removal of the H3K4Me2 epitype prior to meiosis by SPR-5 in *Caenorhabditis elegans* is essential to maintaining an immortal germline [77, 78]. Within two-dozen generations of worms acquiring the recessive null genotype these *spr-5* mutants have a brood size several-fold lower than wild type, with 70% of the worms being fully sterile. Homologs of LSD1 (SPR-5) are found throughout the four eukaryotic kingdoms and a number of these genes are known to be essential to normal organismal development [79-81]. The unmodified H3K4 epitype is essential and retention of the histone PTM causes aberrant development. However, there is as yet little evidence that this particular histone PTM epitype is normally preserved through meiosis or that genotype plays any role in determining the H3K4Me2 epitype at any particular locus (**Table A1.1**).

(8) *Inheritance of quantitative epigenetic trait loci.* Two separate genome-wide epigenetic studies demonstrate that multi-generational inheritance of complex traits such as flowering-time, plant height, biomass, and bacterial pathogen resistance behave as quantitative epigenetic trait loci in *Arabidopsis thaliana* [22, 82, 83]. These studies used two independently derived sets of recombinant inbred lines (RILs), where one of the founding parents was a recessive null for one of two genes necessary for DNA cytosine methylation. For example, one study begins with a fourth generation plant homozygous defective *ddm1/ddm1* that is highly compromised in a number of phenotypic traits due to DNA hypomethylation. *DECREASED DNA METHYLATION1* (DDM1) is a Swi2/Snf2-like DNA-dependent ATPase chromatin remodeler required for most DNA cytosine methylation. The *ddm1/ddm1* line was backcrossed to wild type and this heterozygous F1 *ddm1/DDM1* was backcrossed to wild type again and screened to obtain hundreds of separate DDM1/DDM1 lines. These lines were selfed to establish hundreds of epiallelic Recombinant Inbred plant Lines (epiRILs) [22]. For several generations, approximately 30% of the *DDM1/DDM1* epiRILs displayed aberrant morphological phenotypes affecting flowering time and plant height, among other phenotypes. They assayed 22 epiRILs for the methylation of 11 candidate genes that are normally cytosine hypermethylated, but are hypomethylated in *ddm1*. Six alleles showed partial remethylation and five alleles were completely remethylated producing the identical complex epitype for this later gene set to wild type. Control genes that were previously unmethylated remained unmethylated.

In one particular example, Johannes et al [22] followed the methylation sensitive *FWA* gene, for which the ectopic expression of the hypomethylated epiallele in *ddm1*

parental plants produces strong late flowering phenotypes [84]. All of the 22 randomly selected epiRILs were now normally methylated at *FWA* and flowered at normal times. However, when they examined three extremely late flowering lines from among the population of hundreds of epiRILs (i.e., plants that flowered after more than 48 days vs 33 days to flowering in wild type) these epiRILs were almost completely hypomethylated at *FWA* and expressed high levels of *FWA* transcripts, accounting for their phenotype. Hence, out of hundreds only a few of the epiRIL lines escaped from the remethylation of *FWA*, when *DDM1* was restored.

In summary, aberrant DNA methylation epitypes at many loci and the resulting changes in downstream molecular and developmental phenotypes appear to be transgenerationally inherited. Most genes regain WT methylation patterns and phenotypes within a few generations and the restoration appears to be sequence specific. Hence, the genetic machinery necessary for the *de novo* remethylation of these completely unmethylated loci is encoded in the *Arabidopsis* genome and remethylation did not require hemi-methylated DNA templates to be newly inherited. These data suggest that genotype predisposes this global cytosine methylation epitype.

9. *Reprogramming of DNA cytosine methylation by double stranded dsRNAs.* The 5'-methylation of DNA cytosine residues occurs in three sequence contexts 5MeCG, 5MeCHG and 5MeCHH (**Figure A1.1C**). A number of DNA methyl-transferases (DMTs) are known to methylate DNA cytosine in the 5' position. DMT1 efficiently propagates hemimethylated symmetrical CG sequences, and hence, the somatic inheritance of islands of 5MeCG hypermethylation that may lead to gene silencing is not hard to explain. However, DNA methylation of all types is predominantly erased (i.e., 80

to 90% loss of methylation) in germ line cells in the embryos of both plants and animals [85-87]. Hence, the reprogramming of CG, CHG, and CHH methylation and a mechanism for transgenerational inheritance of these epitypes has been of intense interest in recent years [88, 89]. To simplify the discussion of the gene-specific DNA cytosine remethylation and subsequent inheritance of methylation, Eric Richards [90] introduced three working categories: obligate, facilitative, and pure DNA methylation.

Epialleles in heterochromatic DNA that display obligate DNA cytosine methylation always remain methylated due to the presence of large numbers of transposable elements in various orientations producing dsRNA that promote a strong RNA interference response and adjacent target gene remethylation [91]. Genes within or closely adjacent to the centromer are good examples of obligate epialleles. *Axin1Fu* and *AgoutiAy* are typical examples of facilitative epialleles, because the presence of an upstream change in DNA sequence facilitates a seemingly stochastic epigenetic variation in methylation and phenotype. Because the wild type loci for these alleles lack an altered promoter element there is seldom any variation in the cytosine methylation epitype at the wild type loci. Pure epialleles are defined as those showing variation in cytosine methylation without a known genotypic cause and appear to be examples of *de novo* DNA cytosine methylation. If pure epialleles are truly independent of genotype, then they stand as strong evidence against our hypothesis. The well-studied hypermethylation and silencing of wild-type *CNR* and *RYR1* alleles fit the definition of pure epialleles. Schmitz et al [92] examined the complete methylome of 100 *Arabidopsis* lines propagated for 30 generations by single seed descent from a single parent. They observed that CG ↔ 5MeGC Single Methylation Polymorphisms (SMPs) occurred at a 10,000-fold increased frequency per

generation over the DNA base mutation rate, which they also measured (**Figure A1.1D**). While CG SMPs occurred primarily within gene bodies, large numbers of CHG and CHH SMPs occurred in flanking regions. Thus, novel inherited SMPs are generated at high frequencies and if this remethylation is independent of DNA sequence then pure epialleles are common.

One relevant question for this discussion is the following: are ostensibly pure epialleles truly independent of genotype, or are they simply facilitative epialleles for which we have not yet identified the associated cis- or trans-acting genes making dsRNAs that program inherited CG, CHG and CHH methylation epitypes? There is recent evidence supporting the latter interpretation that we now summarize.

Despite being generated through slightly different mechanisms, many classes of small RNAs (e.g., siRNA, miRNA, piRNA) are known to template the remethylation of cytosine in different sequence-specific contexts (**Figure A1.1C**) for the transgenerational inheritance of gene silencing and or activation [89, 93]. This general mechanism for reprogramming using different classes of small RNAs appears ancient in that it is found in all four eukaryotic kingdoms. These RNAs facilitate the remethylation of appropriate CG, CHG, and CHH sequences. But these data began to raise the question - *does remethylation occur on a global genome-wide scale?* To address the scope of remethylation, Teixeira et al [94] examined the remethylation of numerous transposable element loci in DDM1/DDM1 epiRIL plants that had descended from an essentially unmethylated *ddm1/ddm1* plant backcrossed to wild type. Those loci that were remethylated after a few generations in the epiRILs contained cytosine rich gene sequences that were highly complementary to the sequence of small inhibitory (interfering) RNAs

(siRNAs). Those loci with similar cytosine rich composition for which they could not identify complementary siRNAs remained hypomethylated. siRNAs attract RNA interference and DNA methylation machinery to complementary DNA sequences and thereby template sequence-related DNA methylation [95]. This shows that RNAi mechanisms are essential for the proper remethylation of much of the Arabidopsis genome. These and other data make it clear that for a large number of repetitive elements in yeast, plants, and animals, the matching genotypes of structural genes and small RNAs predict a cytosine methylation phenotype. However, the Teixeira et al [94] study raises further questions about the biology, regulation, and timing of cytosine remethylation for both transgenerational and somatically inherited epitypes. Recent evidence suggests that in both plants and animals “nurse cells” may transfer hundreds of undefined small RNAs to adjacent egg or sperm germ cells to reprogram cytosine methylation [88, 89, 93]. For example, in mice in which 80% to 90% of the germline DNA methylation is erased for single copy genes at approximately day 11.5 of embryo development (E11.5), remethylation of sperm DNA occurs by embryonic stage ~E16.5 and is significantly directed by populations of 24 to 30 nt long piRNAs produced in adjacent cells in the prospermatogonia [96-98]. The identities of most of the plant and animal RNAs transferred to developing germ cells are not yet known, but there is the real potential that large populations of RNAs may account for most transgenerational remethylation and perhaps even the apparent de novo methylation described by Schmitz et al [92]. Appropriately positioned target sequences in these epialleles and thousands of expressed small RNAs would have to be inherited together for genotype to predispose the transgenerational inheritance of the global DNA methylation epitype.

10. Reprogramming epitype during somatic cell nuclear transfer. In most of the above examples genotype determines the likelihood, but not the certainty, of particular epitypes and phenotypes being displayed, because the same DNA sequence may be flexibly reprogrammed into many different chromatin conformations. It is fundamental to epigenetics that as cell types differentiate the same DNA sequence may display multiple epitypes and some epitypes may be more or less stable than others. An interesting example of a variety of epitypes descending from one genotype comes from research using somatic cell nuclear transfer (SCNT) to produce identical or genetically modified laboratory and farm animals. SCNT is achieved by transplanting a somatic cell nucleus into a functional embryonic cell capable of forming a viable organism. This technology has met with modest success, generating cloned mice, rabbits, pigs, sheep, cows and more, but the efficiency of obtaining viable healthy offspring is low. Even if genetically modified embryos are established in surrogate mothers, developmental abnormalities and spontaneous abortions are common. A major limitation to obtaining relatively normal full-term development appears to be variations in epigenetic reprogramming of the transplanted nucleus [99-102]. The field of regenerative medicine faces similar problems with epigenetic reprogramming when trying to establish genetically altered lines of induced pluripotent stem cells by SCNT, for example, by transferring a somatic cell nucleus into an oocyte [103, 104]. Without prior knowledge of the successes in producing cloned animals by SCNT, one would not necessarily expect that the new nuclear environment should correctly reprogram the donated nucleus. Known sources of the reprogramming problem in the animal cloning field is that the transferred nucleus frequently loses a significant fraction of its DNA cytosine methylation and nucleosomal

histone side chain methylation and acetylation relative to the more modified epitype of nuclei in native embryonic cells (**Figure A1.1C and F**) [105-109]. However, the surprising fact remains that some relatively healthy animals resembling the nuclear donor are obtained via SCNT and that genetic and epigenetic totipotency of the donor nucleus is reestablished in the viable offspring. For appropriate reprogramming to take place on a genome-wide scale the donor DNA sequence must have the capacity to interact with the embryonic cellular environment and determine, albeit at low frequency, an epitype(s) compatible with full-term development. These results support the idea that during SCNT the donated DNA sequence predisposes much of its own epigenetic reprogramming (**Table A1.1**).

B. Evolutionary co-conservation of DNA sequence and chromatin structure filters out transgenerationally inherited epitypes. If genotype determines epitype then a reasonable corollary is that some transgenerationally-inherited chromatin structures should align with particular DNA sequence motifs and be passed on to duplicate gene copies. In this model, the range of possible epitypes for a sequence would evolve by gene duplication and mutation in parallel with genotype [19, 51]. Rapidly evolving epitypes might only be conserved and identifiable among very recently duplicated genes examined among a limited number of related cell types or when examined statistically in comparisons of large numbers of aligned sequences, while slowly evolving highly conserved epitypes might be found among anciently duplicated genes and descended from a common ancestral protist sequence.

(1) Short DNA sequence repeats such as RRRRRYYYYY determine the bending and positioning of DNA around the nucleosome. More than 30 years ago Trifonov and his

colleagues [110, 111] presented the argument that gene sequence is fundamentally important to nucleosome positioning. He argued that the necessary high degree of bending of DNA as it wraps twice around and binds the nucleosome would be favored by particular 10.5 bp repeat sequences of approximately 5 purines (R) followed by 5 pyrimidines (Y) (*RRRRRYYYYY*) (**Figure A1.1B**), or the inverse of this sequence, *YYYYYRRRRR*. He found a good correlation for 10 bp repetitions of the dinucleotides GG, TA, TG, and TT in the modest compilation of 30,000 bp of DNA sequence from different eukaryotes available at that time¹. The statistical concept was a bit counterintuitive and slow to gain acceptance, because it was hard to reconcile the functional demands of sequences encoding proteins and regulatory regions with the proposed special sequence demands of nucleosome interaction. Recently, with access to nearly unlimited numbers of nucleosome-delimited 147 bp DNA sequence fragments, and more advanced computational methods, it has become very clear that 14 repetitions of the 10.5 bp repeat sequences *Y-RRRRRYYYYY-R* or *R-YYYYYRRRRR-Y* are statistically favored for nucleosome positioning. Regional differences in GC compositions in the genome favor particularly skewed repeats such as *T-AAAAATTTTT-A* or *C-GGGGGCCCCC-G* [112, 113]. These consensus sequences are based on a statistical argument, and at the genome level any one dinucleotide such as *AA* or *GG* is seldom found in a particular position in the 147 bp repeat more than 30% of the time [114]. Because the inward facing helix of any one 147 bp of nucleosomal DNA fragment has 14 chances to contact the core of nucleosomal proteins, this mechanism requires only several correctly positioned dinucleotides contacting the nucleosome to give sequence specificity

¹Trifinov did not have nucleosome specific DNA sequence data available 30 years ago.

to nucleosome positioning. Hence, there is in fact little conflict with conserved coding and regulatory sequences and the sequence constraints of nucleosome positioning. Furthermore, the most common classes of ATP-dependent chromatin remodeling machines, SWI/SNF and ISW2, move DNA in approximately 9 to 11 bp increments over the surface of a nucleosome, consistent with the importance of 10.5 bp repeats in nucleosome binding [115]. These data strongly support a model, where genotype predisposes possible nucleosome position epitypes. More particular support for this argument comes from examining the sequences for subsets of the nucleosomal DNA population binding nucleosomes containing histone variants H2AZ and CENH3.

(2) *The genome-wide positioning of H2AZ nucleosomes.* The histone variant H2AZ and likely other histone variants are inserted into assembled nucleosomes by histone variant exchange complexes (HVE) such as SWR1 (**Figure A1.1E**). Albert et al [116] precisely aligned the sequences of thousands of 147 bp yeast nucleosomal DNA fragments enriched for histone variant H2AZ. Their data show conclusively that H2AZ nucleosome positioning on a genome wide scale is strongly influenced by dinucleotide repeat patterns spaced 10 bp apart in the DNA sequence (**Figure A1.1A and B**). In particular, GC-rich dinucleotides are on the inside as the DNA helix wraps around the nucleosomal protein core, and AT-rich dinucleotides are on the outside. The preference for these nucleotide pairs at each of their 14 possible positions within any 147 bp nucleosomal fragment is only about 2 to 9%, and therefore, any single nucleosomal fragment sequence is likely to vary significantly from the statistical consensus. However, it is clear that the H2AZ nucleosome position is determined by the overall pattern in the

DNA sequence, and hence, H2AZ nucleosome position will be conserved following gene duplication.

Similar results were obtained for genome-wide positioning of all nucleosomes from humans and *Arabidopsis* [117] and subsets of human nucleosomes specific to certain classes of genes [118]. In the total 147 bp nucleosomal fraction from *Arabidopsis* and humans, an AT rich dinucleotide repeat is spaced every 10 bp and out of phase by 5 bp with a GC rich dinucleotide repeat.

Further support for the concept that DNA sequence positions H2AZ nucleosomes comes from a comparison of duplicated genes in *Arabidopsis*. A single peak of H2AZ enriched nucleosome(s) is found at the 5' end of nearly half of all plant, animal, and fungal genes that have been examined [116, 119-121]. In *Arabidopsis*, three related MADS box genes that regulate flowering time require normal H2AZ for full expression. In wild type cells, all three MADS box genes show a striking bimodal distribution of H2AZ deposition, with peaks of H2AZ histone-containing nucleosomes at their 5' and 3' ends [122]. This pattern is quite distinct from the single 5' spike of H2AZ observed for other MADS genes in humans, *Arabidopsis*, and yeast. These three genes are estimated to have diverged from a common gene ancestry in the eudicot plant lineage in the last 100 million years (MY) and stand alone in their own distinct clade, among more than 100 other MADS box genes in *Arabidopsis* that do not have a bimodal distribution of H2AZ nucleosomes. These data are consistent with the bimodal distribution of H2AZ being inherited following gene sequence duplication from an ancestral MADS gene [19].

(3) *The genome-wide positioning of CENH3 centromeric nucleosomes.* Recent experimental evidence demonstrates that CENH3 enriched centromeric nucleosome

positions are determined by DNA sequence. Animal and plant centromeres are composed of a diverse variety of retroelements and repetitive satellites that generally appear unrelated in their DNA sequences. Numerous earlier studies of centromere and neocentromere sequences concluded that a distinct conserved DNA sequence was not essential to centromere activity. However, a very recent analysis of 100,000 centromeric histone CENH3 enriched nucleosomal DNA fragments from maize suggests that a 10 bp repeat of AA or TT dinucleotides contributes to determining the positioning of centromeric nucleosomes [123]. The CENH3 nucleosome specific sequence was not revealed until the 147 bp micrococcal nuclease protected DNA sequences were precisely aligned. The preference for AA or TT nucleotide pairs at each of the 14 positions within a typical 147 bp nucleosomal fragment was statistically significant. The likelihood of finding one of these dinucleotide pairs at any of the potential contact points ranges from 13% to 60% above the frequency at which other dinucleotides are found. Thus, CENH3 enriched nucleosomes are positioned by a variation on what is shown in **Figure A1.1B**, where the inward (IN) facing DNA base pairs that bind are generally AA or TT and would be classified as weak (W) binding. This would indicate that any single centromeric nucleosomal sequence may vary significantly from the statistical consensus for these nucleotide pairs. In this way, a subset of retroelements that are seemingly unrelated in sequence using standard sequence alignment methods may contain suitable sequence repeats that position centromeric nucleosomes.

The human and Arabidopsis genomes each encode more than a dozen histone protein sequence variants for each of three classes of histones, H2A, H2B, and H3. Within each class a few subclass variants are easily identified as predating the divergence

of plants, animals, and fungi from their more recent protist ancestors. Thus, it is reasonable to speculate that distinct DNA sequence patterns evolved in concert with each histone variant subclass to provide complex patterns of nucleosome positioning. If true, then DNA sequence would be responsible for the transgenerational positioning of most classes of nucleosomes.

(4) *Cytosine methylation in the human plasminogen gene family*: In an attempt to show that epitypes and associated phenotypes can evolve by gene duplication and divergence, Cortese et al. [51] compared promoter CG methylation patterns among the four duplicated gene members of the ~35-million-year-old human plasminogen (*PLG*) precursor gene family, encoding blood-clotting factors found only in hominids. Cytosine DNA methylation patterns are well conserved among seven CG sites located -171 to -378 nucleotides upstream from the start of transcription within all four *PLG* gene promoters (similar to **Figure A1.1A and C**). In liver, where transcripts for all four genes are expressed, one allelic copy of each gene pair is almost completely unmethylated at all seven sites. In heart muscle and in skeletal muscle, where the four *PLG* genes are turned off, nearly 100% of the seven sites are fully cytosine methylated on both alleles for all four genes. In other words, promoter cytosine methylation silences all gene copies in the two non-expressing tissues examined, while hypomethylation of one copy of each *PLG* gene activates their expression in liver. The *PLG* data support the generational inheritance and conservation of the cytosine methylation epitype following gene duplication for recently duplicated genes that are co-expressed. Cortese et al [51] also compared promoter CG methylation patterns among several members of the much older human T Box (TBX) gene family in which the most gene duplications date back 300 to

600 million years. No evidence was obtained for conserved CG methylation patterns among any pair-wise comparison of *TBX* genes. Perhaps because the *TBX* genes are differentially expressed and the divergence events between genes are much more ancient, the lack of conserved CG methylation patterns is to be expected.

(5) *Histone side chain modifications in human segmental sequence duplications.*

Barski et al (2007) published a ground-breaking genome-wide study on sequence specific location of 23 histone PTMs and a few other epitopes in purified human CD4+ T cells. From this data set, Zheng [124] examined 14 distinct patterns of histone PTM in nucleosomes from 1,646 relatively recent (i.e., less than approximately 25-million-year-old) segmental chromosome duplications (SD). They found no significant evidence for the inheritance of these histone modifications between the original and derived loci. Specifically, the duplicated copy did not inherit the parental pattern of histone side chain methylation or acetylation (**Figure A1.1F**). Moreover, inheritance appears to be distinctly asymmetric for some of the modifications, such that there is a strong statistical bias toward histone methylation of one gene copy for each SD and not the other copy, beyond what might have occurred at random. Many of the asymmetrical histone modifications correlate with gene activation and repression, suggesting that active genes in the parent sequence are silenced in the duplicated loci, and *visa versa*. These data imply that histone PTM epitopes may not be the direct transgenerationally inherited “cause” of the phenotypes with which they are associated. Thus, these data on histone PTM epitopes at SDs do not support our working hypothesis. If these results are supported by more experimental studies, it will not necessarily mean that histone modifications are not useful epitopes for predicting risk, but that they may be further from the inherited cause

of epigenome induced pathologies than other epitypes such as nucleosome position and cytosine methylation. Histone PTMs are indeed important to somatic inheritance and development [46, 125].

(6) Nucleosome positioning and H3K4Me2 modifications in the HOXD cluster:

There are six genes at the *HOXD* gene cluster (i.e., *HOXD13*, 11, 9, 8, 4, 3) covering ~100,000 bp on human Chromosome 2. In human sperm, there are one or two spikes of general nucleosome occupancy and histone H3 lysine 4 dimethylation (H3K4Me2) enriched nucleosome occupancy within each of the promoters of these genes, whereas the ~100,000 bp of 5' flanking region is relatively free of nucleosomes [126] (**Figure A1.1A and F**). Because nucleosome positioning was performed using microarrays, the sequence specificity of H3K4Me2-enriched nucleosomes among these *HOXD* promoters cannot be determined from these data or compared to the results from Barski et al. [127] in which they did not find sequence specificity for histone H3K4Me2-enriched nucleosome binding. These results showing the conserved positioning of nucleosomes in *HOXD* promoters in human sperm are similar to those for H2AZ-enriched nucleosomes among the *FLC*-related *MADS* genes in Arabidopsis shoot tissue [122].

(7) Higher-order chromatin structures: Genes and regulatory sequences that are narrowly or widely spaced on a chromosome may interact productively through higher-order chromatin structures such as solenoids, small and giant loops, and minibands [128-130]. For example, small concatenated DNA loops may be formed by re-association of the single strands of the poly (CA)-poly (TG) microsatellite at their base [131]. These small loops appear to impact the control of gene expression via binding to HMG-box proteins [131, 132]. There is mounting evidence that interactions of distant intra- and

inter-chromosomal domains provide epigenetic mechanisms to maintain specialized gene expression states [133-135]. Hence, the potential exists that higher order structures contribute to epigenetic control and are determined in part by DNA sequence.

Summary from direct and indirect analyses of epigenetic inheritance: An examination of several examples of the direct transgenerational inheritance of epitype and the epitypes of duplicated and/or conserved DNA sequences revealed the complexities of determining cause-and-effect relationships among genotype, epitype and phenotype. However, in balance, there are robust experimental data supporting the hypothesis that “*genotype predisposes epitype*,” for some epitypes (**Table A1.1**). In particular, it is becoming clear that a large fraction of, if not all, cytosine methylation is determined by gene sequence and the presence of paired sequence-specific complementary small RNAs that direct their transgenerational remethylation. Similarly, based on the sequences of H2AZ and CENH3 enriched nucleosomal fragments, nucleosome position appears strongly influenced by DNA sequence (**Figure A1.1A-C**). However, there is little evidence suggesting that DNA sequence determines the position of any of more than 20 different classes of histone PTM enriched nucleosomes (**Figure A1.1F, Table A1.1**).

Based on this analysis, it is worth ranking the utility of various classes of epitype in estimating epigenetic risk. A risk pyramid linking the relationships of **genotype** and **epitype** with **epigenetic risk phenotype** is shown in **Figure A1.2B**. DNA sequence is placed at the apex, as the primary cause of epigenetic risk. This is followed by nucleosome position that appears to be directly dependent upon 10 bp repeats in DNA sequence and DNA cytosine methylation that is highly dependent upon *cis*-acting CG,

CHG, and CHH sequences in the target gene and the sequence of trans-acting small RNAs. However, while histone PTM may be strongly correlated with epigenetically controlled phenotype, there is no evidence that any histone PTM is causal to transgenerationally-inherited risk. Histone PTM epitypes may represent the effect of other epigenetic and genetic controls and may be principally important to somatic inheritance of epigenetic controls. The clear relationship between novel genotypes and many of the most robustly characterized inherited epitypes of nucleosome position and cytosine methylation is a recurrent theme in the literature of the most thoroughly studied genes under epigenetic control. This suggests that human and animal therapeutic treatments or plant and animal genetic breeding strategies that address harmful meiotically inherited epitypes should consider the possibility that there are genotypic causes predisposing these epitypes. If for example, the environment of a developing somatic tissue (e.g., obesity, stress, nutrients) is influencing RNA sequence directed cytosine remethylation and gene silencing, drugs targeting downstream histone PTM epitypes of that gene may be less effective than ones addressing remethylation. Strategies directed at controlling gene expression by altering histone PTM epitypes may be useful if they target the gene or genes producing the diseases phenotype. Finally, the undeniable influence of genotype on epigenetic controls leading to deleterious phenotypes has to be taken into account in a consideration of epigenetic risk, even if it confounds many current, working definitions of epigenetics.

Defining epigenetics: We've summarized direct and indirect evidence that *genotype predisposes epitype* and that epigenetic controls are strongly influenced by DNA and RNA sequences (**Figures A1.1 and A1.2**). Our hypothesis and these

supporting data may be viewed as contrary to some of the widely stated precepts of epigenetics, which for example, define epigenetics as “*the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence*” [34, 136]. A rephrasing of Riggs et al’s [136] statement as “*the study of mitotically and/or meiotically inherited changes in gene function that cannot be explained by the classical central dogma of molecular genetics*” (**Figure A1.2A**) provides a working definition that is quite consistent with our discussions. In David Nanney’s seminal article describing epigenetic control systems, he states “*The term “epigenetic” is chosen to emphasize the reliance of these systems on the genetic systems*” and goes on to say “*epigenetic systems regulate the expression of the genetically determined potentialities*” [39]. Nanney’s definitions of epigenetics are completely consistent with *genotype predisposing epitype*, and with epitype modifying gene expression and risk phenotype.

The influence of DNA sequence on epigenome-induced pathologies points a way forward: Understanding that genotype predetermines many inherited epitypes suggests a few useful strategies and concerns as we try to address epigenome induced pathologies. **First**, we are in a better technical position than ever before to determine the influence of genotype on epitype. New rapid DNA sequencing and DNA bead array methods for identifying SNPs and 5MeC residues combined with a wide selection of treatments to chromatin (e.g., ChIP, bisulfite, micrococcal nuclease) allow us to quantitatively determine the precise genome-wide sequence-specific positioning of every nucleosome, methylated cytosine residue, and dozens of distinct histone PTMs in a genome. These epitypes may be correlated with the risk of cancer, behavioral disorders,

pathogen susceptibility, or the role of aging and environmental factors on risk, as examples. The lower costs of these genome-wide approaches is enabling the epitypes of larger populations of humans, laboratory animals, and plants to be examined in order to identify the epigenetic causes of complex diseases such as obesity, lupus, or pathogen susceptibility [137-140]. **Second**, we are in a position to develop batteries of gene-specific epigenetic biomarkers for DNA methylation epitypes that are clearly associated with disease risk and may be predictive of the penetrance of pathology. For example, this is currently being done for systemic lupus erythematosus, myeloid leukemia, and breast cancer [138, 141-143]. However, new technologies are needed if we are also to use nucleosome position and histone PTM epitypes as inexpensive epigenetic biomarkers for screening populations. **Third**, because the development of each plant and animal cell type in an organ system is under strong epigenetic control, it is essential that we examine epitypes in distinct cell types within organs. Most current epigenetic studies examine mixed cell types such as are present in whole organs and tissues (e.g., blood, tumor, hippocampus, skeletal muscle, plant shoots or roots), wherein cell type-specific epitypes are blurred due to variation of epitypes among developmentally distinct cell types. For example, several orders of magnitude more statistically significant relationships were obtained between the cytosine methylation epitype of various genes with lupus, when CD4⁺ T cells were examined as compared to the data obtained on from mixed populations of white blood cells [138, 144]. Technologies have been developed to access cell type-specific epitypes, including laser cell capture micro-dissection, fluorescent activated cell sorting (FACS) of dissociated fluorescently tagged cells, and the Isolation of Nuclei TAGged in specific Cell Types (INTACT). These technologies

enable the more precise determination of epitypes within individual cell types such as has been shown for CD4+ T cells, primordial germ cells, ovarian epithelium, retinal cones, and plant root epithelial trichoblasts and atrichoblasts [61, 127, 145-148]. **Fourth**, therapeutic approaches to human epimutations that increase the risk of pathology, or plant breeding strategies to address epigenetic susceptibility to stress or disease, need to consider that molecular mechanisms may be obscurely hidden in DNA sequence motifs and/or the sequences of small RNAs that are imperfectly matched with their target genes (**Figure A1.1**). Current basic research is laying the course for using small RNAs to direct transcriptional gene silencing by promoter DNA methylation for therapeutics and crop improvement. For example, siRNA transgenes have been used for the methylation based transcriptional silencing of the *Heparanase* gene in human cancer cells in culture [149] and to elucidate the mechanisms of small RNA-based transcriptional silencing in plants [150, 151]. Unless we can develop therapeutic approaches, identifying genotypic influences on epigenetic risk may only add more diseases to the list of thousands for which we know the genetic cause, but have no known cure. However, taking the numerous advances in epigenetics research altogether, it is reasonable to propose that during the next two decades effective therapeutic treatments will follow the dissection of the molecular mechanisms by which genotype and epitype interact to produce disease pathologies.

Conclusion. There is substantial evidence that altered epigenetic controls contribute to a variety of diseases ranging from cancer and developmental malformations to susceptibility to various forms of biotic and abiotic stress. We reviewed experimental genetic, epigenetic, cell biological, and biochemical data surrounding the

transgenerational inheritance of several examples of well-studied epigenome-induced pathologies and the contribution of conserved DNA sequence motifs to epitype. The preponderance of evidence suggests that *genotypes predispose epitypes* for most chromatin structures that are transgenerationally-inherited and this relationship contributes to the penetrance of epigenetically controlled diseases. Genotypes influencing inherited epigenetic risk are often obscurely encoded in DNA sequence and small RNAs. Furthermore, the remethylation of DNA cytosine residues may only be reprogrammed at particular times in development and only in particular tissues such that a special effort may be required to identify and characterize these mechanisms. Some of the best-characterized examples that were discussed herein suggest we are only just beginning to understand the molecular biology behind inherited epigenome-induced disorders. Finally, the paths to effective therapeutic development or to lowering epigenetic risk will be easier to trace out once we understand the mechanisms by which *genotype predisposes epitype* for a particular disease.

Acknowledgements

Drs. Roger Deal, Muthugapatti Kandasamy and Mary Anne Della-Fera offered useful editorial comments on the manuscript for which we are grateful. Kip Carter and the UGA's College of Veterinary Medicine's Department of Medical Illustration generously provided the artwork. This study was supported by a grant from the National Institutes of Health (GM36397-25) and the University of Georgia's Research Foundation to RBM and the UGA Graduate Student Association Recruitment Award, the Linton and June Bishop Graduate Fellowship, and the NIH Genetics Training Grant (GM 07103-37) awards to KJM.

REFERENCES

1. Fumagalli M, Cagliani R, Riva S, Pozzoli U, Biasin M, Piacentini L, Comi GP, Bresolin N, Clerici M, Sironi M: **Population genetics of *IFIH1*: ancient population structure, local selection, and implications for susceptibility to type 1 diabetes.** *Mol Biol Evol* 2010, **27**:2555-2566.
2. Zhang B, Fackenthal JD, Niu Q, Huo D, Sveen WE, DeMarco T, Adebamowo CA, Ogundiran T, Olopade OI: **Evidence for an ancient *BRCA1* mutation in breast cancer patients of Yoruban ancestry.** *Fam Cancer* 2009, **8**:15-22.
3. Hayano-Kanashiro C, Calderon-Vazquez C, Ibarra-Laclette E, Herrera-Estrella L, Simpson J: **Analysis of gene expression and physiological responses in three Mexican maize landraces under drought stress and recovery irrigation.** *PLoS One* 2009, **4**:e7531.
4. Zwonitzer JC, Bubeck DM, Bhatramakki D, Goodman MM, Arellano C, Balint-Kurti PJ: **Use of selection with recurrent backcrossing and QTL mapping to identify loci contributing to southern leaf blight resistance in a highly resistant maize line.** *Theor Appl Genet* 2009, **118**:911-925.
5. Pataky JK, Bohn MO, Lutz JD, Richter PM: **Selection for quantitative trait loci associated with resistance to Stewart's wilt in sweet corn.** *Phytopathology* 2008, **98**:469-474.
6. Wang G, Ellendorff U, Kemp B, Mansfield JW, Forsyth A, Mitchell K, Bastas K, Liu CM, Woods-Tor A, Zipfel C, et al: **A genome-wide functional investigation into the roles of receptor-like proteins in Arabidopsis.** *Plant Physiol* 2008, **147**:503-517.
7. van Vliet J, Oates NA, Whitelaw E: **Epigenetic mechanisms in the context of complex diseases.** *Cell Mol Life Sci* 2007, **64**:1531-1538.
8. Stokes TL, Kunkel BN, Richards EJ: **Epigenetic variation in Arabidopsis disease resistance.** *Genes Dev* 2002, **16**:171-182.
9. Hitchins MP: **Inheritance of epigenetic aberrations (constitutional epimutations) in cancer susceptibility.** *Adv Genet* 2010, **70**:201-243.
10. Alvarez ME, Nota F, Cambiagno DA: **Epigenetic control of plant immunity.** *Mol Plant Pathol* 2010, **11**:563-576.
11. Holliday R: **The inheritance of epigenetic defects.** *Science* 1987, **238**:163-170.

12. Nicolson GL: **Cell surface molecules and tumor metastasis. Regulation of metastatic phenotypic diversity.** *Exp Cell Res* 1984, **150**:3-22.
13. Ferluga J: **Possible organ and age-related epigenetic factors in Huntington's disease and colorectal carcinoma.** *Med Hypotheses* 1989, **29**:51-54.
14. Bjornsson HT, Fallin MD, Feinberg AP: **An integrated epigenetic and genetic approach to common human disease.** *Trends Genet* 2004, **20**:350-358.
15. Mehler MF: **Epigenetic principles and mechanisms underlying nervous system functions in health and disease.** *Prog Neurobiol* 2008, **86**:305-341.
16. Graff J, Mansuy IM: **Epigenetic codes in cognition and behaviour.** *Behav Brain Res* 2008, **192**:70-87.
17. Dolinoy DC, Huang D, Jirtle RL: **Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development.** *Proc Natl Acad Sci U S A* 2007, **104**:13056-13061.
18. Cropley JE, Suter CM, Beckman KB, Martin DI: **CpG methylation of a silent controlling element in the murine *Avy* allele is incomplete and unresponsive to methyl donor supplementation.** *PLoS One* 2010, **5**:e9055.
19. Meagher RB: **The evolution of epitype.** *The Plant cell* 2010, **22**:1658-1666.
20. Slatkin M: **Epigenetic inheritance and the missing heritability problem.** *Genetics* 2009, **182**:845-850.
21. Martin C, Zhang Y: **Mechanisms of epigenetic inheritance.** *Curr Opin Cell Biol* 2007, **19**:266-272.
22. Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuissou J, Heredia F, Audigier P, et al: **Assessing the impact of transgenerational epigenetic variation on complex traits.** *PLoS Genet* 2009, **5**:e1000530.
23. Jablonka E, Raz G: **Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution.** *Q Rev Biol* 2009, **84**:131-176.
24. Whitelaw NC, Whitelaw E: **Transgenerational epigenetic inheritance in health and disease.** *Curr Opin Genet Dev* 2008, **18**:273-279.

25. Xing Y, Shi S, Le L, Lee CA, Silver-Morse L, Li WX: **Evidence for transgenerational transmission of epigenetic tumor susceptibility in *Drosophila*.** *PLoS Genet* 2007, **3**:1598-1606.
26. Rakyan V, Whitelaw E: **Transgenerational epigenetic inheritance.** *Curr Biol* 2003, **13**:R6.
27. Morgan DK, Whitelaw E: **The case for transgenerational epigenetic inheritance in humans.** *Mamm Genome* 2008, **19**:394-397.
28. Febbo PG: **Epigenetic events highlight the challenge of validating prognostic biomarkers during the clinical and biologic evolution of prostate cancer.** *J Clin Oncol* 2009, **27**:3088-3090.
29. Haslberger A, Varga F, Karlic H: **Recursive causality in evolution: a model for epigenetic mechanisms in cancer development.** *Med Hypotheses* 2006, **67**:1448-1454.
30. Mills J, Hricik T, Siddiqi S, Matushansky I: **Chromatin structure predicts epigenetic therapy responsiveness in sarcoma.** *Mol Cancer Ther* 2010.
31. King GJ, Amoah S, Kurup S: **Exploring and exploiting epigenetic variation in crops.** *Genome* 2010, **53**:856-868.
32. Long Y, Xia W, Li R, Wang J, Shao M, Feng J, King GJ, Meng J: **Epigenetic QTL mapping in *Brassica napus*.** *Genetics* 2011, **189**:1093-1102.
33. Jullien PE, Berger F: **Gamete-specific epigenetic mechanisms shape genomic imprinting.** *Curr Opin Plant Biol* 2009, **12**:637-642.
34. Haig D: **The (dual) origin of epigenetics.** *Cold Spring Harb Symp Quant Biol* 2004, **69**:67-70.
35. Hoekenga OA, Muszynski MG, Cone KC: **Developmental patterns of chromatin structure and DNA methylation responsible for epigenetic expression of a maize regulatory gene.** *Genetics* 2000, **155**:1889-1902.
36. Waddington CH: *Chapter 2. The Cybernetics of Development. In. The strategy of the genes: A Discussion of Some Aspects of Theoretical Biology.* London: Ruskin House, George Allen & Unwin LTD; 1957.

37. Beisson J, Sonneborn TM: **Cytoplasmic Inheritance of the Organization of the Cell Cortex in Paramecium Aurelia.** *Proc Natl Acad Sci U S A* 1965, **53**:275-282.
38. Verstrepen KJ, Fink GR: **Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi.** *Annu Rev Genet* 2009, **43**:1-24.
39. Nanney DL: **Epigenetic Control Systems.** *Proc Natl Acad Sci U S A* 1958, **44**:712-717.
40. Hardy KM, Kirschmann DA, Seftor EA, Margaryan NV, Postovit LM, Strizzi L, Hendrix MJ: **Regulation of the embryonic morphogen Nodal by Notch4 facilitates manifestation of the aggressive melanoma phenotype.** *Cancer Res* 2010, **70**:10340-10350.
41. Daxinger L, Whitelaw E: **Transgenerational epigenetic inheritance: more questions than answers.** *Genome Res* 2010, **20**:1623-1628.
42. Menon DU, Meller VH: **Germ line imprinting in Drosophila: Epigenetics in search of function.** *Fly (Austin)* 2010, **4**:48-52.
43. Lange UC, Schneider R: **What an epigenome remembers.** *Bioessays* 2010, **32**:659-668.
44. Ho DH, Burggren WW: **Epigenetics and transgenerational transfer: a physiological perspective.** *J Exp Biol* 2010, **213**:3-16.
45. Youngson NA, Whitelaw E: **Transgenerational epigenetic effects.** *Annu Rev Genomics Hum Genet* 2008, **9**:233-257.
46. Margueron R, Reinberg D: **Chromatin structure and the inheritance of epigenetic information.** *Nat Rev Genet* 2010, **11**:285-296.
47. Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP: **Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells.** *Nat Genet* 2009, **41**:246-250.
48. Sass GL, Pannuti A, Lucchesi JC: **Male-specific lethal complex of Drosophila targets activated regions of the X chromosome for chromatin remodeling.** *Proc Natl Acad Sci U S A* 2003, **100**:8287-8291.
49. Kang JK, Park KW, Chung YG, You JS, Kim YK, Lee SH, Hong SP, Choi KM, Heo KN, Seol JG, et al: **Coordinated change of a ratio of methylated H3-lysine**

- 4 or acetylated H3 to acetylated H4 and DNA methylation is associated with tissue-specific gene expression in cloned pig.** *Exp Mol Med* 2007, **39**:84-96.
50. Morgan DK, Whitelaw E: **The role of epigenetics in mediating environmental effects on phenotype.** *Nestle Nutr Workshop Ser Pediatr Program* 2009, **63**:109-117; discussion 117-109, 259-168.
51. Cortese R, Krispin M, Weiss G, Berlin K, Eckhardt F: **DNA methylation profiling of pseudogene-parental gene pairs and two gene families.** *Genomics* 2008, **91**:492-502.
52. Clarke NF, Waddell LB, Cooper ST, Perry M, Smith RL, Kornberg AJ, Muntoni F, Lillis S, Straub V, Bushby K, et al: **Recessive mutations in *RYR1* are a common cause of congenital fiber type disproportion.** *Hum Mutat* 2010, **31**:E1544-1550.
53. Robinson RL, Carpenter D, Halsall PJ, Iles DE, Booms P, Steele D, Hopkins PM, Shaw MA: **Epigenetic allele silencing and variable penetrance of malignant hyperthermia susceptibility.** *Br J Anaesth* 2009, **103**:220-225.
54. Wilmshurst JM, Lillis S, Zhou H, Pillay K, Henderson H, Kress W, Muller CR, Ndong A, Cloke V, Cullup T, et al: ***RYR1* mutations are a common cause of congenital myopathies with central nuclei.** *Ann Neurol* 2010, **68**:717-726.
55. Zhou H, Lillis S, Loy RE, Ghassemi F, Rose MR, Norwood F, Mills K, Al-Sarraj S, Lane RJ, Feng L, et al: **Multi-minicore disease and atypical periodic paralysis associated with novel mutations in the skeletal muscle ryanodine receptor (*RYR1*) gene.** *Neuromuscul Disord* 2010, **20**:166-173.
56. Zhou H, Brockington M, Jungbluth H, Monk D, Stanier P, Sewry CA, Moore GE, Muntoni F: **Epigenetic allele silencing unveils recessive *RYR1* mutations in core myopathies.** *Am J Hum Genet* 2006, **79**:859-868.
57. Chen H, Taylor NP, Sotamaa KM, Mutch DG, Powell MA, Schmidt AP, Feng S, Hampel HL, de la Chapelle A, Goodfellow PJ: **Evidence for heritable predisposition to epigenetic silencing of *MLH1*.** *Int J Cancer* 2007, **120**:1684-1688.
58. Hitchins MP, Wong JJ, Suthers G, Suter CM, Martin DI, Hawkins NJ, Ward RL: **Inheritance of a cancer-associated *MLH1* germ-line epimutation.** *N Engl J Med* 2007, **356**:697-705.

59. Vilar E, Scaltriti M, Balmana J, Saura C, Guzman M, Arribas J, Baselga J, Tabernero J: **Microsatellite instability due to *hMLH1* deficiency is associated with increased cytotoxicity to irinotecan in human colorectal cancer cell lines.** *Br J Cancer* 2008, **99**:1607-1612.
60. Lin JC, Jeong S, Liang G, Takai D, Fatemi M, Tsai YC, Egger G, Gal-Yam EN, Jones PA: **Role of nucleosomal occupancy in the epigenetic silencing of the *MLH1* CpG island.** *Cancer Cell* 2007, **12**:432-444.
61. Ren F, Wang D, Jiang Y: **Epigenetic inactivation of *hMLH1* in the malignant transformation of ovarian endometriosis.** *Arch Gynecol Obstet* 2012, **285**:215-221.
62. Wu PY, Fan YM, Wang YP: **Germ-line epimutations and human cancer.** *Ai Zheng* 2009, **28**:1236-1242.
63. Morgan HD, Sutherland HG, Martin DI, Whitelaw E: **Epigenetic inheritance at the *agouti* locus in the mouse.** *Nat Genet* 1999, **23**:314-318.
64. Martin DI, Cropley JE, Suter CM: **Environmental influence on epigenetic inheritance at the *Ary* allele.** *Nutr Rev* 2008, **66 Suppl 1**:S12-14.
65. Morgan HD, Jin XL, Li A, Whitelaw E, O'Neill C: **The culture of zygotes to the blastocyst stage changes the postnatal expression of an epigenetically labile allele, *agouti viable yellow*, in mice.** *Biol Reprod* 2008, **79**:618-623.
66. Perry WL, Copeland NG, Jenkins NA: **The molecular basis for dominant yellow *agouti* coat color mutations.** *Bioessays* 1994, **16**:705-707.
67. Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, Woychik RP: **Differential expression of a new dominant *agouti* allele (*Aiapy*) is correlated with methylation state and is influenced by parental lineage.** *Genes Dev* 1994, **8**:1463-1472.
68. Ruvinsky A, Flood WD, Zhang T, Costantini F: **Unusual inheritance of the *AxinFu* mutation in mice is associated with widespread rearrangements in the proximal region of chromosome 17.** *Genet Res* 2000, **76**:135-147.
69. Vasicek TJ, Zeng L, Guan XJ, Zhang T, Costantini F, Tilghman SM: **Two dominant mutations in the mouse fused gene are the result of transposon insertions.** *Genetics* 1997, **147**:777-786.

70. Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV, Whitelaw E: **Transgenerational inheritance of epigenetic states at the murine *Axin(Fu)* allele occurs after maternal and paternal transmission.** *Proc Natl Acad Sci U S A* 2003, **100**:2538-2543.
71. Waterland RA, Dolinoy DC, Lin JR, Smith CA, Shi X, Tahiliani KG: **Maternal methyl supplements increase offspring DNA methylation at *Axin Fused*.** *Genesis* 2006, **44**:401-406.
72. Fernandez-Gonzalez R, Ramirez MA, Pericuesta E, Calle S, Gutierrez-Adan A: **Histone Modifications at the Blastocyst *Axin/IFU* Locus Mark the Heritability of In Vitro Culture-Induced Epigenetic Alterations in Mice.** *Biol Reprod* 2010.
73. Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB: **A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening.** *Nat Genet* 2006, **38**:948-952.
74. Gustafsson A: ***Linnaeus' peloria*: the history of a monster.** *TAG* 1979, **54**:241-248.
75. Cubas P, Vincent C, Coen E: **An epigenetic mutation responsible for natural variation in floral symmetry.** *Nature* 1999, **401**:157-161.
76. Li B, Carey M, Workman JL: **The role of chromatin during transcription.** *Cell* 2007, **128**:707-719.
77. Katz DJ, Edwards TM, Reinke V, Kelly WG: **A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory.** *Cell* 2009, **137**:308-320.
78. Arico JK, Katz DJ, van der Vlag J, Kelly WG: **Epigenetic patterns maintained in early *Caenorhabditis elegans* embryos can be established by gene activity in the parental germ cells.** *PLoS Genet* 2011, **7**:e1001391.
79. Agger K, Christensen J, Cloos PA, Helin K: **The emerging functions of histone demethylases.** *Curr Opin Genet Dev* 2008, **18**:159-168.
80. Zhou X, Ma H: **Evolutionary history of histone demethylase families: distinct evolutionary patterns suggest functional divergence.** *BMC Evol Biol* 2008, **8**:294.

81. Jeong JH, Song HR, Ko JH, Jeong YM, Kwon YE, Seol JH, Amasino RM, Noh B, Noh YS: **Repression of *FLOWERING LOCUS T* chromatin by functionally redundant histone H3 lysine 4 demethylases in Arabidopsis.** *PLoS One* 2009, **4**:e8033.
82. Reinders J, Wulff BB, Mirouze M, Mari-Ordonez A, Dapp M, Rozhon W, Bucher E, Theiler G, Paszkowski J: **Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes.** *Genes Dev* 2009, **23**:939-950.
83. Richards EJ: **Quantitative epigenetics: DNA sequence variation need not apply.** *Genes Dev* 2009, **23**:1601-1605.
84. Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJ: **The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene.** *Mol Cell* 2000, **6**:791-802.
85. Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA: **Epigenetic reprogramming in mouse primordial germ cells.** *Mech Dev* 2002, **117**:15-23.
86. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W: **Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency.** *Nature* 2010, **463**:1101-1105.
87. Monk M, Boubelik M, Lehnert S: **Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development.** *Development* 1987, **99**:371-382.
88. Feng S, Jacobsen SE, Reik W: **Epigenetic reprogramming in plant and animal development.** *Science* 2010, **330**:622-627.
89. Bourc'his D, Voinnet O: **A small-RNA perspective on gametogenesis, fertilization, and early zygotic development.** *Science* 2010, **330**:617-622.
90. Richards EJ: **Inherited epigenetic variation--revisiting soft inheritance.** *Nat Rev Genet* 2006, **7**:395-401.
91. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al: **Role of transposable elements in heterochromatin and epigenetic control.** *Nature* 2004, **430**:471-476.

92. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR: **Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants.** *Science* 2011.
93. Lejeune E, Allshire RC: **Common ground: small RNA programming and chromatin modifications.** *Curr Opin Cell Biol* 2011, **23**:258-265.
94. Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, et al: **A role for RNAi in the selective correction of DNA methylation defects.** *Science* 2009, **323**:1600-1604.
95. Verdel A, Vavasseur A, Le Gorrec M, Touat-Todeschini L: **Common themes in siRNA-mediated epigenetic silencing pathways.** *Int J Dev Biol* 2009, **53**:245-257.
96. Hajkova P, Ancelin K, Waldmann T, Lacoste N, Lange UC, Cesari F, Lee C, Almouzni G, Schneider R, Surani MA: **Chromatin dynamics during epigenetic reprogramming in the mouse germ line.** *Nature* 2008, **452**:877-881.
97. van der Heijden GW, Castaneda J, Bortvin A: **Bodies of evidence - compartmentalization of the piRNA pathway in mouse fetal prospermatogonia.** *Curr Opin Cell Biol* 2010, **22**:752-757.
98. Hajkova P: **Epigenetic reprogramming in the germline: towards the ground state of the epigenome.** *Philos Trans R Soc Lond B Biol Sci* 2011, **366**:2266-2273.
99. Boiani M, Eckardt S, Leu NA, Scholer HR, McLaughlin KJ: **Pluripotency deficit in clones overcome by clone-clone aggregation: epigenetic complementation?** *EMBO J* 2003, **22**:5304-5312.
100. Whitworth KM, Prather RS: **Somatic cell nuclear transfer efficiency: how can it be improved through nuclear remodeling and reprogramming?** *Mol Reprod Dev* 2010, **77**:1001-1015.
101. Balbach ST, Esteves TC, Houghton FD, Siatkowski M, Pfeiffer MJ, Tsurumi C, Kanzler B, Fuellen G, Boiani M: **Nuclear reprogramming: kinetics of cell cycle and metabolic progression as determinants of success.** *PLoS One* 2012, **7**:e35322.
102. Hosseini SM, Hajian M, Forouzanfar M, Moulavi F, Abedi P, Asgari V, Tanhaei S, Abbasi H, Jafarpour F, Ostadhosseini S, et al: **Enucleated ovine oocyte**

- supports human somatic cells reprogramming back to the embryonic stage.** *Cell Reprogram* 2012, **14**:155-163.
103. Noggle S, Fung HL, Gore A, Martinez H, Satriani KC, Prosser R, Oum K, Paull D, Druckenmiller S, Freeby M, et al: **Human oocytes reprogram somatic cells to a pluripotent state.** *Nature* 2011, **478**:70-75.
 104. Pan G, Wang T, Yao H, Pei D: **Somatic cell reprogramming for regenerative medicine: SCNT vs. iPS cells.** *Bioessays* 2012, **34**:472-476.
 105. Niemann H, Carnwath JW, Herrmann D, Wiczorek G, Lemme E, Lucas-Hahn A, Olek S: **DNA methylation patterns reflect epigenetic reprogramming in bovine embryos.** *Cell Reprogram* 2010, **12**:33-42.
 106. Cui XS, Zhang DX, Ko YG, Kim NH: **Aberrant epigenetic reprogramming of imprinted microRNA-127 and Rtl1 in cloned mouse embryos.** *Biochem Biophys Res Commun* 2009, **379**:390-394.
 107. Su J, Wang Y, Li R, Peng H, Hua S, Li Q, Quan F, Guo Z, Zhang Y: **Oocytes Selected Using BCB Staining Enhance Nuclear Reprogramming and the In Vivo Development of SCNT Embryos in Cattle.** *PLoS One* 2012, **7**:e36181.
 108. Wang F, Kou Z, Zhang Y, Gao S: **Dynamic reprogramming of histone acetylation and methylation in the first cell cycle of cloned mouse embryos.** *Biol Reprod* 2007, **77**:1007-1016.
 109. Jullien J, Pasque V, Halley-Stott RP, Miyamoto K, Gurdon JB: **Mechanisms of nuclear reprogramming by eggs and oocytes: a deterministic process?** *Nat Rev Mol Cell Biol* 2011, **12**:453-459.
 110. Trifonov EN: **Sequence-dependent deformational anisotropy of chromatin DNA.** *Nucleic Acids Res* 1980, **8**:4041-4053.
 111. Trifonov EN, Sussman JL: **The pitch of chromatin DNA is reflected in its nucleotide sequence.** *Proc Natl Acad Sci U S A* 1980, **77**:3816-3820.
 112. Frenkel ZM, Bettecken T, Trifonov EN: **Nucleosome DNA sequence structure of isochores.** *BMC Genomics* 2011, **12**:203.
 113. Trifonov EN: **Nucleosome positioning by sequence, state of the art and apparent finale.** *J Biomol Struct Dyn* 2010, **27**:741-746.

114. Rapoport AE, Frenkel ZM, Trifonov EN: **Nucleosome positioning pattern derived from oligonucleotide compositions of genomic sequences.** *J Biomol Struct Dyn* 2011, **28**:567-574.
115. Zofall M, Persinger J, Kassabov SR, Bartholomew B: **Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome.** *Nat Struct Mol Biol* 2006.
116. Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF: **Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome.** *Nature* 2007, **446**:572-576.
117. Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, et al: **Relationship between nucleosome positioning and DNA methylation.** *Nature* 2010, **466**:388-392.
118. Cui F, Zhurkin VB: **Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro.** *J Biomol Struct Dyn* 2010, **27**:821-841.
119. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K: **Dynamic regulation of nucleosome positioning in the human genome.** *Cell* 2008, **132**:887-898.
120. Meneghini MD, Wu M, Madhani HD: **Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin.** *Cell* 2003, **112**:725-736.
121. Li B, Pattenden SG, Lee D, Gutierrez J, Chen J, Seidel C, Gerton J, Workman JL: **Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling.** *Proc Natl Acad Sci U S A* 2005, **102**:18385-18390.
122. Deal RB, Topp CN, McKinney EC, Meagher RB: **Repression of flowering in *Arabidopsis* requires activation of *FLOWERING LOCUS C* expression by the histone variant H2A.Z.** *The Plant cell* 2007, **19**:74-83.
123. Gent JI, Schneider KL, Topp CN, Rodriguez C, Presting GG, Dawe RK: **Distinct influences of tandem repeats and retrotransposons on CENH3 nucleosome positioning.** *Epigenetics Chromatin* 2011, **4**:3.
124. Zheng D: **Asymmetric histone modifications between the original and derived loci of human segmental duplications.** *Genome Biol* 2008, **9**:R105.

125. Margueron R, Trojer P, Reinberg D: **The key to development: interpreting the histone code?** *Curr Opin Genet Dev* 2005, **15**:163-176.
126. Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR: **Distinctive chromatin in human sperm packages genes for embryo development.** *Nature* 2009, **460**:473-478.
127. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129**:823-837.
128. Dillon N: **Gene regulation and large-scale chromatin organization in the nucleus.** *Chromosome Res* 2006, **14**:117-126.
129. Murrell A, Heeson S, Reik W: **Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops.** *Nat Genet* 2004, **36**:889-893.
130. Teller K, Solovei I, Buiting K, Horsthemke B, Cremer T: **Maintenance of imprinting and nuclear architecture in cycling cells.** *Proc Natl Acad Sci U S A* 2007, **104**:14970-14975.
131. Gaillard C, Borde C, Gozlan J, Marechal V, Strauss F: **A high-sensitivity method for detection and measurement of HMGB1 protein concentration by high-affinity binding to DNA hemicatenanes.** *PLoS One* 2008, **3**:e2855.
132. Jaouen S, de Koning L, Gaillard C, Muselikova-Polanska E, Stros M, Strauss F: **Determinants of specific binding of HMGB1 protein to hemicatenated DNA loops.** *J Mol Biol* 2005, **353**:822-837.
133. Gondor A, Ohlsson R: **Chromosome crosstalk in three dimensions.** *Nature* 2009, **461**:212-217.
134. Deng W, Blobel GA: **Do chromatin loops provide epigenetic gene expression states?** *Curr Opin Genet Dev* 2010, **20**:548-554.
135. Sumida N, Ohlsson R: **Chromosomal networks as mediators of epigenetic states: the maternal genome connection.** *Epigenetics* 2011, **5**:297-300.
136. Riggs AD, Martienssen RA, Russo VEA: **Introduction.** In *Epigenetic mechanisms of gene regulation*. Edited by Russo VEA. Cold Spring Harbor, NY Press: Cold Spring Harbor Laboratory Press; 1996.

137. Wang X, Zhu H, Snieder H, Su S, Munn D, Harshfield G, Maria BL, Dong Y, Treiber F, Gutin B, Shi H: **Obesity related methylation changes in DNA of peripheral blood leukocytes.** *BMC Med* 2010, **8**:87.
138. Jeffries MA, Dozmorov M, Tang Y, Merrill JT, Wren JD, Sawalha AH: **Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus.** *Epigenetics* 2011, **6**:593-601.
139. Luna E, Bruce TJ, Roberts MR, Flors V, Ton J: **Next-generation systemic acquired resistance.** *Plant Physiol* 2012, **158**:844-853.
140. Yazbek SN, Spiezio SH, Nadeau JH, Buchner DA: **Ancestral paternal genotype controls body weight and food intake for multiple generations.** *Hum Mol Genet* 2010.
141. McDevitt MA: **Clinical applications of epigenetic markers and epigenetic profiling in myeloid malignancies.** *Semin Oncol* 2012, **39**:109-122.
142. Radpour R, Kohler C, Haghghi MM, Fan AX, Holzgreve W, Zhong XY: **Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array.** *Oncogene* 2009, **28**:2969-2978.
143. Deng D, Liu Z, Du Y: **Epigenetic alterations as cancer diagnostic, prognostic, and predictive biomarkers.** *Adv Genet* 2010, **71**:125-176.
144. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O'Hanlon TP, Rider LG, et al: **Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus.** *Genome Res* 2010, **20**:170-179.
145. Tilgner K, Atkinson SP, Yung S, Golebiewska A, Stojkovic M, Moreno R, Lako M, Armstrong L: **Expression of GFP under the control of the RNA helicase VASA permits fluorescence-activated cell sorting isolation of human primordial germ cells.** *Stem Cells* 2010, **28**:84-92.
146. Merbs SL, Khan MA, Hackler L, Jr., Oliver VF, Wan J, Qian J, Zack DJ: **Cell-specific DNA methylation patterns of retina-specific genes.** *PLoS One* 2012, **7**:e32602.
147. Deal RB, Henikoff S: **A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue.** *Dev Cell* 2010, **18**:1030-1040.

148. Deal RB, Henikoff S: **The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*.** *Nat Protoc* 2011, **6**:56-68.
149. Jiang G, Zheng L, Pu J, Mei H, Zhao J, Huang K, Zeng F, Tong Q: **Small RNAs Targeting Transcription Start Site Induce Heparanase Silencing through Interference with Transcription Initiation in Human Cancer Cells.** *PLoS One* 2012, **7**:e31379.
150. Jauvion V, Rivard M, Bouteiller N, Elmayan T, Vaucheret H: ***RDR2* partially antagonizes the production of *RDR6*-dependent siRNA in sense transgene-mediated PTGS.** *PLoS One* 2012, **7**:e29785.
151. Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ: **RNA-mediated chromatin-based silencing in plants.** *Curr Opin Cell Biol* 2009, **21**:367-376.

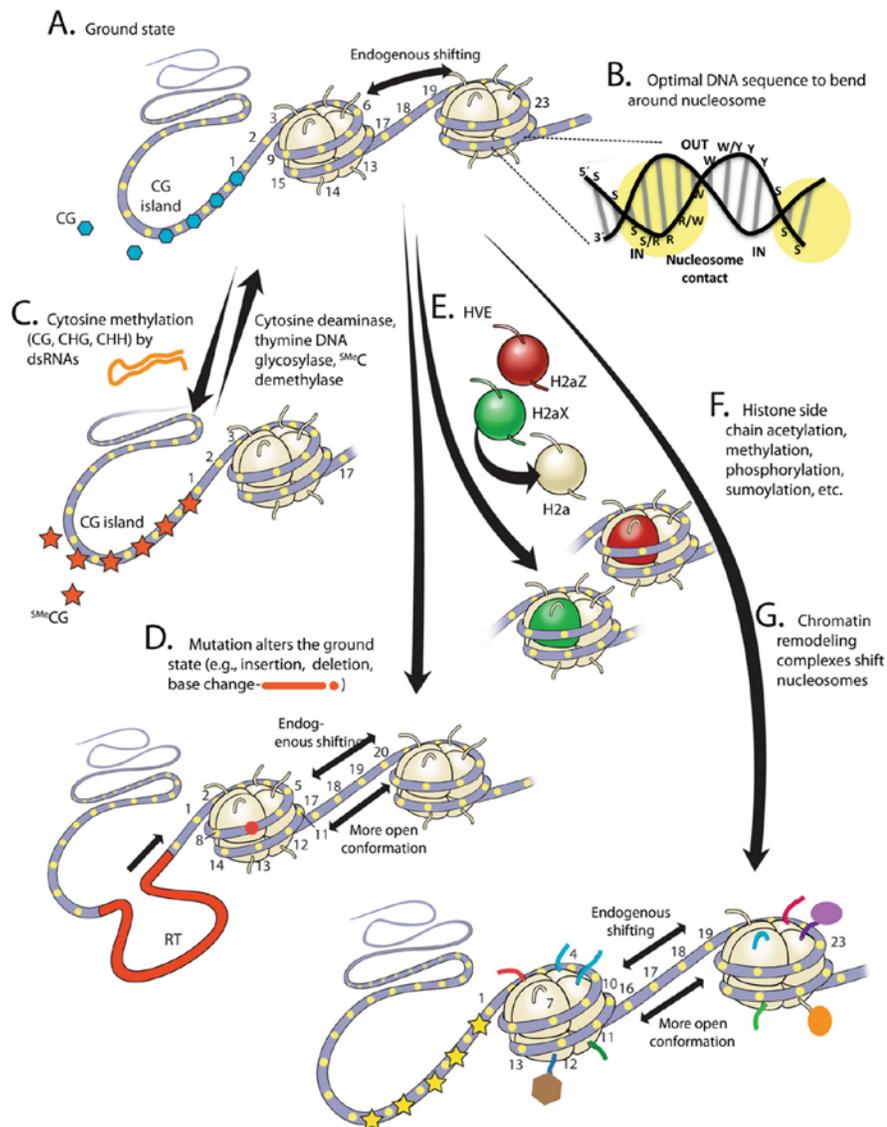


Figure A1.1: Summary of relationship between epitype and DNA sequence. A.

Theoretical ground state for a chromatin structure comprised of naked DNA bound to two nucleosomes and an unbound upstream DNA region. Every 10 bp the approximately two base pairs of inward facing surface of the DNA helix has the potential to contact and bind nucleosomal histones (e.g., yellow ovals numbered 1 to 23 for region surrounding one nucleosome, see **B**). Each nucleosome has the potential to bind 14 such two base pair

regions. **B.** One 10 bp region of the DNA helix with the consensus ((Y)RRRRRYYYYY(R) provides a bend for optimal nucleosome binding. Nucleotides that provide strong or weak nucleosome binding are indicated (S = strong binding to G or C nucleotides, W = weak binding to A or T nucleotides, R = purine, Y = pyrimidine, IN identifies the surface facing the nucleosome, and OUT the surface facing away from the nucleosome). The strength of nucleosome binding and positioning to 147 bp stretches of DNA appears to be determined by the sum of affinities for 14 small sequences (yellow ovals, same as in **A**). **C.** Double stranded (ds) RNAs (e.g., siRNA, piRNA, miRNA) program cytosine methylation for transgenerational inheritance and somatic inheritance in different tissues, while various enzymes remove 5MeC. **D.** Mutations such as single nucleotide polymorphisms (SNPs, red dot) and inserted retrotransposons (RT, red line) may alter nucleosome binding and the stochastic movement of nucleosomes. **E.** Histone variant exchange (HVE) by a subset of chromatin remodeling complexes (e.g., SWR1) replaces common core histones (e.g., H2A) with specialized protein sequence variants (e.g., H2AZ, H2AX). **F.** A variety of histone PTMs of primarily lysine and arginine residues at the N- and C- termini of core histones produce a diverse “histone code” for different nucleosomes. **G.** A large number of chromatin remodeling machines (e.g., SWI/SNF, INO80) control nucleosome positioning, often moving nucleosomes in approximately 10 bp increments. Not shown is that the individual epitypes interact with each other to produce complex epitypes. For example, a subset of individuals may contain in their genome a retrotransposon that is targeted by small RNAs, which cause the hypermethylation or hypomethylation of adjacent sequences and alters gene expression (i.e., the interaction of **C** and **D**).

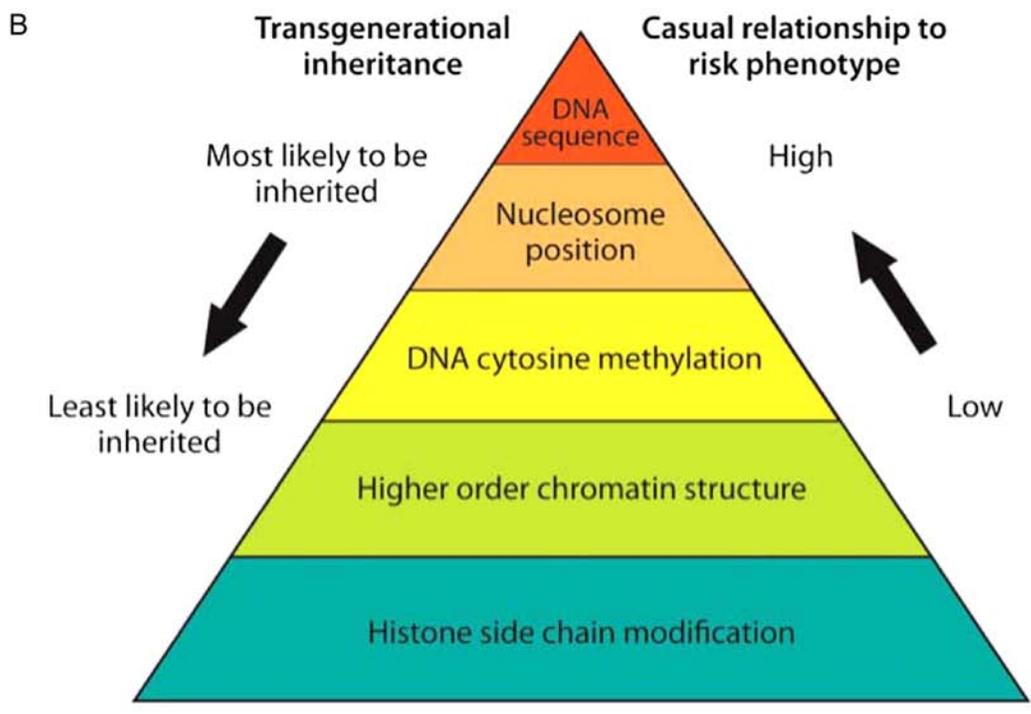
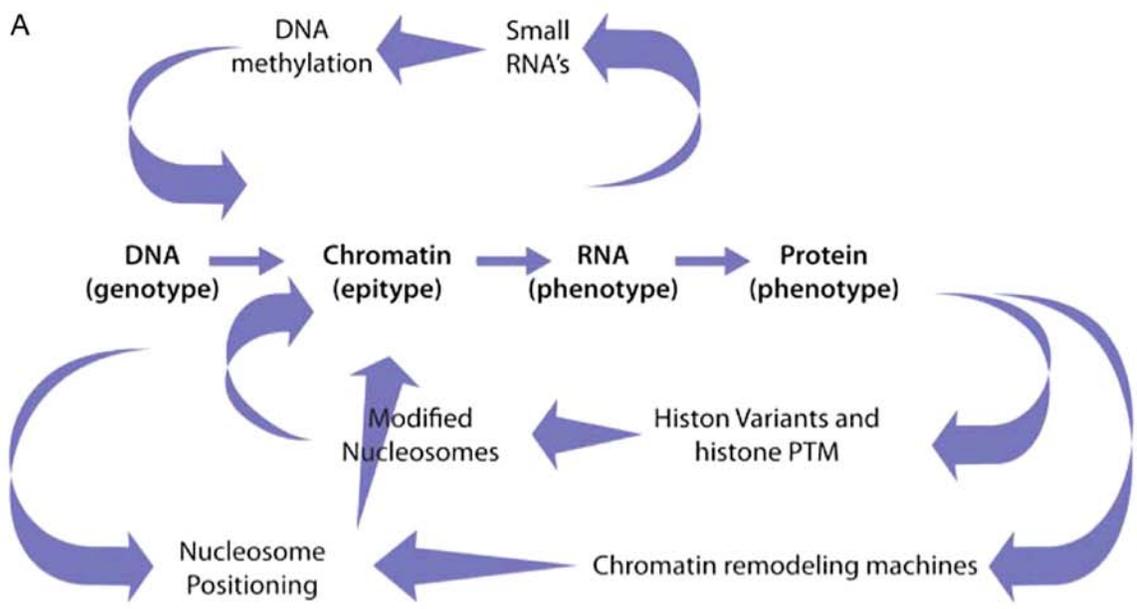


Figure A1.2: The relationships among genotype, epitype, and phenotype. **A.** The informational relationship and interaction of genotype, epitype and phenotype described in the context of the central dogma of molecular genetics. **B.** A pyramid illustrating the likelihood of different classes of epitypes being transgenerationally inherited and ranking the relative causal relationships of these epitypes to the risk of an aberrant phenotype.

Genes/Sequences affected	Genotype contribution	Known aberrant epitype/gene expression	Phenotype of epimutation or epigenetic change	Species	Supports/rejects hypothesis
A. Direct analysis of trans-generational inheritance					
1. <i>RYR1</i> ryanodine-receptor	Unknown	Cytosine hypermethylation/silenced	hyperthermia, core myopathies	human	Rejects
2. <i>MLH1</i> (Homolog of mismatch repair protein MutL)	Allele specific silencing	Cytosine hypermethylation/silenced	Colorectal or endometrial cancers	human	Weakly supports
3. <i>AGOUT1</i> (paracrine signaling peptide)	Alleles with retrotransposon	Cytosine hypomethylation/activation	Yellow, obese	mouse	Supports
4. <i>AXIN1-FUSED</i>	Alleles with retrotransposon	Cytosine hypomethylation, histone acetylation/activation	Axin-fused kinked tail	mouse	Supports
5. <i>CNR</i> Colorless Non-Ripening	Native CpG rich region	Cytosine hypermethylation/silenced	Carotenoid synthesis	tomato	Rejects
6. <i>CYC</i> – <i>cycloidea</i> (transcription factor)	Native CpG rich region and possible genotype difference	Cytosine hypermethylation/silenced	Floral morphology	<i>Linaria vulgaris</i>	Likely supports
7. H3K4Me2 demethylase	None identified	Histone H3 lysine4 dimethylation retained causing gene activation	Germ line immortality	<i>Caenorhabditis elegans</i>	Likely rejects
8. Quantitative epigenetic trait loci (for example, many loci)	DNA DEMETHYLATION1 ddm1/ddm1 restored to DDM1/DDM1	Cytosine re-methylation and re-silencing	Flowering time and plant height	<i>Arabidopsis thaliana</i>	Supports
9. Reprogramming of ^{5mC} by dsRNA	siRNA, miRNA, piRNA, and other dsRNAs	Cytosine re-methylation and re-silencing	Complex, molecular, and developmental	Arabidopsis, mice	Supports
10. Somatic cell nuclear transfer	Genome-wide	Cytosine re-methylation and histone modifications	Embryonic and fetal development	Mice, sheep, pigs, cows	Mostly supports
B. Indirect analysis using sequence conservation and gene duplication					
1. RRRRRYYYYY repeat throughout the genome	10.5 bp repeats position most nucleosomes	N.M.	N.M.	Diverse animal species	Supports
2a. Histone H2AZ in > 1,000 nucleosomes	10 bp repeat of G+C and A+T rich dinucleotides	Histone H2AZ variant positioning	Potentiated for expression, N.M.	Yeast, human, Arabidopsis	Supports
2b. H2AZ in <i>FLC</i> , <i>MAF4</i> , <i>MAF5</i>	Subfamily of three recently duplicated MADS box genes	Bimodal distribution of H2AZ enriched nucleosomes/activated	Altered flowering time and gene expression	Arabidopsis	Supports
3. Histone CenH3 in ~100,000 nucleosomes	10 bp repeat of AA or TT dinucleotides	Histone CenH3 variant positioning	Essential for chromosomal segregation, N.M.	maize	Supports
4. Blood plasminogen genes (PMGs)	Cytosine methylation in 208 bp region upstream of four PMG genes	N.M.	Demethylation activates four linked PMG alleles genes in liver. Methylation silences in other organs.	human	Supports
5. 1600 segmental duplications	Duplicated gene sequences	Several different histone side chain modifications	Duplicate alleles generally silenced relative to active parental allele. N.M.	human	Rejects
6. <i>HoxD</i> cluster	Five gene duplicated HOXD genes	Modestly conserved nucleosomal and H3K4Me2 patterns	N.M.	human	Supports
7. DNA loops and microsatellites	Concatenated DNA loops and trans-chromosomal contacts	Binding by HMG box proteins to control gene expression	N.M.	mammals	Modestly supports

ND., no data; N.M., not based on a mutational study.

Table A1.1: Examples of genes and specific sequences that support or reject the hypothesis that genotype predisposes transgenerationally inherited epitype and phenotype