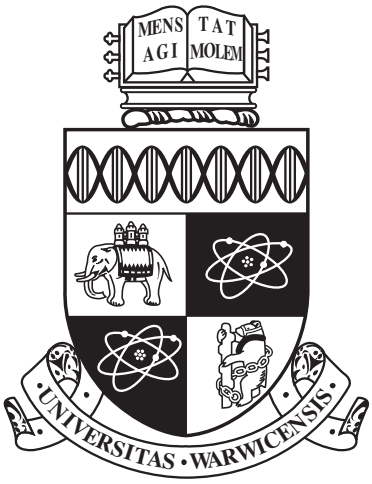


Functional differences among H3K27me3 histone demethylases in *Arabidopsis thaliana*

by

Lorenzo Pellegrini

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Life Sciences

School of Life Sciences

April 2023

THE UNIVERSITY OF
WARWICK

Abstract

Histone demethylases REF6 and ELF6 share strikingly similar protein structures, yet they exhibit opposing phenotypes in plants. The molecular basis for this paradox has remained elusive. In this study, we investigated the evolutionary divergence and functional diversification of REF6 and ELF6 to better understand their contrasting roles in plant development and regeneration. Our phylogenetic analysis revealed that these demethylases diverged during the emergence of flowering plants, concomitant with the appearance of distinct roles in gene regulation. Further investigation through domain swapping experiments demonstrated that the Jmj catalytic domains of REF6 and ELF6 confer distinct functional properties, leading to differential effects on gene expression and chromatin dynamics in plants. Surprisingly, when expressed in human cells, the differences between REF6 and ELF6 became even more pronounced, emphasising the functional divergence between these paralogs. Moreover, our study revealed contrasting effects of REF6 and ELF6 on plant regeneration, providing new insights into their roles in tissue differentiation and organogenesis. Collectively, our findings unveil the molecular underpinnings of the opposing phenotypes of REF6 and ELF6, shedding light on their evolutionary divergence, functional diversification, and unique roles in plant development and regeneration.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Javier Antunez-Sanchez (University of Warwick, United Kingdom) Assisting with RNA-seq analysis and R scripts
(Chapter 4)

Dr Jing Yuan (University of Warwick, United Kingdom) Producing human cell lines for REF6, ELF6 and JMJ13 and FACS analysis (Chapter 4)

Acknowledgements

I would like to express my deep appreciation to Dr. Jose Gutierrez-Marcos, my supervisor, for providing me with the opportunity to become a member of his research team and for his assistance and guidance throughout my research project. I am also grateful to the Midlands Integrative Biosciences Training Partnership for providing the necessary funding for this project without which it would not have been feasible. Javier deserves special thanks for all of his assistance and counsel. I want to thank my family, my fiancée Yuka and my friend Andrew for their invaluable support throughout the years.

Abbreviations

ACT2	ACTIN
AGL15	AGAMOUS-LIKE 15
ARF	AUXIN RESPONSE FACTORS
BBM	BABY BOOM
bp	Base pair
C2H2-ZnF	Cys2His2 zinc finger
CaMV 35S	Cucumber mosaic virus 35S promotor
CenH3	Centromeric H3
CLF	CURLY LEAF
DEX	Dexamethasone
DNA	DeoxyriboNucleic Acid
E(z)	ENHANCER OF ZESTE
ELF6	EARLY FLOWERING 6
EMF2	EMBRYONIC FLOWER2
Esc	EXTRA SEX COMBS
FIE	FERTILIZATION INDEPENDENT ENDOSPERM
FIS2	FERTILIZATION INDEPENDENT SEED 2
FLC	FLOWERING LOCUS C
GFP	Green Fluorescent Protein
H2A	Histone 2A
H2AK119	Lysine 119 on histone 2A
H2B	Histone 2B
H3	Histone 3
H3K27me3	Tri-methylation at lysine 27 on histone H3
H3K4	Lysine 4 on histone 3
H3K9	Lysine 9 on histone 3
H3K9me3	Tri-methylation at lysine 9 on histone 3
H4	Histone 4
HDA6	HISTONE DEACETYLASE 6
inRKD4ox	Arabidopsis line with inducible RKD4 over expression
JMJ	JUMONJI
JMJ13	JUMONJI13
JmJ-C	Jumonji-C domain containing protein
K	Lysine
LEC2	LEAFY COTYLEDON 2
LHP1	LIKE HETEROCHROMATIN PROTEIN 1
LSD1-4	LYSINE SPECIFIC DEMETHYLASE 1-4
me1/me2/me3	Mono-/di-/trimethylated
PcG	Polycomb Group
PKL	PICKLE

PRC	Polycomb Repressive Complex
PRC1/2	Polycomb repressive complex 1/2
PREs	Polycomb Response Elements
R	Arginine
REF6	RELATIVE OF EARLY FLOWERING 6
RKD4	RWP-RK DOMAIN-CONTAINING 4
RKD4ox	RKD4 over expression
RNA	RiboNucleic Acid
SE	Somatic Embryogenesis
SNP	Single Nucleotide Polymorphism
Su(z)12	Suppressor of Zeste
SWN	SWINGER
TF	Transcription Factors
VRN2	VERNALIZATION 2
μl	Micro litre
μm	Micro metre
WOX	WUSCHEL-RELATED HOM

Contents

List of Figures

1 Introduction

1.1 Chromatin control of gene expression

Diverse cellular phenotypes arise from identical genetic codes, with gene expression being selectively modulated in response to temporal and environmental cues (Alberts et al., 2014). This extraordinary regulation is facilitated by the chromatin structure, a highly organised system that incorporates DNA, structural proteins, RNA molecules, and DNA modifications (Bannister and Kouzarides, 2011).

The fundamental unit of chromatin is the nucleosome, which consists of a 146 base pair (bp) segment of DNA double helix wrapped around a histone protein dimer (Luger et al., 1997). Histones serve as spools around which the DNA strand winds, with their binding facilitated by electrochemical forces (Alberts et al., 2014). Each histone is an octamer containing two copies of histone proteins H2A, H2B, H3, and H4 (Luger et al., 1997). Histone H1, a fifth histone protein, secures the DNA wrapped around the nucleosome, binding at the DNA entry and exit sites and contributing to nucleosome stability and higher-order chromatin folding (Happel and Doenecke, 2009). Histones H2A, H2B, H3, and H4 comprise two regions: the histone "fold" and the histone "tail" (Alberts et al., 2014). These tails are crucial for regulating nucleosome core stability and facilitating the formation of higher-order chromatin architecture (Zhao et al., 2005). The carboxyl terminal end (C-terminal) enables interactions between histones and DNA, while the amino N-terminal tail contains residues targeted for post-translational modification (Strahl and Allis, 2000).

1.1.1 Histone modifications and chromatin states

Numerous residues within histones undergo modifications. Among the most prevalent of these are methylation, acetylation, phosphorylation, and ubiquitination. These modifications manifest in diverse residues and exhibit distinct patterns (Figure 1.1).

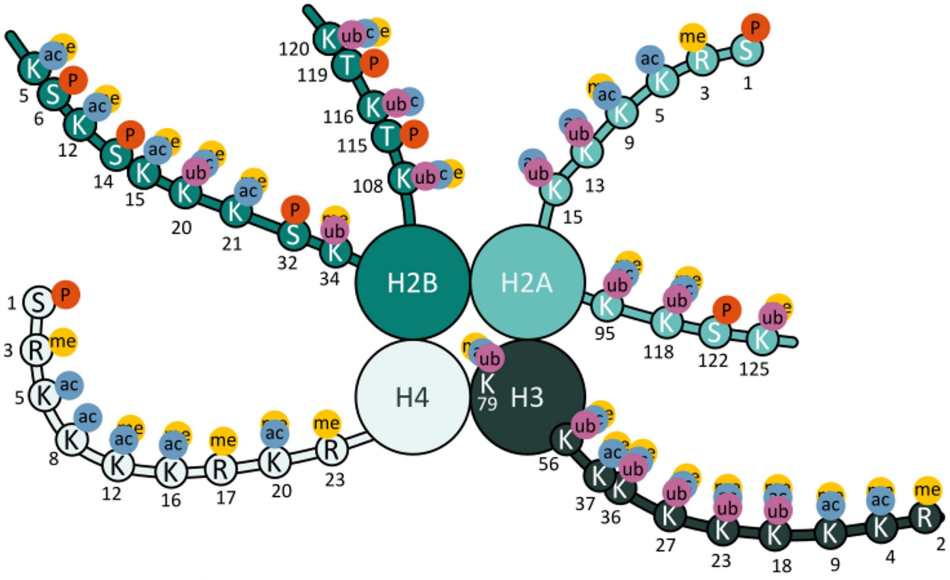


Figure 1.1: Examples for histone modifications of selected amino acid positions in histones H2A, H2B, H3, and H4. Figure adapted from Strahl and Allis (2000).

Such patterns, often referred to as the 'histone code', function as intricate genomic 'beacons' or 'docking stations' (Jenuwein and Allis, 2001). They orchestrate the recruitment of particular protein sets, which in turn modulate chromatin structure and thus regulate gene expression. The degree of chromatin compaction is intrinsically linked to gene expression; regions with a greater compaction show diminished accessibility in comparison to those with a more relaxed structure (Misteli, 2007). Constitutive heterochromatin, with its consistent compact nature, is vital for maintaining structural integrity and is predominantly located in specific areas such as centromeres and telomeres (Allshire and Madhani, 2018). This type of chromatin is rich in specific histone marks, one notable example being the trimethylation of histone H3 at lysine 9 (H3K9me3) (Franz et al., 2009). Beyond H3K9me3, constitutive heterochromatin also displays histone hypoacetylation and cytosine methylation (Casas-Delucchi et al., 2012; Suzuki et al., 2011). These modifications accentuate the condensed character of DNA in these domains. Conversely, euchromatin provides a more fluid structure that permits increased accessibility to the machinery essential for gene expression, thereby acting as a linchpin for gene regula-

tion (Cavalli and Misteli, 2013). Histone modifications frequently seen in association with gene activation within euchromatin include trimethylation at lysine 4 of histone H3 (H3K4me3), which commonly locates near the transcriptional start sites of active genes, and trimethylation at lysine 36 of histone H3 (H3K36me3), linked with transcriptional elongation (Li et al., 2008). Moreover, histone acetylation, particularly of H3 and H4, is generally associated with gene activation (Yan et al., 2003). In contrast, trimethylation at lysine 27 on histone H3 (H3K27me3) is typically correlated with gene repression within euchromatin, leading to a condensed chromatin configuration known as facultative heterochromatin (Bernstein et al., 2006). Importantly, these modifications are not isolated events; they collaborate synergistically to orchestrate gene expression. This collaborative essence of multiple histone alterations forms the backbone of the 'histone code'.

1.1.2 Reversible nature of marks and mechanisms for mark deposition

These histone marks are inherently dynamic and reversible, bestowing upon the cell the adaptability to refine its gene expression in alignment with environmental changes and throughout developmental stages (Wang et al., 2004; Jenuwein and Allis, 2001).

The nuanced balance of these marks is preserved by specific cellular machineries adept in introducing (writing), deciphering (reading), or negating (erasing) them, safeguarding their reversible nature.

1. Writers: These are enzymes tasked with appending modifications onto histone tails. A paradigmatic instance is the histone acetyltransferase (HAT) family, which appends acetyl groups to lysine residues, leading to a more relaxed chromatin framework conducive to transcriptional activation (Marmorstein, 2004). Conversely, histone methyltransferases, including SUV39H1 and EZH2, incorporate methyl groups onto lysine or arginine residues of histones. The impact on transcription—either promotion or suppression—is dictated by the specificity of the residue altered and the number of methyl groups deposited (Cardoso et al., 2000; Rea et al., 2000).

2. Readers: These proteins are equipped with domains tailored to discern and adhere to distinct histone marks, thereby translating the chromatin modification code into a cellular directive. Proteins endowed with bromodomains, for instance, detect and bind acetylated lysine residues (Tamkun et al., 1992). Chromodomains, by contrast, frequently affiliate with methylated lysines. Through these nuanced interactions, reader proteins can usher in supplementary factors to modified nucleosomes, orchestrating processes such as transcriptional initiation or curtailment (Koonin et al., 1995).

3. Erasers: The erasers, pivotal for the ephemerality of histone marks, are enzymes designated to annul these PTMs. Histone deacetylases (HDACs), serving as antitheses to HATs, are engaged in excising acetyl groups, typically culminating in a denser chromatin configuration and transcriptional inhibition (Torok and Grant, 2004). In the realm of methylation, enzymes such as the histone demethylases LSD1 and the Jumonji C-domain-containing proteins (JMJDs) play the role of custodians, ensuring the fluid balance of histone methylation states remains intact (Yang and Chou, 1999; Klose et al., 2006).

1.1.3 Epigenetic memory

The dynamic homeostasis of histone PTMs mediated by the writer, reader, and eraser establishes a refined epigenetic language, guiding chromatin structure and transcriptional outcomes. A subset of these dynamic chromatin alterations have been demonstrated to be passed on during mitosis and occasionally during meiosis (Almouzni and Cedar, 2016; Felsenfeld, 2014). These inheritable chromatin indicators are termed epigenetic marks. They can extend the influence of temporary environmental triggers, developmental signals, and cellular metabolic conditions on gene activity well beyond the initial exposure (Almouzni and Cedar, 2016; Felsenfeld, 2014). Various processes have developed to ensure that these epigenetic changes are replicated and preserved through cell divisions, consequently establishing cellular records that uphold distinct differentiation stages

The term "epigenetic" was introduced by Conrad Waddington in 1942 to describe the processes leading to adult organism development from a

zygote (Waddington, 1953). He portrayed cell differentiation as a ball on an "epigenetic landscape" rolling towards mature cell state valleys (Figure 1.2).

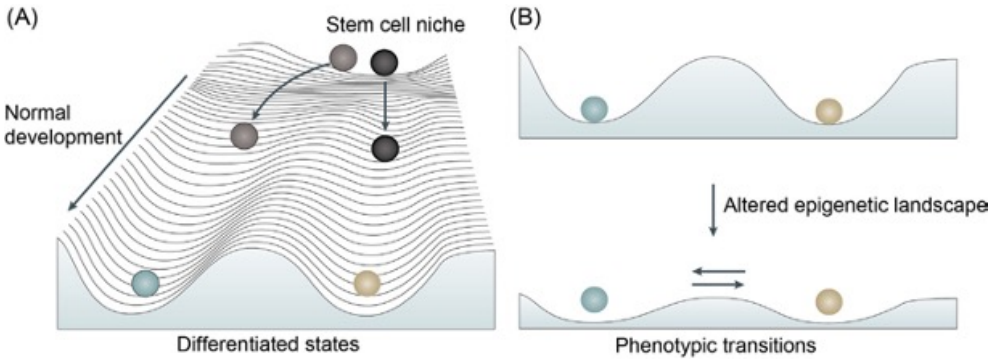


Figure 1.2: (A) Waddington's developmental landscape metaphor describes cell differentiation as balls navigating an "epigenetic landscape" towards valleys, representing mature cell states. Within this concept, "canalisation" ensures stable developmental outcomes regardless of perturbations, and the emergence of diverse phenotypes from a singular genotype indicates "phenotypic plasticity". As differentiation progresses, stem cells see a diminishing developmental potential. (B) This differentiation and restriction in developmental potential can be reversed. External or internal signals can reshape the epigenetic landscape, either easing trans-differentiation barriers or influencing dedifferentiation (as illustrated by the depth of the valleys). This suggests that, given the right cues, cell fates can be interchangeable, allowing cells to transform into varied types, emphasising the dynamic nature of cellular development. Figure adapted from Feinberg et al. (2016).

Within this landscape, "canalisation" represents stable developmental results despite environmental changes, whereas "developmental plasticity" indicates multiple cellular phenotypes from one genotype. The idea of a regulatory interaction layer with both genotype and environment predates the discovery of chromatin-based gene regulation (Noble, 2015). Modern epigenetics, building on this knowledge, refers to mitotically or meiotically stable gene function changes without DNA sequence alterations (Riggs and Porter, 1996). Epigenetic phenomena provide cellular chromatin state memories, crucial for maintaining cell type-specific gene expression and normal development. Since histone PTMs directly impact chromatin states, exploring their influence on chromatin compaction and gene expression is pivotal in shedding light on their roles in development and cellular identity.

1.1.4 Histone acetylation

Histone acetylation predominantly takes place on lysine residues, neutralising their positive charge. This neutralisation impedes electrostatic interactions between the DNA and the histones, leading to chromatin decompaction. Moreover, histone acetylation can inhibit interactions between adjacent nucleosomes, resulting in less compact genomic regions with high acetylation levels (Bannister and Kouzarides, 2011; Kouzarides, 2007). Histone acetylation is strongly connected to transcription and is enriched at the transcription start site of active genes. Acetylation provides a binding site for proteins containing an acetyl-lysine binding domain, known as "bromodomain". These proteins can recruit additional bromodomain-containing proteins, establishing a positive feedback loop and promoting further acetylation (Fujisawa and Filippakopoulos, 2017). Histone acetylation can also offer binding sites for other epigenetic effector complexes, such as the chromatin remodeling complex SWI/SNF, which utilises its bromodomain to bind to acetylated regions of damaged DNA and initiate DNA repair mechanisms (Tessarz and Kouzarides, 2014).

Acetylation is dynamically regulated by opposing enzyme families: histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Seto and Yoshida, 2014). Hyperacetylation of histone H3 and H4 is considered an active mark frequently associated with transcription. Genome-wide experiments suggest that multiple HATs are recruited simultaneously, acetylating multiple lysines at a given locus (Blasi et al., 2016). It is believed that multiple lysine charges must be altered to significantly affect chromatin structure. Among the most studied acetylation marks are the acetylation of lysines 4, 9, and 27 on histone H3, all promoting active gene expression (Tessarz and Kouzarides, 2014). However, histone acetylation alone is insufficient to trigger transcriptional activation (Choudhary et al., 2014).

1.1.5 Histone phosphorylation

Histone phosphorylation adds a negative charge to histones, similar to acetylation, and results in a more open chromatin conformation. It is

typically associated with gene expression and DNA damage repair (Rossetto et al., 2012). Histone phosphorylation can occur on serine, threonine, and tyrosine residues, playing an essential role in the "histone code" or combinatory function of post-translational modifications (PTMs) on chromatin (Rossetto et al., 2012). In plants, histone phosphorylation has been observed only on histone H3 and H2A, and has been associated with developmental processes such as flowering time regulation, chromosome condensation, and cell apoptosis (Bergmüller et al., 2007; Houben et al., 2011).

1.1.6 Histone ubiquitination

Histone ubiquitination differs from other PTMs as it involves the covalent binding of a large 76-amino acid protein, rather than small chemical groups. Histone ubiquitination plays a crucial role in various DNA-related processes, including DNA replication, transcription, and repair (Weake and Workman, 2008). In humans, it mainly occurs on histone H2A at lysine 119 and histone H2B at lysine 120, catalysed by the sequential action of E1, E2, and E3 ligase enzymes (Bergink et al., 2006). The substrates can be mono-ubiquitinated or poly-ubiquitinated. Poly-ubiquitination generates an irreversible signal for proteasomal-mediated degradation, while mono-ubiquitination creates a regulatory signal that can be reversed by ubiquitin-specific proteases (USPs) (Zuin et al., 2014). In plants, H2A mono-ubiquitination (H2Aub1) is associated with transcription repression mediated by the Polycomb pathway, whereas H2Aub1 is involved in transcription activation (Bonnet et al., 2022). H2Aub1 is implicated in regulating flowering time through transcriptional activation of FLC (FLOWERING LOCUS C) (Woodson et al., 2015). This mark is also associated with the expression of dormancy-related genes and circadian clock-related genes (Zhou et al., 2017).

1.1.7 Histone methylation

Histone methylation, unlike acetylation, does not alter the charge of histone tails. Instead, it can either enhance or suppress gene expression,

contingent on the specific residues being methylated and the number of methyl groups attached (Allis and Jenuwein, 2016). Histone methylation has been implicated in numerous biological processes, such as DNA repair, cell cycle, stress response, transcription, development, and differentiation. The dynamic nature of this mark can either be stable throughout a cell's life or inherited across generations (Zhao and Fernald, 2005).

In certain biological contexts, specific methylations must be stably maintained, such as those involved in the inheritance of a silenced chromatin state through mitosis. In other cases, these marks may be more susceptible to change, as when cells differentiate in response to environmental stimuli. Histone methylation is particularly complex, with up to three possible methylation states for lysines and two different types of mono- or di-methylation for arginines: symmetric and asymmetric (Black et al., 2012).

1.1.7.1 Lysine histone methylation

Lysine histone methylation has garnered significant attention in recent years due to its role in regulating various nuclear processes, including transcription and maintenance of genome integrity. Lysine methylation does not alter the net charge of the modified residues but increases hydrophobicity, potentially generating new binding sites for reader proteins that specifically recognise the methylated domain (Greer and Shi, 2012).

In *Arabidopsis*, histone lysine methylation primarily occurs at positions K4, K9, K27, and K36 of histone H3. H3K9 and H3K27 methylation are typically associated with silenced regions, whereas H3K4 and H3K36 methylation are linked to active genes (Liu et al., 2010). H3K27 methylation has been extensively studied for its importance in genome architecture and regulation. This specific lysine can harbour up to three methylation states or even acetylation, which is associated with active transcription.

1.1.7.2 H3K27me3

H3K27me1 is linked to transcription promotion in eukaryotes and seems to accumulate in transcribed genes (Berr et al., 2011). Loss of H3K27me1

leads to heterochromatin decondensation and release of silencing of transposable elements. H3K27me3, a highly characterised histone mark, is a hallmark of silenced gene expression conserved across the animal and plant kingdoms. In *Arabidopsis*, H3K27me3 represses between 15-60% of protein-coding genes (Berr et al., 2009). This mark exerts its repressive role in euchromatic regions and is preferentially associated with the whole transcribed region of inactive genes.

H3K27me3 plays a dual role in repressing transcription and heterochromatin formation, and selectively silencing genes involved in plant development (Berr et al., 2009). H3K27me3 often interacts with the active mark H3K4me3 in bivalent domains, which are crucial for proper cell differentiation (Bernatavichute et al., 2008). These bivalent domains are typically found in embryonic cells and are fundamental for proper cell differentiation. The dual-mark is present in key developmental genes requiring fine-tuned regulation, placing these genes in a "poised state". H3K27me3's role in developmental regulation provides a cellular memory to maintain the repressed transcription states of target genes during cell division (Zheng et al., 2016). In animals, the catalysation of H3K27 tri-methylation is mediated by Enhancer of Zeste (E(z)), a SET domain histone methyltransferase. This domain is part of a larger complex, the Polycomb repressive complex 2 (PRC2), which operates in tandem with PRC1 (Margueron and Reinberg, 2011). Together, these two complexes form the Polycomb group, essential for gene regulation and developmental processes.

1.2 Polycomb group proteins and their functions in plants

Polycomb group (PcG) proteins were initially identified in Drosophila melanogaster as regulators of homeotic genes during development, but it has become evident that PcG proteins are conserved in all multicellular organisms (Lewis, 1978; Alkema et al., 1997). In *Arabidopsis*, PcG proteins participate in various developmental processes, such as floral transition, fertilisation, transgenerational epigenetic memory, hormone signalling transduction, and stress responses (Goodrich et al., 1997; Gross-

niklaus et al., 1998). PcG proteins can be divided into two complexes: Polycomb repressive complex 2 (PRC2) and PRC1. PRC2 is responsible for H3K27me3 deposition, and its core components are Extra sex combs (Esc), the methyltransferase enhancer of zeste (E(z)), Suppressor of zeste 12 (Su(z)12), and p55 (Margueron and Reinberg, 2011). The PRC1 complex recognises and stabilises H3K27me3, introduces the repressive ubiquitination of H2AK121ub, and mediates chromatin compaction for gene silencing (Tavares et al., 2012). H2AK121ub is catalysed by two different PRC1 complex components, RING finger proteins BMI1s (B lymphoma Mo-MLV insertion region 1 homolog) and RINGs. The chromatin compaction is conducted by another PRC1 component, EMF1 (EMBRYONIC FLOWER 1) and its homologs in higher plants (Merini and Calonje, 2015).

1.2.1 PcG proteins' roles in *Arabidopsis*

In *Arabidopsis*, PcG proteins are involved in a broad array of biological processes. PRC1 and PRC2 are crucial for growth and differentiation throughout the life cycle. This is evidenced by mutants affecting the different components of the PcG complex, which display pleiotropic phenotypes of varying severity, including loss of vegetative growth, early flowering, abnormal flower formation, and sterility (Chanvivattana et al., 2004; Bouyer et al., 2011). Transcriptome analysis of mutant seedlings revealed enrichment for activation of genes involved in transcriptional regulation, developmental transition, growth, and differentiation (Wang et al., 2016a). PcG plays a fundamental role in major phase transition programs (Figure 1.3).

During germination, genes involved in seed-specific pathways, such as seed dormancy, desiccation tolerance, and seed storage compound accumulation, need to be repressed to allow growth and differentiation of the developing plant (Bouyer et al., 2011). In germinating seedlings, PcG components FERTILIZATION-INDEPENDENT ENDOSPERM (EMF1) and FIE repress seed regulatory genes ABSCISIC ACID INSENSITIVE 3 (ABI3) and LEAFY COTYLEDON 2 (LEC2), major regulatory genes which promote embryo development and maturation (Yang et al., 2016).

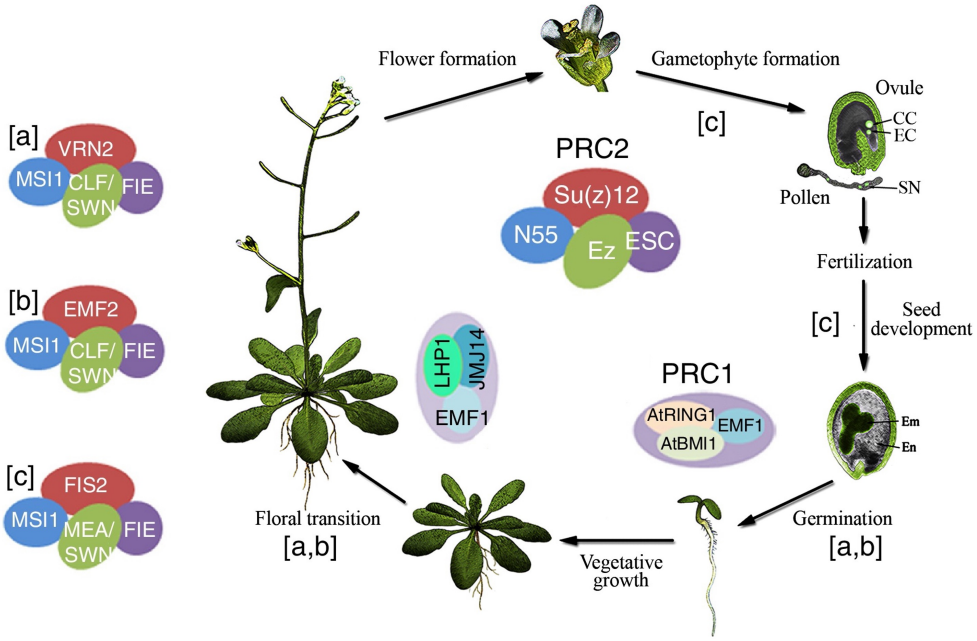


Figure 1.3: Polycomb-mediated suppression is critical for the establishment and preservation of cell identity in Arabidopsis. PcG protein assemblies are pivotal in managing transitions in plant development. Distinct PRC2 complexes, namely VRN-PRC2, EMF-PRC2, and FIS-PRC2, are instrumental at various stages of developmental phase transitions, as denoted alongside the arrows. Additionally, separate PRC1 complexes are implicated in the regulation of seed germination and the shift from vegetative growth to reproductive maturity. Figure adapted from Xiao and Wagner (2015) and Butenko and Ohad (2011).

Mutants impaired in the PRC1 and PRC2 components AtRING1A/B, AtBMI1A/B, CURLY LEAF (CLF) and SWINGER (SWN), EMF1, EMF2, or FIE show upregulation of these regulatory genes and of the downstream seed storage genes that produce starch, lipid, and proteins, including LIPID TRANSFER PROTEIN 3 (LTP3), OLEOSIN 2 and SEED STORAGE ALBUMIN1 (SESA1) (Bratzel et al., 2010; Bouyer et al., 2011). PcG proteins are also involved in multiple hormone signalling pathways. During germination, FIE and EMF1 repress genes that are positive regulators of abscisic acid, which promotes seed development, and repress negative regulators of gibberellic acid, involved in germination (Suzuki et al., 2007; Pu et al., 2013). Therefore, PcG proteins mediate the transition from seed to seedling by selectively repressing hormone production and signalling genes, as well as master seed regulatory and seed storage genes.

Seedlings need to suppress reproductive development to allow vegetative growth and the formation of photosynthetic leaves. Floral MADS box genes are direct targets of both EMF1 and FIE which act in concert to repress the floral program in seedlings (Calonje et al., 2008). Deregulation of multiple floral MADS box genes leads to a premature transition from vegetative to reproductive development (Hennig et al., 2003). The PcG complex not only affects the transition to the flowering phase by repressing floral genes during the vegetative phase but also promotes it when the time is appropriate by repressing the flowering repressor FLC (Yang et al., 2016). In *Arabidopsis*, FLC is activated in vegetative development to prevent premature flower transition. FLC expression is elevated by active histone marks such as the COMPASS-like complex-deposited H3K4me3, the EARLY FLOWERING IN SHORT days (EFS) catalysed H3K36me3, and histone H3K27 acetylation (Jiang et al., 2011). Meanwhile, PcG proteins deposit H3K27me3 at the FLC locus, making the FLC chromatin region in a bivalent state (Farrona et al., 2004). This bivalent chromatin state allows for precise control of developmental transitions, ensuring proper timing for flowering in response to environmental cues and plant development.

1.3 Histone demethylases as a counterbalance to silencing

During the life cycle of a plant, developmental programs that have been established may need to be erased in response to specific environmental or developmental signals, such as during gametogenesis (Hisanaga et al., 2019). Consequently, gene networks previously silenced by PRC2 complexes and H3K27me3 may need to be reactivated.

There are several types of enzymes that can remove specific methyl groups from histone proteins, leading to changes in chromatin structure and gene expression. In plants, there are two main classes of histone demethylases: the lysine-specific demethylase 1 (LSD1) homologs and the Jumonji C (JMJC) domain-containing proteins. LSD1 homologs are FAD-dependent enzymes that specifically target mono- and dimethylated lysine residues on histones. In *Arabidopsis*, the LSD1 homologs are called FLD and are involved in demethylating H3K4me1/2 and H3K9me1/2 (Yang and Chou, 1999). LSD1 homologs have been shown to play crucial roles in regulating gene expression and developmental processes in plants, such as flowering and seed development (Jiang et al., 2007; Hu et al., 2014). In contrast, JMJC domain-containing proteins are a diverse group of histone demethylases that possess a conserved JMJC domain, which binds to methylated lysine residues on histone proteins. This binding leads to the enzymatic removal of the methyl group, resulting in a change in chromatin structure (Klose et al., 2006). JMJC domain-containing proteins have been shown to demethylate various histone methylation marks, such as H3K4me3 and H3K27me3, depending on the specific protein (Lu et al., 2011b; Bernstein et al., 2006). Presented below is a comprehensive overview of histone modifications, delineating their genomic loci, impact on chromatin states, and their influence on euchromatic gene transcription, in *Arabidopsis*. Additionally, the table catalogues the enzymes responsible for both initiating and reversing each modification (Figure 1.4).

1.3.1 JmjC Histone demethylases in *Arabidopsis*

In *Arabidopsis*,a total of 21 JmjC proteins have been identified and can be classified into five groups according to phylogenetic information and pro-

Modification	Location	Effect	Transcription	Established by	Removed by
Ubiquitination					
H2Aub1	Eu-chromatin	Open	Activation	PcG complex	Deubiquitinases UBP12, UBP13
Methylation					
H3K4me1	Eu-chromatin	Open			
H3K4me2	Eu-chromatin	Open	Activation	ATX1-5	JMJ14
H3K4me3	Eu-chromatin	Open	Activation	ATX1, SDG8, ATX4	JMJ14
H3K9me1	Hetero-chromatin	Closed		SUVH2, SUVH4, SUVH6, SUVH5	
H3K9me2	Hetero-chromatin	Closed		SUVH2, SUVH4, SUVH6, SUVR4	
H3K9me3	Eu-chromatin	Closed	Repression	SUVH2, SUVH4, SUVH6, SUVR4	Histone demethylases, JMJD2 family
H3K27me1	Hetero-chromatin	Closed		ATXR5, ATXR6	
H3K27me2	Hetero-chromatin	Closed			Histone demethylases, JmjC-domain
H3K27me3	Eu-chromatin	Closed	Repression	CLF, SWN, MEA (PRC2)	Histone demethylases, JmjC-domain
H3K36me1	Eu-chromatin	Open		SDG8	
H3K36me2	Both	Open	Activation	SDG8	
H3K36me3	Eu-chromatin	Open		SETD2	
Acetylation					
H3K*ac/H4K*ac	Both	Open	Activation	Histone acetyltransferases, GNAT family: GCN5/HAG1, CBP/p300 family: HAC1, HAC5, HAC12, TAFII family: HAF2/TAF1	Histone deacetylases, RDP3 family: HDA19, HDA6, HDA1 family: HDA18, HD2 family: HD2A, HD2B, HD2C
Phosphorylation					
H3S10ph, H3S28ph, H2T11ph	Both	Open	Activation	Kinases	Phosphatases

Figure 1.4: Table of histone modification types, their genomic loci, and their subsequent effects on chromatin states (either open or closed) and euchromatic gene transcription, in *Arabidopsis*. Enzymes involved in the initiation and reversal of each modification are also catalogued. Adapted from Pfluger et al., 2007.

tein domain architecture. In mammals, two main proteins, UTX/KDM6A and JMJD3/KDM6B, act as histone demethylases for the H3K27 mark (Agger et al., 2007). These proteins have diverse roles in regulating

gene expression during embryonic development, cellular reprogramming, immune diseases, and cancer (Chang et al., 2019; Chen et al., 2023). However, these proteins are not conserved in plants. Instead, *Arabidopsis* has up to five proteins reported to have H3K27me3 demethylase activity, including JMJ30 (JUMONJI30), JMJ32, ELF6 (EARLY FLOWERING 6), REF6 (RELATIVE OF EARLY FLOWERING 6), and JMJ13 (Gan et al., 2014; Crevillén et al., 2014; Lu et al., 2011a; Zheng et al., 2019). JMJ30 and JMJ32 are JmjC domain-only proteins, while the other three proteins contain a zinc-finger domain (Figure 1.5). All these proteins are involved in regulating gene expression and developmental processes in plants.

The *Arabidopsis* JmjC domain-only group includes two H3K27me3 demethylases, JMJ30 and JMJ32, which belong to the conserved subfamily PKMD12 found in plants and animals. JMJ30 is known to regulate the circadian clock by modulating the expression of (CIRCADIAN CLOCK ASSOCIATED 1) and LHY (LATE ELONGATED HYPOCOTYL), and it is also involved in several developmental processes such as callus formation, growth arrest, and heat acclimation (Lu et al., 2011b). JMJ30 interacts with EFM (EARLY FLOWERING MYB PROTEIN) to regulate the expression of the floral integrator FT and participates in the flowering thermosensory pathway with JMJ32 (Yan et al., 2014). However, the specificity of the JmjC domain-only demethylases is not completely understood, as some studies suggest that JMJ30 can demethylate H3K36me3 or alter H3K9me3 levels at target genes (Lee et al., 2018).

The PKMD9 subfamily is unique to plants and consists of three *Arabidopsis* JmjC domain proteins that contain ZnFn motifs, ELF6, REF6 and JMJ13. ELF6 and REF6 are plant-specific homologs that have four C2H2-type ZnFn motifs in their N-terminal domains. The C2H2 zinc finger domain is characterised by a structure containing two cysteine and two histidine residues that coordinate a zinc ion. JMJ13 is related to ELF6 and REF6 but carries instead a C4HCHC-type helical-zinc finger cassette fused to the JmjC domain (Zheng et al., 2019). This arrangement also coordinates a zinc ion but in a different configuration. While C2H2 zinc fingers are primarily involved in DNA binding, but can also

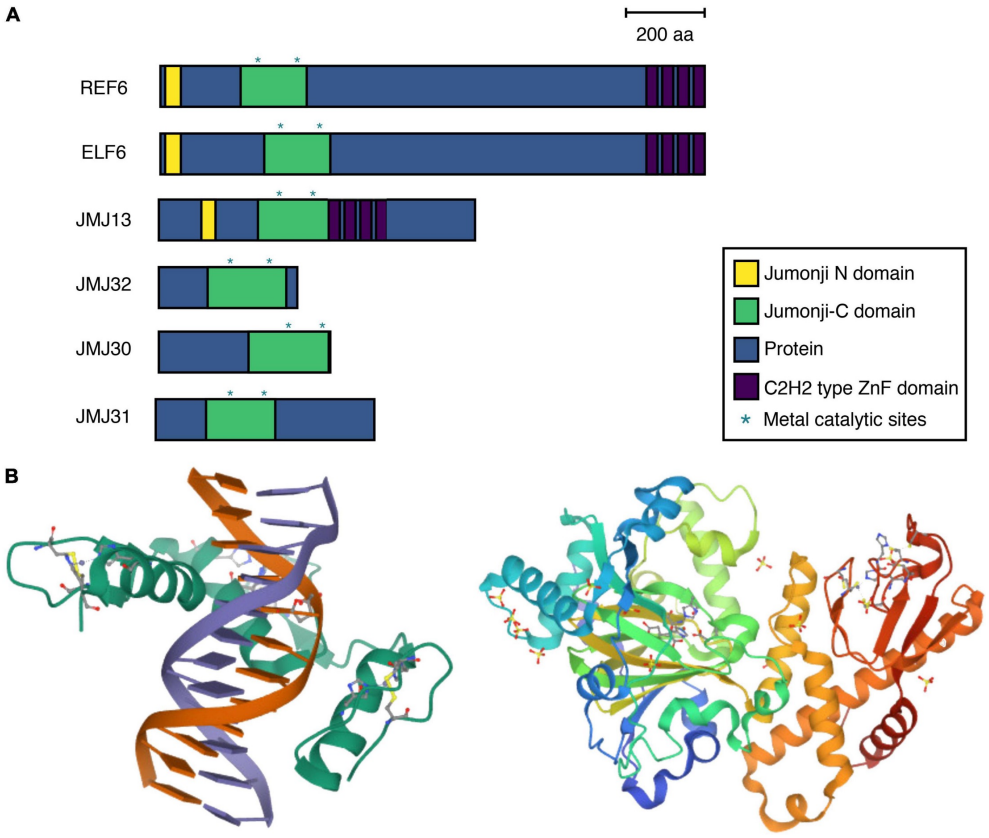


Figure 1.5: Conserved motifs, and three-dimensional structure of H3K27me3 demethylases in Arabidopsis thaliana. Domain architecture of Arabidopsis thaliana Jumonji domain–containing proteins (JMJs), highlighting the Jumonji N and C domains, protein domains, and C2H2-type zinc-finger domains in yellow, green, blue, and purple, respectively, with metal catalytic sites marked by asterisks. (B) Three-dimensional structures of JMJ proteins. On the left, REF6's structure with ZnF and double-stranded DNA, illustrates the interaction between protein and DNA. REF6 protein is shown in green, while dsDNA is shown in orange and purple. On the right, the JMJ13 catalytic domain complexed with AKG, depicted in blue and green for JMJ domains and orange and red for helical and ZnF domains, demonstrates substrate engagement. Structures derived from Protein Data Bank data. Figure adapted from Yamaguchi (2021).

interact with RNA and proteins, C4HCHC zinc fingers, are often associated with high-affinity binding to single-stranded nucleic acids, especially single-stranded RNAs (Garton et al., 2015; Wang et al., 2021).

A recent study demonstrated that JMJ13 is the most abundant H3K27 demethylase in pollen sperm cells, suggesting that it may play a role in resetting paternal H3K27me3 (Borg et al., 2020). Based on this and other evidence, the authors of the study have proposed that H3K27me3 demethylases have a crucial role in a complex, genome-wide epigenetic reprogramming mechanism that ensures that epigenetic marks are erased and not inherited by the next generation (Borg et al., 2021). Contrarily, a different study questions the prevailing ideas about epigenetic reprogramming in the germline, demonstrating that only a selective reset of H3K27me3 takes place at developmental regulators in sperm (Zhu et al., 2023). This research implies that a full reset of the parental epigenome might not be essential to ensure a correct developmental program for the offspring.

REF6 acts indirectly as a repressor of the FLC gene, which controls flowering time, and interacts with thousands of putative targets through its DNA binding domain (Gan et al., 2014; Cui et al., 2016b). It is widely expressed throughout the plant and plays several important roles, such as regulating flowering time and modulating gene expression. REF6 is also involved in the plant response to abiotic stresses, such as drought and salt stress (Sani et al., 2013; Pan et al., 2021).

ELF6 is the main H3K27 demethylase regulating the floral repressor FLC (Crevillén et al., 2014). Mutations in the ELF6 gene result in an early flowering phenotype, which is caused by increased expression of the floral integrators FLOWERING LOCUS T (FT) and SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1) (Crevillén et al., 2014). ELF6 also participates in the reprogramming and resetting of the epigenetic state of the FLC gene during gametogenesis, which is crucial for the correct timing of flowering (Crevillén et al., 2014). However, the precise role of ELF6 beyond FLC and flowering time regulation remains largely unexplored. Figure 1.6 illustrates the distinct roles of REF6, ELF6, and

JMJ13 in the different flowering pathways and their respective impacts on the regulation of flowering in *Arabidopsis*.

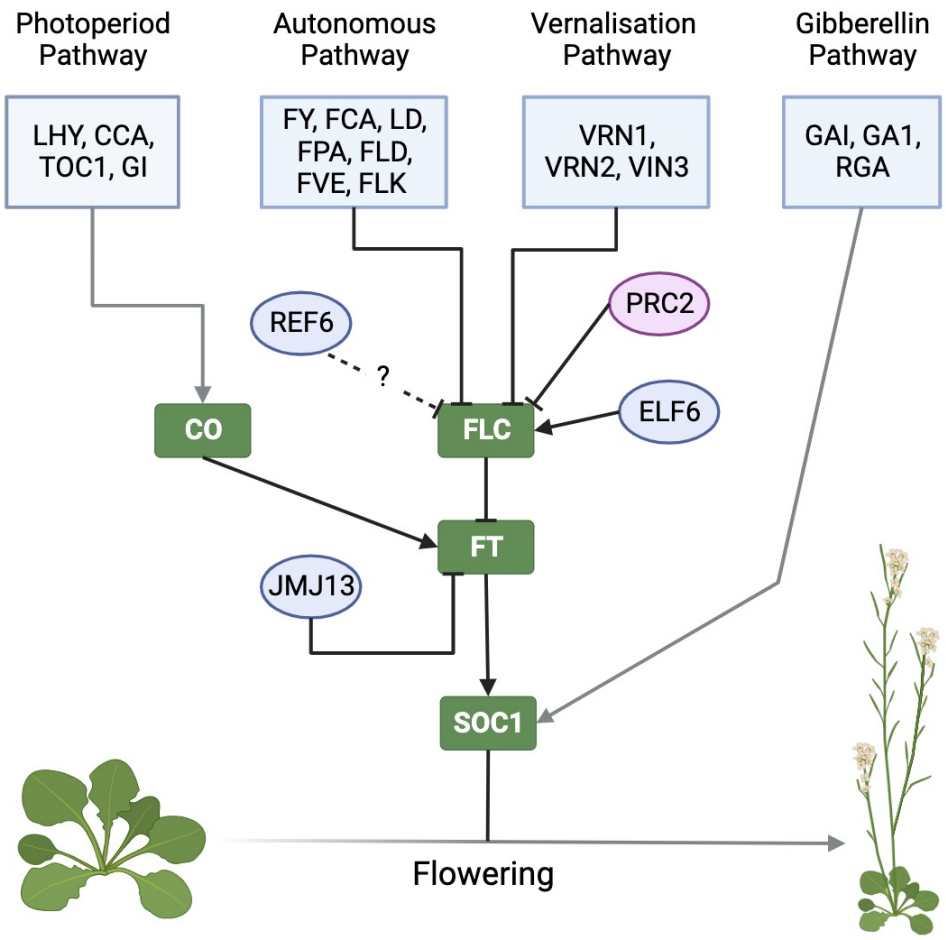


Figure 1.6: A simplified chart showing the flowering pathways in *Arabidopsis*, emphasising the functions of the H3K27me3 histone methyltransferase PRC2 (encircled in red) and the H3K27me3 histone demethylases REF6, ELF6, and JMJ13 (encircled in blue). Arrows indicate positive regulation and bars negative regulation.

1.4 Chromatin modifications in developmental programming and regeneration

Chromatin modifications are crucial for developmental programming and regeneration in plants, as they regulate DNA accessibility for transcriptional machinery and coordinate gene expression patterns. Chromatin-modifying factors guide epigenome reprogramming to alter gene chromatin states, particularly during phase transitions throughout the plant's life cycle. The seed-to-seedling transition is of particular importance for understanding in-vitro embryogenesis, a process used in plant biology to generate large numbers of identical plants through somatic embryogenesis (SE).

The transition from seed to seedling requires specific chromatin modifications to repress embryonic gene expression, primarily involving histone methylation and acetylation. Loss-of-function mutants in chromatin modification proteins controlling this phase transition undergo spontaneous SE due to deregulation of subsets of SE-inducing genes like WIND3 and LEC2 (Ikeuchi et al., 2015). H3K27 trimethylation and histone deacetylation are particularly crucial for the transcriptional repression of embryo genes (Figure 1.7).

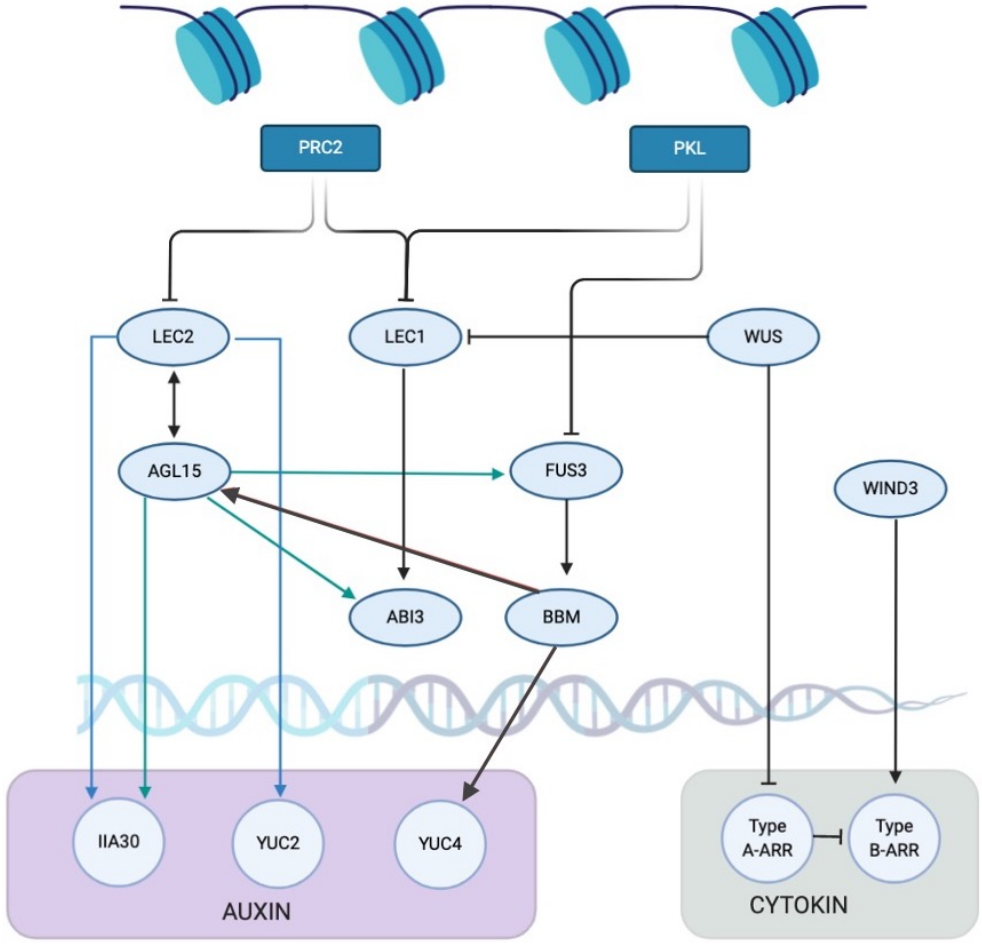


Figure 1.7: Schematic overview of the molecular regulation of Arabidopsis somatic embryogenesis. Chromatin-modifying proteins (blue rectangles at the top) repress or restrict expression of TFs (light blue ovals beneath) during Arabidopsis development. These TFs regulate each other's expression, as well as expression of common target genes involved in the auxin and cytokinin pathways (lower purple and grey boxes respectively).

PRC2 plays a role in repressing pluripotency and stem cell identity-associated genes, contributing to differentiation and preventing unregulated growth (He et al., 2012; Chanvivattana et al., 2004; Bouyer et al., 2011). PRC1 also participates in the process by working synergistically with PRC2 to repress embryonic gene expression during germination through histone H2A ubiquitination (H2Aub) (Chen et al., 2010). VAL proteins, transcription factors capable of binding chromatin, are necessary for PRC1-mediated H2Aub deposition at seed maturation genes (Yang et al., 2013). PRC1 is likely recruited to its target genes by VAL proteins, as the VAL-binding motif is enriched in promoters of PRC1 target genes (Merini et al., 2017). After the initial silencing by PRC1/VAL, repression is maintained by PRC2-mediated H3K27me3 deposition.

The role of histone deacetylation in embryogenesis has been investigated, with *hda6/hda19* HDAC mutants showing upregulated embryonic markers and somatic embryo formation on leaves (Tanaka et al., 2008). HDA19 specifically interacts with VAL2 (VIVIPAROUS1/ABI3-LIKE2) (Zhou et al., 2013), while HDA6 interacts with VAL1 (Chhun et al., 2016), and both VAL1 and VAL2 interact with the repressive CDK8 (CYCLIN-DEPENDENT KINASE E1) module of the Mediator complex and the chromatin remodeler PICKLE (PKL) (Liang et al., 2022). VAL proteins seem to recruit both Polycomb group proteins and HDACs to achieve repression of SE-related genes.

Histone methylation (H3K27me3) and histone acetylation play important roles in SE by regulating the expression of embryonic and meristematic genes. The extent of chromatin reprogramming required before SE likely depends on the explant's nature. Younger plant tissues, such as immature zygotic embryos or young leaves, have a higher proportion of undifferentiated or less differentiated cells, which exhibit more open chromatin and a greater potential for transcriptional reprogramming. This permissive chromatin state in younger plant cells enables them to respond more effectively to developmental cues and hormone treatments, resulting in a higher rate of somatic embryogenesis (Gaj, 2004).

1.5 Rationale and Aims for Research

In the realm of plant biology, while a great deal of research has delved into the nuances of proteins like REF6 and ELF6, significant knowledge gaps remain. A deep understanding of the structural and functional differences between these two proteins is still elusive. Moreover, the full extent of the enzymatic activities of these proteins and their implications for chromatin structure is yet to be completely elucidated. Importantly, while we acknowledge the significance of chromatin marks in embryogenesis, the development of a reliable epigenetic editing approach to optimise somatic embryogenesis remains a challenge.

With these gaps in mind, the overarching aim of this thesis is to dissect the functional differences between REF6 and ELF6, further validate their enzymatic activities, and explore the potential of epigenetic editing in improving somatic embryogenesis.

To address this aim, the following objectives have been established:

1. Understanding Structural and Evolutionary Differences: The first chapter will delve into the undiscovered differences between REF6 and ELF6 phenotypes. By conducting a comprehensive analysis of the structural domains of these proteins and investigating the reasons for their evolutionary divergence, we aim to illuminate their functional distinctions. This will be supported by a phylogenetic analysis of the proteins and an in-depth exploration of the functional domain disparities.

2. Validating Enzymatic Activities: In the second chapter, while previous studies have assessed the enzymatic activities of REF6 and ELF6, our research seeks to provide additional confirmation. We will evaluate the catalytic activities of these proteins through chromatin data in mutants and subsequently assess the expression levels in plants where the protein's catalytic domains have been swapped.

3. Epigenetic Editing for Enhanced Somatic Embryogenesis: The third chapter recognises the existing understanding of the impact of chromatin marks on embryogenesis. We will pioneer an approach to test the effect of directing histone demethylases to SE-related transcription factors with the goal of inducing cell dedifferentiation. Through meticulous modification

of the chromatin structure, we aspire to either activate or repress specific genes, potentially enhancing the success rate of somatic embryogenesis.

By undertaking these objectives, this thesis strives to contribute valuable insights and methodologies to the field, potentially paving the way for advanced applications in plant biology.

2 Methods

2.1 Cloning reactions

Cloning procedures were performed following the manufacturer's guidelines (Invitrogen), with 5 µl of the reaction mixture used to transform ccdB survival E. coli competent cells (TOP 10). A volume of 100-200 µl of the transformed culture was then spread onto appropriate selection media plates and incubated at 37°C overnight.

2.1.1 Golden Gate cloning

Golden Gate cloning was performed using a standard protocol (Engler et al., 2008) with slight modifications. The sequences of the two proteins AtELF6 and AtREF6 were obtained. A breaking point was selected for both proteins to separate the Jmj domain and the Znf domain at amino acid 491 for ELF6 and at amino acid 433 for REF6. Four DNA sequences were obtained (ELF6 Jmj, REF6 Jmj, ELF6 Znf and REF6 Znf). Those sequences were codon-optimised for *Arabidopsis thaliana*, and Golden Gate cloning sites were added at their ends. The four parts were synthetised by Integrated DNA Technologies. The four DNA sequences were inserted in the universal Level 0 Golden Gate vector (pAGM9121) containing the appropriate type IIS restriction enzyme recognition sites for BpiI (Figure 2.1). All the sequences were confirmed via Sanger sequencing. Level 1 Golden Gate vectors were assembled using type IIS restriction enzyme BsaI. Every Level 1 vectors consists of six Level 0, a pACTIN2 promoter (pICH87644), a Jmj domain, a Znf domain, a GFP-tag (pICSL50008) and a Terminator (pICH44300). Four assemblies were created: pACT2::ELF6 Jmj::REF6 ZnF::GFP (AER), pACT2::REF6 Jmj::ELF6 ZnF::GFP (ARE), pACT2::ELF6 Jmj::ELF6 ZnF::GFP (AEE) and pACT2::REF6 Jmj::REF6 ZnF::GFP (ARR) (Figure 2.2). All the Level 1 sequences were confirmed via Sanger sequencing. The four Level 1s were transformed in the destination vector Level 2 (pICSL4723) along with the red seed selectable marker cassette FAST-R (pICSL70008) and a linker (pICH41744). The Level 2 plasmid was suitable for Agrobacterium

transformation and these reactions employed the type IIS restriction enzyme BpiI. All the Level 2 sequences were confirmed via Sanger sequencing.

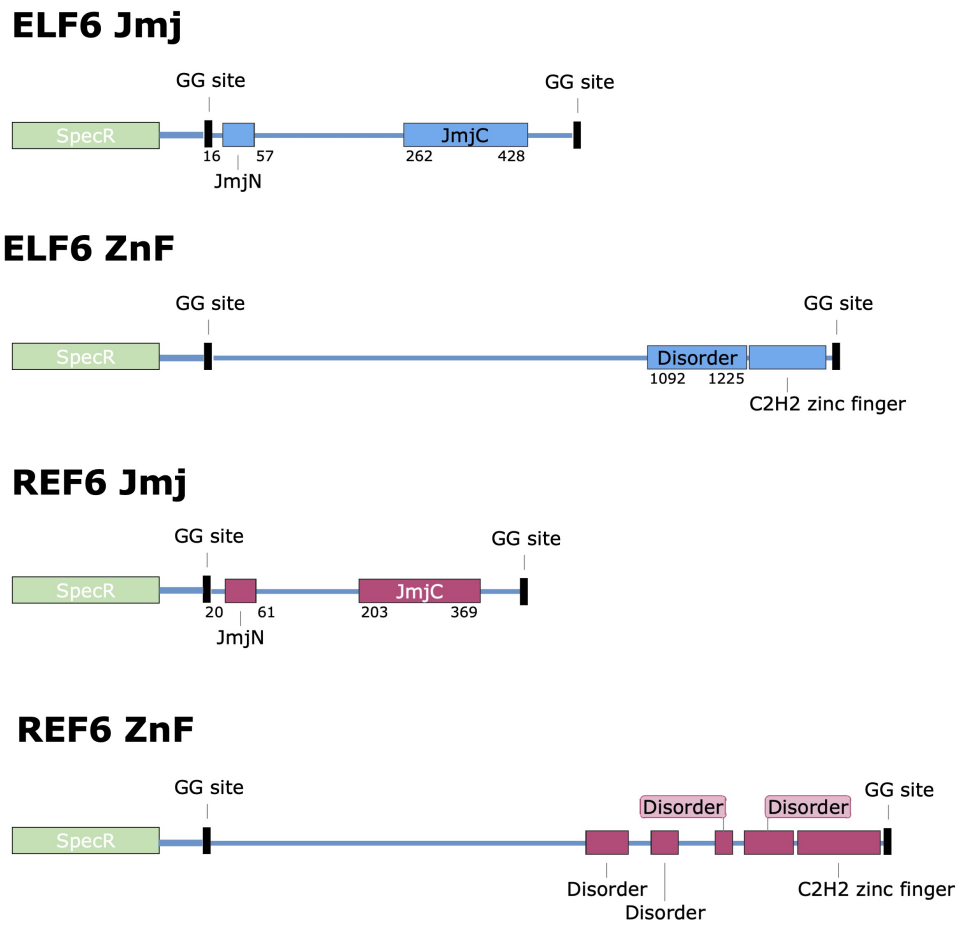


Figure 2.1: Schematic representation of Level 0 plasmids for ELF6 Jmj, REF6 Jmj, ELF6 Znf and REF6 Znf.

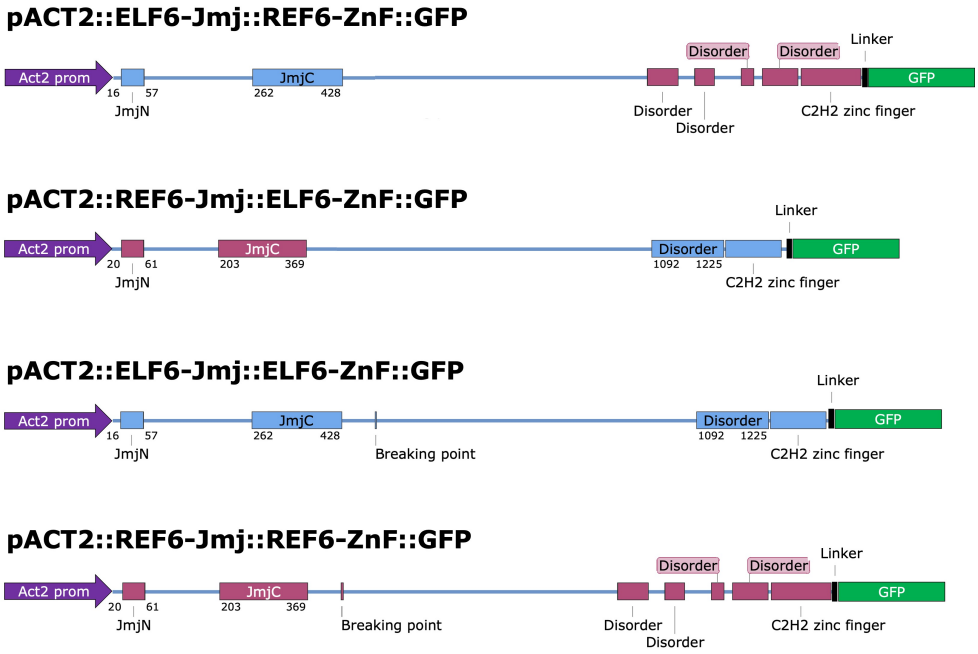
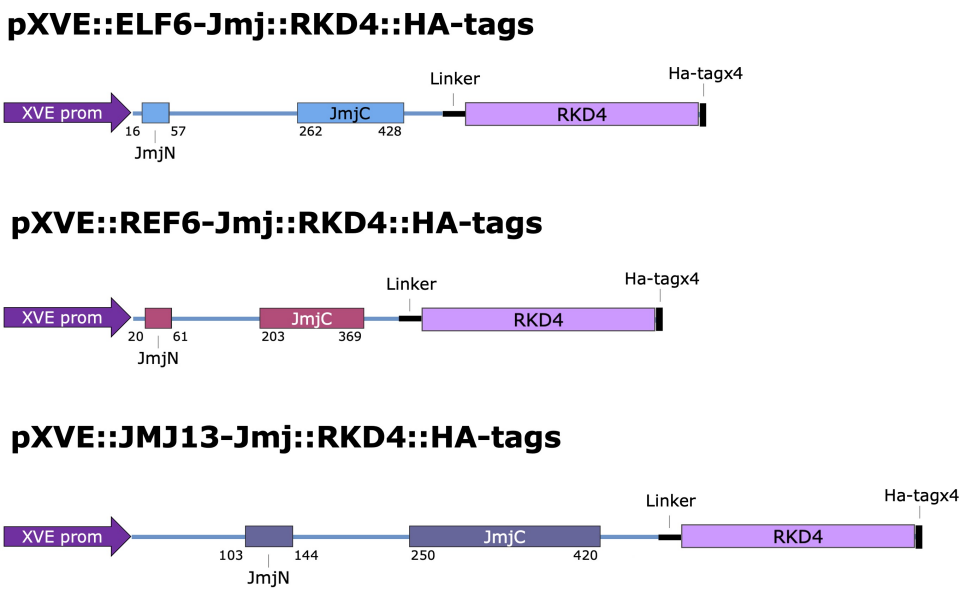


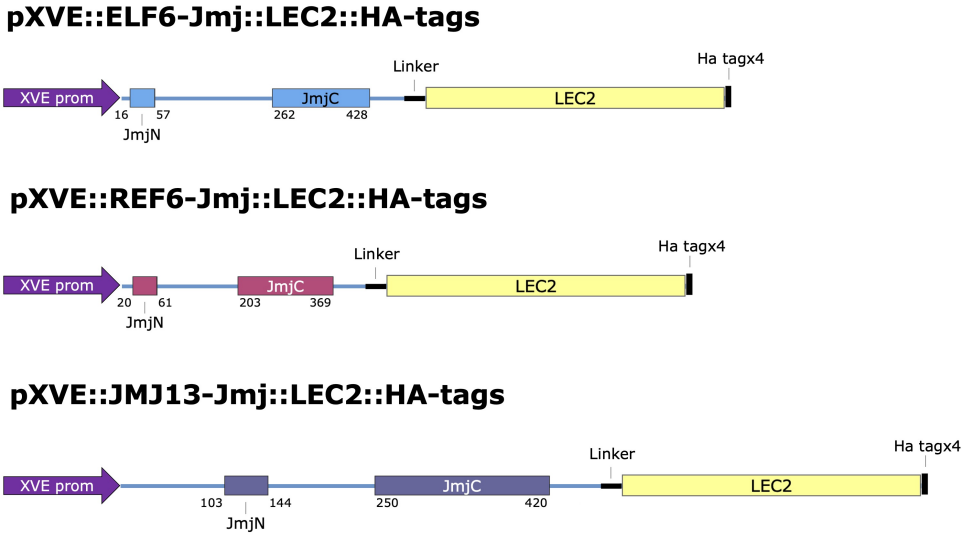
Figure 2.2: Schematic representation of Level 1 plasmids for pACT2::ELF6 Jmj::REF6 ZnF::GFP (AER), pACT2::REF6 Jmj::ELF6 ZnF::GFP (ARE), pACT2::ELF6 Jmj::ELF6 ZnF::GFP (AEE) and pACT2::REF6 Jmj::REF6 ZnF::GFP (ARR).

2.1.2 Gateway Cloning

Gateway cloning was performed using a standard protocol (Thermo Fisher Scientific) with slight modifications. Jmj domains of ELF6 and REF6 from the previous experiments were adapted adding Gateway cloning sites at their end. The sequence of AtJMJ13 was also obtained. A breaking point was selected to separate the Jmj domain from the rest of the protein at amino acid 475. JMJ13 Jmj domain was codon-optimised for *Arabidopsis thaliana*, and Gateway cloning sites were added at their ends. The sequences of two transcription factors AtRKD4 and AtLEC2 were obtained. Also in this case those sequences were codon-optimised for *Arabidopsis thaliana*. A linker was added at the N-terminal of the transcription factors and four HA-tags at the C-terminal. After that, Gateway cloning sites were added at their ends. A PCR-fusion method was used to fuse the two domains. The sequences were cloned in a pGEMT-Easy entry vector with a BP reaction. All the sequences were confirmed via Sanger sequencing. The six plasmid combinations obtained were then transformed in the destination vector pMDC7-OLE1 by LR cloning. Six constructs were obtained, each featuring the pXVE promoter, an estradiol-inducible promoter: (pXVE::ELF6 Jmj::RKD4::HA-tags(ELF6-RKD4), pXVE::REF6 Jmj::RKD4::HA-tags (REF6-RKD4), pXVE::REF6 Jmj::RKD4::HA-tags (JMJ13-RKD4), pXVE::ELF6 Jmj::LEC2::HA-tags (ELF6-LEC2), pXVE::REF6 Jmj::LEC2::HA-tags (REF6-LEC2) and pXVE::REF6 Jmj::LEC2::HA-tags (JMJ13-LEC2)(Figure 2.3).



(a) RKD4 constructs.



(b) LEC2 constructs.

Figure 2.3: Schematic representation of Jmj-TF constructs. (a) shows RKD4 constructs (pXVE::ELF6-Jmj::RKD4::HA-tags; pXVE::REF6-Jmj::RKD4::HA-tags and pXVE::JMJ13-Jmj::RKD4::HA-tags) and (b) shows LEC2 constructs (pXVE::ELF6-Jmj::LEC2::HA-tags; pXVE::REF6-Jmj::LEC2::HA-tags and pXVE::JMJ13-Jmj::LEC2::HA-tags).

2.2 Primers used

The primers used are shown in Figures 2.4, 2.5, and 2.6.

2.3 Plasmid DNA Extraction

Individual bacterial colonies were selected and inoculated into 5ml of LB with appropriate antibiotic selection, followed by overnight incubation at 37°C. Bacterial cells were then centrifuged at 5,000 g for 5 minutes, and plasmid DNA was extracted using the Qiaprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Extracted plasmid DNA was stored at -20°C for future use.

2.4 Escherichia coli Transformation

Escherichia coli TOP10 cells were transformed using heat shock at 42°C for 45 seconds, then plated onto agar medium containing the appropriate selection antibiotic and incubated overnight at 37°C. Colonies growing on the selection medium were screened using colony PCR and sequencing.

2.5 Colony PCR

For colony PCR, multiple transformants from each construct were selected and re-suspended in 20 µl of sterile water. 2 µl of the suspension was then used as a DNA template in a 25 µl PCR mixture. The PCR reaction mixture included 0.15 units of Kapa Taq (KAPA Biosystems), 2.5 µl of 10X PCR buffer, 0.3 µl of 10 mM dNTP mixture, and 1 µl of 5 µM of each primer. The PCR program began with an initial denaturation step at 95°C for 2 minutes, followed by 30 cycles of 95°C for 30 seconds, 58°C for 30 seconds, and 72°C with a variable extension time based on the target sequence length (1 kb per minute). A final extension step was performed at 72°C for 5 minutes. The amplified fragments were analysed using gel electrophoresis.

2.6 DNA Sanger Sequencing

To verify the plasmid sequence, Sanger sequencing LIGHTRUN (GATC Biotech) was performed according to the supplier's instructions. A total of 400 ng of DNA and 1 µM plasmid in a 10 µl volume were sent to the supplier for analysis.

2.7 Plant Material and Growth Conditions

All plant materials utilised in this study were derived from *Arabidopsis thaliana* (Col-0 accession). The *ref6-5* mutant (GABI 705E03) was obtained from the GABI-Kat collection (Kleinboelting et al., 2012). The *elf6-C* mutant was generated using CRISPR-Cas9 directed mutagenesis (Durr et al., 2018). The double mutant *ref6-5 / elf6-C* was created by genetically crossing *ref6-5* and *elf6-C*. The *ref6-5*, *elf6-C*, and *ref6-5 / elf6-C* mutants used in this study contained an RKD4 inducible system (Waki et al., 2011). Seeds for plate-grown experiments were surface sterilised using 10% Sodium hypochlorite (VWR) with shaking for 10 minutes, followed by six washes with sterile H2O. Seeds were then dispersed in sterile 0.1% Agarose and placed on base media consisting of 1 x Murashige and Skoog (MS) salts (Duchefa Biochemie), pH 5.7, and 1% sucrose (Sigma-Aldrich), solidified with 0.8% phytoagar (Duchefa Biochemie). For soil-grown experiments, seeds were dispersed in 0.1% Agarose before being sown individually on the surface of a John Innes and Perlite soil mix. *Arabidopsis* seeds were stratified for two days at 4°C in all experiments. Seeds were germinated and grown in either a light cabinet (Sigma) or a growth chamber (Conviron), with long day (16 h day, 8 h night) or short day (10 h day, 14 h night) conditions maintained, depending on the experiment. Temperature was held constant at 22°C and light intensity was set at 100 µmol/sec/m2.

2.8 Agrobacterium tumefaciens Transformation

Agrobacterium cells were transformed with the binary vector carrying the DNA fragment of interest via electroporation. The procedure was performed in a 0.1 cm cuvette at 2.2 kV using a Bio-Rad micropulser. A

40 µl aliquot of electro-competent Agrobacterium tumefaciens GV3131 cells was gently mixed with the binary vector. Following electroporation, transformed competent cells were transferred to low-salt liquid LB medium and incubated for 1 hour at 28°C. Cells were then spread on LB agar plates containing suitable antibiotics and incubated at 28°C for 48 hours.

2.9 *Arabidopsis thaliana* Stable Plant Transformation

A single colony of transformed Agrobacterium tumefaciens strain carrying each DNA construct in a binary vector was inoculated into 5 ml of low-salt LB containing appropriate antibiotics. The overnight culture was used to inoculate 500 ml low-salt liquid culture containing appropriate antibiotics and grown overnight at 28°C. *Arabidopsis* plants were transformed using the floral dip method as previously described (Zhang et al., 2006). T0 seeds (RFP positive) were selected and grown to the T1 generation. Subsequent lines exhibiting a 3:1 ratio of RFP+ to non-RFP seeds were chosen as potential single-insertion lines and grown to the T2 generation. In the T2 progeny, homozygous lines (lines with 100% RFP+ seeds) were identified, analysed, and selected as stably expressing lines for further study.

2.10 Confocal microscopy imaging

Images were obtained using a Zeiss Laser Scanning Microscope (LSM) 710 (Carl Zeiss Ltd; Cambridge, UK), and processed (including the addition of scale bars) using the Zeiss 2011 software (Zeiss). The 10x or the 20x objective lenses were used.

2.11 RNA extraction

Samples were collected, flash-frozen in liquid nitrogen, and pulverised using a pestle and a mortar. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) according to manufacturer instructions. The extracted RNA was stored at -80°C until further use. RNA concentration and purity were estimated using a NanoDrop spectrophotometer.

2.12 Reverse Transcription (cDNA Synthesis)

Extracted RNA was treated with TURBO DNA-freeTM (Promega, Madison, WI). For each sample, 1 µg of total RNA was reverse transcribed into complementary DNA (cDNA) using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific). The reverse transcription reaction was carried out in a 20 µL volume, following the manufacturer's protocol. The resulting cDNA was diluted 1:10 with nuclease-free water in order to reduce negative effects on the PCR from possible inhibitors carried over from the initial RNA isolation step.

2.13 Primer Design and Validation

Gene-specific primers were designed using Primer-BLAST (NCBI). The primers were synthesised by a commercial supplier (Integrated DNA Technologies) and validated by performing standard PCR on the synthesised cDNA. The amplified products were resolved on a 1% agarose gel and visualised using GelRed nucleic acid stain to confirm the expected amplicon size.

2.14 RT-PCR Analysis

All RT-qPCR analyses were performed using a MyiQ System (BIO-RAD) with the MESA Blue qPCR MasterMix Plus reagent (Eurogentec Headquarters). PCR fragments were analysed using a dissociation protocol to ensure that each amplicon was a single product. All RT-qPCRs were performed using two biological replicates in a final volume of 25 µl containing 5µl of cDNA template (diluted beforehand 1:10), 0.2 µM of each primer, and 12.5 µl of 35 µl 2xMESA Blue qPCR MasterMix (Eurogentec Headquarters). Each reaction was run in duplicate (technical replicates). Negative controls included in each run were a reaction conducted in the absence of reverse transcriptase and a reaction with no template (2 µl of nuclease-free water instead of 2 µl of cDNA). Analysis of expression data was performed according to the $\delta\delta$CT method (Livak and Schmittgen, 2001) and normalised using AtGADPH as a reference gene (AT3G26650).

2.15 RNA sequencing analysis

RNA-Seq libraries were prepared by the sequencing company (Novogene; Cambridge, UK) and the RNA samples were sequenced. Raw sequencing reads were subjected to quality control using FastQC (Babraham Bioinformatics). Low-quality reads and adapter sequences were trimmed using Trimmomatic. The high-quality trimmed reads were mapped to the *Arabidopsis thaliana* TAIR10 reference genome using STAR. Gene expression quantification was performed using htseq-count (from the HTSeq package) by counting the number of reads mapped to each gene. The raw read counts were then normalised using methods Reads Per Kilobase of transcript per Million mapped reads (RPKM) to account for differences in gene length and sequencing depth. Differential gene expression analysis was carried out using the DESeq2 R package. Raw read counts were input into the chosen package, and normalisation factors were calculated to account for library size differences. The differential expression analysis was then performed using a negative binomial generalised linear model. Differentially expressed genes (DEGs) were identified based on the FDR-adjusted p-value of < 0.05 and an absolute log2FC of > 1. Principal Component Analysis (PCA) was conducted on the variance-stabilised transformed data using the 'prcomp' function. DEGs were categorised as upregulated or downregulated based on the direction of the log2FC. Heatmaps, volcano plots, and other visualisations were created using R packages ggplot2 and pheatmap.

2.16 ß-estradiol induction

Seedlings of transgenic lines harbouring the inducible XVE cassette, along with controls were grown in 1/2 MS media and subjected to ß-estradiol induction for seven days. The concentrations used for the experiments were 5µm and 30µm.

2.17 Root measurements and statistical analysis

Approximately 20 seedlings were placed on one plate (10 for the experiment and 10 for the control). Digital images were obtained from each plate and

a ruler was included in each scan for calibration purposes. The digital images were imported into ImageJ software (National Institutes of Health) for root length measurements. The images were first calibrated using the ruler included in the scans. The root systems were then traced using the segmented line tool. Root length measurements were recorded for each plant, and the mean root length was calculated for each experimental group. The normality of the data was assessed by visual inspection of histograms. For statistical analysis, a T-test was employed with a significance threshold set at a p-value < 0.05

2.18 Protoplast isolation

Protoplast isolation was carried out according to the protocol published by Yoo et al (Yoo et al., 2007). *Arabidopsis thaliana* col-0 plants were grown for four weeks at 20-22°C under short day conditions (12-hour light). Well-expanded leaves were picked and sliced into 1mm wide strips. The finely cut leaf strips were transferred into 5ml of enzyme solution (20mM MES pH 5.7, 0.4M mannitol, 20mM KCl, 1.5% (w/v) cellulose R10, 0.4% (w/v) macerozyme R10, 10mM CaCl2 and 0.1% BSA) in a petri dish, with vacuum infiltration of leaf strips twice for 5 minutes. The leaf material was left at 25°C for 2.5 hours. The tissue was then filtered through a 70 μm nylon cell strainer and rinsed with the enzyme/protoplast solution with an equal volume of W5 solution (2mM MES pH 5.7, 154 mM NaCl, 125mM CaCl2, 5mM KCl). The flow-through was centrifuged at 100g for 1-2 minutes, and the pellet was washed twice with W5 solution. The cells were resuspended in 5 ml of MMG solution (4 mM MES pH 5.7, 0.4 M mannitol, and 15 mM MgCl2) and counted using a Fuchs-Rosenthal haemocytometer. The volume was adjusted by adding MMG at 400,000 protoplasts/ml, and cells were kept on ice.

2.19 Transfection of protoplasts

Approximately 80,000 cells were used for transfection. For 200 µl of leaf protoplasts, 8 µg of each plasmid was added. A total of 216 µl of freshly prepared PEG solution (40% v/w PEG4000, 0.2M mannitol, 100

mM CaCl2) was added to each reaction and mixed gently by flicking the bottom of the tube with fingertips. Protoplasts were then incubated at room temperature for 30 minutes for transfection. 500 µl of W5 was added to stop the transfection process. The samples were mixed gently and centrifuged at 100 g for 2 minutes. The supernatant was removed, and 500 µl of WI solution (4mM MES pH 5.7, 0.5 M mannitol, 20 mM KCl) was added. Transfected protoplasts were analysed for editing efficiency.

2.20 Flowering time phenotypic and statistical analysis

A total of 15 plants per line, along with 15 positive control plants (Col0) and 15 plants each for two negative controls (*ref6-5* and *elf6-C*), were grown in a randomised arrangement in a growth chamber (Conviron) under short day conditions. The plants were monitored for two phenotypic measurements: days after bolting and the number of leaves. To identify significant differences between the experimental lines and negative controls, statistical analyses were performed using the Tukey test. This post hoc test was conducted after a one-way analysis of variance (ANOVA) to compare multiple group means and account for multiple comparisons, so enabling the identification of lines with phenotypes significantly different from the negative controls.

2.21 3D Protein structure modelling and visualisation

Deep learning-based algorithm AlphaFold2 (v2.3.2 accessed on January 2023) from DeepMind was employed to generate the predicted models of the proteins of interest. The AlphaFold2-predicted protein models were then visualised and analysed using the ChimeraX software (UCSF Resource for Biocomputing) for evaluating protein folding and identifying key functional domains.

2.22 Protein sequence alignment and phylogenetic analysis

A total of 152 sequences were collected from various species representing distinct clades within the plant kingdom, including green algae, Charophytes, Bryophytes, Pteridophytes, Tracheophytes, Gymnosperms, and

Angiosperms. The dataset comprised 12 REF6 and ELF6 shared homologs from basal plants, 3 REF6 homologs from Gymnosperms, 3 ELF6 homologs from Gymnosperms, 65 REF6 homologs from Angiosperms, and 69 ELF6 homologs from Angiosperms. These sequences were aligned using the Clustal Omega multiple sequence alignment tool. The conservation of amino acids across the aligned sequences was visualised using the Geneious software. The phylogenetic tree was constructed using the Jukes-Cantor genetic distance model and the neighbor-joining tree building method was employed. Bootstrapping was performed with 100 replicates to assess the reliability of the branches in the tree.

2.23 Protein features prediction

Protein disorder regions and polar groups were identified in the proteins of interest using the MobiDataBase v5.0 (DSB, Padova) which combines data from experimental annotations, indirect disorder evidence from structural data and disorder predictions from protein sequences (using ESpritz and IUPred predictors). Molecular recognition features (MoRFs) were predicted by the MoRFPred tool, which uses machine learning model trained on known MoRFs to predict predict the likelihood of each region in the protein being a MoRF (Disfani et al., 2012). Phosphorilation sites were predicted by the deep-learning framework MusiteDeep which uses convolutional neural networks implemented on library Keras and Theano backend (Wang et al., 2020a). RCSB PDB strucmotif-search (Bittrich et al., 2020) and Foldseek (van Kempen et al., 2023) were used to search for the structure of a query protein domains against a database. Foldseek works by translating the 3D structure of proteins into a '3D-interaction alphabet' based on amino acid proximity and orientation. RCSB PDB StrucMotif-Search focuses on identifying local structural motifs in proteins, irrespective of their sequence or overall structural similarity. This tool retrieves occurrences of specific 3D motifs, which could be functionally significant, from a database of available structures.

2.24 Cell culture

Human embryonic kidney 293FT (HEK293FT) cells were cultured in complete growth medium consisting of Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin solution (PenStrep). Cells were maintained at 37°C in a humidified CO2 incubator with 5% CO2 and 95% air.

2.25 Production of lentivirus

Lentivirus production was achieved through the transfection of HEK293FT cells with the necessary plasmids to produce virus particles carrying the gene of interest. The required plasmids included envelope and packaging expression plasmids for the development of lentiviral particles and a backbone plasmid containing the sequence of interest. HEK293FT cells were plated to reach 60-80% confluence on the day of transfection. The pCMVR8.74 (packaging expression plasmid), pMD2.G (envelope plasmid), and the backbone plasmid were used at a 4:2:1 ratio (in µg). For TransIT®-Lenti (Mirus) transfection reagent use, a 3:1 ratio of TransIT®-Lenti (µl) to DNA (µg) was used. The TransIT®-Lenti-DNA complex was prepared using reduced serum medium (OPTI-MEM® I) and incubated at room temperature for 20 minutes. The mixture was then added to the HEK293FT cells, and lentivirus was harvested and filtered after 48 hours of incubation.

2.26 Transduction cells with Lentivirus

Target cells were plated 24 hours before infection to reach approximately 50% confluence on the day of transduction. The medium was removed, and 1 ml of lentiviral particle solution containing 8 µg polybrene (MERCK) was added to the cells. The cells were incubated at 37°C and 5% CO2 for six hours. After incubation, the medium was replaced with fresh medium without a selection antibiotic. The mCherry expression was monitored using an image reader cytation-3 (BioTek Instruments, Inc.) two days after transduction. After between three and four days, the medium was replaced with a selection antibiotic. Non-targeted cells were eliminated by

a selection antibiotic during passaging for two weeks after transduction. Transduced cells were re-fed with fresh media and antibiotics every two days. Cells could later be maintained without a selection antibiotic or at a reduced concentration.

2.27 Antibiotic kill curve test of cell lines and cell sorting

To determine the optimal dosage of the G418 antibiotic (MERCK) for cell line selection, a kill curve test was performed. Cells were plated in 6-well tissue plates with 70-80% confluence and subjected to increasing concentrations of G418 for two weeks. Fresh medium and antibiotics were replaced every two days. A concentration of 600 µg/ml of G418 was used for HEK293 cell selection. Cells were cultured in a T175 flask until 80% confluence before fluorescence-activated cell sorting (FACS). Cells were treated with 1 µg/ml doxycycline (DOX) at 37°C and 5% CO2 for 48 hours. For each inducible cell line, cells exhibiting high mCherry expression efficiency were specifically sorted into 6-well tissue plates, and single cells were sorted into 96-well tissue plates.

2.28 Fluorescence-Activated Cell Sorting (FACS)

Following 48 hours of transfection, cells were harvested by trypsinisation and resuspended in phosphate-buffered saline (PBS). Transfected cells were analysed using FACS based on the mCherry fluorescent protein expressed by the transfected plasmid to identify cells expressing the gene of interest. Positively transfected cells were collected and cultured in a new tissue culture dish with complete growth medium. RNA was extracted using the RNeasy kit (Qiagen).

Primer	Sequence (5'-3')	Used for
ELF6 GG FW (GG sites for promoter)	gaagacCAACTCAggagTGTCTCAGTGTTCCCAGCCC	Golden Gate sites creation
ELF6 GG RV (GG sites for promoter)	gaagacAACTCGcattCTTAAATTCCCAATTCCGTAAGACCC	Golden Gate sites creation
ELF6 JmJ 1201-1222 FWD 5	GCCGTCTAGTCCAAAACCCGGG	Golden Gate plasmids sequencing
ELF6 JmJ 1407-1428 FWD 7	GTAGACTGCGAGATCGACAGCG	Golden Gate plasmids sequencing
ELF6 JmJ 704-725 FWD 4	TCCCTCTGGGAAAAATGGAGAA	Golden Gate plasmids sequencing
ELF6 Jmj RV 178-200	ACCAAAAGCTGACGCCTCCTTCT	Golden Gate plasmids sequencing
ELF6 prom new mut FW	ttGAAGACaaGTGTTCAAGCTCCGCTCACTC	Golden Gate sites creation
ELF6 prom new mut RV	ttGAAGACaaACACCGGAGATATTCTCGTTTTCTGCC	Golden Gate sites creation
ELF6 promoter 899-920 FWD 1	CGTTTCTCCGTCGTTGCGCAAC	Golden Gate plasmids sequencing
ELF6 Znf 1195-1219 FWD 4	GGGGATAGTGGAAGAGTGATCGGCG	Golden Gate plasmids sequencing
ELF6 Znf 13-35 AGGT site	AGGTTCGCGACTCGTTATGTGGG	Golden Gate plasmids sequencing
ELF6 Znf 1693-1717 FWD 5	GCAGAGGCCGTATCAACTAGAGAGA	Golden Gate plasmids sequencing
ELF6 Znf 2030-2055 FWD 6	ACGACAACCCAGAATCATCAATTGGT	Golden Gate plasmids sequencing
ELF6 Znf 731-754 FWD 3	ACGAACACAGTGTAGATTGGACCA	Golden Gate plasmids sequencing
ELF6 Znf FW 2400-2422	TGAGTGTTCTTGGAAGGGATGCT	Golden Gate plasmids sequencing
ELF6 Znf RV 178-200	AGAGGCTGAGTTCTTCAAGGAGA	Golden Gate plasmids sequencing
ELF6_JmJ_FW	TTGAAGACcaactcaAATGGGGAATGTCGAGATTCCCAACTGG	Golden Gate sites creation
ELF6_JmJ_RV	TTGAAGACaactcgACCTGGTTCTCTTAATAGcgaACTTAAG	Golden Gate sites creation
ELF6_promoter_FW	GGCCTGTCTCAGTGTTCCCAG	Golden Gate plasmids sequencing
ELF6_promoter_mutagenesis_FW	ATATCTCCGGTCTTCAAGCTCC	Golden Gate plasmids sequencing
ELF6_promoter_mutagenesis_RV	GGAGCTTGAAGACCGGAGATAT	Golden Gate plasmids sequencing
ELF6_promoter_RV	GGATCTGCAAATTCAGTGTCTGTAG	Golden Gate plasmids sequencing
REF prom new mut FW	ttGAAGACaaAGTCCTAAGAAACGAGATCAAC	Golden Gate sites creation
REF6 GG FW (GG sites for promoter)	gaagacCAACTCAggagTCATTGTCCGGTTACACTCGAGC	Golden Gate sites creation
REF6 GG RV (GG sites for promoter)	gaagacAACTCGcattATCTCTCTCTCTCTCTCACACACACAGGG	Golden Gate sites creation
REF6 Jmj FW 1216-1237	GAAGCCGAGAAGCTCACGTCTA	Golden Gate plasmids sequencing
REF6 Jmj RV 168-189	GGCGGTAATGGCGGGAGGATTT	Golden Gate plasmids sequencing
REF6 prom new mut RV	ttGAAGACaaGACTCTCTGTTTTTATTGGATTAAAG	Golden Gate plasmids sequencing
REF6 promoter 788-812 FWD 2	TGTGTGTTAGAAATGGCAGAATCGT	Golden Gate plasmids sequencing
REF6 Znf 13-35 AGGT site	AGGTAAGGGTTCACCGGTGGCAT	Golden Gate plasmids sequencing
REF6 Znf 1321-1342 FWD 6	GCGTCCTCTCCTTCAAAGCCGA	Golden Gate plasmids sequencing
REF6 Znf 1860-1881 FWD 7	TACTGTTAGGGCGTGTGCAGCT	Golden Gate plasmids sequencing
REF6 Znf 2357-2378 FWD 1	TCGGGAAGAAACGATCAGGCAA	Golden Gate plasmids sequencing
REF6 Znf 818-843 FWD 5	ACGGGCGTACTAGCGACTTTAACTGT	Golden Gate plasmids sequencing
REF6 Znf FW 2569-2590	CCACAGCGACGATCGACCTCTT	Golden Gate plasmids sequencing
REF6 Znf RV 304-325	CGCAGCCGTAGCGCTTAGTGAT	Golden Gate plasmids sequencing
REF6_JmJ_FW	TTGAAGACcaactcaAATGGCTGTTTCTGAACAGTCCCAAGAC	Golden Gate plasmids sequencing
REF6_JmJ_RV	ttGAAGACaactcgACCTAGcgaACTTAACAAATAATGTTCTG	Golden Gate sites creation
REF6_promoter_FW	GGTTCTTGATTTGTTTCAGGTTTCC	Golden Gate sites creation
REF6_promoter_mutagenesis_FW	AAAAACAGAGAGACCTAAGAAACGAGATCAAC	Golden Gate plasmids sequencing
REF6_promoter_mutagenesis_RV	GTTGATCTCGTTTCTTAGGTCTCTCTGTTTTT	Golden Gate plasmids sequencing
REF6_promoter_RV	CACATCTTGACTCTGCTCTGAAAC	Golden Gate plasmids sequencing
L2-fw to L1P1	ACCTCTGACTTGAGCGTCGA	Golden Gate plasmids sequencing
TER-internal_fw	TAGCTCTGAGTGATCGAATTG	Golden Gate plasmids sequencing
FASTR-internal_rev	CTGCCGATGATATCGTGATG	Golden Gate plasmids sequencing
FASTR-internal_fw	CATCTACAACGTCAAGATCAGA	Golden Gate plasmids sequencing
L2-rev to linker	TATATCCTGTCAAACACTGATAG	Golden Gate plasmids sequencing
Level-0_UA Fwd	TTACGGTTCCTGCACTCTGTG	Golden Gate plasmids sequencing
Level-0_UA Rv	GCTTATGTCCACTGGGTTCGT	Golden Gate plasmids sequencing
LVL1 FW	CTGGTGGCAGGATATATTGTGGTG	Golden Gate plasmids sequencing
LV1 RV	GAACCCTGTGGTTGGCATGCACATAC	Golden Gate plasmids sequencing
REF6 prom 107-130 RV	CTGGTGGCAGGATATATTGTGGTG	Golden Gate plasmids sequencing
GFP FW	CACTCACGGCATGGACGAGCTG	Golden Gate plasmids sequencing
GFP RV 5586-5607	CACGCCGTAGCTGAAGGTGGTC	Golden Gate plasmids sequencing

Figure 2.4: Primers used for Golden Gate cloning.

Primer	Sequence (5'-3')	Used for
AT5G53750.1_RT_FW	CGCATCTGTCTCTGTTCTTCT	RT-qPCR of REF6 targets
AT5G53750.1_RT_RV	CAATCAGATTCTGTGCTCCTTTG	RT-qPCR of REF6 targets
AT1G64380_RT_FW	CGGATTCTTCGTGTTTGTCATC	RT-qPCR of REF6 targets
AT1G64380_RT_RV	CGAAACTGATACCGGAGAACTAT	RT-qPCR of REF6 targets
AT1G49160.3_RT_FW	GTGAGTGTAGAAACTCTGCTCAA	RT-qPCR of REF6 targets
AT1G49160.3_RT_RV	GCTGGCAATAAACATTTCTCGAT	RT-qPCR of REF6 targets
AT5G61640.2_RT_FW	GTTCTGGTCTAAGCATGATCCC	RT-qPCR of REF6 targets
AT5G61640.2_RT_RV	GTTGTAGAAATATATCCCTGAACGGTA	RT-qPCR of REF6 targets
AT3G51910.1_RT_FW	TGAGGATAATGAGTCAGCAAAGA	RT-qPCR of REF6 targets
AT3G51910.1_RT_RV	TCATCCAACTCTCTCTCTACCA	RT-qPCR of REF6 targets
AT5G52310_RT_FW	CCGATAACGTTGGAGGAAGAG	RT-qPCR of REF6 targets
AT5G52310_RT_RV	TCAAATTGTCCTGGCTTCTGA	RT-qPCR of REF6 targets
GFP1FW	GAAGCGCGATCACATGGTC	RT-qPCR GFP detection
GFP1RV	CTTGTAGTTGCCGTCGTCCT	RT-qPCR GFP detection
GFP2FW	GACCACTACCAGCAGAACAC	RT-qPCR GFP detection
GFP2RV	TGCTCAGGTAGTGGTTGTCG	RT-qPCR GFP detection
GFP3FW	AGAACGGCATCAAGGTGAACT	RT-qPCR GFP detection
GFP3RV	GAACTCCAGCAGGACCATGTG	RT-qPCR GFP detection
LEC2_1FW	GGACTCTAAGCGACAGCTAATG	RT-PCR LEC2 fusion proteins
LEC2_1RV	TCGCCTGGTGATGTTGTAAG	RT-PCR LEC2 fusion proteins
LEC2_2FW	CAGGTAGTGAGTTCGGTTCTTT	RT-PCR LEC2 fusion proteins
LEC2_2RV	GGGTAGTGGTCGTTTAGGTTAC	RT-PCR LEC2 fusion proteins
LEC2_3FW	ACTTCCTTCCTTTCCCATCTTC	RT-PCR LEC2 fusion proteins
LEC2_3RV	CACAGAGAACCCTCAACTTCTT	RT-PCR LEC2 fusion proteins

Figure 2.5: Primers used for RT-PCR reactions.

Primer	Sequence (5'-3')	Used for
JMJ13-linker 1188-1209 FWD 8	CCGGCCAAAGGTCCATAAAGAC	GateWay plasmids sequencing
JMJ13-linker 1286-1306 REV 8	GAATTCCCCTGAACCAGCCGC	GateWay plasmids sequencing
JMJ13-linker 18-38 FWD 5	AAAAAGCAGGCTgaacaATGG	GateWay plasmids sequencing
JMJ13-linker 215-236 FWD 6	AAGATCGTATCCCCACTGACCG	GateWay plasmids sequencing
JMJ13-linker 589-610 REV 6	CGTTGACTTTGGCAGACGAGAC	GateWay plasmids sequencing
JMJ13-linker 687-709 FWD 7	CGTGGCATGTAGAAGACCACTAC	GateWay plasmids sequencing
JMJ13-linker 87-108 REV 5	TCGGTCCATTTTAGGTCGTCCG	GateWay plasmids sequencing
linker-BBM 102-123 REV 2	ACGGGTAGTACTTGAATCCACG	GateWay plasmids sequencing
linker-BBM 1098-1119 FWD 8	TGCGGCTATTAAGTTCAGGGGA	GateWay plasmids sequencing
linker-BBM 1106-1127 REV 11	GCGGAGAGTCCCCTGAACTTAA	GateWay plasmids sequencing
linker-BBM 1274-1294 FWD 12	CTAATAACGTGAGCGGGTGGC	GateWay plasmids sequencing
linker-BBM 129-150 REV 2	CTGATGATGATCGTAGGTAGGG	GateWay plasmids sequencing
linker-BBM 1494-1515 REV 8	AACGGTTACTGAGTCGTCGGAC	GateWay plasmids sequencing
linker-BBM 1609-1630 FWD 9	GCCTACAATGCGAGGAACCACT	GateWay plasmids sequencing
linker-BBM 1641-1662 REV 12	CTGTTGTTGCTGTTGATGCTGC	GateWay plasmids sequencing
linker-BBM 1752-1773 FWD 13	GGGCGCACCAACATTTTCAGTA	GateWay plasmids sequencing
linker-BBM 1863-1884 REV 13	ACAAGAAAGCTGGGTCTATGCA	GateWay plasmids sequencing
linker-BBM 1863-1884 REV 9	ACAAGAAAGCTGGGTCTATGCA	GateWay plasmids sequencing
linker-BBM 269-290 FWD 10	GGGACATCAATGGTGGTGCTTG	GateWay plasmids sequencing
linker-BBM 492-513 REV 6	TGCCCCATTACCGTTATCCTGA	GateWay plasmids sequencing
linker-BBM 608-629 FWD 7	TAGTGGAGACTACACCCAAAAA	GateWay plasmids sequencing
linker-BBM 644-667 REV 10	GGTATATAGACGTTCGTTGCCCGA	GateWay plasmids sequencing
linker-BBM 754-775 FWD 11	GGCAGGCAAGTTTATTTGGGGG	GateWay plasmids sequencing
linker-BBM 93-114 FWD 6	TCGAACCGACGTGGATTCAAGT	GateWay plasmids sequencing
linker-BBM 981-1002 REV 7	CCCGTGCTGGTGATGTCTAGTT	GateWay plasmids sequencing
linker-LEC2 1200-1221 REV 6	ACAAGAAAGCTGGGTCTATGCA	GateWay plasmids sequencing
linker-LEC2 129-150 REV 2	CTGATGATGATCGTAGGTAGGG	GateWay plasmids sequencing
linker-LEC2 18-38 FWD 4	GGCTGGTTCAGGGGAATTCGA	GateWay plasmids sequencing
linker-LEC2 313-334 REV 4	CTAAAAACCCGCCACGCTCTTG	GateWay plasmids sequencing
linker-LEC2 427-448 FWD 5	TTGGTGGACTCTAAGCGACAGC	GateWay plasmids sequencing
linker-LEC2 775-796 REV 5	TGAGGAAGTCCCCTATTTCCGC	GateWay plasmids sequencing
linker-LEC2 937-958 FWD 6	GAAGAGGCGTCTATTGCGATGC	GateWay plasmids sequencing
linker-RKD4 116-137 REV 2	GGACTGTCCAAGTGGTGTAAGG	GateWay plasmids sequencing
linker-RKD4 238-263 FWD 4	GCAGGCCTTCATTTTCAGTTTCACTA	GateWay plasmids sequencing
linker-RKD4 645-666 REV 4	AACTTCCTCCTCCATCCCTACA	GateWay plasmids sequencing
linker-RKD4 771-792 FWD 5	ACGACGTAAAAGTCTCGGAGAC	GateWay plasmids sequencing
REF6-linker 1133-1154 REV 7 11	TAGCAGCGTCTTTTGCCATTCG	GateWay plasmids sequencing
REF6-linker 1258-1279 FWD 8 12	GAAGCCGAGAAGCTCACGTCTA	GateWay plasmids sequencing
REF6-linker 136-157 REV 5 9	TGGGTCCTGAAACTCTGCAAGT	GateWay plasmids sequencing
REF6-linker 1394-1414 REV 8 12	GAATTCCCCTGAACCAGCCGC	GateWay plasmids sequencing
REF6-linker 18-38 FWD 5 9	AAAAAGCAGGCTgaacaATGG	GateWay plasmids sequencing
REF6-linker 259-280 FWD 6 10	TCTCAATAGAAGTCTGGCGGCA	GateWay plasmids sequencing
REF6-linker 651-671 REV 6 10	CAGTTTCCCCGACTGTCCCAC	GateWay plasmids sequencing
REF6-linker 776-797 FWD 7 11	TTTAGTTGGTTCGCTTGGCACG	GateWay plasmids sequencing

Figure 2.6: Primers used for Gateway cloning.

3 Structural analysis of two Jmj-C zinc finger containing demethylases

3.1 Introduction

Flowering is a crucial developmental switch in plants that marks the transition from vegetative to reproductive growth. As flowering time has a significant impact on a plant's reproductive success, each species has evolved its own strategies to regulate the timing of flowering.

In *Arabidopsis thaliana*, various genetic pathways control flowering under specific growth conditions. Recent findings indicate that these pathways are interconnected within a complex, multilevel regulatory network (Quiroz et al., 2021). Earlier studies proposed that multiple flowering pathways converge on key transcription factors (TFs) such as FLOWERING LOCUS C (FLC), FLOWERING LOCUS T (FT), SUPPRESSOR OF OVER EXPRESSION OF CONSTANS (SOC1), and LEAFY (LFY), which function as integrators of flowering time. However, a more recent and intricate model suggests that a genetic network involving approximately 300 genes regulates the transition to flowering in *Arabidopsis* (Quiroz et al., 2021). Through the years, forward genetic analyses have been conducted on Arabidopsis mutants to identify components involved in bolting time, with these mutants exhibiting contrasting flowering times under different environments.

3.1.1 ELF6 and REF6 are Regulators of Flowering

An early-flowering mutant, *elf6-1*, was identified through this process (Noh and Amasino, 2003; Noh et al., 2004). The *elf6-1* mutant demonstrated an earlier floral transition compared to the wild-type under both long and short days. Similarity searches of *Arabidopsis* genome databases using ELF6 revealed a gene with high similarity to ELF6, subsequently named RELATIVE OF EARLY FLOWERING 6 (REF6). To address the function of REF6 the gene was mutated and all three *ref6* mutants showed a surprising recessive late-flowering phenotypes both in long and short day, an opposite phenotype to *elf6* mutant (Noh et al., 2004).

The initial documented H3K27me3 demethylase in plants was REF6 which is involved in multiple developmental pathways. Although the *ref6* mutant was initially identified as late flowering due to high FLC expression levels (Noh et al., 2004), the mechanism by which REF6 regulates floral transition is still unclear since it doesn't bind to the FLC locus (Cui et al., 2016b). The floral development is affected by REF6 in various ways, such as its interaction with Nuclear Factor Y proteins (NF-Y) which control SOC1 expression in inflorescences. This interaction further facilitates the recruitment of REF6 to regulate downstream targets, including FRUITFUL (FUL) and TARGET OF FLC AND SVP 1 (TFS1), with the aid of SOC1 (Hou et al., 2014; Hyun et al., 2016; Richter et al., 2019). During floral development, REF6 also activates KNUCKLES (KNU), a repressor of the stem cell pool regulator WUSCHEL (WUS) (Zhou et al., 2018). This evidence highlights the significance of REF6 in coordinating the removal of the H3K27me3 mark in silenced genes throughout the development of floral meristems. Consequently, plants overexpressing REF6 exhibit Polycomb mutant-like traits and display various developmental abnormalities, such as small plant size and wrinkled, curled leaves, due to increased expression of floral homeotic genes in seedlings (Lu et al., 2011a).

ELF6 has also been recognised as an H3K27me3 demethylase (Crevillén et al., 2014), although its functions appear to differ from those of REF6. In contrast to REF6, the ELF6 protein has a direct association with the FLC locus. In this instance, ELF6 (an H3K27me3 demethylase) interacts with the H3K36me3 methyltransferase EARLY FLOWERING IN SHORT DAYS, EFS (Yang et al., 2016). The connection between EFS and ELF6 links H3K36me3 activation to H3K27 demethylation at the FLC locus. EFS directly links to the Pol II transcriptional machinery, promoting transcriptional initiation and elongation. All of these factors counteract PRC2 activity, promoting a high expression state at FLC (Yang et al., 2016).

Further research has investigated ELF6's role in FLC regulation. Certain *Arabidopsis* ecotypes, such as those from Northern Europe, possess a functional FRIGIDA (FRI) allele, a powerful FLC expression activator.

These ecotypes typically exhibit a late-flowering phenotype, which is characterised by a prolonged vegetative phase and delayed onset of flowering under long-day photoperiods. Plants carrying this allele require an extended period of cold exposure (vernalisation), to epigenetically silence FLC expression (Costa and Dean, 2019). The silencing of FLC by H3K27me3 is reset between generations, ensuring that each new generation needs to undergo vernalisation anew. Research suggests that the silenced FLC gene is reactivated during male gametogenesis and also early in embryogenesis (Sheldon et al., 2008). Notably, hypomorphic mutation in the ELF6 protein hampers the reactivation of FLC in reproductive tissues, leading to the inheritance of a partially vernalised state, thus revealing that ELF6 is necessary for the epigenetic reprogramming of H3K27m3 at the FLC locus following vernalisation (Crevillén et al., 2014). Moreover, it has been suggested that ELF6 could potentially bind with the FT (FLOWERING LOCUS T) locus (Jeong et al., 2009). Although ELF6 has been associated with the FLC gene and potentially the FT locus, no direct targets for ELF6 have been conclusively identified. Consequently, the exact role of ELF6 beyond its involvement in FLC regulation and flowering time control is still mostly undiscovered.

3.1.2 REF6 and ELF6 Further Roles in Plant Development

Both REF6 and ELF6 have been found involved in brassinosteroid signalling. Brassinosteroids (BRs) are a class of plant hormones that are involved in various developmental processes, including flowering. BR signalling pathway regulates the expression of several flowering-related genes, which can influence the timing and regulation of flowering. Both REF6 and ELF6 proteins were found to interact with of BRI1-EMS-SUPPRESSOR 1 (BES1), a key downstream component of the BR signalling pathway (Yu et al., 2008). Additionally, research has demonstrated that REF6 directly controls the expression of several brassinosteroid (BR)-responsive genes, such as TOUCH 4, which participate in cell wall synthesis and modification, and CEL3, a transmembrane protein implicated in cuticle membrane and wax production (Cui et al., 2016b).

Within the BR signalling pathway, ELF6, in addition to its interaction with BES1, is also recruited by BZR1 (BRASSINAZOLE-RESISTANT 1), a homolog of the transcription factor BES1, to a specific FLC intronic regulatory region, which aids in the activation of the FLC locus (Li et al., 2018b).

In contrast to ELF6, whose role appears to be confined to flowering regulation, REF6 is involved in various other developmental processes. REF6 mutants showed significant downregulation of three major ethylene signalling genes EIN2, ORE1 and NAP. It was demonstrated that REF6 directly promotes the transcription of its target senescence-associated genes by reducing their H3K27me3 level (Wang et al., 2019b). The same authors discovered that REF6 promotes lateral root primordium initiation and lateral root emergence by directly binding to the chromatin of PIN1/3/7, key polar auxin transporter genes (Wang et al., 2019a). Additionally, it has been shown that REF6 is involved in thermomemory by establishing a heritable feedback loop heat-induced HEAT SHOCK TRANSCRIPTION FACTOR A2 (HSFA2)(Liu et al., 2019). REF6 also acts as a positive regulator of light-initiated seed germination as the loss of function of REF6 in *Arabidopsis* inhibits Phytocrome B-dependent seed germination (Wang et al., 2023).

3.1.3 Complex Interplay of REF6 and ELF6 in Regulating Gene Expression

Recent studies have shed light on the mechanisms employed by REF6 to enable the identification of all its target regulatory regions. The results of ChIP-seq analyses have demonstrated that REF6 protein exhibits direct binding to an extensive number of genomics regions in both seedlings (Cui et al., 2016b) and flowers (Yan et al., 2018). Biochemical investigations suggest that the REF6 Znf domain is capable of exhibiting sequence-specific binding to DNA. Analysis of the genomic DNA motif revealed that REF6 binds to the CTCTGYTY motif specifically (Cui et al., 2016b; Li et al., 2018a). This DNA-protein interaction has been characterised as a new half-cross braced Znf-type domain (Tian et al., 2020) Notably, DNA methylation hinders REF6 binding to DNA, providing an explanation for

the depletion of REF6 from heterochromatin regions within the genome (Tian et al., 2020).

Interestingly, the REF6 protein lacking the C-terminal C2H2-Znf domain continues to bind to over 7,000 sites when expressed in a *ref6 elf6 jmj13* triple mutant background (Yan et al., 2018). The binding sites of the truncated REF6 δZnf protein show a preferential co-localisation with genomic regions bound by several transcription factors, including MADS factors such as SEPALLATA3 (SEP3) (Yan et al., 2018).

Information about ELF6 targets is scarce since no immunoprecipitation assays have been conducted on its protein thus far. Additionally, the motif of its DNA binding zinc finger domain remains unknown, although ChIP-seq data from its mutant have uncovered indirect targets (internal data).

3.1.4 Redundant Functionalities in H3K27 Demethylases

As H3K27 demethylases, REF6 and ELF6 function to regulate gene expression by reactivating H3K27me3 Polycomb-repressed genes in response to both internal and external cues. Mutations in REF6 result in thousands of hypermethylated H3K27me3 regions throughout the genome, emphasising its significance in gene regulation (Antunez-Sanchez et al., 2020). The number of hypermethylated regions is reduced in epigenomic analyses of *elf6* mutants and even less pronounced in *jmj13* mutants (a Znf-containing JmjC H3K27 demethylase), suggesting some functional redundancy among these proteins (Zheng et al., 2019; Antunez-Sanchez et al., 2020). H3K27me3 genomic analyses of the triple mutant *elf6 ref6 jmj13* reveal that these proteins play a crucial role in preventing H3K27me3 spreading and defining Polycomb-silenced regions. The binding patterns of REF6 and Polycomb complex do not overlap, as REF6 binding sites are located at the boundaries of H3K27me3 regions, and Polycomb proteins cover H3K27me3-marked areas (Yan et al., 2018). Therefore, REF6 not only counteracts Polycomb silencing but also serves as a barrier that restricts H3K27me3 spreading and regulates gene expression.

While H3K27me demethylases REF6, ELF6 and JMJ13 each have distinct roles in various developmental processes and plant responses to biotic and abiotic stresses, there is indeed some functional overlap among them.

The *elf6 jmj13* double mutant in FRI background, accelerated flowering and reduced FLC expression compared with *elf6*. During vernalisation, *elf6 jmj13* showed lower relative FLC expression than *elf6* suggesting that ELF6 and JMJ13 may have partially redundant roles in flowering time control (Yang et al., 2016). Recent evidence suggests that REF6 and ELF6 may play independent and partially redundant functions in regulating genomic H3K27me3 profiles (Antunez-Sanchez et al., 2020). In fact, double *elf6 ref6* knockout mutants exhibit dwarf phenotype and pleiotropic leaf morphology defects that are absent in single mutant plants (Antunez-Sanchez et al., 2020).

The *elf6 ref6 jmj13* triple mutant exhibits a more pronounced pleiotropic phenotype, characterised by a dwarf appearance, altered leaf morphology, shortened internodes, and flowers with a partial loss of determinacy, resulting in severely bent siliques (Yan et al., 2018). When combined with *jmj30-2* and *jmj32-1*, two mutations affecting genes encoding more distantly related JMJ-type H3K27me3 demethylases, the quintuple mutant does not display further enhancement of the phenotype. This finding suggests that these two families of plant histone demethylases perform independent functions and do not have overlapping roles in the observed phenotypic changes (Yan et al., 2018).

This redundant function of histone demethylases may conceal unforeseen new functions. Recent research suggests, for instance, that reduced ELF6 and REF6 activity during sexual reproduction may result in the transgenerational inheritance of ectopic H3K27me3 marks (Antunez-Sanchez et al., 2020). Some of the H3K27me3 imprints observed in epimutants were initially established in *elf6-C/ref6-5* and were stably transmitted across five generations, even after the restoration of wild-type function for these histone demethylases (Antunez-Sanchez et al., 2020).

3.1.5 Experimental rationale

REF6 promotes late flowering and ELF6 early flowering, despite their protein structure similarities. These two proteins regulate flowering differently, raising intriguing questions about the molecular basis of their functional divergence and their contribution to other developmental processes in *Arabidopsis thaliana*.

To further elucidate the distinct roles of these two proteins, we will investigate their protein domains in detail. This will involve examining the functional and structural characteristics of various domains, such as the DNA-binding domain, the catalytic domain, and any regulatory or interaction domains.

In addition, phylogenetic analyses will help to trace the evolutionary divergence of these proteins.

3.2 Results

To investigate the differences between REF6 and ELF6, we began by analysing the evolutionary relationships between the two histone demethylases. Given the high structural similarity between these proteins, we aimed to conduct a comparative phylogenetic analysis to determine whether they originated from a common ancestor. By identifying their closest homologs across diverse plant species, we hoped to gain insights into the diversification of their functions over the course of evolution.

To this end, we collected sequences from 73 species representing various clades within the plant kingdom, including green algae, Charophytes, Bryophytes, Pteridophytes, Tracheophytes, Gymnosperms and Angiosperms (the list of genes is shown in the Appendix, figures 6.1, 6.2 and 6.3). This comprehensive dataset allowed us to perform a thorough phylogenetic analysis that encompassed the full breadth of plant diversity, providing a robust framework for investigating the evolutionary history of REF6 and ELF6 and their potential shared ancestry.

Our phylogenetic analysis revealed a distinct separation between seed plants and basal plants, providing evidence for the evolutionary divergence of these two major plant groups (Figure 3.1).

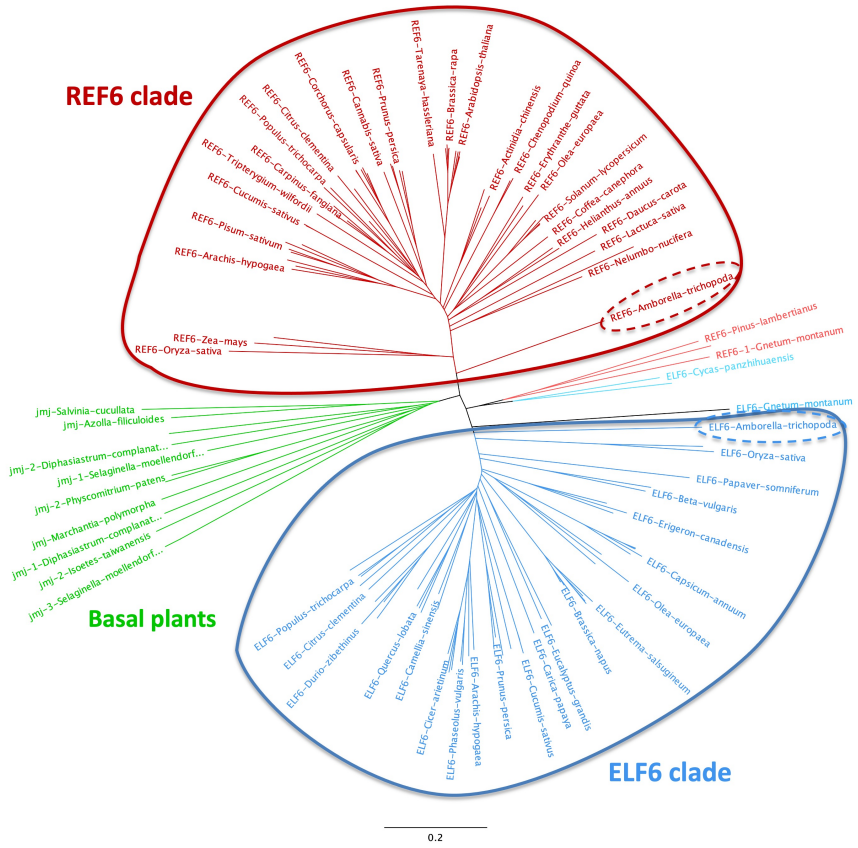


Figure 3.1: Phylogenetic tree illustrating the relationships among REF6 and ELF6 homologs across 73 plant species. Distinct lineages are represented by varying colours. The green branch encompasses green algae, Charophytes, Bryophytes, Pteridophytes, and Tracheophytes. Dark red and dark blue branches correspond to Angiosperm species associated with REF6 and ELF6, respectively, while light red and light blue branches represent Gymnosperm species linked to REF6 and ELF6, respectively.

This result is consistent with the well-established classification of plants, further validating our phylogenetic approach and data quality.

Interestingly, the phylogenetic tree also displayed a clear branching between REF6 and ELF6, indicating that these two histone demethylases have evolved independently and may have undergone functional divergence. The first species identified as possessing distinct REF6 and ELF6 genes was Amborella trichopoda, a member of the family Amborellaceae. This finding is particularly noteworthy, as Amborella trichopoda is considered to represent the earliest diverging branch among extant angiosperms. The presence of differentiated REF6 and ELF6 genes in this

basal angiosperm lineage suggests that the functional divergence of these histone demethylases may have occurred early in the evolution of flowering plants, potentially reflecting the diversification of epigenetic regulatory mechanisms associated with the emergence of angiosperms.

Intriguingly, our analysis indicates that the divergence between REF6 and ELF6 happened not only in angiosperms but also in gymnosperms—albeit via a separate evolutionary event. Specifically, in two gymnosperm species—Gnetus montanum and Cycas panzhihuanensis—we found that their REF6 and ELF6 genes also diverged, but through an evolutionary event distinct from that in angiosperms. This finding raises the possibility that REF6 and ELF6 may have diverged independently in both gymnosperms and angiosperms, hinting at either convergent evolution or parallel paths of functional diversification across these plant lineages. Further research is needed to clarify the evolutionary mechanisms that led to these separate divergences in the two plant groups.

To further understand the differences between REF6 and ELF6 we compared their conserved domains (Figure 3.2).

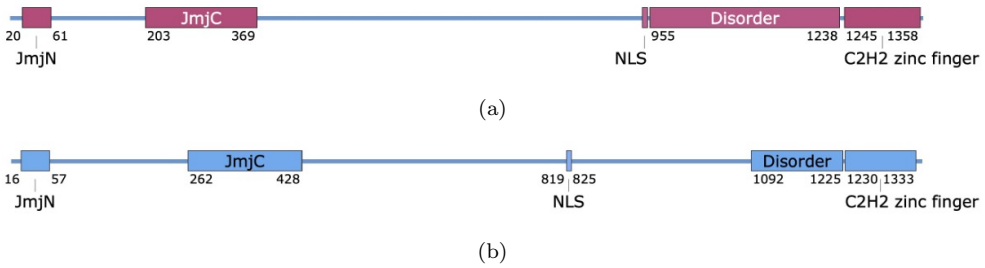


Figure 3.2: Diagram illustrating the domain annotations for (a)AtREF6 and (b)AtELF6 proteins. Both proteins possess an N-terminal Jmj domain, composed of JmjN and JmjC subdomains, and a C-terminal C2H2 DNA-binding zinc finger domain. The intrinsically disordered regions located upstream of the zinc fingers exhibit distinct characteristics in REF6 and ELF6.

Both REF6 and ELF6 proteins share a common architecture, with an N-terminal Jmj catalytic domain consisting of two subdomains, JmjN and JmjC. Additionally, both proteins possess a zinc-finger DNA-binding domain situated at the C-terminus. Despite these similarities, the intrinsically disordered regions (IDRs) between the structured domains display a lack of conservation between REF6 and ELF6. Notably, REF6 contains

a higher number of IDRs compared to ELF6. As nuclear proteins, both REF6 and ELF6 possess a nuclear localisation signal (NLS) situated in a similar location.

We started our comparative study by investigating the zinc-finger domains of REF6 and ELF6. Both of these proteins possess four analogous tandem C2H2 zinc-finger motifs that specialise in identifying and binding to distinct DNA targets. We gathered and aligned the sequences of the zinc-fingers from REF6 and ELF6 homologs in nine representative species, including liverwort, mosses, monocotyledons and dicotyledons, and compared them to AtREF6 and AtELF6 (Figure 3.3).

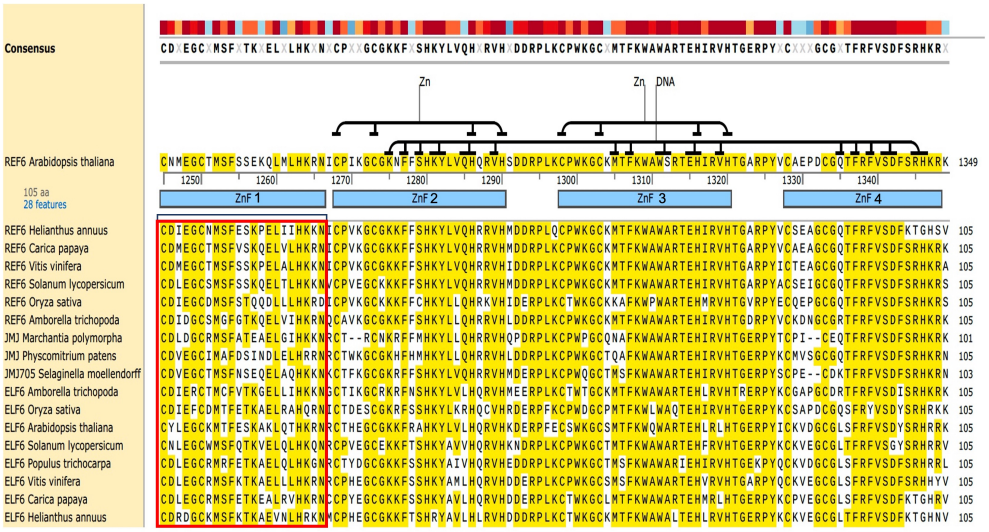


Figure 3.3: Sequence alignment of REF6 and ELF6 zinc-finger homologs from nine plant species. The positions of zinc molecules and zinc finger contact points with DNA are indicated above the alignment. The red-highlighted box illustrates the lower conservation level of ZnF1 compared to the other zinc fingers.

The sequence alignment revealed high conservation between the REF6 and ELF6 zinc-fingers across these species, particularly in the last three zinc-fingers (Znf2, Znf3 and Znf4). In contrast, Znf1 displayed a lower level of conservation. Notably, the amino acids responsible for DNA interactions in the last three zinc-fingers were highly conserved among all sequences compared. We conducted a phylogenetic analysis on the zinc-fingers to determine if the observed differences were indicative of their association with either REF6 or ELF6. Our data demonstrated that

the zinc-fingers of REF6 and ELF6 are evolutionarily separated in the selected species, with the closest matches in basal plants (liverwort and mosses) branching separately.

To further explore the implications of these differences at the structural level, we extracted the 3D protein zinc-finger structures of *Arabidopsis* REF6 and ELF6 using AlphaFold2 and superimposed them to investigate spatial differences. We observed that the last three zinc-finger domains—Znf2, Znf3, and Znf4—demonstrated an almost perfect overlap, a finding that underscores their pronounced sequence conservation across species; however, the first zinc-finger domains exhibited distinct angles. We hypothesised that this conserved difference between the zinc-fingers could potentially influence gene target specificity, thereby contributing to the functional divergence of REF6 and ELF6. It's worth noting that while the crystal structure of the REF6 zinc-finger has been empirically determined (Tian et al., 2020), the structure of the ELF6 zinc-finger has not yet been scientifically validated and is based solely on an AlphaFold2 prediction. Nonetheless, this prediction carries a high degree of reliability, with a Local Distance Difference Test (LDDT) confidence score exceeding 90%, thereby lending considerable credence to the predicted structure. Further studies are required to elucidate the functional consequences of these structural differences in the zinc fingers (Figure 3.4). For example introducing targeted mutations into the zinc finger domains that are differentially conserved to assess how these changes affect function. This could involve either site-directed mutagenesis or random mutagenesis.

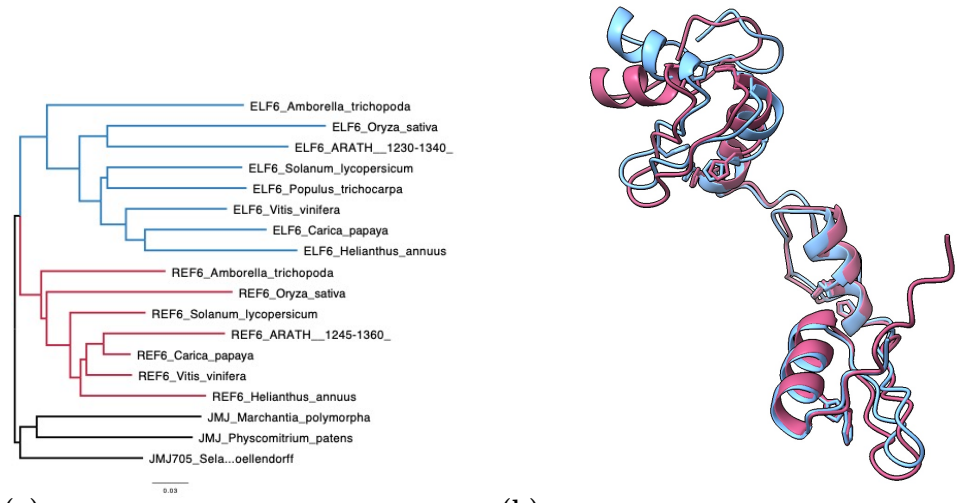


Figure 3.4: Phylogenetic tree of Znf domains of REF6 and ELF6 homologs from nine plant species (a). Red and blue branches correspond to REF6 and ELF6 proteins, respectively; the black branch represents homologs of REF6 and ELF6 in basal plants. Superposition of the AlphaFold2-predicted 3D structure of the Znf domain in AtREF6 and AtELF6, showcases the similarities and differences between the two proteins (b).

Another aspect that could contribute to the differences between REF6 and ELF6 is their interactions with other proteins. A possible source of variation in protein-protein interactions could stem from the analysis of their intrinsically disordered regions (IDRs). These regions are characterised by high conformational flexibility, allowing them to interact with multiple binding partners. We hypothesised that the significant differences in genomic targets between REF6 and ELF6 might be due to differences in these regions, leading to interactions with distinct transcription factors. Both REF6 and ELF6 share a conserved IDR upstream of their zinc-finger domains, but display divergent arrangements of IDRs elsewhere. ELF6 has a unique IDR between its Jmj subdomains, while REF6 features four exclusive IDRs, three of which are upstream of the shared IDR adjacent to the zinc finger domain.

IDRs often contain molecular recognition features (MoRFs), which are short, linear amino acid sequences that undergo a disorder-to-order transition upon binding to a partner protein. The presence of MoRFs in a protein can indicate potential protein-protein interactions. In plants,

MoRFs have been identified in a range of proteins. For example, in species such as *Arabidopsis* and rice, Ethylene Response Factors (ERFs) feature a well-conserved DNA-binding domain (AP2 domain), where MoRFs have been associated with established functional characteristics (Sun et al., 2019). We used the machine-learning molecular recognition feature predictor MoRFPred (Disfani et al., 2012) to identify MoRFs in REF6 and ELF6, and found that most MoRFs are located at the C-terminus of both proteins, upstream of the zinc finger domain, spanning 400–500 amino acids (Figure 3.5).

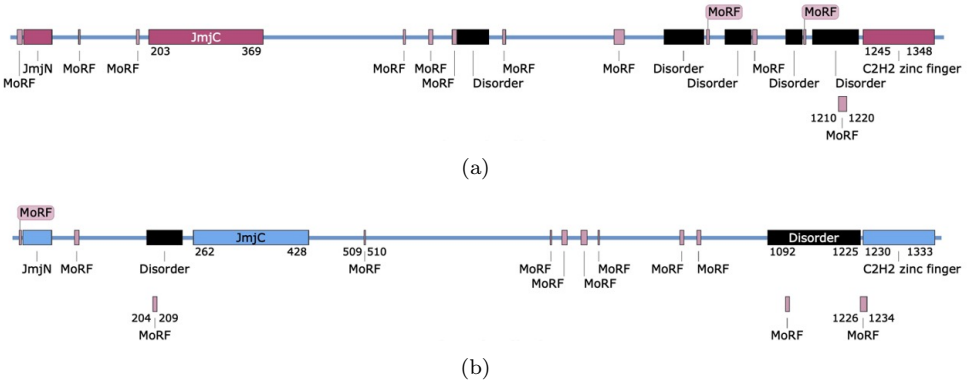
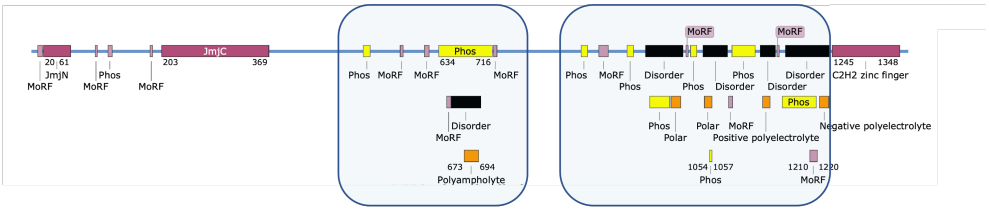


Figure 3.5: Predicted locations of MoRFs in (a)REF6 and (b)ELF6.

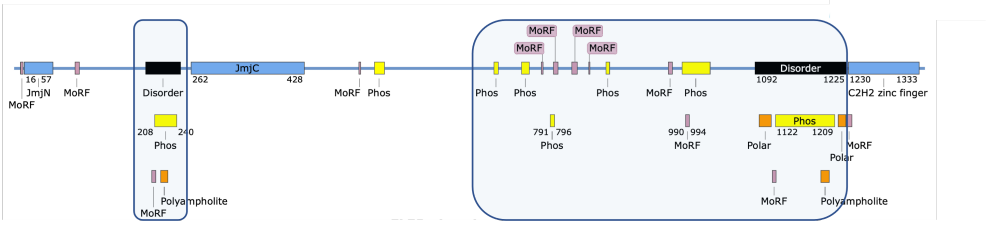
Additional MoRFs were found between the JmjN and JmjC domains. Surprisingly, REF6 exhibited an "island" of 150 amino acids containing several MoRFs, flanking an IDR absent in ELF6. To corroborate these observations, we investigated other markers of protein interaction.

Phosphorylation is a common post-translational modification that can impact a protein's activity, stability, localisation, or ability to interact with other proteins by adding a phosphate group to specific amino acid residues (usually serine, threonine, or tyrosine). Phosphorylation can induce conformational changes in IDRs, either by directly affecting the IDR's structure or by creating or disrupting specific binding sites. Deep-learning frameworks, which use convolutional neural networks trained on empirical evidence, have been developed to predict phosphorylation sites effectively. Using MusiteDeep (Wang et al., 2020a), we predicted the locations of phosphorylated serines and threonines in REF6 and

ELF6. We found that these sites accumulated in regions with MoRFs and predominantly overlapped with IDRs (Figure 3.6). Two distinct large phosphorylated site regions were identified, each exclusive to one protein: one region is found in ELF6, located between the Jmj subdomains; the other is in REF6, where it overlaps with the large intrinsically disordered region (IDR) that is absent in ELF6.



(a) Predicted features in REF6.



(b) Predicted features in ELF6.

Figure 3.6: Predicted locations of phosphorylation sites, polar groups, and MoRFs in REF6 and ELF6. The boxes emphasise distinct regions abundant in features. Both proteins share a feature-rich region upstream of their zinc fingers. REF6 has a unique feature-rich region at the center of the protein, while ELF6 has one between the JmjN and JmjC domains.

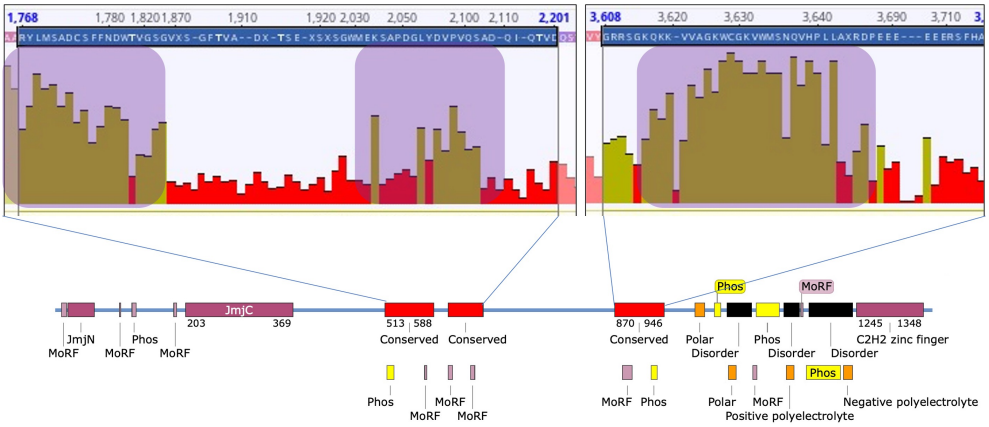
Another marker for investigating protein binding sites is the presence of polar groups. The polar groups in IDRs contribute to molecular recognition and interaction, as they often participate in protein-protein, protein-nucleic acid, or protein-ligand interactions. Several polar groups were found in REF6 and ELF6, all located within IDRs (Figure 3.6). Most polar groups are located at the C-terminal of these proteins, except for two exclusive polyampholyte groups: one in the region between ELF6 Jmj subdomains and one within the exclusive REF6 IDR in the center of the protein. We also observed the presence of two charged groups at the C-terminal of REF6, absent in ELF6: one positive and one negative electrolyte group. These charged residues can have various roles, such as

modulating IDR conformation or mediating protein-protein interactions by participating in electrostatic interactions with charged residues on interacting proteins.

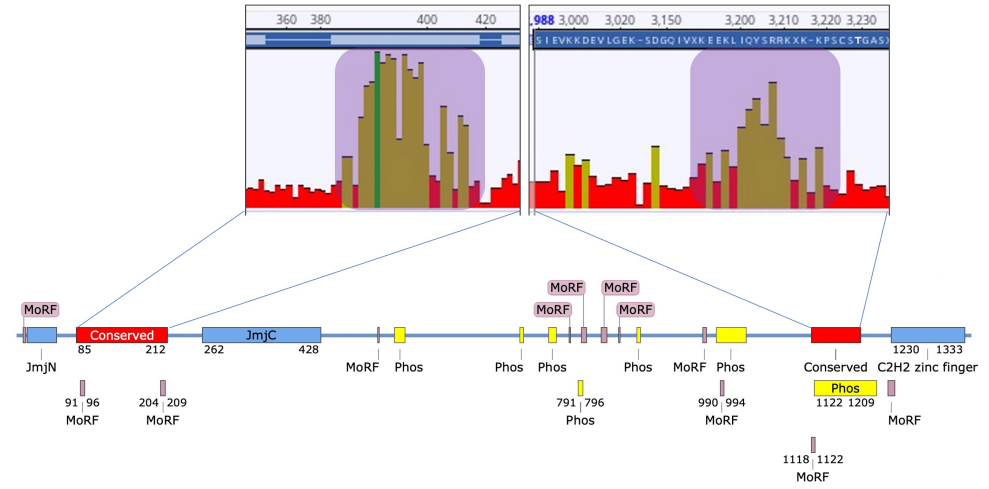
Collectively, our findings indicate that REF6 and ELF6 share many signatures associated with protein-protein interaction at their C-terminal region, upstream of their zinc-finger domains. However, several differences were observed. ELF6 contains an exclusive IDR between its Jmj subdomains with a high number of predicted phosphorylated sites, while REF6 has an exclusive IDR in the middle of the protein, featuring a high accumulation of these features, possibly hinting at a differential protein interaction site. Moreover, REF6 possesses exclusive positively and negatively charged amino acid residues at the C-terminal, which are absent in ELF6.

These differences in IDRs, MoRFs, phosphorylation sites, and charged residues between REF6 and ELF6 could potentially contribute to distinct protein-protein interactions and, subsequently, differences in their genomic targets and functions. It is essential to note that these findings are based on predictions, and further experimental validation is necessary to support these conclusions.

We hypothesised that protein-binding regions would exhibit conservation across various plant species. To test this hypothesis, we aligned 71 AtREF6 homologs precedently used for the phylogenetic study (shown in the Appendix, figures 6.1, 6.2 and 6.3) and analysed their sequences to identify conserved sites. We applied a similar approach for AtELF6, using 79 homologs for the analysis. By examining these sequences, we aimed to uncover evidence of conserved regions, which could potentially support our hypothesis. Our analysis revealed that several conserved regions corresponded to protein-interaction features identified in our previous analysis (Figure 3.7).



(a)



(b)

Figure 3.7: Conserved locations in (a) AtREF6 and (b) AtELF6 homologs. In the graphs, peaks correspond to areas with highly conserved nucleotides, signifying sequence conservation, whereas valleys represent regions with lower conservation. These conserved peaks are highlighted in purple, often corresponding to regions enriched with MoRFs and sites of phosphorylation.

For example, in REF6, three distinct peaks, each representing a group of neighbouring highly conserved amino acids, not associated with any identified domain coincided with the predicted MoRFs. Similarly, in ELF6, two peaks were located in close proximity to phosphorylated sites and MoRFs. These findings suggest that the conserved regions in REF6 and ELF6 may play essential roles in the proteins' functions and their interactions with other proteins. The high conservation of these regions

across different plant species highlights their potential importance in the histone demethylating activity and other regulatory processes.

To conduct a comprehensive analysis of shared conserved regions, we combined the alignments of REF6 and ELF6 homologs and assessed the frequency of conserved amino acids. By pooling these homologs together, our aim was to gain a better understanding of the common evolutionary traits and possible functional similarities or differences between these proteins. This analysis revealed that several other conserved regions exist among different plant species for both REF6 and ELF6, beside the known functional domains (JmjN, JmjC, and zinc fingers).

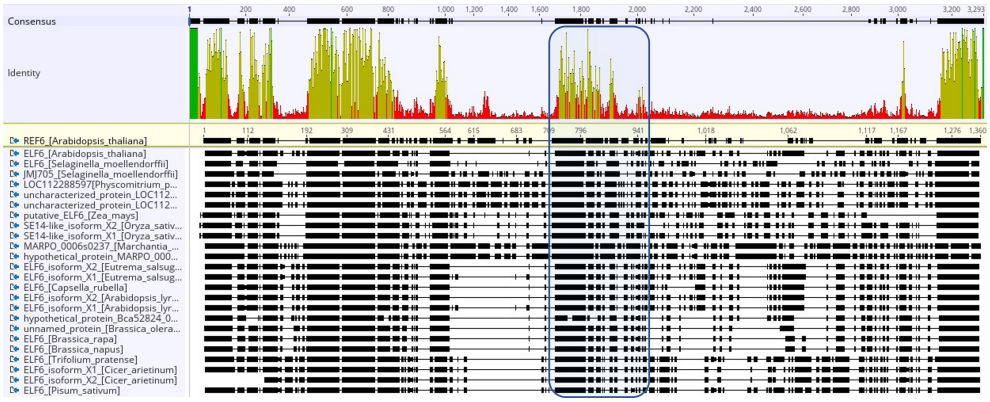
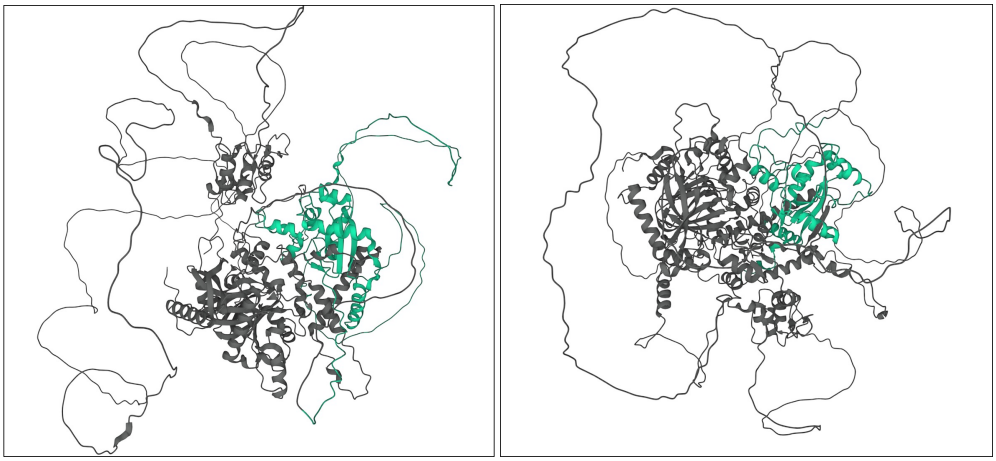


Figure 3.8: Sequence comparison of 62 angiosperm REF6, 59 angiosperm ELF6, 3 gymnosperm REF6, 3 gymnosperm ELF6, and 12 basal plant homologs. The boxed region emphasises a highly conserved sequence shared among all considered homologs.

Interestingly, we found a highly conserved region in the center of both proteins (206 amino acids in ELF6 and 189 amino acids in REF6) that lacked any known annotation or protein interaction features (Figure 3.8).

By examining the predicted structures of REF6 and ELF6 using AlphaFold2, we discovered that this conserved region formed a structured domain with high confidence (Figure 3.9). Despite the presence of an IDR in REF6 that splits the region into two parts, the domains in both proteins displayed high similarity, comprising six alpha helices and two beta sheets (Figure 3.10).



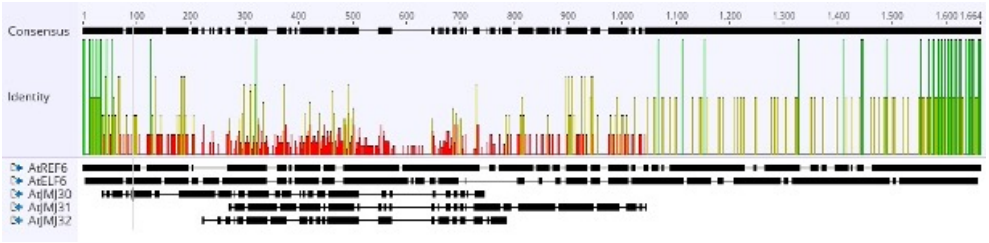
(a) **(b)**

Figure 3.9: Predicted 3D structure of (a) REF6 and (b) ELF6 proteins using Alphafold2. The newly identified domain is colored green in both proteins.

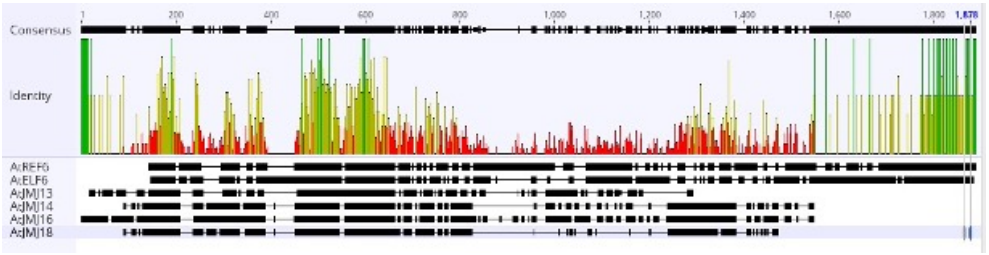
These structured domains were found near the Jmj catalytic domain, leading us to speculate that they may have a conserved function in the histone demethylating activity. We used different software tools available for comparing protein domains based on their structure. RCSB PDB StrucMotif-Search (Bittrich et al., 2020) and Foldseek (van Kempen et al., 2023) both align the structure of a query protein against a database by describing tertiary amino acid interactions within proteins as sequences over a structural alphabet. However, as no similarities were found with characterised domains, we could not determine their specific function.

Figure 3.10: Superposition of the AlphaFold2-predicted 3D structure of the newly identified domain in AtREF6 and AtELF6. Both structures exhibit a high degree of similarity, with the REF6 domain featuring an additional unstructured region.

We proceeded to investigate the significance of the newly discovered structured domain in terms of catalytic functionality. If this domain is crucial for proper enzymatic function, it should be conserved among other proteins within the Jumonji family. We compared REF6 and ELF6 sequences with JmjC domain-only H3K27me3 demethylases Jmj30, Jmj31, and Jmj32 (Figure 3.11a). Apart from the highly conserved JmjN and JmjC regions, the new domain was not conserved. In addition, we conducted a comparison between REF6 and ELF6 and the JmjC H3K4 demethylases Jmj14, Jmj16, and Jmj18 (Figure 3.11b). The latter enzymes are part of the KDM4/JHDM3 group, which exhibit different substrate specificities but are evolutionarily related to the KDM5 group where REF6 and ELF6 are classified.



(a)



(b)

Figure 3.11: Sequence comparisons between AtREF6 and AtELF6 with other Arabidopsis histone demethylases. (a) Comparison between AtREF6 and AtELF6 and other H3K27 histone demethylases AtJMJ30, AtJMJ31 and AtJMJ32.
(b) Comparison between AtREF6 and AtELF6 and H3K4 histone demethylases AtJMJ14, AtJMJ16 and AtJMJ18.

In this case too we did not observe conservation of the new domain, but we did find conservation of 180 amino acids downstream of the JmjC domain. The reason for the conservation of this region is unknown, as no annotations or other features were identified. From these results we can infer that the newly discovered structured domain in REF6 and ELF6 may not be essential for the catalytic activity of other Jumonji family members, as it is not conserved in the compared JmjC domain-only H3K27me3 demethylases (Jmj30, Jmj31, and Jmj32) or JmjC H3K4 demethylases (Jmj14, Jmj16, and Jmj18). However, the conservation of 180 amino acids downstream of the JmjC domain in the H3K4 demethylases suggests there might be some functional significance for this region, even though its specific role remains unclear due to the lack of annotations or other features.

Lastly we sought to determine whether the divergence of REF6 and ELF6 could be attributed to the loss of specific domains. This process, known as functional protein divergence is a common phenomenon in evol-

ution and often leads to proteins acquiring different specialised functions (Lopez-Bigas et al., 2008). To investigate this possibility, we examined regions that were present in basal Jmj proteins but were lost either in REF6 or ELF6 in angiosperms.

Our analysis revealed that ELF6 proteins had lost the intrinsically disordered region (IDR) found in the center of REF6, although remnants of this region were still observed in the early angiosperm Amborella and in monocots (Figure 3.12). Conversely, REF6 proteins in angiosperms lacked the IDR between the JmjN and JmjC domains, a feature present in most of the basal plants we considered. Intriguingly, the loss of this region is the distinguishing factor between REF6 and ELF6 in Gymnosperms (Figure 3.12).

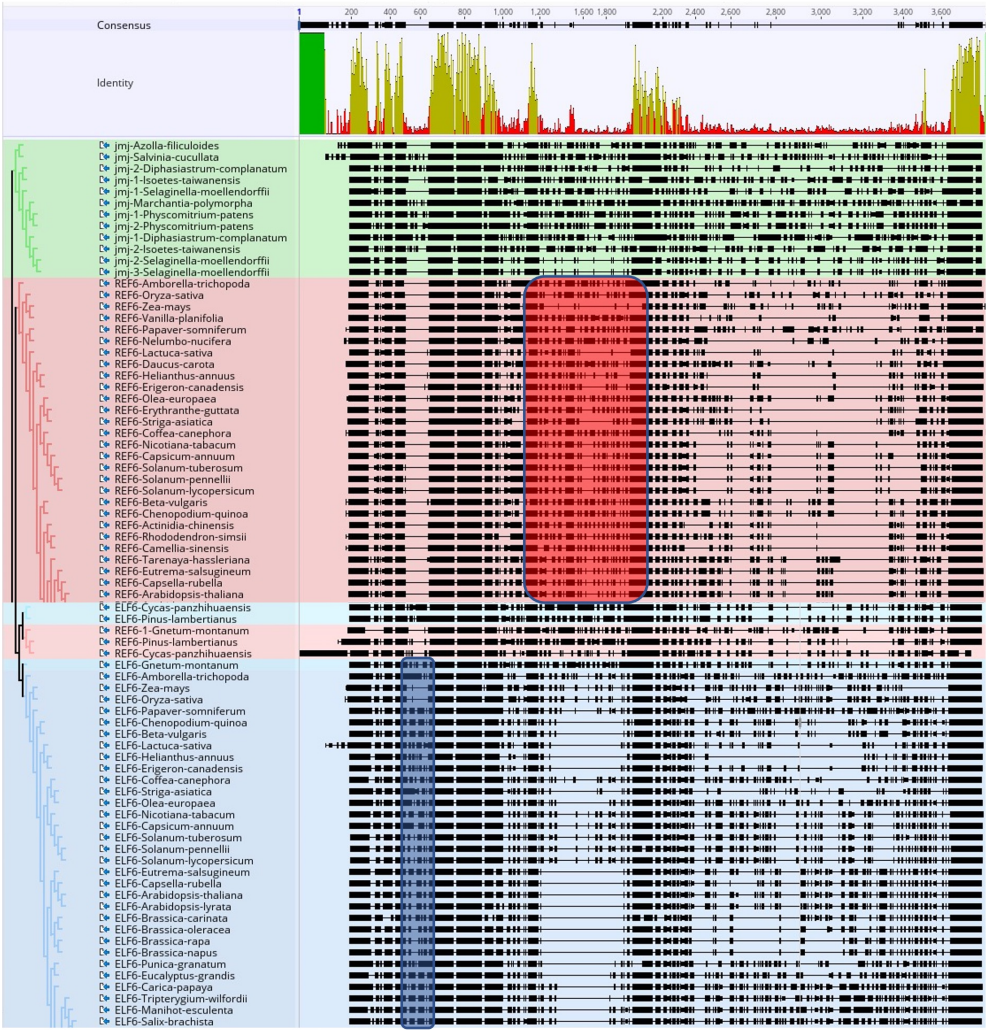


Figure 3.12: Sequence comparison across 62 angiosperm REF6, 59 angiosperm ELF6, 3 gymnosperm REF6, 3 gymnosperm ELF6, and 12 basal plant homologs, illustrating evolutionary conservation and divergence patterns. The red box highlights a region conserved in basal plants and REF6 but lost in ELF6; the blue box highlights a region conserved in basal plants and ELF6 but lost in REF6.

From these observations, we can infer that as REF6 and ELF6 diverged and their functions became more specialised, they underwent domain loss events. The specific loss of certain domains in each protein lineage may have contributed to the functional diversification of REF6 and ELF6, allowing them to adopt distinct roles in the epigenetic regulation of gene expression in plants.

3.3 Discussion

In this study, we aimed to investigate the differences between two histone demethylases, REF6 and ELF6 and their potential shared ancestry, by conducting a comparative phylogenetic analysis. Phylogenetic analysis has long been a valuable tool to reveal the evolutionary relationships among proteins and their associated functions (Joseph, 2004). Our analysis used sequences from a diverse range of plant species, providing a comprehensive dataset to shed light on the evolutionary history of these proteins.

Our findings indicated a clear branching between REF6 and ELF6, suggesting that these histone demethylases have evolved independently and may have undergone functional divergence. This observation is supported by the presence of distinct REF6 and ELF6 genes in Amborella trichopoda, an early diverging angiosperm, hinting at the early divergence of these proteins in the evolution of flowering plants (Albert et al., 2013). The functional diversification of REF6 and ELF6 might be related to the emergence of novel epigenetic regulatory mechanisms associated with the rise of angiosperms. This notion aligns with the idea that epigenetic mechanisms, such as histone modifications, have played a critical role in the adaptation and diversification of plants (Zhao et al., 2019; Ma et al., 2022).

The discovery of a separate evolutionary event in gymnosperms seems to support the theory of parallel functional diversification of REF6 and ELF6 in these two plant lineages (Wickett et al., 2014). Further investigation into the evolutionary forces driving the independent divergence of REF6 and ELF6 in gymnosperms and angiosperms is required.

In addition to the phylogenetic analysis, we compared the structures and domain annotations of REF6 and ELF6, focusing on their zinc-finger domains, intrinsically disordered regions (IDRs) and other conserved regions. Structural comparisons can be instrumental in revealing functional similarities and differences among proteins. Our analysis demonstrated high conservation in the zinc-finger domains of REF6 and ELF6, particularly in their last three zinc-fingers, which are responsible for DNA interactions. This finding is consistent with previous studies showing that

direct binding with the dsDNA strand involves insertion of only the ZnF3 and ZnF4 helices into the major groove (Tian et al., 2020). The ZnF1 domain neighbours the acidic region and its surface is less electropositive than the other ZnFs. Moreover, the acidic region repels DNA. Taken together, this explains why the ZnF1 domain is not directly involved in DNA binding and therefore why its sequence is not as well conserved as the others.

We also examined the IDRs of REF6 and ELF6, which are known to be involved in protein-protein interactions and may contribute to the functional differences between these proteins (Pazos et al., 2013; Wright and Dyson, 2015). Our analysis identified unique IDR arrangements in REF6 and ELF6, accompanied by variations in molecular recognition features (MoRFs), phosphorylation sites, and charged residues in two exclusive regions, one specific to REF6 and another to ELF6. These differences could potentially impact protein-protein interactions (Katuwawala et al., 2019; Bah and Forman-Kay, 2016; Wright and Dyson, 2015) and subsequently, the genomic targets and functions of REF6 and ELF6. However, it is important to note that these findings are based on predictions, and experimental validation is necessary to support these conclusions.

The discovery of a conserved structured domain in REF6 and ELF6, not found in other Jumonji family members, is intriguing. Although the specific function of this domain remains unknown, its conservation across different plant species suggests that it may play an essential role in the histone demethylating activity of REF6 and ELF6. Further studies are needed to elucidate the precise function of this domain and its implications in the catalytic activity of these proteins. For example phenotypic analyses of mutants of REF6 and ELF6 lacking the unique IDRs or the shared conserved domain could reveal altered developmental or biochemical pathways, thereby shedding light on the specific roles these domains play in ensuring the correct functionality of these proteins.

Our investigation of domain conservation in the evolution of REF6 and ELF6 revealed that both proteins underwent domain loss as they diverged and specialised in function. This observation is consistent with the idea that domain loss can drive functional diversification among

proteins (Nasir et al., 2014). Notably, the regions lost were the same as those we identified earlier while analysing protein interaction features. We hypothesise that the loss of these particular domains in each protein lineage may have played a role in the functional differentiation of REF6 and ELF6, enabling them to assume distinct roles in epigenetic regulation through varying interactions with different protein sets.

In conclusion, phylogenetic and structural analyses of REF6 and ELF6 conducted provides valuable insights into the evolutionary history and functional divergence of these histone demethylases in plants. The identified differences in IDRs, MoRFs, phosphorylation sites, and charged residues between REF6 and ELF6 suggest that these proteins may have distinct protein-protein interactions, potentially contributing to differences in their genomic targets and functions. Moreover, the conserved regions in REF6 and ELF6, which are not associated with known functional domains or protein interaction features, raise intriguing questions about their roles in the histone demethylating activity and other regulatory processes. The domain loss events observed in REF6 and ELF6 during their divergence and functional specialisation provide additional evidence for the functional diversification of these proteins.

Future studies should focus on experimentally validating the predicted protein-protein interactions, elucidating the specific roles of the conserved regions in the function of REF6, ELF6 and other Jumonji family members, and determining the precise mechanisms underlying the domain loss events and their consequences for the functions of REF6 and ELF6. A better understanding of the evolutionary forces driving the independent divergence of REF6 and ELF6 in gymnosperms and angiosperms would also be of great interest.

3.4 Summary

Our comparative analysis of REF6 and ELF6 has revealed several conserved regions and putative protein interaction features that may contribute to their distinct and specialised functions in the epigenetic regulation of gene expression. Further investigation into these features and domains

will provide a more comprehensive understanding of the evolutionary divergence and functional diversification of these histone demethylases in plants.

4 Investigation on JmjC domain activity

4.1 Introduction

Histone modifications are crucial post-translational modifications that regulate chromatin structure and transcription. For several decades, histone methylation was thought to be a stable and irreversible modification (Thomas et al., 1972; Byvoet et al., 1972). This notion persisted due to the relatively slow turnover rate of methyl groups on histone proteins compared to other histone modifications such as acetylation (Honda et al., 1975). The discovery of two families of enzymes capable of demethylating histones has shifted our understanding of histone modifications. These two evolutionarily conserved families of histone demethylases employ distinct reaction mechanisms for demethylation, showcasing the complexity and diversity of epigenetic regulation.

4.1.0.1 LSD and JmjC demethylases

The first family is called LSD demethylases after LSD1, the first reported histone demethylase, which catalyses demethylation of H3K4me1 and H3K4me2 and can also demethylate H3K9me1 and H3K9me2 (Shi et al., 2004). The other LSD family member, LSD2 (also known as KDM1B) has so far been shown to only demethylate H3K4me and H3K4me2 (Karytinos et al., 2009). Both LSD1 and LSD2 use a flavin adenine dinucleotide (FAD)-dependent amine oxidation reaction to catalyse the demethylation of their substrate. Owing to the requirement of a free electron pair at the methylated lysine residue, LSD1 can only demethylate mono- and dimethylated lysine residues but not tri-methylated lysine residues (Lee et al., 2007; Stavropoulos et al., 2006).

The identification of the JmjC family of histone demethylases occurred in 2006, when two distinct research teams independently discovered JHDM1A (now referred to as KDM2A) and JHDM2A (now referred to as KDM3A), both of which contain the JmjC domain (Tsukada et al., 2006; Whetstine et al., 2006). These enzymes were shown to possess the capability of demethylating lysine residues in a manner reliant on the JmjC

domain. Within the JmjC family of proteins, demethylases with activity towards H3K4, H3K9, H3K27, H3K36 and H4K20 have been identified (Lee et al., 2007; Tsukada et al., 2006; De Santa et al., 2007; Tahiliani et al., 2007). JmjC histone demethylases employ an α-ketoglutarate- and Fe(II)-dependent dioxygenase mechanism which allows the JmjC family of enzymes, in contrast to the LSD family, to demethylate trimethylated lysine residues. The histone demethylases have specific substrate specificities that can discriminate between different lysine residues and their degree of methylation. Clues as to what confers their substrate specificities have come from crystallographic studies (Hopkinson, 2013; Tsai et al., 2014).

4.1.0.2 Structural features of 2-OG enzymes

The JmjC domain features a double-stranded β-helical (DSBH) fold which consists of eight β-strands forming a β-sandwich structure composed of two four-stranded antiparallel β-sheets (Clifton et al., 2006). Most 2-OG oxygenases (the enzyme family to which histone demethylases belong) also contain extra secondary structure elements surrounding the DSBH, which define the various subfamilies. These additional elements include extra β-strands that extend the DSBH, at least one helix on the C-terminal side of the DSBH, and inserts between the fourth and fifth β-strand. The DSBH core fold offers a rigid scaffold for the cosubstrates 2-OG and Fe2+, which are situated in the more open end of the barrel. The iron is coordinated by two histidine residues along with a glutamate or aspartate residue in a conserved HxE/DxH motif found among 2-OG oxygenases (Figure 4.1).

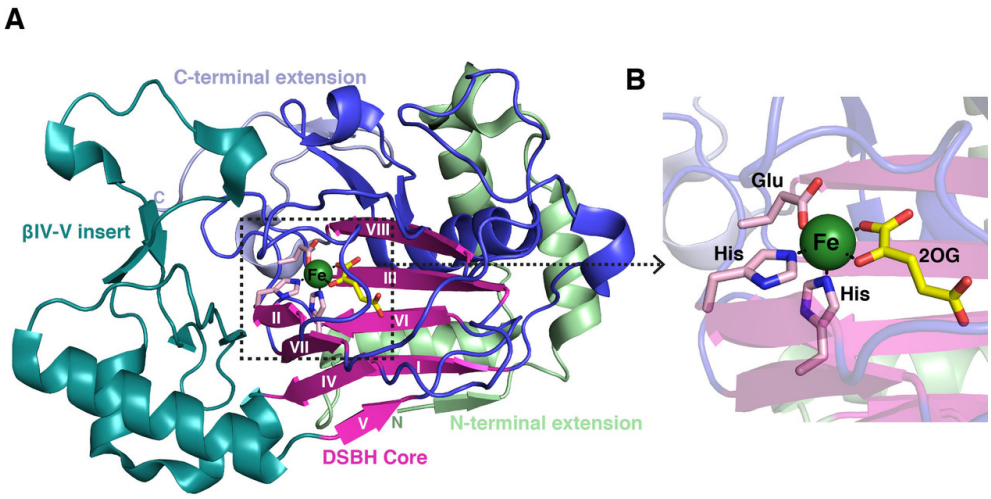


Figure 4.1: Overview of the canonical structure of 2OG oxygenases. (A) Representation of the 2OG oxygenase fold. The core distorted double-stranded beta-helix (DSBH) fold is in pink (b-strand I–VIII). Elements surrounding this core are subfamily characteristic.(B) Insert showing the coordination of Fe(II) by a conserved triad of residues and the cosubstrate 2OG. Figure adapted from Walport and Schofield (2018).

4.1.0.3 Substrate specificity

Substrate binding varies significantly among the different demethylase subfamilies. Crystallographic studies have shown that it often involves residues from the first and second β-strand, in addition to strands and loops that extend the DSBH. Other protein domain modules, such as plant homeodomain (PHD), Tudor, AT-rich interaction domain (ARID), or CxxC zinc finger motifs, which differ between the various subfamilies, are typically required for interaction with DNA/RNA or chromatin protein substrates (Clifton et al., 2006; McDonough et al., 2010).

For example, different members of the JMJD2 family (a subfamily of the JmjC demethylases) have different substrate preferences. JMJD2D can demethylate H3K9me2 and H3K9me3 but not H3K36me2 and H3K36me3. Sequence alignment of JMJD2A and JMJD2D shows that a variable Ser/Ala position in the binding pocket is responsible for this difference, and by generating S288A and A291S mutations in JMJD2A and JMJD2D respectively, it is possible to switch the substrate specificity of the two enzymes (Couture et al., 2007). Substrate specificity is however, not

exclusively determined by the structure of the catalytic domain. Within the PHF2–PHF8–KIAA1718 (also known as JHDM1D) subfamily of JmjC enzymes, PHF8 targets H3K9me1 and H3K9me2, whereas its close relative KIAA1718 catalyses demethylation of both mono- and dimethylated H3K9 and H3K27 (Horton et al., 2010; Huang et al., 2010). This difference in substrate specificity has been ascribed to different distances between the JmjC and plant homeodomain (PHD) finger domains. Both enzymes associate with H3K4me3 through their PHD domain, but PHF8 has a shorter, more flexible linker that assumes a bent conformation, allowing its JmjC domain to interact with and demethylate H3K9me1 and H3K9me2. For KIAA1718, the linker is longer and more rigid, resulting in an extended conformation that renders it inactive towards H3K9me1 and H3K9me2 in the presence of nearby H3K4me3, and leads to selectivity towards H3K27me1 and H3K27me2 in vitro (Horton et al., 2010).

4.1.0.4 H3K27me3 demethylases in plants

Plant histone demethylases display unique traits and divergent evolutionary relationships relative to those found in animals (Lu et al., 2008). A notable distinction is the absence of the H3K27me3 demethylase KDM6 subfamily in plants. Instead, the two plant H3K27me3 demethylases REF6 and ELF6 show sequence similarities to the human H3K9me3/H3K36me3 bi-specific KDM4 subfamily demethylases (Lu et al., 2008; Mosammaparast and Shi, 2010). These two proteins do not yet have crystal structures, whereas two other *Arabidopsis* histone demethylases have been characterised: JMJ13 is an H3K27me3 demethylase belonging to the KDM4 subfamily and JMJ14 is an H3K4me3 demethylase belonging to the KDM5 subfamily. Structural analyses of these two plant demethylases have revealed that the JmjC active site displays characteristics typical of α-KG-dependent oxygenases observed for other jumonji domain histone demethylases (Zheng et al., 2019; Yang et al., 2018).

Comparing the structures of H3K27me3 demethylases JMJ13 and KDM6 provided insight into the factors that determine substrate specificity. The recognition of the histone residue H3P30 is essential for differentiating between H3K27me3 and H3K9me3, suggesting that JMJ13

and KDM6 have independently developed specific stacking interactions with H3P30 to maintain substrate specificity (Zheng et al., 2019). Comparisons between JMJ13 and JMJ14 have identified key residues implicated in the recognition of H3K4me3 by JMJ14 and H3K27me3 by JMJ13. These essential residues are not conserved, despite the fact that the JMJ13 and JMJ14 have comparable overall structures. This implies that the preserved jumonji-helical-zinc finger module could function as a universal scaffold for histone demethylases, while unique attributes evolve within individual demethylases to enable differential substrate specificity (Zheng et al., 2019).

Although the crystal structures of REF6 and ELF6 are unavailable, molecular and genetic analyses indicate that H3K27me3 is their substrate (Lu et al., 2011a; Crevillén et al., 2014). Overexpression of REF6 decreases the H3K27me3 and H3K27me2 signals but not the H3K27me1 signal (Lu et al., 2011a). The effects of REF6 overexpression on H3K27me3/2 levels were abolished when the histidine at position 246, a conserved iron-binding amino acid, was replaced by an alanine. It has been reported that recombinant REF6 can demethylate H3K4me3/2 and H3K36me3/2 (Ko et al., 2010) but the different activities observed may be the result of the presence of different co-factors, as demonstrated for LSD1 (Lan et al., 2008). Additionally, ChIP-Seq analysis using H3K27me3 antibodies in wild-type Col and the *ref6-1* mutant revealed an increase in H3K27me3 in the *ref6-1* mutant by more than threefold (Lu et al., 2011a). In the case of ELF6, an enzymatic activity assay revealed that transient expression of AtELF6 decreased the levels of H3K27me2 and H3K27me3 in tobacco (Nicotiniana benthamiana) leaves (Crevillén et al., 2014); no changes were observed in the levels of H3K27me1, H3K4me3, H3K9me2 and H3K36me3. In the *elf6-5* mutant, an alanine is replaced with a valine (amino acid 424) at the carboxy-terminal end of the JmjC domain. This amino acid is conserved between ELF6 and the human H3K27me3 demethylases KDM6. The mutation of this highly conserved residue reduced H3K27 demethylase activity and increased H3K27me3 levels by about twofold to fourfold in vernalised seedlings (Crevillén et al., 2014).

However, despite widespread agreement that REF6 and ELF6 function as main erasers of H3K27me2/3, examination of ChIP-seq data, originally published in a prior study by our laboratory, (Antunez-Sanchez et al., 2020) on *ref6*, *elf6*, and *elf6 ref6* double mutants calls into question the prevailing assumption that ELF6 lacks the ability to demethylate H3K27me1.

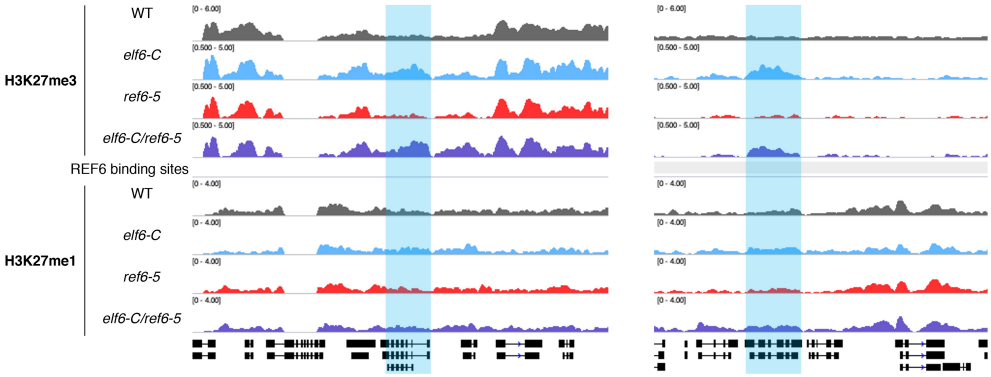
4.1.1 Experimental rationale

The purpose of this chapter was to investigate the differences in catalytic activity by first analysing the data that we collected from chromatin and transcriptome analyses, and then by generating chimeric proteins that have features from both REF6 and ELF6 demethylases and expressing those proteins in different systems. This was done in order to determine whether or not the two different proteins exhibited any discernible behavioural differences.

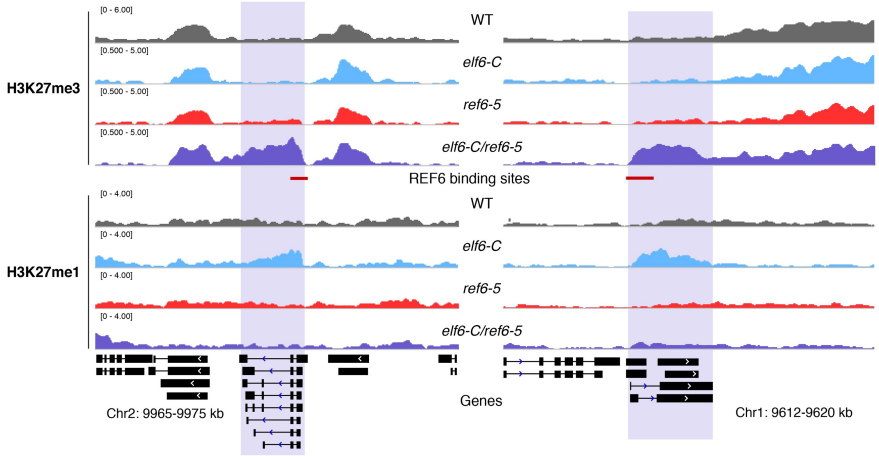
4.2 Results

We aimed to investigate if REF6 and ELF6 exhibit identical catalytic effects on their common target, H3K27me3, by examining their respective profiles of histone modification. To this end, we scrutinised ChIP-seq data from a prior publication by our research group (Antunez-Sanchez et al., 2020), focusing on *elf6-C*, *ref6-5*, and *elf6-C/ref6-5* double mutants, as well as the wild type, specifically looking at histone modifications H3K27me3 and H3K27me1. According to existing literature, both REF6 and ELF6 can demethylate the repressive H3K27me3 mark while preserving the expression-permissive H3K27me1 mark in euchromatin (Lu et al., 2011a; Crevillén et al., 2014). Due to a lack of research on ELF6's direct targets, we identified potential indirect targets based on the presence of the H3K27me3 mark in *elf6-C* and *elf6-C/ref6-5* mutants, which was absent in the *ref6-5* mutant (Figure 4.2). Contrary to current models suggesting that ELF6 can reduce H3K27me3 no further than H3K27me1 (Crevillén et al., 2014), we found that most of the putative ELF6 targets displayed no peaks for H3K27me1.

We next examined genomic locations jointly targeted by REF6 and ELF6. Targets were identified based on the appearance of the H3K27me3 mark exclusively in the *elf6-C/ref6-5* double mutant, whilst being absent in the individual *elf6-C* or *ref6-5* mutants. Again examining the H3K27me1 profile, we observe two distinct characteristics. In accordance with the previously described catalytic activity of REF6, there is a peak for this marker in the *elf6-C* mutant in which REF6 is active. Subsequently, we observe that the peak is absent in the *ref6-5* mutant, where ELF6 is active (Figure 4.2). This provides additional evidence that the catalytic domains of the two demethylases may result in distinct activities.



(a)



(b)

Figure 4.2: Genome browser views of background subtracted ChIP-seq signals for H3K27me3 and H3K27me1 as normalised reads per genomic content (RPGC) in wild-type (WT) and histone demethylase mutants (*elf6-C*,*ref6-5* and *elf6-C/ref6-5*). On the top (a), shaded boxes show ELF6 target genes which gain H3K27me3 methylation in elf6 and *elf6-C/ref6-5* but not in *ref6-5*; concurrently, these genes do not show H3K27me1 in WT. On the bottom (b), shaded boxes show shared REF6 and ELF6 target genes which only gain H3K27me3 methylation in *elf6-C/ref6-5*. The H3K27me1 mark is present only in *elf6-C*.

We developed two hypotheses that could explain the observed data. The first hypothesis suggested that ELF6 would remove one of the three methylations found at H3K27me3, effectively converting the mark from me3 to me2. Absence of antibodies specific to the H3K27me2 mark prevented the collection of experimental data to test this hypothesis. The second hypothesis was that ELF6 is capable of completely removing

tri-methylation from H3K27. In order to determine the validity of either of the two hypotheses, we created a predictive model representing what we would observe in the four distinct backgrounds (*elf6-C*, *ref6-5*, *elf6-C/ref6-5* double mutant and wild type) on gene targets of both REF6 and ELF6, based on the two hypotheses described previously (Figure 4.3).

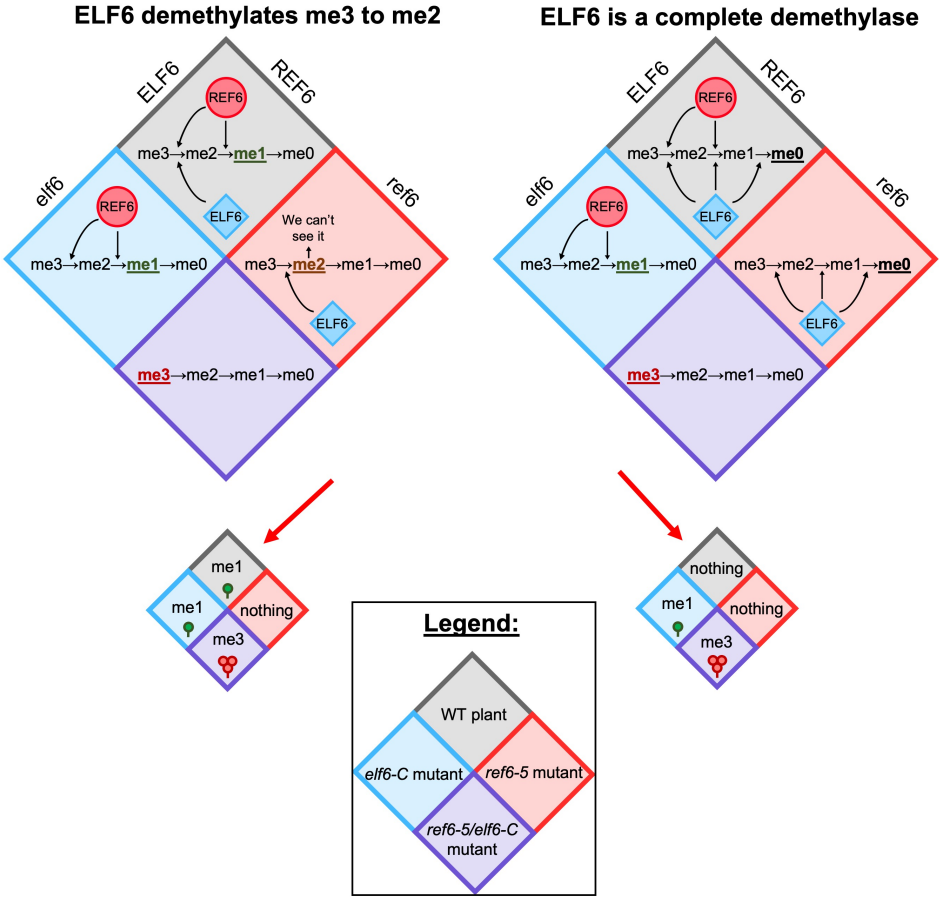


Figure 4.3: Two hypotheses for the histone demethylating activity of ELF6. We show the varying degrees of histone methylation which can be observed in the different backgrounds (WT, elf6-C, ref6-5 and elf6-C/ref6-5). The matrices on the left illustrate the possible combinations of histone methylation for the first hypothesis. The matrices on the right are in support of the second hypothesis.

Our results are consistent with the view that ELF6 is a full H3K27 demethylase. It's important to highlight that the demethylase JMJ13 possesses the capability to remove H3K27me3, which may suggest a

potential influence on our observed results. Nonetheless, existing evidence demonstrates that in vivo, JMJ13 can only remove a single methylation (Zheng et al., 2019), indicating it wouldn't have altered our outcomes. Furthermore, when examining exclusive ELF6 targets, the absence of enrichment of H3K27me3 in the *elf6-C* mutants would be expected if JMJ13 targeted those sites. As this is not the case, we can confidently dismiss the idea that JMJ13 had any interference with our findings.

We then sought to determine what will occur at the level of gene expression in genes targeted by both REF6 and ELF6 in all four conditions (WT, *elf6-C*, *ref6-5*, and *elf6-C/ref6-5* double mutant).
We selected a group of target genes common to both REF6 and ELF6 based on the absence of the H3K27me3 mark in the WT and single mutants, but its presence in the double mutant. We then analysed their transcriptomic data. There were two replicates for each condition, and they were clustered accordingly (Figure 4.4). In the WT, where both demethylases are active, the repressive H3K27me3 mark is removed from target genes, allowing for higher expression. In contrast, the double mutant, where the repressive mark isn't removed, manifests the lowest expression level. Intriguingly, the *ref6-5* mutant clustered with the WT and has significantly higher expression than the *elf6-C* mutant. These observations indicate that the expression levels of genes targeted by both demethylases are higher when ELF6 is present and REF6 is absent, and lower when REF6 is present and ELF6 is absent. Following our initial hypothesis that ELF6 can remove all three methylations from H3K27, whereas REF6 can only remove two, we can hypothesise that genes with H3K27me1 are less expressed than genes that have lost all three methylations from H3K27.

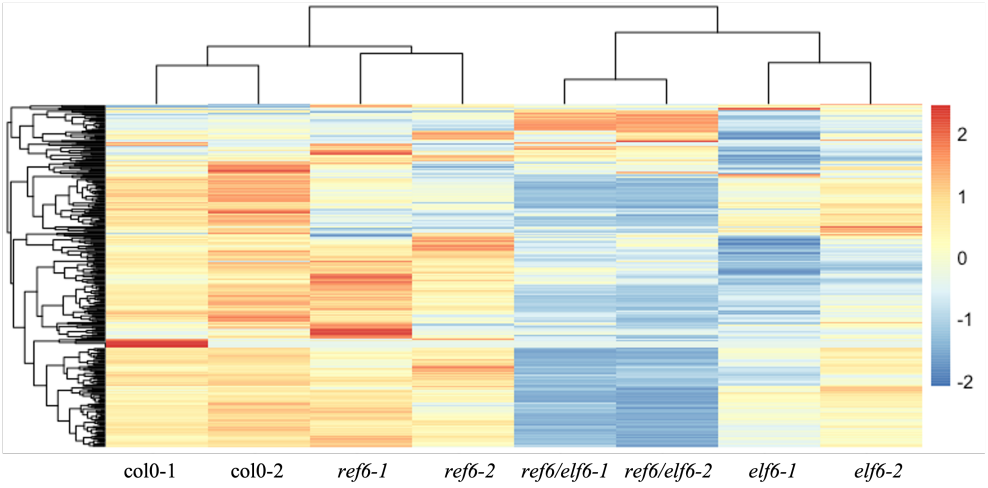


Figure 4.4: Heatmap showing scaled expression levels of genomic targets of both REF6 and ELF6 in wild-type (col0), *elf6*, *ref6* and *ref6/elf6*.

We devised a proof-of-concept experiment to test the hypothesised difference in catalytic activity between REF6 and ELF6. It has been demonstrated that REF6 has more than 5000 direct targets (Cui et al., 2016b), whereas ELF6 activity is restricted to a small subset of genes, most of which can also be targeted by REF6 (Antunez-Sanchez et al., 2020). We hypothesised that ELF6 could completely remove the three methyl groups from the H3K27me3 mark. To investigate this, we replaced the Jmj domain of REF6 with the Jmj domain of ELF6 by creating a chimeric protein that combined the catalytic activity of ELF6 with the ability to target REF6's genomic sites. This would help us study the impact of targeting ELF6 to a large number of known REF6 target sites. In order to test our hypothesis, we generated two constructs: pACT2::ELF6-Jmj::REF6-Znf::GFP (AER), which is the chimeric protein, and pACT2::REF6-Jmj::REF6-Znf::GFP (ARR), serving as a full-length REF6 protein and functioning as a positive control (Figure 4.5). All the constructs were assembled using the Golden Gate cloning system and the primers used for the assembly are shown in Figure 2.4.

We transformed AER and ARR into a *ref6-5* mutant background, ensuring that the native REF6 demethylase activity at its genomic targets would not introduce confounding effects. The *ref6-5* line employed in our

study was originally crossed with another line designed to express RKD4 in an inducible manner for a separate, unrelated experiment. RKD4, a member of the RKD family of transcription factors, known to be involved in early embryogenesis, was not induced in our experiments. For clarity, despite the background capability for inducible RKD4 expression, this feature was not utilised in our study, and thus we will refer to the plant line simply as *ref6-5* throughout our analysis.Moreover, we aimed to examine whether ELF6 could initiate methylation removal from H3K27me1, based on our hypothesis. To assess this, we transformed the AER construct into the wild-type col0 plant, in which the native REF6 protein is active and its targets would exhibit H3K27me1. Exploring the expression and chromatin changes in REF6 target genes would provide insights into ELF6's enzymatic potential.

Conversely, we sought to investigate whether REF6 would influence ELF6 targets if its catalytic activity were attenuated. Following the same rationale, we generated two additional constructs: pACT2::REF6-Jmj::ELF6-Znf::GFP (ARE), containing the REF6 Jmj domain and the ELF6 Znf domain, and pACT2::ELF6-Jmj::ELF6-Znf::GFP (AEE), which is a full-length ELF6 protein serving as a positive control (Figure 4.5).

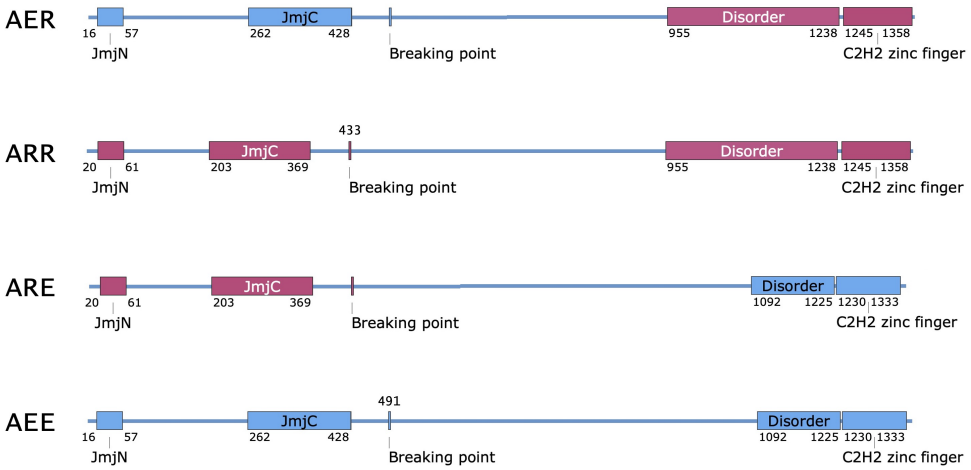


Figure 4.5: Construction of the chimeric protein ELF6-Jmj and REF6 protein binding and Znf. The native domain structure of REF6 and ELF6 and their relative features are shown at the top. The chimeric protein containing ELF6-Jmj and REF6 protein binding and Znf is shown at the bottom with its predicted protein structure. The two separated protein origins are shown in different colours, blue for REF6 and red for ELF6.

Both of these constructs were transformed into the *elf6-C* mutant background. To test our constructs, we infiltrated tobacco leaves with Agrobacterium tumefaciens carrying the plasmids ARR and AEE, in addition to positive control (ACT2:GFP) and negative control (mock infiltration). As expected, the positive control and our constructs expressed GFP, whereas the negative control lacked fluorescence (Figure 4.6).

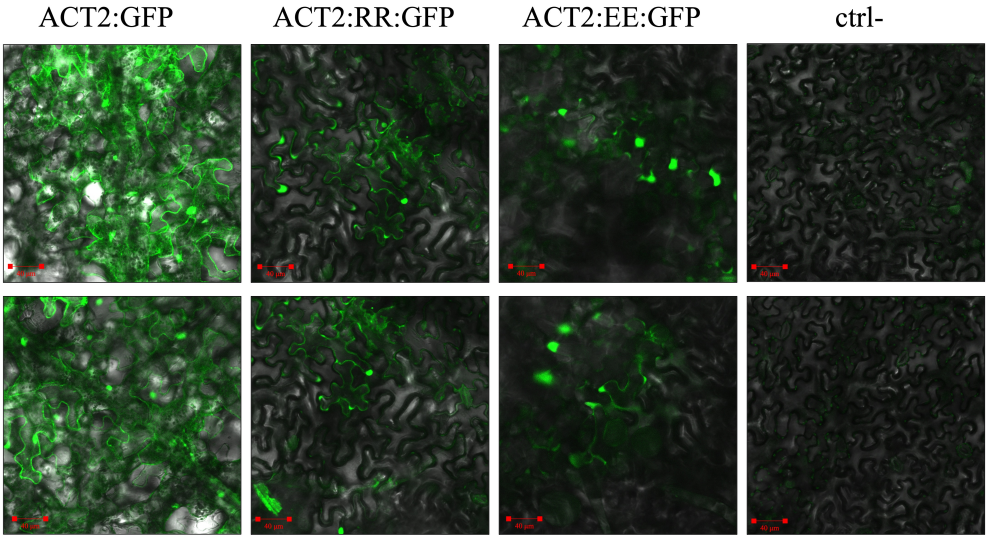


Figure 4.6: Transient expression of GFP-fusion proteins in agroinfiltrated tobacco leaves. Representative tobacco epidermal cells are shown by confocal microscopy 2 days after the agroinfiltration of Agrobacterium tumefaciens that harbored the plasmid expressing ACT2::GFP, pACT2::REF6-Jmj::REF6-Znf::GFP, pACT2::ELF6-Jmj::ELF6-Znf::GFP and col0 as the negative control.

After confirming the functionality of our constructs, we transformed the vectors into the different *Arabidopsis* mutant backgrounds, as described previously. We obtained approximately 20 single transgene copy lines per construct that were propagated to generate homozygous T3 lines (Figure 4.7).

Background	Construct	Line number
ref6-5	**ARR**	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
	AER	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
col0	**AER**	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
elf6-C	**AEE**	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
	ARE	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

GFP positive 4 days after germination

Figure 4.7: Table of T3 homozygous lines generated in the different *Arabidopsis* mutant backgrounds. Lines which have GFP expression in four-day plantlets are highlighted in green.

Seeds from these lines were germinated in plates for four days in 1/2 MS media in a controlled conditions growth chamber (see section 2.7 for details) and 10 plantlets per line were analysed using a confocal microscope to detect GFP expression. At least five lines per construct resulted positive for GFP expression four days after germination.

As anticipated, fluorescence was detected in the nuclei of all the lines, as both REF6 and ELF6 nuclear proteins (Figure 4.8). We repeated the GFP screen seven days after germination and observed a significant decrease in fluorescence in all of the examined lines (Figure 4.9). This phenomenon may be explained by the silencing of the transgene, which is frequently mediated by epigenetic mechanisms including DNA methylation, histone modifications, and RNA interference. RNA silencing is one of the most common forms of transgene silencing, which can induce an RNA silencing response, resulting in the down-regulation or suppression of the transgene. Regrettably, in the course of this study, it was discovered that images for the positive control ACT:GFP and the negative control col0 were not available. This unavailability poses a limitation in our experimental documentation; however, the results and conclusions drawn are based on the consistent trends observed in the experimental lines, which were rigorously validated through our screening process.

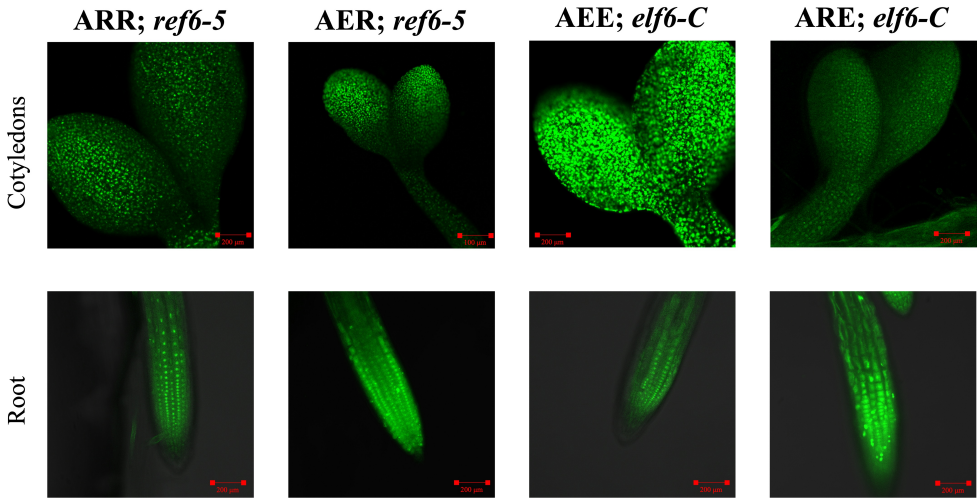


Figure 4.8: Confocal images of four-day old *Arabidopsis* seedlings expressing ARR and AER in *ref6-5* background and AEE and ARE in *elf6-C* background. The panels on top show nuclear GFP expression in cotyledons; the panels on the bottom show nuclear GFP expression in the primary roots. All the lines were observed in T2 generation. Representative microscopic images were reproducibly observed in more than 10 different transgenic lines.

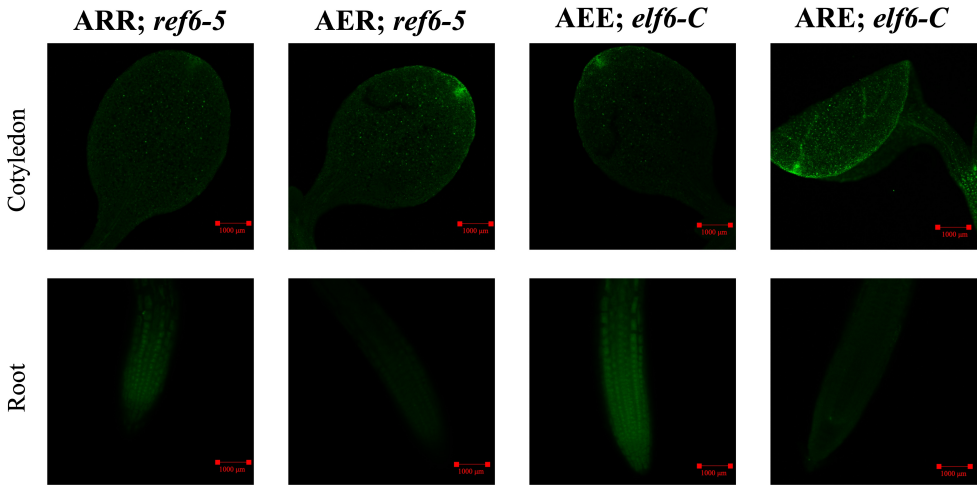


Figure 4.9: Confocal images of seven-day old *Arabidopsis* seedlings expressing ARR and AER in *ref6-5* background and AEE and ARE in *elf6-C* background. The panels on top show reduced nuclear GFP expression in cotyledons; the panels on the bottom show reduced nuclear GFP expression in the primary roots. All the lines were observed in T2 generation. Representative microscopic images were reproducibly observed in more than 10 different transgenic lines.

Using RT-qPCR, we verified the GFP expression at the molecular level in order to determine the extent of transgene silencing. We designed three sets of primers (shown in Figure 2.5) targeting the GFP sequence and selected three T2 lines at two time points, four days after germination (DAG) and fourteen DAG, for the ARR;*ref6-5* and AER;*ref6-5* constructs. Both col0 and *ref6-5* served as negative controls. Figure 4.10 illustrates the GFP expression level for the lines considered relative to GAPDH (two more primers combinations tested are shown in the Appendix, figure 6.4 and 6.5). For all tested primer combinations, the results indicate that: Firstly, for all lines considered, the GFP expression levels decrease significantly (two to fivefold) between the samples at four DAG and fourteen DAG, confirming the confocal microscope observations. Secondly, we observed a distinct difference between the two constructs: ARR;*ref6-5* GFP expression ranged between 10 and 20 times that of AER;*ref6-5*.

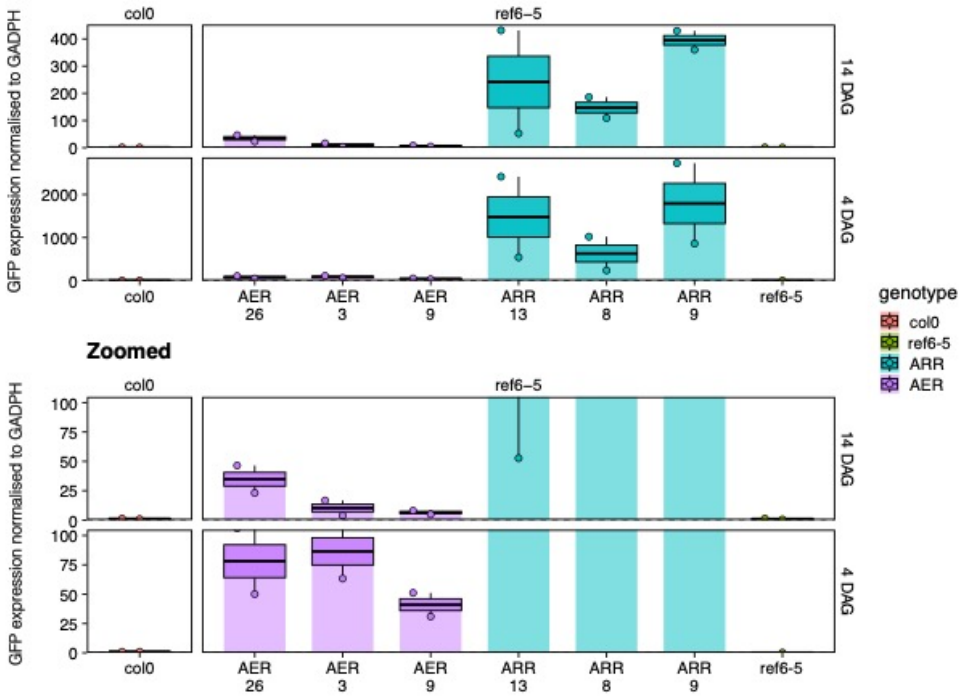


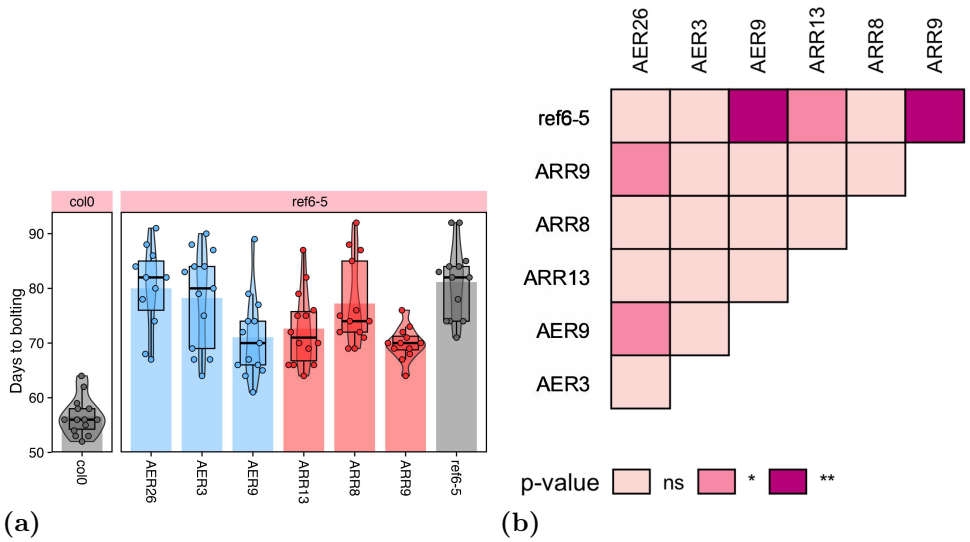
Figure 4.10: GFP expression normalised to GAPDH in selected lines. Three independent lines per genotype (ARR and AER) were screened for GFP expression by qPCR analysis, with two technical replicates per sample. Col0 and *ref6* were used as negative control for GFP expression. The lines were tested at four days after germination (DAG) and 14 DAG. The results were normalised using GAPDH as housekeeping gene. The figure on top shows the GFP expression relative to the ARR lines whereas the figure at the bottom shows the GFP expression relative to the AER lines.

As both REF6 and ELF6 play a role in the regulation of flowering time and their mutants exhibit contrasting phenotypes, we sought to determine whether the expression of our proteins affects flowering time. We anticipated that the ARR;*ref6-5* lines would complement the defects caused by *ref6-5*. If AER possessed the same functionality as ARR, we hypothesised that a flowering phenotype similar to that of ARR would be observed. Nevertheless, if AER had a different effect than ARR, we would observe a different reproductive phenotype. RT-qPCR quantitative data revealed that the expression level of GFP, and by extension our proteins, decreases four days after germination. However, it is still detectable fourteen days after germination. Based on these observations, we assumed for this experiment that our proteins were still active in

some manner after fourteen days, or at least had downstream effects on transcription factors affecting the flowering time. This study did not include a comparison of transgene expression levels with those of the native protein, a factor that could be addressed in future research for a more comprehensive understanding.

A total of 20 heterozygous ARR;*ref6-5* and 22 AER;*ref6-5* lines were grown under short-day conditions until flowering, with the aim of observing variations in flowering times and identifying lines that displayed distinct flowering phenotype compared to the negative control, *ref6-5*. Three lines of ARR;*ref6-5* and AER;*ref6-5* were chosen for propagation to T2 and subsequent testing and analysis. As for the other constructs (AER:col0, AEE;*elf6-C* and ARE;*elf6-C*), three lines each were selected based on the most consistent GFP expression and propagated to T2 After obtaining homozygous seeds for all of the constructs, we proceeded with another flower assay, this time employing all obtained lines (ARR;*ref6-5*, AER;*ref6-5*, AER:col0, AEE;*elf6-C* and ARE;*elf6-C*). 15 plants per line were grown in short day condition. As positive controls, we utilised col0 plants, and as negative controls, *ref6-5* and *elf6-C* were utilised. The days before bolting and the number of leaves at bolting were measured.

The flowering time of the ARR;*ref6-5* and AER;*ref6-5* lines compared to positive control col0 and negative control *ref6-5* is shown in Figure 4.11a. The number of leaves at bolting for the specified lines, shown in Figure 6.6 in the Appendix, is consistent with the observed flowering time data.



(a) **(b)**

Figure 4.11: (a) Flowering time in ARR;*ref6-5* and AER;*ref6-5* under short day conditions. Box plots of col0, *ref6-5*, ARR and AER lines grown in soil until bolting. The boxplot shows the days before bolting.(b) Tukey test for multiple comparison between and within lines and controls, relative to days before bolting.

Tukey test statistical analysis was performed to detect the differences between each line against each other (Figure 4.11b). We observed that neither ARR;*ref6-5* nor AER;*ref6-5* were capable of fully complementing the mutation and displaying the same flowering phenotype as col0. However, two of the ARR;*ref6-5* lines (9 and 13) and one of the AER;*ref6-5* lines (9) displayed substantial deviations from their respective background *ref6-5*, suggesting a partial complementation.

The flowering times of the AER:col0 lines compared to positive control col0 and negative control *ref6-5* is shown in Figure 4.12a. The number of leaves at bolting for the mentioned lines is shown in figure 6.7 in the Appendix.

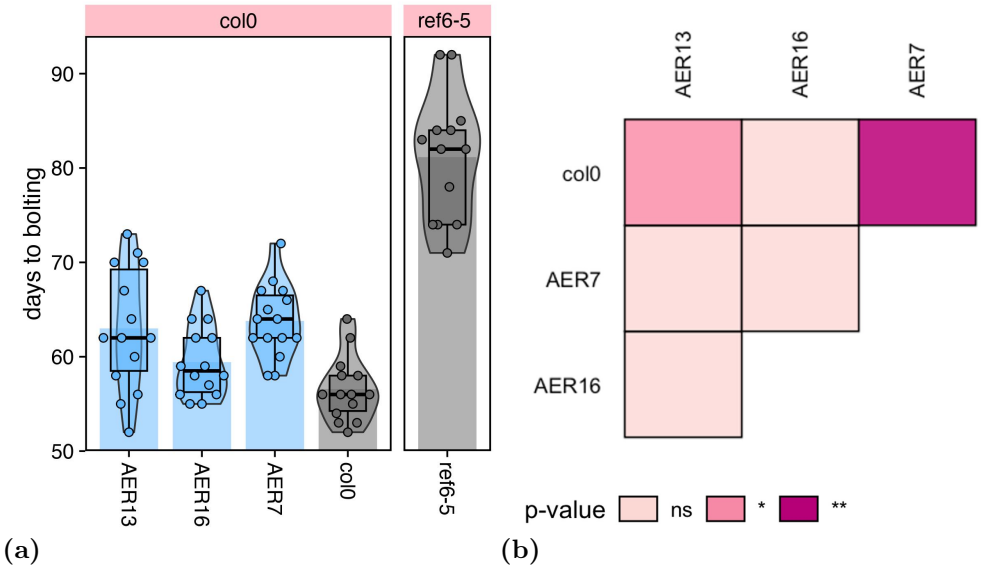


Figure 4.12: (a) Flowering time in AER;col0 under short day conditions. Box plots of col0, *ref6-5* and AER:col0 lines grown in soil until bolting. The boxplot shows the days before bolting. (b) Tukey test for multiple comparison between and within AER;col0 lines and WT control, relative to days before bolting.

We noted that the AER:col0 lines generally exhibited a delayed flowering compared to the background, although not as delayed as *ref6-5*. Our findings revealed that two AER:col0 lines tested were statistically different from the col0 background (Figure 4.12b). In contrast to AER;*ref6-5*, AER:col0 lines 7 and 13 exhibited a late flowering phenotype relative to the WT col0.

The flowering time of the ARE;*elf6-C* and AEE;*elf6-C* lines compared to positive control col0 and negative control *elf6-C* are shown in Figure 4.13a. The number of leaves at bolting for the mentioned lines are shown in figure 6.8 in the Appendix.

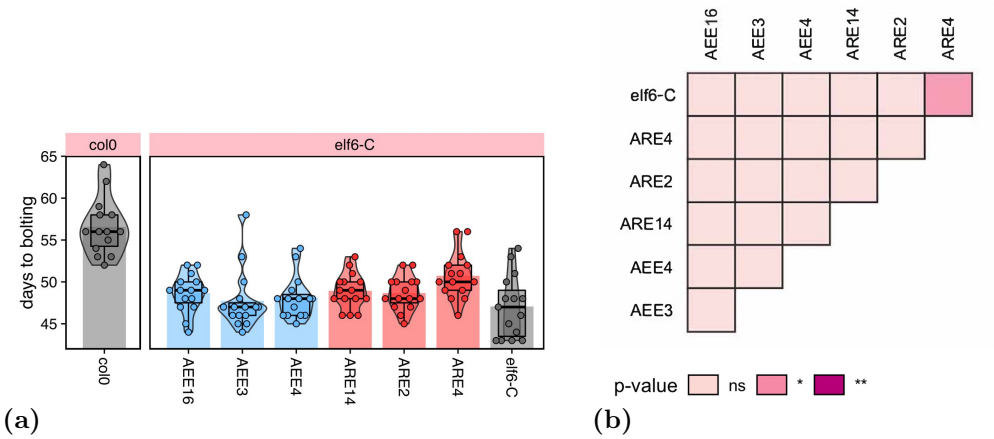


Figure 4.13: (a) Flowering time in ARE;*elf6-C* and AEE;*elf6-C* under short day conditions. Box plots of col0, *elf6-C*, ARE;*elf6-C* and AEE;*elf6-C* lines grown in soil until bolting. The boxplot shows the days before bolting. (b) Tukey test for multiple comparison between and within ARE and AEE lines and controls, relative to days before bolting.

The Tukey test did not reveal any statistically significant differences between our lines and the *elf6* background for these lines (Figure 4.13b), so the AEE;*elf6-C* lines did not complement the flowering phenotype of col0. Only line 4 of ARE;*elf6-C* demonstrates a significant difference in flowering time, showing a later flowering time than *elf6-C*.

The flowering time experiment revealed that two ARR lines and one AER line showed statistically significant differences from the *ref6-5* background, suggesting that the transgenes may have influenced the induction of flowering. To confirm these changes at the molecular level, we examined the expression levels of REF6 target genes in our lines using RT-qPCR. In order to select the REF6 target that would be most informative for our analysis, we selected multiple conditions based on ChIP-seq data for col0 and *ref6-5*: low H3K27me3 in col0 and high H3K27me3 in *ref6-5*; and high H3K27me1 in col0 and low H3K27me1 in *ref6-5*. As for the levels of expression in our chosen targets, they should be high in col0 and low in *ref6-5* (Figure 4.14).

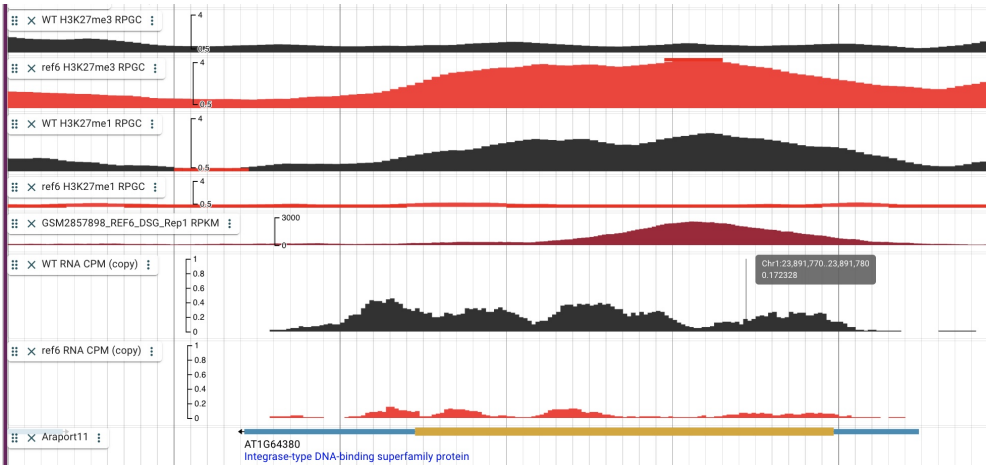
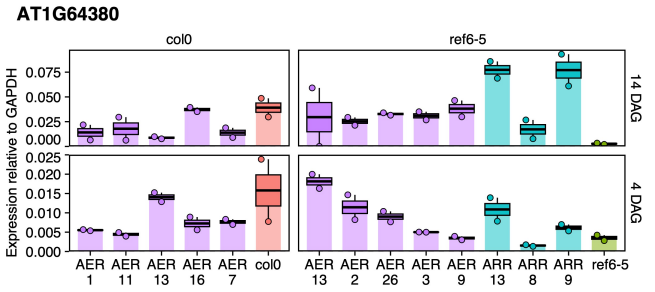


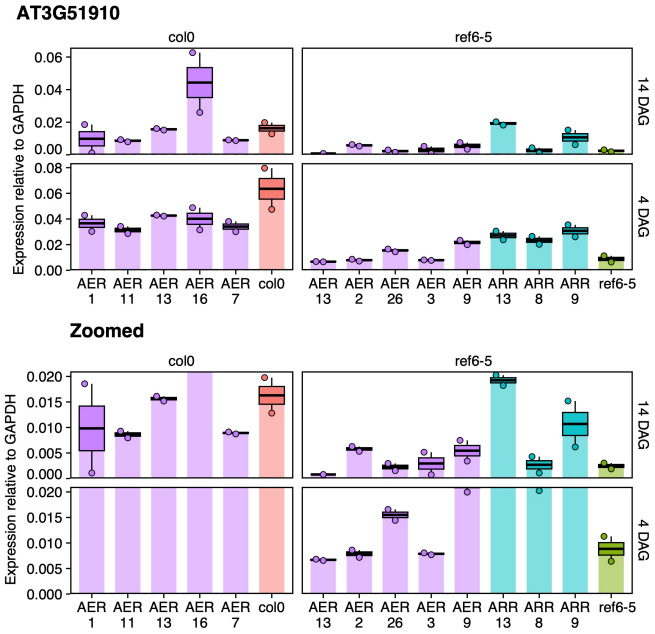
Figure 4.14: Genomic view of a sample REF6 target gene. The first and third tracks show the presence on a REF6 target gene location (AT1G64380) of chromatin marks in WT background, H3K27me3 and H3K27me1 respectively. The second and fourth tracks display the same combination of chromatin marks in *ref6-5* background. The fifth and sixth tracks represent expression data for WT and *ref6-5* respectively. Chromatin marks data were obtained from ChiP-seq analysis on WT and *ref6-5* on the specified histone modifications (Antunez-Sanchez et al., 2020); expression data were acquired from RNA-seq analysis of WT and *ref6-5* (Lu et al., 2011a).

This would allow us to determine if ARR;*ref6-5*, AER;*ref6-5* and AER;col0 were effectively targeting REF6 targets, as they would increase gene expression in our lines relative to the negative control *ref6-5*. We designed primer pairs to specifically measure the expression of REF6 targets (The primers used are shown inf Figure 2.5). We chose three T2 lines for the ARR;*ref6-5*, AER;*ref6-5*, and AER:col0 constructs at two time points, four and fourteen DAG. Both col0 and *ref6-5* served as negative controls. Only three of the six combinations of primers we selected were chosen as they exhibited the expected expression levels for the *ref6-5* negative control and the col0 positive control (AT1G64380, AT3G51910 and AT5G52310). The relative expression levels of the target genes in the ARR;*ref6-5*, AER;*ref6-5*, and AER:col0 lines relative to GAPDH is shown in Figures 4.15a, 4.15b and 4.16a. At four and fourteen DAG, we observed that the expression levels of the ARR and AER lines were, on average, greater than those of the *ref6-5* background. The expression levels of AER:col0 were on average lower than those of the col0 background, indicating a distinct behaviour. A further observation

was that the expression level at fourteen DAG was greater than that at four DAG for two of the three tested genes, possibly indicating that the transgene could induce downstream effects even later than the four days when the silencing is starting to occur.

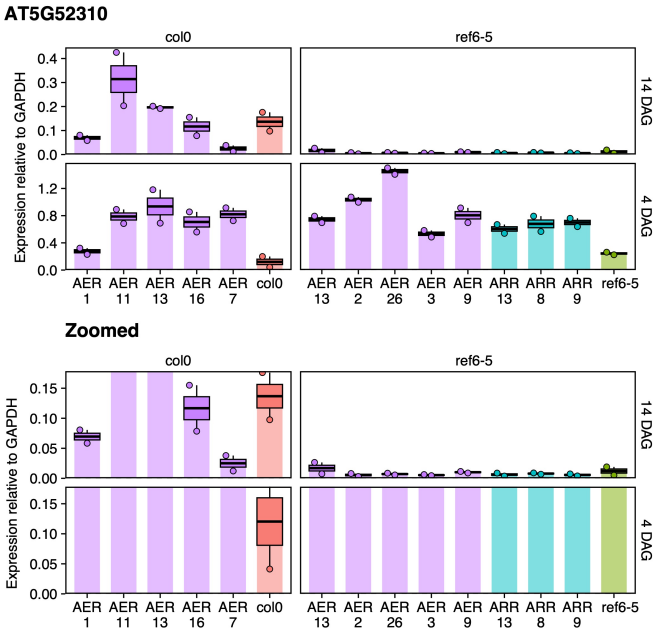


(a)



(b)

Figure 4.15: Expression analysis of selected lines for two REF6 gene targets: (a) AT1G64380 and (b)AT3G51910. Five independent AER genotype lines were screened for each of the two different backgrounds, RKD4 (indicated as col0) and *ref6-5*. Three independent lines were tested for the ARR genotype in ref6 background. The lines were tested at four days after germination (DAG) and fourteen DAG.The results were normalised using GAPDH as housekeeping gene. The figures show the up-regulation of the selected lines compared to *ref6-5* negative control at both four and fourteen DAG.



(a)

Figure 4.16: Expression analysis of selected line for one REF6 gene target: AT5G52310. Five independent AER genotype lines were screened for each of the two different backgrounds, RKD4 (indicated as col0) and *ref6-5*. Three independent lines were tested for the ARR genotype in ref6 background. The lines were tested at four days after germination (DAG) and fourteen DAG.The results were normalised using GAPDH as housekeeping gene. The figure shows the up-regulation of the selected lines compared to *ref6-5* negative control at both four and fourteen DAG.

To acquire a comprehensive understanding of the expression level changes, an RNA-seq analysis was performed (see section 2.15 for details). RNA was extracted from four-day-old plantlets. Based on the GFP expression and on the previous expression assay we selected three lines each from ARR;*ref6-5* and AER;*ref6-5* to use as biological replicates and used three biological replicates each for WT RKD4 and *ref6-5* as positive and negative control, respectively. To evaluate the consistency of gene expression patterns among biological replicates, we initially examined the correlations between replicates to confirm their similarity in expression profiles (see Figure 6.9 in the Appendix). In general, replicates within each genotype exhibited comparable expression characteristics, with the exception of the third replicate in the AER;*ref6-5* line, which deviated from the expected trend. A principal component analysis (PCA) was

performed (see section 2.15 for details) to identify potential outliers and verify if biological replicates clustered together (Figure 4.17).

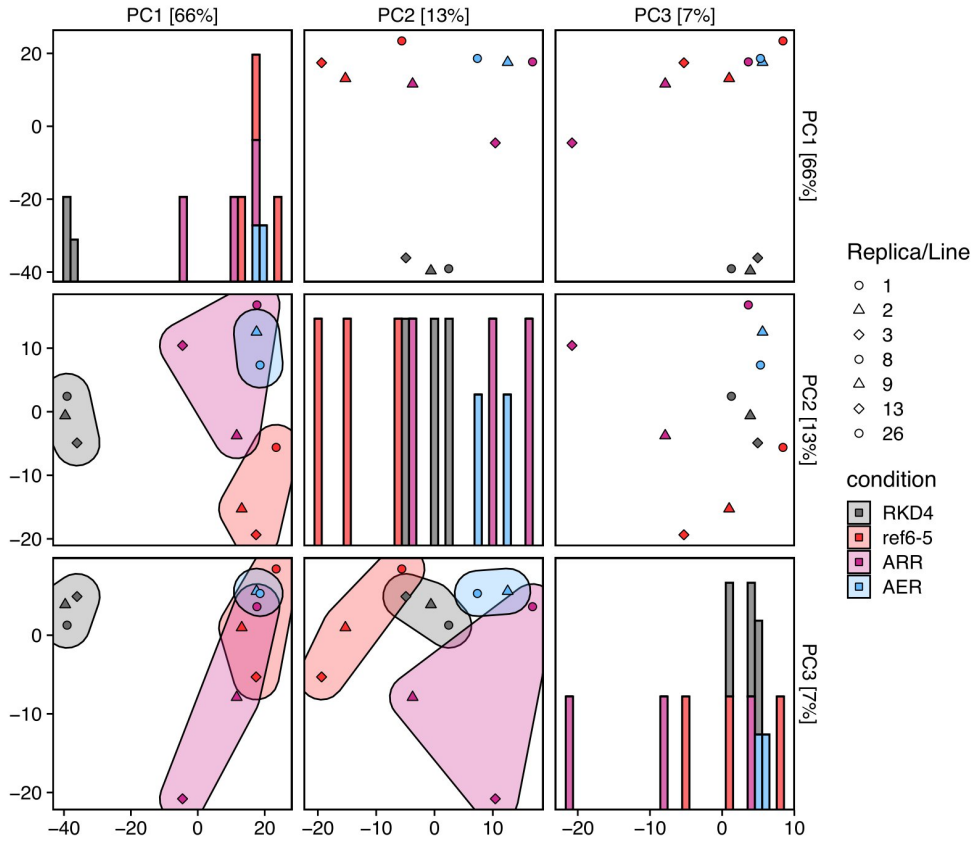


Figure 4.17: Matrix of Principal components (PCs) for variance stabilised transformed counts. The PCA shows sample distances between WT (RKD4), *ref6-5*, ARR;*ref6-5* and AER;*ref6-5* lines grown for four days and analysed via RNA Seq. Line 3 sample from AER;*ref6-5* genotype was removed due to low RNA quality.

Again the line 3 sample of the AER;*ref6-5* did not cluster with the group to which it belonged, and it also received a low score in the quality control test so we decided to exclude it from the PCA. In the first principal component, the WT is observed to cluster separately from the other categories. ARR and AER clustered more closely with ref6, but they are still distinct clusters. The second PC distinguished ref6 from the other groups. ARR and AER clustered more closely with the WT in this instance, despite the variance explained being significantly less

than in the first case. For both PCAs, the ARR sample cluster was more dispersed than the AER sample cluster.

The lists of differentially expressed genes were then analysed (Figure 4.18).

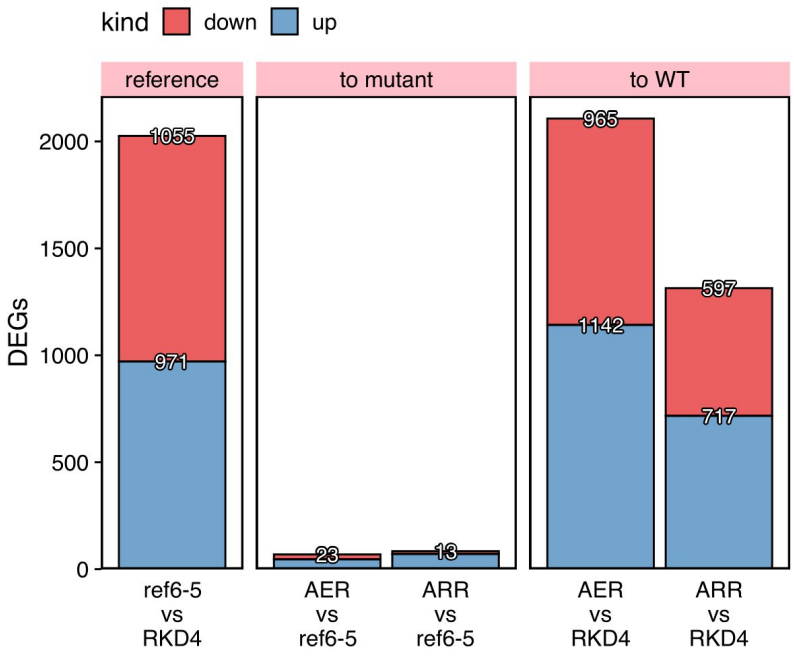


Figure 4.18: Identification of differentially expressed genes (DEGs) by RNA sequencing analysis. Number of DEGs adjusted p-value (p-adj) < 0.01 for *ref6-5* vs WT (RKD4); AER vs *ref6-5*; ARR vs *ref6-5*; AER vs WT (RKD4) and ARR vs WT (RKD4).

We compared the WT positive control to the ref6 negative control as a baseline and we found 971 up-regulated genes and 1055 down-regulated genes. Then we compared the ARR;*ref6-5* and AER;*ref6-5* to their respective background *ref6-5*. Few genes were discovered to be differentially regulated in both cases (84 and 69, respectively). However, comparing the DEGs between our lines and col0, we found that ARR;*ref6-5* had 717 up-regulated and 597 down-regulated genes, whereas AER;*ref6-5* had 1142 up-regulated and 965 down-regulated genes (Volcano plots are shown in the Appendix, Figures 6.10). In this comparison ARR;*ref6-5* shows one-third fewer DEGs than AER;*ref6-5* indicating that the ARR lines were at least partially complementing for the ref6 mutation at the

transcription level. For all comparisons considered, the proportion of up- and down-regulated genes was roughly equivalent.

We then sought to evaluate the expression levels of identified DEGS in our lines (Figures 4.19 and 4.20).

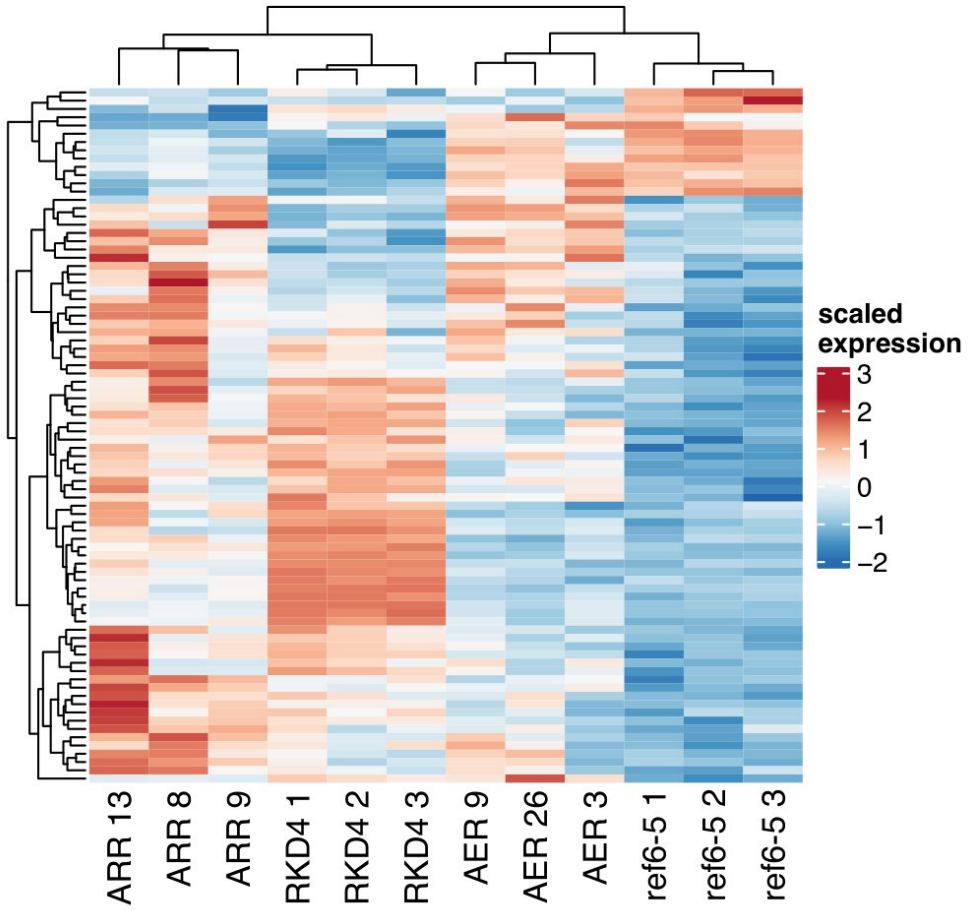


Figure 4.19: Heatmap showing scaled expression levels of differentially expressed genes(DEGs) between ARR;*ref6-5* and *ref6-5*. The heatmap is clustering the log2 fold change (log2(FC)) of each gene.

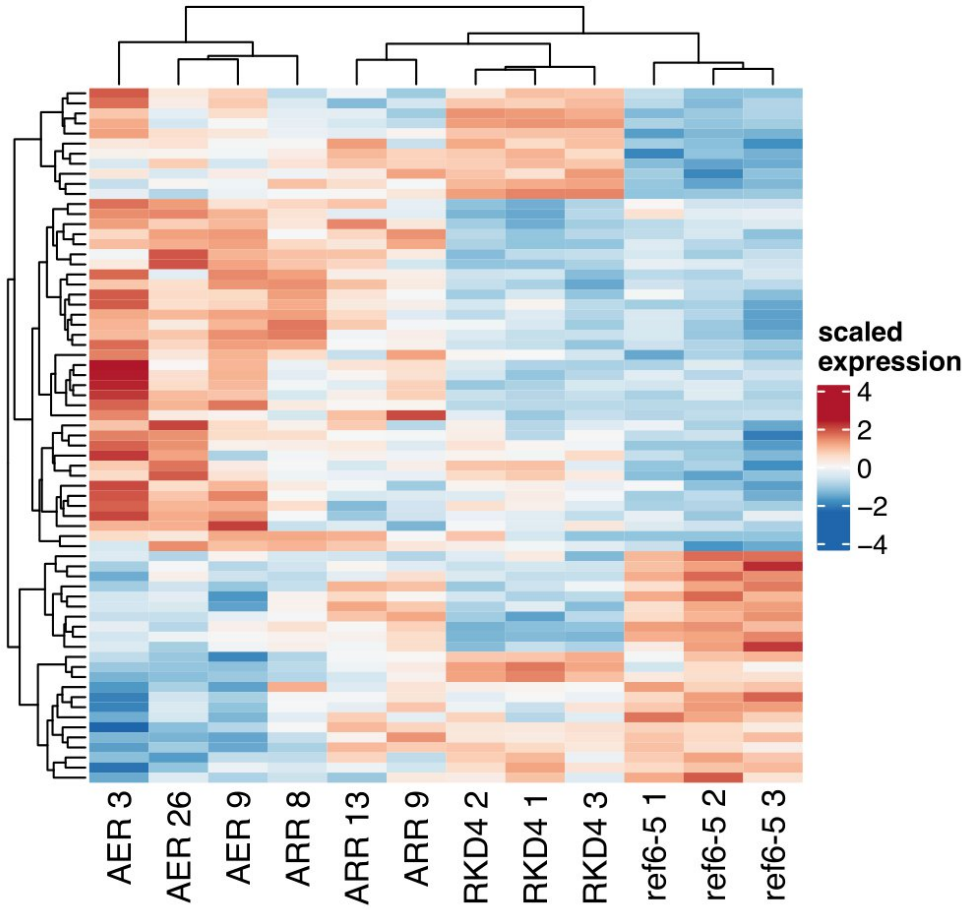


Figure 4.20: Heatmap showing scaled expression levels of differentially expressed genes (DEGs) between AER;*ref6-5* and *ref6-5*. The heatmap is clustering the log2 fold change (log2(FC)) of each gene.

We observed 84 DEGs in the comparison between ARR and ref6 and for these genes, the three samples of the each genotype clustered together.

A cluster of the genes that are highly expressed in ARR relative to *ref6-5* were also overexpressed in WT, suggesting that for a subset of genes there was a compensatory effect (Figure 4.19). We obtained 69 DEGs in the comparison between AER and ref6 and for these genes as well, the three samples of the each genotype clustered together. Surprisingly, the genes which were up-regulated in AER, were not up-regulated in the WT,

suggesting that those genes were not associated with the complementation of the *ref6-5* mutant background.

To examine the data from a different perspective, we chose a set of genes that are down-regulated between *ref6-5* and WT and filtered based on the gain of H3K27me3 in *ref6-5*. We compiled a list of 142 REF6 targets (see the list in Appendix, Figure 6.11) and analysed the levels of expression in ARR and AER. Overall both ARR and AER lines exhibited a higher level of gene expression than the background *ref6-5* but markedly lower than the WT (Figure 4.21). Specifically, the ARR lines appeared to have a higher level of expression compared to the AER lines, with line 13 exhibiting the highest level of activation and clustering, together with the WT. The other ARR lines cluster together with the AER lines (apart from the outlier line 3) which cluster separately from the *ref6-5* background. As previously observed both ARR and AER are not fully capable of complementing the WT, but they are partially effective at activating the REF6 targets.

The fact that AER seems to be less able to activate REF6 targets could be explained by the fact that the fusion of the Jmj catalytic domain of ELF6 with the rest of the REF6 protein might have affected the proper folding or function of the chimeric protein. This could have lead to reduced or altered activity, resulting in only partial complementation of the *ref6-5* mutant. Also the lower expression level of AER detected in the qPCR assay compared to ARR could have affected the extent of the complementation.

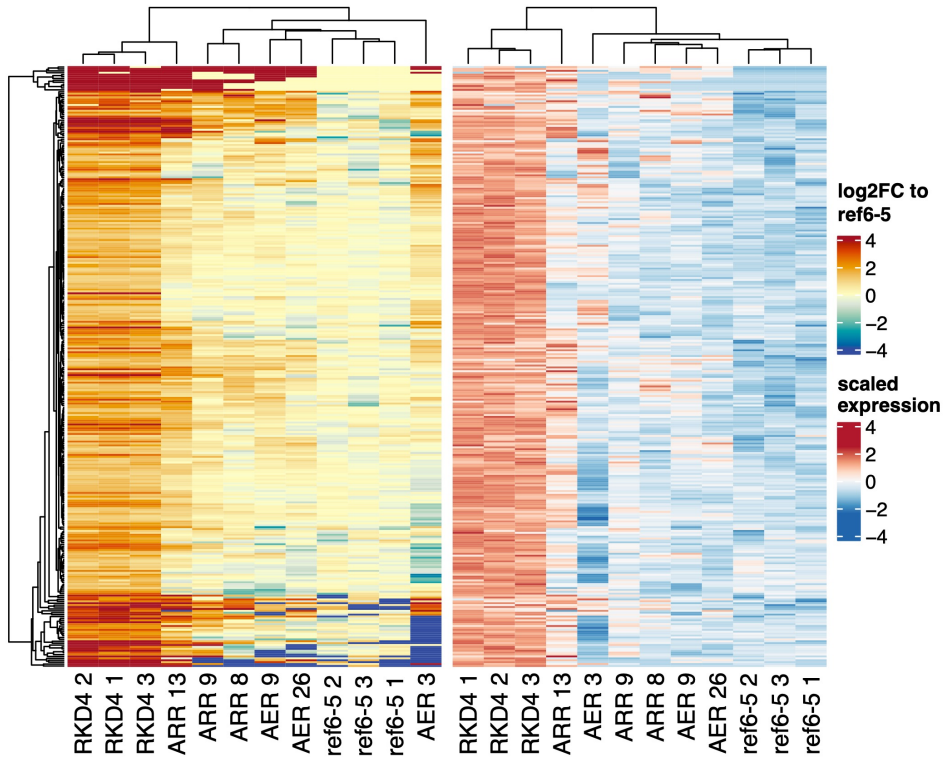


Figure 4.21: Heatmap showing scaled expression levels of differentially expressed genes (DEGs) H3K27me3-methylated in *ref6-5*. Z-score, on the left. For those genes Z-scores are computed and showed on the matrix on the right.

To assess the differential activity of the H3K27me3 demethylases, we took a distinct approach as we sought to examine the behaviour of the JmjC catalytic domain of REF6, ELF6 and JMJ13 in a mammalian system. As previously described, in humans there are two H3K27me3 demethylases, KDM6A and KDM6B, which have distinct characteristics and functions, despite both containing the JmjC domain responsible for their demethylase activity. At the C-terminal of both proteins there is a GATA-like zinc finger domain which is found to have a DNA binding function and plays an important role in the regulatory regions of genes. This form of zinc finger is also found in plants. Hence we sought to investigate the effects of fusing a plant Jmj catalytic domain to the zinc finger domain of human H3K27me3 demethylases, specifically exploring whether plant demethylases could

exert functional influence in a heterologous organism, such as humans. Furthermore, we aimed to determine if these chimeric proteins, derived from the three distinct plant demethylases REF6, ELF6 and JMJ13, would induce different alterations at the transcriptional level. We decided to use the zinc finger domain of KDM6B instead of KDM6A one. Our choice was motivated by the fact that unlike KDM6B, KDM6A contains on the N-terminal region, seven Tetratricopeptide repeat (TPR) domains which mediate interactions with other proteins and allosterically regulate its enzymatic activity. KDM6B lacks the TPR motifs, which simplifies the structure of the resulting chimeric proteins and minimises the possibility of undesired interactions or allosteric effect which can potentially hinder the functionality of the plant JmjC domains. Following this rationale three constructs containing the plant Jmj demethylase domains followed by the KDM6B DNA zinc finger were assembled: REF6-Jmj:KDM6B-Znf; ELF6-Jmj:KDM6B-Znf; JMJ13-Jmj:KDM6B-Znf (Figure 4.22). All the constructs had a mCherry tag at the C-terminal.

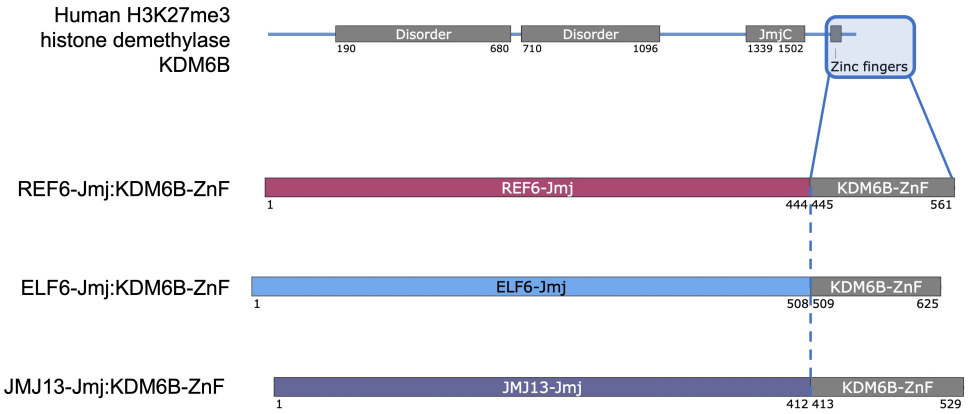


Figure 4.22: Schematic representation of the human KDM6B gene with its functional domains (on top). Below the schematic representation of ELF6, REF6 and JMJ13 Jmj domains fused to the KDM6B Znf domain.

Our constructs REF6-Jmj:KDM6B-Znf; ELF6-Jmj:KDM6B-Znf; JMJ13-Jmj:KDM6B-Znf were transduced into replicative human cells (see section 2.24 for details). FACS was used to select cells with high expression of

mCherry (Figure 6.12 in the Appendix)(see section 2.28 for details), RNA was extracted at three and six hours after induction with doxycycline and sequenced. Cells transduced with an empty vector were used as a negative control. The first PCA (Figure 4.23) demonstrates, as expected, that the WT at distinct time points clusters separately from the experiments. REF6 and JMJ13 cluster together, even at various time periods, whereas ELF6 clusters independently. Specifically the ELF6 with 6h induction clusters in each PC individually.

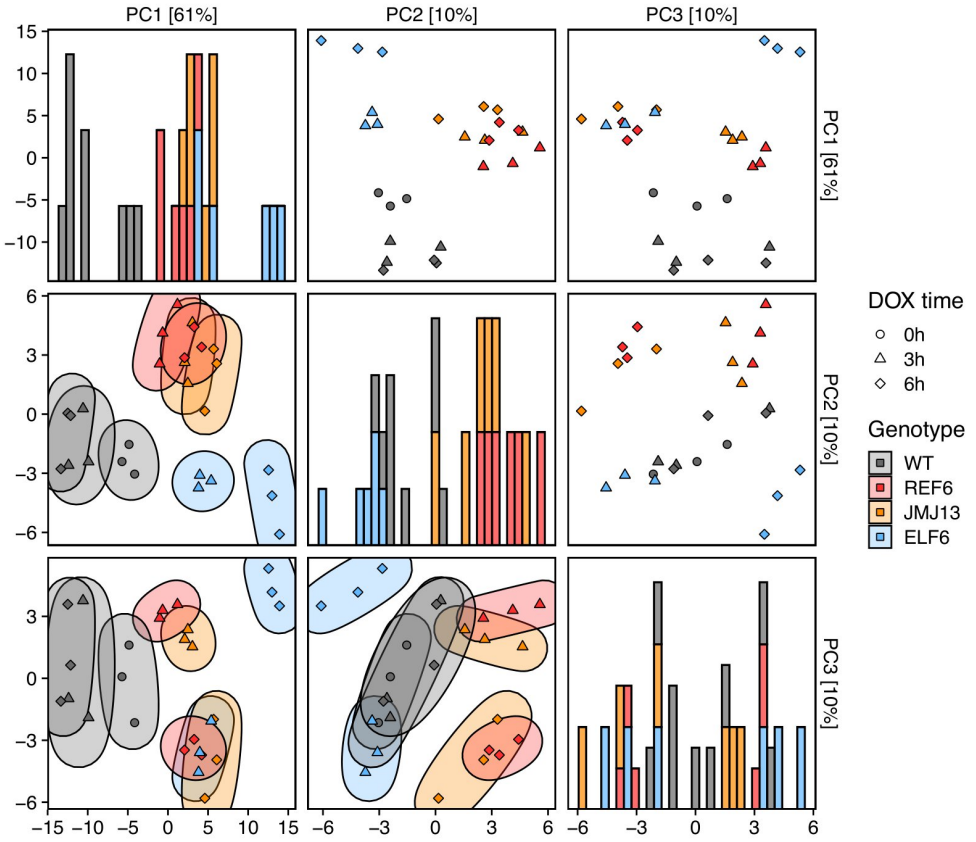


Figure 4.23: Principal component analysis of WT, REF6, ELF6 and JMJ13 expressed in mammalian cell lines. Principal component analysis (PCA) plot visualising sample distances between WT, REF6, ELF6 and JMJ13 in mammalian cells before induction, 3 hours after induction and 6 hours after induction.

The number of DEGS of the three demethylases relative to the WT at 3h and 6h following induction are shown in Figure 4.24 (Volcano plots are shown in the Appendix, Figure 6.13, 6.14 and 6.15). At the three

hour time point, we observed that both JMJ13 and REF6 exhibited a comparable number of DEGs, approximately 200, whereas ELF6 displayed a three-fold higher count, with 614 DEGs. At this time point, the number of down-regulated genes for all three demethylases was slightly greater than that of up-regulated genes, representing between 30 and 40% of the total. Upon reaching the six hour time point, the DEGs for JMJ13 and REF6 more than doubled compared to the three hour mark, with JMJ13 displaying 718 DEGs and REF6 showing 482 DEGs. In contrast, ELF6 experienced a more substantial increase, with its DEGs surging more than three-fold compared to the 3-hour mark, reaching a total of 1,728 DEGs. Notably, at the six hour time point, the number of down-regulated genes significantly exceeded the number of up-regulated genes across all three demethylases, accounting for 80% of the total DEGs in JMJ13, REF6, and ELF6.

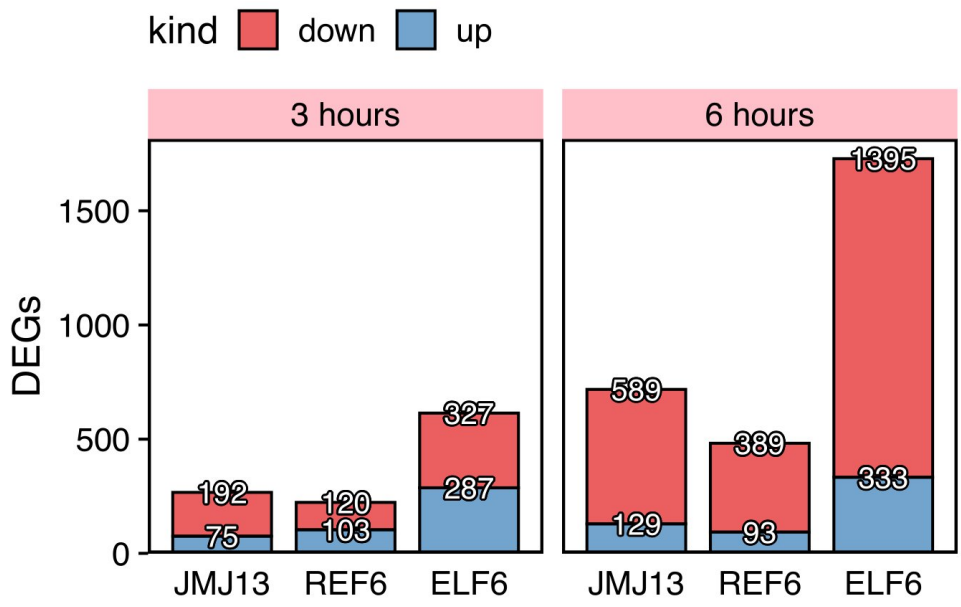


Figure 4.24: Identification of differentially expressed genes (DEGs) by RNA sequencing analysis. Number of DEGs in the RNA-seq data, adjusted p-value (p-adj) < 0.01 for JMJ13-KDM6, REF6-KDM6 and ELF6-KDM6 at 3h and 6h.

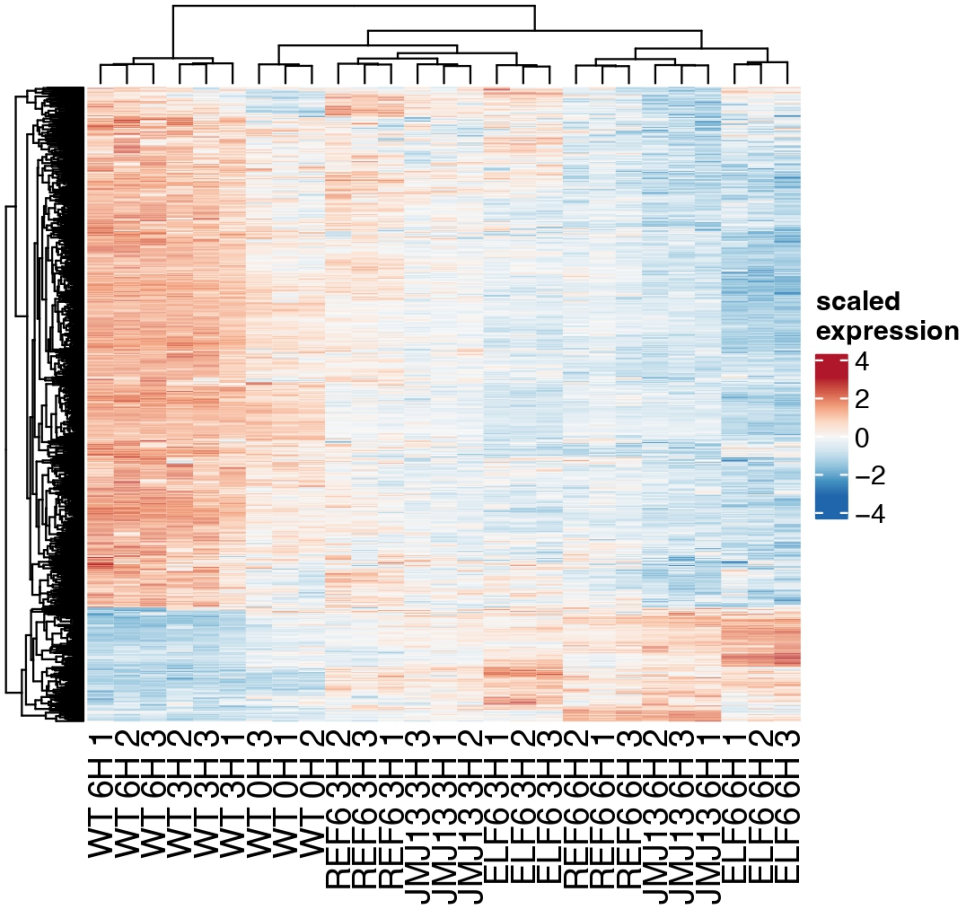


Figure 4.25: Heatmap of differentially expressed genes (DEGs) from RNA-seq data for JMJ13-KDM6. The heatmap displays the hierarchical clustering of DEGs (rows) and samples (columns) based on their normalised, log2-transformed gene expression levels. Each cell represents the expression level of DEGS between JMJ13-KDM6 and WT at 6h from induction, with colours indicating relative expression: high expression (red), average expression (white), and low expression (blue). DEGs were identified using a log2 fold change threshold of > 1 and an adjusted p-value (FDR) < 0.05.

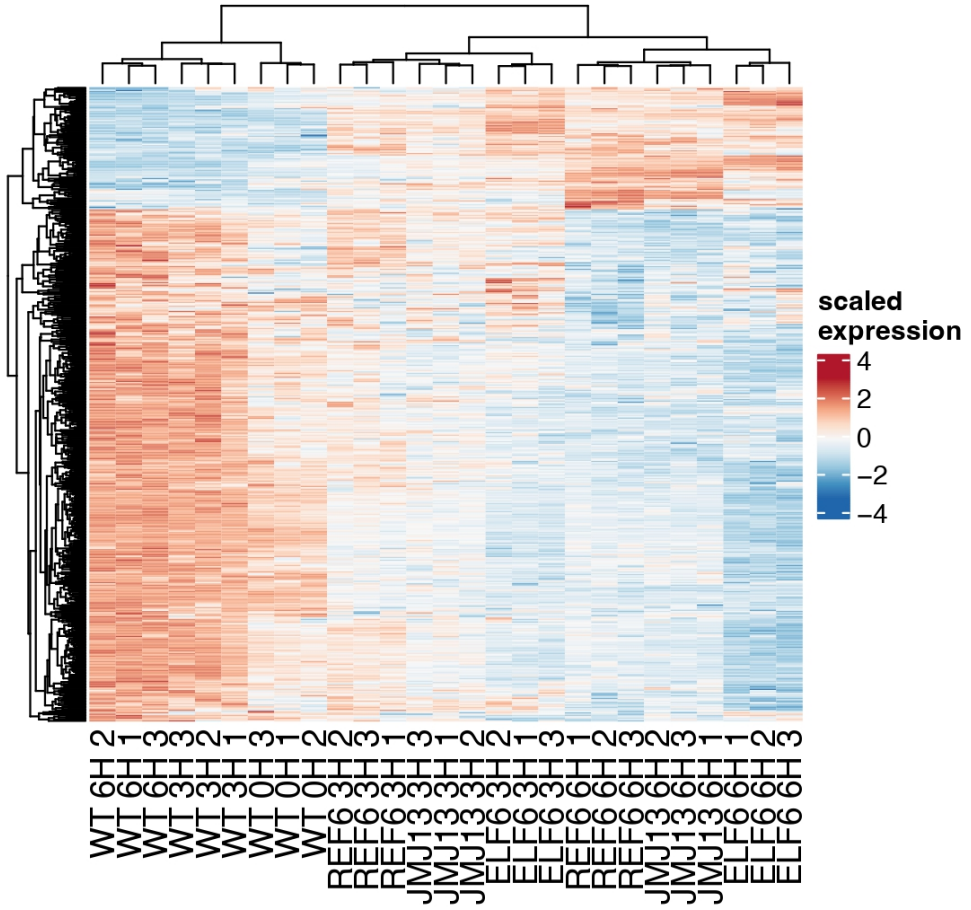


Figure 4.26: Heatmap of differentially expressed genes (DEGs) from RNA-seq data for REF6-KDM6. The heatmap displays the hierarchical clustering of DEGs (rows) and samples (columns) based on their normalised, log2-transformed gene expression levels. Each cell represents the expression level of DEGS between REF6-KDM6 and WT at 6h from induction, with colours indicating relative expression: high expression (red), average expression (white), and low expression (blue). DEGs were identified using a log2 fold change threshold of >1 and an adjusted p-value (FDR) < 0.05.

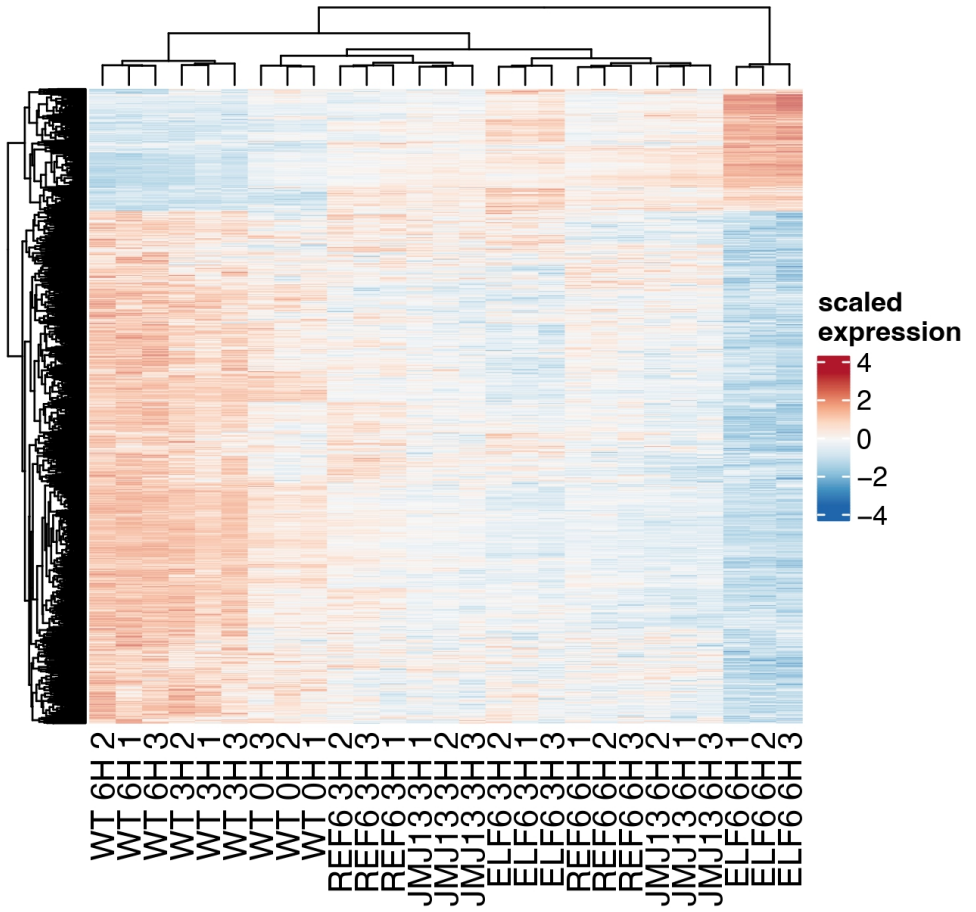


Figure 4.27: Heatmap of differentially expressed genes (DEGs) from RNA-seq data for ELF6-KDM6. The heatmap displays the hierarchical clustering of DEGs (rows) and samples (columns) based on their normalised, log2-transformed gene expression levels. Each cell represents the expression level of DEGS between ELF6-KDM6 and WT at 6h from induction, with colours indicating relative expression: high expression (red), average expression (white), and low expression (blue). DEGs were identified using a log2 fold change threshold of >1 and an adjusted p-value (FDR) < 0.05.

We focused our analysis on the six hour time point, where a greater number of DEGs were identified (heatmaps for the three hour time point are shown in the Appendix, figure 6.16, 6.17 and 6.18). Upon comparing the expression of DEGs for JMJ13, we observed that these same DEGs displayed either increased up-regulation or reduced down-regulation in ELF6 at both three hour and six hour time points (Figure 4.25). A more pronounced example of this trend was observed with REF6 DEGs,

where ELF6 again exhibited stronger expression in up-regulated genes and weaker expression in down-regulated genes at both time points (Figure 4.26). However, when we examined the ELF6 DEGs, we found that their expression levels in REF6 and JMJ13 were not substantially different from those in the wild-type control (Figure 4.27). This observation suggests that while JMJ13 and REF6 DEGs at the six hour time point share commonalities with ELF6, which displays a more pronounced variation in expression compared to the other two demethylases, the DEGs unique to ELF6 are not shared with JMJ13 and REF6 and appear to be more specific to ELF6.

In this study, we generated chimeric proteins by fusing plant Jmj domains from JMJ13, REF6, and ELF6 to the human KDM6B zinc finger domain and introduced them into a human cell line. Surprisingly, these plant Jmj domains were active in human cells, resulting in distinct gene expression patterns. The chimeric proteins demonstrated unique regulatory profiles, potentially due to differences in enzymatic activity, protein-protein interactions, or allosteric effects.

4.3 Discussion

JmjC domain-containing proteins are a diverse family of enzymes that play a critical role in epigenetic regulation of gene expression by removing specific chemical modifications from histones. Different JmjC domain-containing proteins have been shown to have specific preferences for certain histone modifications. In animals, high-resolution crystal structures of several JmjC-containing demethylases helped to elucidate the substrate specificity and demethylation mechanism, which has implications for chromatin function and gene regulation (Kooistra and Helin, 2012). However, the mechanisms that determine substrate specificity and enable these enzymes to discriminate between differential degrees of methylation on the same lysine residue remain largely unclear.

In plants in particular the knowledge about how plant demethylases interact with their peptide substrates is limited. So far only two JmjC proteins have been characterised structurally in plants, so for the majority of them the mechanisms underlying substrate specificity are still poorly understood.

In the course of our investigation into the substrate specificity of REF6 and ELF6, evidence from chromatin data has led to the hypothesis that ELF6 can remove the H3K27me1 mark. This hypothesis contradicts the current consensus on ELF6 substrate specificity, as investigations on ELF6 have shown that it can remove H3K27me3 and H3K27me2, but not H3K27me1 (Crevillén et al., 2014). In general, demethylases belonging to the same family are implicated in targeting similar histone modifications, most likely because they share a conserved catalytic domain and may interact with the same cofactors and other components of the histone modification machinery. However, there are instances in which proteins from the same group can target distinct targets. The KDM4 family of JmjC domain-containing proteins for instance, consists of KDM4A, KDM4B, KDM4C and KDM4D. While all four members of this family have been shown to demethylate H3K9me3, KDM4A and KDM4C have also been shown to demethylate H3K23me1 and H3K23me2 (Berry and Janknecht, 2013). A similar situation occurs in the KDM5 family, which

consists of KDM5A, KDM5B, KDM5C, and KDM5D. While all four family members have been shown to demethylate H3K4me3, KDM5A and KDM5B also demethylate H3K36me3 (Lan and Shi, 2009). There are also instances of substrate specificity related to the number of targetable methylations on the same histone residue. In *Arabidopsis*, one example is the KDM4/JHDM3 subfamily demethylase JMJ13, a relative of REF6 and ELF6. As a result of the resolution of its crystal structure (Zheng et al., 2019) we have a clear understanding of its substrate specificity and peptide binding mechanisms. JMJ13 can demethylate H3K27me3, but it is unable to remove H3K27me2 or H3K27me1. This behaviour within the same subfamily demonstrates that even closely related proteins can have distinct substrate specificities. However, it should be noted that the structures of REF6 and ELF6 are extremely similar, whereas JMJ13 is quite dissimilar. In our research, we generated transgenic plant lines with exchanged Jmj domains between REF6 and ELF6 to investigate potential phenotypic and molecular differences resulting from this domain swap. We measured flowering time for AER;*ref6-5* and found that it only partially compensated for the *ref6-5* mutation, which was similar to what we observed in the ARR;*ref6-5* positive control. Interestingly, when AER was transformed into the col0 background, a delayed flowering phenotype was observed compared to the wild type (WT). It has been demonstrated that REF6 overexpression accelerates flowering in *Arabidopsis* seedlings due to the upregulation of FT, SOC1, and floral homeotic genes (Lu et al., 2011a). If ELF6 possessed the same catalytic activity as REF6 an early flowering phenotype was to be anticipated. In light of the possibility that ELF6 demethylates H3K27me1, the additional demethylation of REF6 targets may have resulted in overcompensation by the mechanism that controls flowering time. Another hypothesis can be that as REF6 acts on floral development at several levels, the expression of our chimeric protein may have perturbed the regular dynamics of the native REF6, either by competing for its role in the many complexes REF6 is involved in or by acting directly on its targets with an unknown effect. This may have resulted in an effective down-regulation and consequently a delay in flowering time.

RNA-seq analysis revealed minor differences in gene expression between AER;*ref6-5* and ARR;*ref6-5* and their respective *ref6-5* backgrounds. However, ARR;*ref6-5* exhibited a smaller difference in the number of DEGs compared to WT, suggesting a higher level of complementation, compared to AER;*ref6-5* which had DEGs had a similar number to *ref6-5*. This was further supported by the expression patterns of ARR;*ref6-5*, which appeared to partially restore the WT expression profile. In contrast, differentially expressed genes in AER;*ref6-5* did not show differential expression between the WT and the *ref6-5* mutant, indicating the presence of novel targets. Additionally, when examining the expression levels of ARR;*ref6-5* and AER;*ref6-5* on REF6 targets, we found that ARR;*ref6-5* lines were able to partially demethylate these targets while AER;*ref6-5* only achieved this to a lesser extent.

The observed partial complementation for both ARR and AER might be attributable to the potential lethality of these constructs when expressed at high levels, given that we employed the constitutive Actin2 promoter. T2 generation seed propagation may have selected lines with weaker expression. Specifically, in the case of AER which displayed lower complementation compared to ARR, qPCR assays revealed that AER expression was 20 times lower than ARR. This suggests that overexpression of REF6 with an ELF6-Jmj domain may be more detrimental, possibly due to differences in catalytic activity. Future experiments should explore the use of alternative promoters to circumvent the emergence of gene silencing or selection for weak gene expression lines in order to further elucidate the effects of exchanging Jmj domains between REF6 and ELF6.

The overexpression of exogenous plant demethylases in human cells has demonstrated that plant demethylases can influence their expression levels. In addition, our studies demonstrated that distinct plant demethylases have distinct effects on human cells, as evidenced by their different expression patterns. Interestingly, ELF6 had a notable effect on KDM6B-target down-regulation, compared to JMJ13 and REF6.

Several factors could account for the observed differences in gene expression patterns. Firstly, intrinsic differences in enzymatic activity or

substrate specificities among the Jmj domains may lead to variations in their ability to modulate H3K27me3 levels at target loci, which in turn results in differential gene expression patterns. Secondly, the plant Jmj domains may have different affinities for interacting with human proteins when fused to the zinc finger domain of KDM6B. These interactions can influence the recruitment of the chimeric proteins to target loci or their association with other chromatin modifiers, thereby impacting gene expression regulation. Additionally, allosteric effects resulting from fusing the plant Jmj domains to the human zinc finger domains could lead to structural changes that affect the catalytic activity, substrate specificity, or protein-protein interactions of the chimeric proteins, ultimately leading to differential gene expression profiles.

The observation that the majority of the differentially expressed genes (DEGs) found for JMJ13, REF6 and ELF6 are down-regulated could be attributed to a combination of factors. One possible explanation is that the chimeric proteins containing the plant Jmj domains preferentially target specific genomic loci, leading to a reduction in H3K27me3 levels primarily at genes that are negatively regulated by this histone modification (Margueron and Reinberg, 2011). As a result, the removal of the repressive mark would yield a higher proportion of down-regulated genes.

Another possibility is that the chimeric proteins indirectly influence the expression of transcriptional repressors, either by modulating their activity or by altering their expression levels, which in turn results in a higher proportion of down-regulated genes among the DEGs (Blackledge et al., 2014). The human cell line used in the study may also contribute to this observation, as its specific transcriptional landscape and endogenous factors, chromatin structure, or interactions with other regulatory elements may lead to the chimeric proteins predominantly affecting the down-regulation of genes (Bannister and Kouzarides, 2011).

Lastly, the chimeric proteins might exert indirect effects on gene expression by altering the activity or expression of other chromatin modifiers or transcription factors, consequently resulting in a higher proportion of down-regulated genes (Sims III et al., 2003).

To better comprehend the underlying causes of these observations, further experiments could be conducted, including ChIP-seq analysis of H3K27me3 marks, can be conducted to pinpoint the genomic loci targeted by the chimeric proteins and evaluate the modification levels at these sites..

4.4 Summary

In this study, the activity of the catalytic domain of the histone demethylases REF6 and ELF6 was investigated. To gain insight into the enzyme activity of the two proteins, we created chimeric constructs and expressed them both in plant and human cell systems. Transgene silencing had a significant impact on results in plants, so we were unable to confirm the validity of our hypothesis. However, the expression of our constructs in human cells revealed disparities in REF6 and ELF6 expression. Additional research is necessary to clarify the enzymatic activity of the two demethylases.

5 Potential role of chromatin modifications in somatic embryogenesis

5.1 Introduction

Somatic embryogenesis (SE) is an artificial process by which an embryo or a plant is developed from single somatic cells. Different kinds of stress such as wounding, high concentrations of plant growth regulators (PGRs) and specific culture media conditions are able to induce SE in cultured explants (Fehér, 2015). These stresses can cause mature plant cells to activate transcriptional cascades which promote cell fate reprogramming and can initiate new developmental programs (Ikeda-Iwai et al., 2003). By undergoing these processes adult cells can de-differentiate into a stem cell-like state that is fundamental for the acquisition of embryogenic potential (Yang and Zhang, 2010). The capacity of inducing the formation of embryos starting from vegetative tissue has great value for both academic and industrial laboratories and the advantages as a plant propagation tool are many (Pais, 2019; Egertsdotter et al., 2019). For example, *in vitro* SE notably shortens the process of clonal propagation over other techniques such as shoot and root regeneration because it allows plantlets to be harvested continuously instead of waiting for new organs to develop. In doing so, SE preserves both the genotype and the ploidy of the selected plants. It is also very effective for regenerating genetically modified plants after transformation. Greater production efficiency and germplasm uniformity conservation have made this technique widely used (Lelu-Walter et al., 2013).

5.1.1 TFs involved in SE

Understanding the physiological and molecular mechanisms by which the induction of SE occurs is therefore a crucial step for its manipulation. Over the past few decades numerous experimental systems have been developed to study various modes of regeneration in *Arabidopsis*. Crucial to this is a knowledge of how cells perceive and respond to inductive cues, as well as how these stimuli modify developmental programs to

reform tissues and organs. Although much progress has been made to understand the molecular basis of SE, the mechanisms underlying early somatic embryo development remain largely mysterious. However, a number of studies on the analysis of proteomes and transcriptomes have led to the molecular identification and functional characterisation of many genes involved in the initiation and development of somatic embryos (Horstman et al., 2017). Most of these genes belong to one of the following four categories: proteins that act in the cell cycle, biosynthesis of PGRs, proteins involved in the signalling pathway, or transcription factors (TFs). Transcription factors in particular play critical roles in regulating the process of somatic embryogenesis. Ectopic expression of specific single TFs can trigger somatic embryogenesis. Representative transcription factors include AP2-domain PLETHORA (PLT) TFs, NF-Y (nuclear factor of the Y box) TF LEAFY COTYLEDON1 (LEC1), B3 TF LEC2, RWP-RK DOMAIN-CONTAINING4 (RKD4)/GROUNDED (GRD), AT-HOOK MOTIF CONTAINING NUCLEAR LOCALIZED15 (AHL15), and WUSCHEL (WUS). All of these TFs can be classified as totipotency-related given their ability to induce SE. Among those, one of the most characterised TFs is LEAFY COTYLEDON 2 (LEC2).

LEC2 is a plant-specific transcription factor belonging to the B3 DNA-binding domain family. LEC2 plays a crucial role in regulating the processes of embryogenesis and seed maturation, as well as the transition from embryonic to post-embryonic development (Lotan et al., 1998). The LEC2 gene is predominantly expressed during embryogenesis, with expression decreasing significantly after germination (Lotan et al., 1998). LEC2 has been shown to regulate various aspects of embryonic development, including embryo patterning, cotyledon identity, and maturation (Stone et al., 2001). In addition, LEC2 controls the accumulation of storage compounds, such as lipids and proteins in seeds (Mendoza et al., 2005). LEC2 acts together with other key regulators of embryogenesis, such as LEAFY COTYLEDON 1 (LEC1), FUSCA3 (FUS3), and ABSCISIC ACID INSENSITIVE 3 (ABI3), to form a complex regulatory network governing the process (Braybrook and Harada, 2008). LEC2 has been found to directly activate the expression of FUS3 and indirectly influence

ABI3 expression (Kroj et al., 2003). Regarding its role in the induction of SE, LEC2 over-expression in various plant species, including *Arabidopsis* and tobacco, has been shown to induce somatic embryo formation, even in the absence of exogenous auxin (Stone et al., 2001; Guo et al., 2013).

Another important TF that has been identified to play a role in somatic embryogenesis in *Arabidopsis* is RWP-RK DOMAIN-CONTAINING PROTEIN 4 (RKD4). RKD4 is a transcription factor belonging to the RWP-RK family, which is involved in the regulation of nitrogen (N)-responsive genes and nitrogen signalling in plants (Guan et al., 2017). RKD4, specifically, has been implicated in playing a role in plant embryogenesis (Waki et al., 2011) and is conserved in plant evolution (Koi et al., 2016).

While much less is known about RKD4 compared to LEC2, it has been reported that overexpression of RKD4 in *Arabidopsis* can induce somatic embryogenesis, similar to LEC2. Moreover, the RKD4 gene is expressed in the egg cell, zygote and early embryos, suggesting its role in early embryonic development (Waki et al., 2011).

Those researchers have also shown that RKD4 overexpression can promote cell proliferation in certain regions of the plant, such as the root meristem and young leaf primordia. In the root, the proliferation induced by RKD4 overexpression occurs in most cell types, differing from the response to hormones, which is confined to the pericycle. However, compared to LEC2, RKD4 appears to act differently in that progression of SE only occurs once transient RKD4 overexpression has been released, whereas continuous overexpression locks the cells into an early embryonic state.

5.1.2 Histone modifications and regeneration

Another layer of regulation has recently emerged as a critical factor in SE is chromatin remodelling. Epigenetic modifications are an essential part of the signalling pathway that leads to changes in the genetic program of the cells and the development of somatic embryos. For instance, some chromatin-modifying enzymes, such as histone deacetylases (HDACs) and histone methyltransferases (HMTs), have been shown to influence somatic

embryogenesis in plants (Tanaka et al., 2008; Gan et al., 2015). Increasing evidence indicates that the deposition of the repressive H3K27me3 marker by PRC2 is one of the epigenetic barriers to the auxin-induced acquisition of embryogenic competence in differentiated somatic cells (Ikeuchi et al., 2015; Mozgová et al., 2017). Interference with this process in *Arabidopsis* results in the formation of callus on the shoot apex or somatic embryos on root hairs. Chemical perturbation or genetic disruption of H3K27me3 deposition is likely to induce somatic embryo formation by de-repressing totipotency-related TF gene loci at the molecular level. In fact, mutations in these PRC2 components have been shown to affect somatic embryogenesis.

For example, CURLY LEAF (CLF) and SWINGER (SWN) mutants exhibit ectopic activation of embryonic genes in seedlings, leading to somatic embryo formation on the shoot apical meristem in the absence of exogenous hormones (Chanvivattana et al., 2004; Bouyer et al., 2011). Similarly, mutations in EMF2 result in the ectopic expression of embryonic genes and the formation of somatic embryos on the shoot apex (Yoshida et al., 2001). These observations suggest that PRC2-mediated gene silencing is crucial for suppressing embryonic traits during post-embryonic development.

Moreover, PRC2 has been implicated in the regulation of genes involved in hormone biosynthesis and signalling pathways which are essential for somatic embryogenesis. For instance, PRC2 components CLF and SWN have been shown to repress the expression of YUCCA genes which are involved in auxin biosynthesis, and auxin-responsive genes, such as PLTs, during post-embryonic development (Wang et al., 2016b).

Researchers have recently proposed, based on joint profiling of chromatin accessibility and gene expression, that the loss of regenerative competence for SE during seed germination is likely to be due to the developmentally regulated removal of the permissive chromatin signature at totipotency-related TF gene loci (Wang et al., 2020b). These transcription factors are more readily re-induced by 2,4-D (a synthetic auxin used in tissue culture to induce SE) in immature zygotic embryos than in post-embryonic explants. This finding explains why immature

zygotic embryos are frequently used as explants for SE. The deeper understanding of the chromatin state dynamics during seed germination and early plant development offers novel perspectives for future research aimed at enhancing somatic embryogenesis induction. By identifying and manipulating key chromatin regulators or epigenetic marks involved in the establishment and maintenance of totipotency, it may be possible to modulate the regenerative competence of plant cells and improve the efficiency of somatic embryogenesis. Furthermore, this knowledge could also facilitate the development of targeted strategies for reprogramming the chromatin state of post-embryonic explants, allowing them to regain totipotency and increase their responsiveness to SE-inducing conditions.

5.1.3 Experiment rationale

The ectopic expression of specific TFs is sufficient to stimulate SE, but the formation of new embryos is restricted to regions in close proximity to the root and shoot meristems. To convert an explant into a somatic embryo, more reprogramming is required the further away the explant is from the zygotic embryo stage. As cells mature, the chromatin associated with embryogenesis-related genes becomes more compact and their expression is strongly repressed. This occurs as a result of chromatin-modifying factors that regulate epigenetic reprogramming and alter gene chromatin states. As mentioned in the previous section, histone methylation on H3K27me3 is primarily responsible for chromatin modifications that inhibit embryo gene expression.

We hypothesised that demethylating H3K27me3 repressed embryonic targets with specific H3K27me3 demethylases would relax the chromatin structure as a result of the removal of the H3K27me3 mark, thereby allowing the transcription of SE-related transcripts. To realise this hypothesis, we proposed a strategy wherein the catalytic domains of three specific H3K27me3 histone demethylases — JMJ13, REF6, and ELF6 — would be fused to embryonic transcription factors. This fusion is anticipated to guide these demethylases directly to the embryonic genes, potentially facilitating their transcription by targeting and relaxing the chromatin structure. In the context of cell reprogramming, histone de-

methylation and transcription factor binding are complex processes that can be interactive and simultaneous (Benveniste et al., 2014; Miyajima et al., 2022). For example, TFs like Oct4 and Sox2, which are important for pluripotency and reprogramming in humans, can bind together to a regulatory element in a cooperative way (Sinha et al., 2023). While our understanding of the synergistic action between histone demethylation and transcription factor binding has advanced significantly in mammalian systems, it remains largely uncharted territory in plants. Nevertheless, the fundamental principles governing gene regulation and chromatin dynamics may well be conserved across kingdoms. Thus, by drawing parallels from human systems, we anticipate that similar mechanisms could be at play in plants, providing a compelling rationale for our experimental approach in this chapter.

5.2 Results

From preceding chapters, we posited that ELF6 and REF6 might exhibit distinct catalytic activities, though both can evidently remove at least two methyl groups. Literature suggests that JMJ13 removes only one methylation (Zheng et al., 2019). Our aim was to assess the influence of these three H3K27me3 demethylases when directed to embryonic gene targets, leveraging the guidance of specific TFs related to embryogenesis. We wished to discern potential phenotypic variations across these demethylation events. For this purpose, we selected two TFs known to stimulate somatic embryogenesis upon overexpression.

For this purpose we fused the Jmj domains of H3K27me3 demethylases REF6, ELF6, and JMJ13 with the coding domains of SE-inducing transcription factors LEC2 and RKD4. We created six fusion constructs containing all possible combinations of Jmj domains and TFs. In constructing our fusion proteins we opted for a long flexible linker to ensure spatial freedom and minimise steric hindrance, thereby preserving the independent functionality of both proteins. Long linkers rich in small and polar amino acids like Glycine and Serine have been shown to provide the necessary flexibility without encouraging aggregation (Klein et al., 2014; Li et al., 2016). As a negative control, we also generated constructs containing only TFs. These constructs were designed with a nuclear localisation signal at the N-terminus for protein import into the nucleus and three HA-tags at the C-terminus for protein detection. All the fusion proteins were under a β-estradiol inducible system. The sequences employed in these experiments were codon-optimised for *Arabidopsis* (Figure 5.1). All the constructs were assembled using the Gateway cloning system and the primers used for the assembly are shown in Figure 2.6.

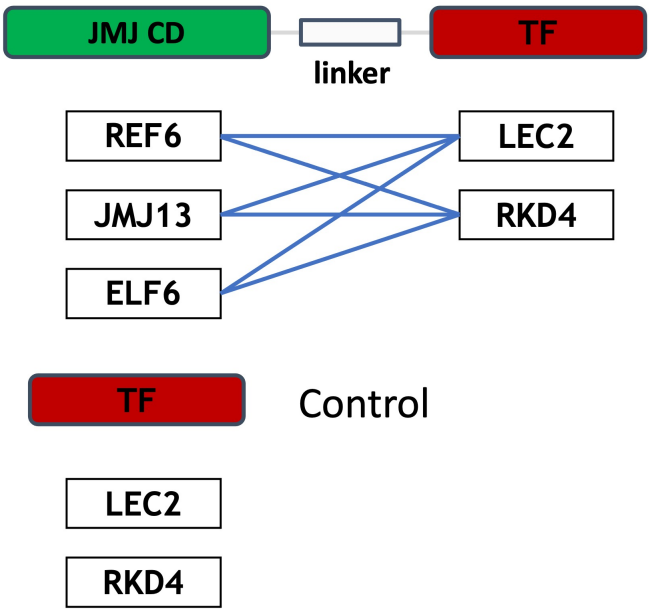


Figure 5.1: Schematic representation of the construct design. The Jmj domains of histone demethylases REF6, ELF6 and JMJ13 were fused with SE-related TFs LEC2 and RKD4 in all combinations. The two domains were connected via a short linker. Constructs containing only the domain of the SE-related TFs were created as controls.

We anticipated that the expression of a single TF (LEC2 or RKD4) would induce callus formation and SE in plant tissue closer to meristems (both apical and root) (Stone et al., 2001; Waki et al., 2011). We hypothesised that when plants express chimeric proteins, even maturely differentiated cells will be able to de-differentiate and form callus. As the activity of the demethylase loosens the severe chromatin repression of SE-related genes, the TFs will strongly promote the transcription of the aforementioned genes (Figure 5.2).

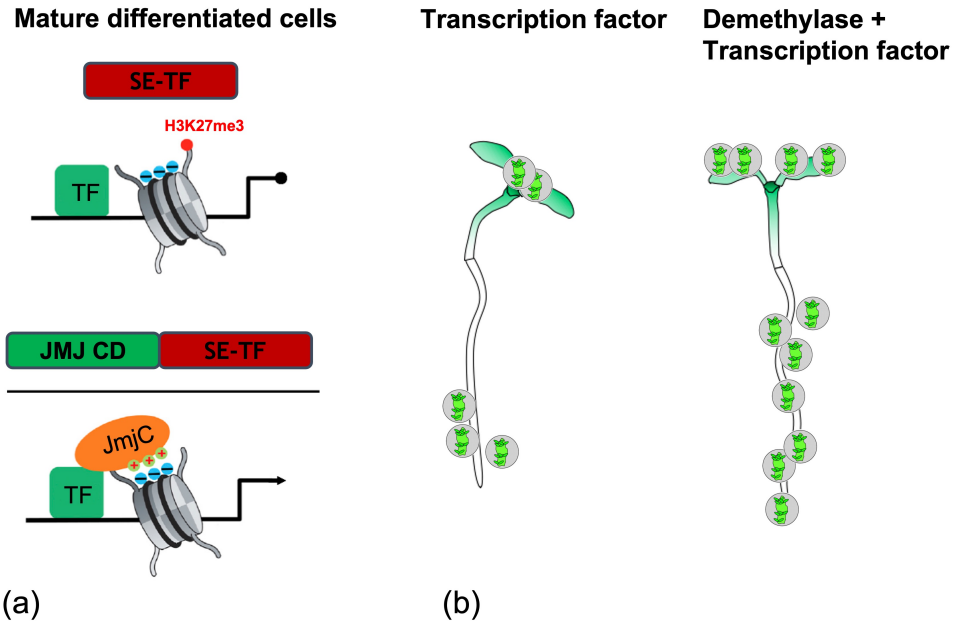


Figure 5.2: Outcomes anticipated for the expression of our fusion constructs (Jmj domain + TF) versus the expression of the TFs alone. (a) illustrates our hypothesis at the DNA level in mature, differentiated cells under both conditions (overexpression of TFs alone on the left and overexpression of Jmj + TF on the right). (b) depicts the expected phenotype of plants under both the above-mentioned conditions. The overexpression of a single SE-TF promotes callus formation and somatic embryogenesis in plant tissue near meristems. When the TF is fused to the demethylase, this phenotype would extend to more mature tissue.

To evaluate the validity of the experimental design, we transfected our constructs into *Arabidopsis* leaf mesophyll protoplasts (see sections 2.18 and 2.19 for details). All of the constructs used for protoplast transfection contained a Ubi-1 constitutive promoter with a high level of expression. The protoplasts were derived from a transgenic line of *Arabidopsis* expressing YFP under the control of the LEC1 promoter (Figure 5.3).

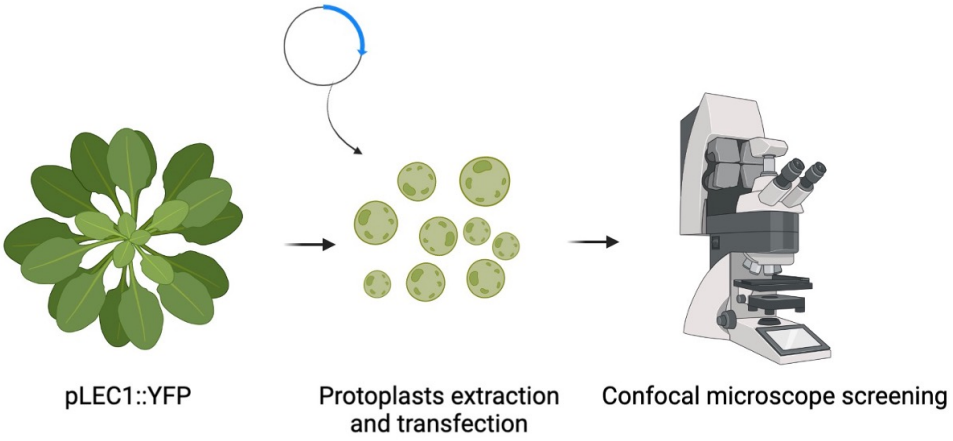


Figure 5.3: Diagram illustrating the experimental method used to examine the expression of our constructs in *Arabidopsis* protoplasts. pLEC1::Transgenic YFP plants were used to obtain protoplasts, which were then transfected with experiment and negative control constructs. With a confocal microscope YFP expression was examined. The expression of YFP was examined using a confocal microscope.

LEC1 is a key regulator of the genes involved in seed development during embryogenesis. Overexpression of LEC2 and RKD4 would activate the network of genes implicated in embryogenesis, including the expression of LEC1 (Boulard et al., 2018; Waki et al., 2011). As a preliminary test to determine the validity of this method, we transfected protoplasts with LEC2 and RKD4 simultaneously. Theoretically, this would result in a high level of LEC1 protein and reporter gene expression. As a control to confirm the efficacy of the protoplast transfection, we transfected the protoplast with a red fluorescent protein-expressing mCherry vector. Confocal microscopy observations detected a fluorescent signal from protoplasts for mCherry expression, but not for YFP. (Figure 5.4). We evaluated each construct and observed similar expression patterns, but no YFP signal was detected.

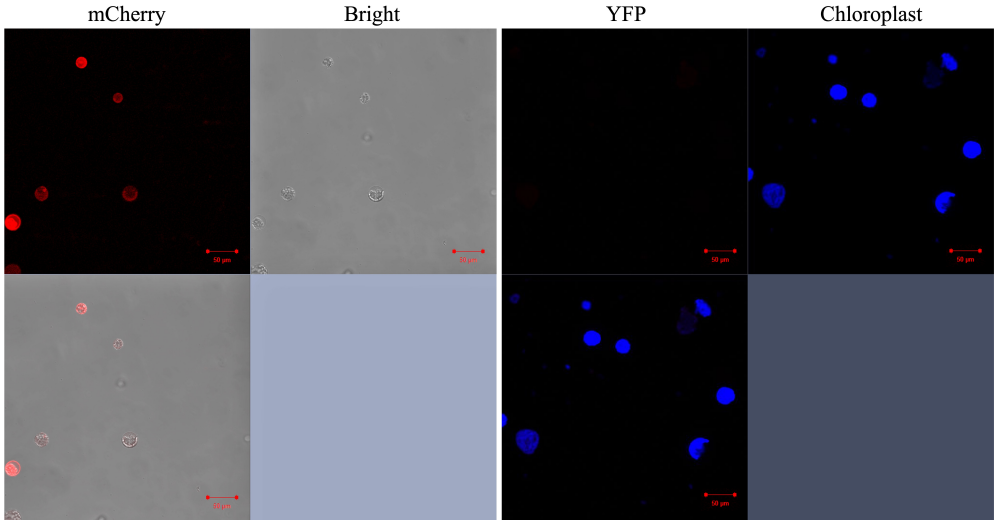


Figure 5.4: Confocal microscope images of protoplasts obtained from *Arabidopsis* leaf mesophyll pLEC1::YFP transgenic line. On the left, protoplasts transfected with mCherry exhibit red fluorescence, demonstrating successful transfection. The mCherry fluorescence serves as a marker to assess the efficiency of the transfection process. On the right, protoplasts transfected with LEC2 and RKD4 do not exhibit YFP fluorescence, suggesting that these conditions do not induce the expression of the LEC1::YFP reporter gene.

Although the preliminary experiment on protoplasts did not produce the expected results—indicative of our transgenes having cascade effects on the embryogenesis network TFs— we proceeded to test the fusion proteins in plants to observe the eventual induction of somatic embryogenesis at the phenotypic level. All the previously obtained construct combinations (fusion proteins and TFs) were introduced into binary vectors, transformed in *Arabidopsis* col0, and propagated across generations (Figure 6.19 in the Appendix).

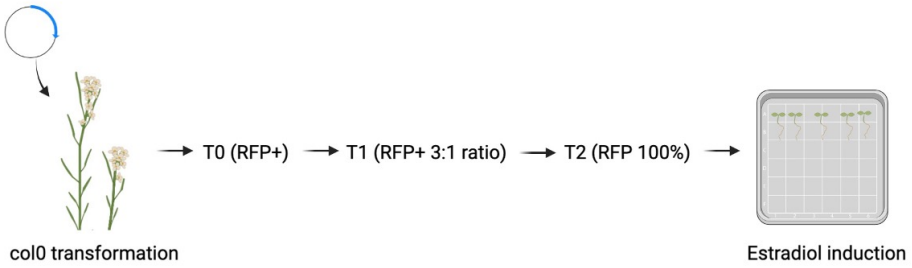


Figure 5.5: Experimental design to examine the impact of our in-planta-based constructs. After obtaining homozygous lines for all constructs, each line was tested in sterile plates with 5 ul of estradiol induction along with a col0 negative control.

Before testing our lines, we tested col0 plants with different concentrations of ß-estradiol on 1/2 MS agar plates and determined that 5 µM was the highest concentration at which the wild-type could develop normally.

20 homozygous lines per construct (LEC2, JMJ13-LEC2, REF6-LEC2, ELF6-LEC2, RKD4, JMJ13-RKD4, REF6-RKD4, and ELF6-RKD4) were then tested. For each line, 10 plants were grown alongside 10 plants of col0 as negative controls on the same plate. The plants were cultivated for seven days in growth chambers under controlled conditions in media containing 5 µm of ß-estradiol (Figure 5.5). We did not observe callus formation or other indications of dedifferentiation in any of the tested lines following transgene induction. However, differences in root length were observed between some of our lines and their respective controls.

To determine whether the observed differences were the result of the induction of the constructs, we analysed the expression levels of our transgenes. We designed two sets of primers for LEC2 and performed RT-PCR on three samples for the LEC2, JMJ13-LEC2, REF6-LEC2, and ELF6-LEC2 constructs (see section 2.2 for details). At least one line per construct expressed the transgene, with ELF6 lines having the highest expression among the samples considered (Figure 5.6). In order to avoid confounding effects from the native LEC2, our primers were designed for the codon-optimised version of LEC2 that we used in our construct, confirming the expression of our transgenes.

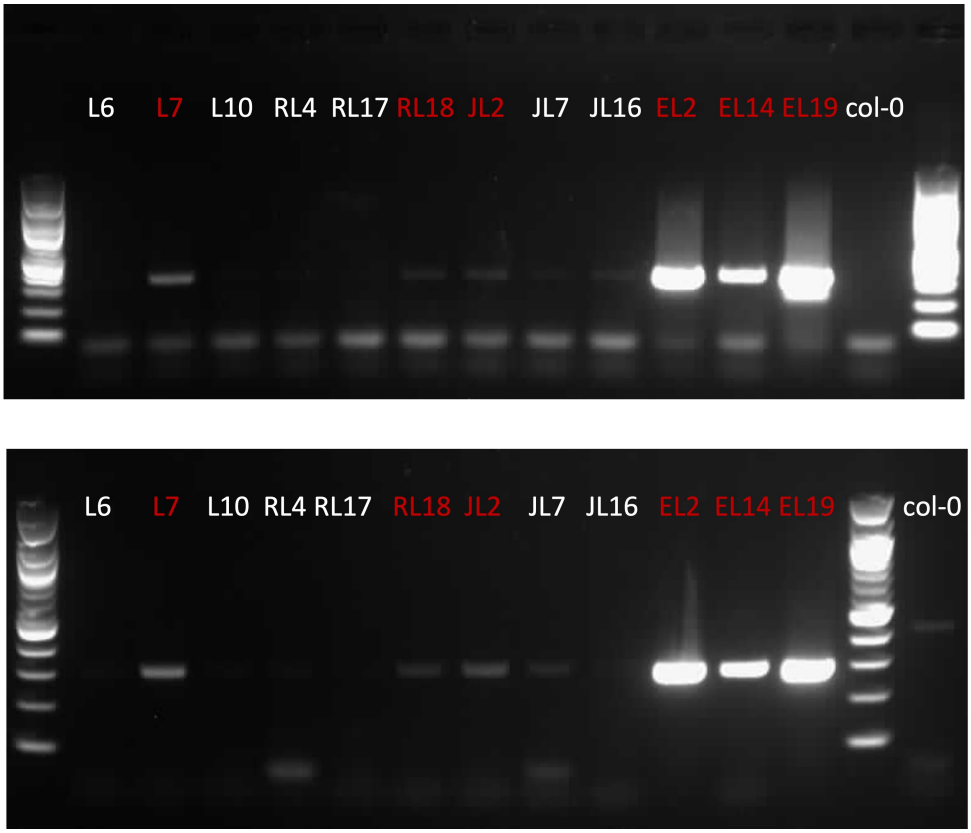


Figure 5.6: Gene expression by RT-PCR assay for LEC2(L); REF6-LEC2 (RL); JMJ13-LEC2 (JL) and ELF6-LEC2 (EL).

We then measured the root length of all the examined plants (experiments and controls) at two time intervals (three and seven days). A comparison of the REF6-RKD4 lines with the WT control at day three is shown in Figure 5.7 (Barplots for all the tested lines are shown in the Appendix, figures 6.20 to 6.27).

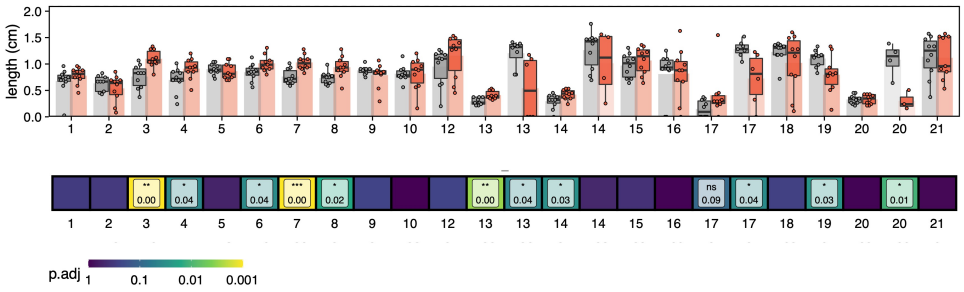


Figure 5.7: The barplot illustrates a comparison of the root length at day three for the REF6-RKD4 lines versus the control col0. On the x-axis, the numbers signify individual lines. For each plate, 10 plants of the test line and 10 of the control col0 were grown Bars in orange represent the respective lines, while those in grey depict the negative control col0. Beneath the plot, p-values are displayed for lines that exhibit a statistically significant difference from the control, as determined by a T-test with a threshold of p-value < 0.05.

A T-test was conducted to identify differences between the experimental and control groups on day three after germination and P-values were determined for each line. REF6-RKD4 was the only RKD4 construct (among JMJ13-RKD4, REF6-RKD4, and ELF6-RKD4) that differed significantly from the TF alone RKD4. In contrast, in the LEC2 constructs (JMJ13-LEC2, REF6-LEC2, and ELF6-LEC2), JMJ13-LEC2 and ELF6-LEC2 had a greater number of lines with a significant p-value (Figure 5.8). Surprisingly, within the RKD4 constructs (RKD4, JMJ13-RKD4, REF6-RKD4, and ELF6-RKD4), the lines that differed significantly from the WT had longer roots than the transcription factor alone. With the LEC2 lines (LEC2, JMJ13-LEC2, REF6-LEC2, and ELF6-LEC2), the root length of the lines with significant p-values was typically shorter than that of the WT. It's worth noting that there was no observed correlation between the root phenotype and the expression level of the transgene.

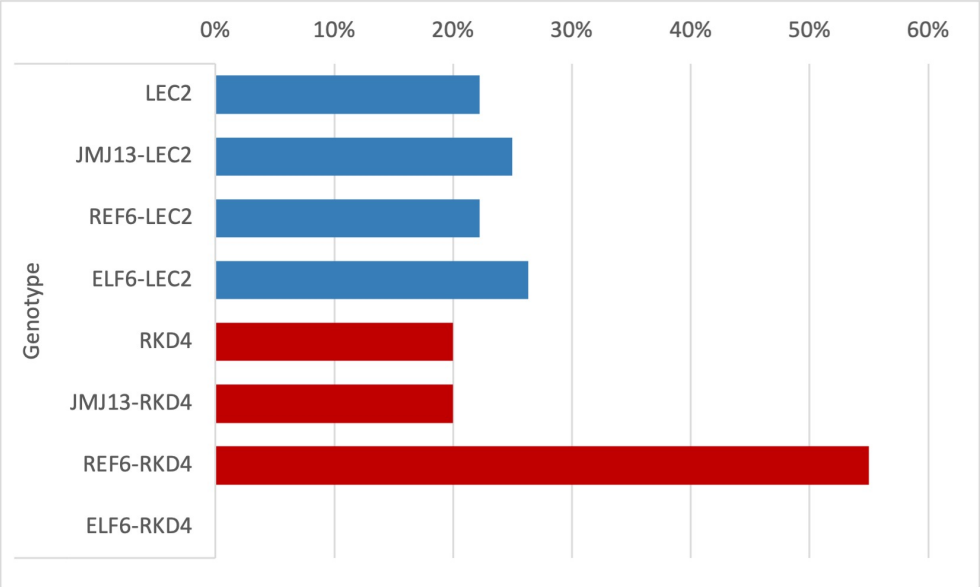


Figure 5.8: The chart shows the percentages of significantly different lines over the total number of lines per construct. LEC2 and RKD4 served as controls against their relative fusion proteins. T-test statistics were used with p-value < 0.05.

After confirming that our transgenes were being expressed, we hypothesised that the quantity of inducer was insufficient to induce SE. The ß-estradiol induction was then repeated in 10 selected lines per construct, with the concentration increased to 30um. We observed that within the RKD4 constructs, the lines that had longer roots than the negative control in a previous assay with an inducer concentration of 5 um now had significantly shorter roots than the WT (Figures 6.20 to 6.23 in the Appendix). Within LEC2 constructs, the p-values of significantly different lines were substantially lower, indicating that the higher inducer concentration exacerbated the short root phenotype (Figures 6.24 to 6.27 in the Appendix). Despite not observing callus formation nor cell de-differentiation in our experiment, we were able to confirm that our constructs had a dose-dependent effect on the phenotypes of the plants, as different concentrations of the inducer contributed to opposite phenotypes in the case of the RKD4 lines and to the accentuation of the phenotype in the case of the LEC2 lines.

5.3 Discussion

Somatic embryogenesis, the process of asexual embryonic development from somatic cells, has garnered significant interest in plant biotechnology due to its potential applications in plant propagation, genetic transformation and germplasm conservation. Somatic embryogenesis has several advantages over other plant propagation methods, such as its high efficiency, rapid clonal multiplication, and the potential to produce genetically uniform populations. These advantages have made somatic embryogenesis an attractive tool for plant biotechnology and agricultural research.

A major breakthrough for somatic embryogenesis would be the development of a reliable and efficient method for inducing somatic embryogenesis in a wide range of plant species and tissues. This would allow for the production of genetically stable clonal plantlets, including those that are currently recalcitrant to somatic embryogenesis.

In recent years *Arabidopsis* protoplasts have been found to be a useful tool for studying the effects of various histone modifications on gene expression in the light of somatic embryogenesis. Despite the fact that we were unable to obtain significant results from our experiment on histone demethylases on protoplasts, other laboratories have used this system to investigate the effects of other histone modifications on cell reprogramming. In a recent study, Sakamoto (Sakamoto et al., 2022) demonstrated that inhibiting the activity of histone acetyltransferases results in a significant decrease in the formation of calluses from *Arabidopsis* leaf mesophyll protoplasts. Histone acetylation enables the transcriptional activation of PLETHORAs, resulting in the induction of their downstream YUCCA1 gene encoding an auxin biosynthesis enzyme. Through the activation of G2/M phase genes mediated by MYB DOMAIN PROTEIN 3-RELATED PROTEINS, Auxin biosynthesis is in turn required for initial cell division. Using a similar approach, Choi (Choi et al., 2023) tested the effects of HDAC inhibitors on cell division and callus proliferation from mesophyll protoplasts of lettuce and tobacco. They speculate that genes that should be blocked by histone deacetylation during the early stage of cell division

are likely to help in later callus formation and shoot regeneration and they suggest that the key mechanism for promoting callus formation and plant regeneration is facilitation of cell division.

Expressing fusion proteins of H3K27me3 demethylase and transcription factors related to somatic embryogenesis we expected various alterations in the regulation of somatic embryogenesis, as both the transcription factors and the demethylase have roles in modulating gene expression and chromatin structure. We anticipated that that the fusion of the H3K27me3 demethylase domain to LEC2 and RKD4 might change the target gene specificity of these transcription factors, potentially affecting the activation of different sets of target genes involved in somatic embryogenesis. In our study, unfortunately we were unsuccessful in stimulating dedifferentiation of the cells expressing the fusion proteins or the transcription factors used on their own. It is plausible that the problem was caused by the ß-estradiol-inducible promoter. Estradiol concentrations between 5 and 10 µM appear to be optimal for inducing gene expression in *Arabidopsis*, according to previous research (Lau et al., 2014; Zuo et al., 2000). As suggested by other studies (Ohira et al., 2017) we also tested a higher concentration of estradiol increased to 30 µM, but we did not observe de-differentiation in our cell lines. We did not, however, test concentrations below 5 µM. A study on the overexpression of the AGAMOUS (AG) gene (Itoh et al., 2000) observed that estradiol concentrations as low as 1 µM induced gene expression. It is conceivable that excessive inducer concentrations inhibited cell de-differentiation. Furthermore, there might have been complexities related to the fusion protein itself. Issues with protein structure, interference between domains, the potential requirement for a linker, or even the need to swap the order of domains could have influenced the results. Further research could investigate the use of lower estradiol concentrations, delve into these protein-related challenges, or explore the use of alternative promoters to improve the efficacy.

Even though we did not observe cell dedifferentiation, we did observe a root length phenotype. The RKD4 constructs (RKD4, JMJ13-RKD4, REF6-RKD4, and ELF6-RKD4) that significantly diverged from the WT had longer roots than the negative control. The root length of the LEC2

lines with significant p-values (LEC2, JMJ13-LEC2, REF6-LEC2, and ELF6-LEC2) was typically shorter than that of the WT. These results indicate that the RKD4 and LEC2 constructs have distinct effects on root length, possibly due to their distinct roles in regulating gene expression. Overexpression of LEC2 and RKD4 transcription factors can influence root development in plants due to their involvement in various aspects of plant growth and embryogenesis.

LEC2 has been shown to modulate auxin and gibberellin biosynthesis and signalling pathways which play crucial roles in root development (Stone et al., 2001). Overexpression of LEC2 may lead to alterations in the balance of hormone signalling, ultimately affecting root growth and development. For instance, overexpressing LEC2 in *Arabidopsis* can cause an increase in auxin levels leading to pleiotropic phenotypes, including changes in root architecture (Braybrook and Harada, 2008). As we observed in our experiments, LEC2 ectopic expression in *Arabidopsis* causes seedlings to have short hypocotyls and unextended roots according to a 2001 study (Stone et al., 2001). In that study they also observed callus-like growth on the cotyledon surfaces of seedlings in that study, which we did not observe.

In another study examining the effects of overexpression of LEC2 in tobacco plants (Guo et al., 2013), it was observed that the transgenic seedlings had significantly longer roots compared to those of the wild type control in contrast to the results of our own study. However, these measurements were taken after a period of 20 days after sowing. The researchers also observed that transgenic plants demonstrated a diminished hypocotyl length. Some studies also observed that LEC2 is involved in lateral root formation by activating the expression of the auxin biosynthesis gene YUCCA4 (YUC4), which in turn promotes the generation of lateral roots in *Arabidopsis* (Tang et al., 2017). However, in our study we had not observed significant difference in lateral roots compared to the wild type control.

In contrast, RKD4 is involved in various developmental processes, including embryogenesis and meristem maintenance. Overexpression of RKD4 has been reported to promote cell differentiation and embryo

development in *Arabidopsis* (Waki et al., 2011). Notably, it has been demonstrated that seedlings overexpressing RKD4 exhibit enhanced cell proliferation in regions typically rich in cycling cells, such as the root meristem and young leaf primordia. In the root, proliferation in response to RKD4-overexpression occurs in most cell types, which differs from the proliferation response to hormones that is confined to the pericycle (Waki et al., 2011). These findings suggest that RKD4 overexpression can directly impact root development by stimulating cell proliferation in the root meristem and other proliferative regions.

In summary, overexpression of LEC2 and RKD4 transcription factors can potentially influence root development through their roles in embryogenesis and plant growth regulation. While LEC2 overexpression has been shown to affect root development in a species-specific manner, as demonstrated in transgenic tobacco plants (Guo et al., 2013), and influence lateral root formation in *Arabidopsis* (Tang et al., 2017), RKD4 overexpression has been shown to promote cell proliferation in the root meristem and other proliferative regions thus affecting root growth.

While the effects of overexpressing the LEC2 and RKD4 transcription factors on root development are evident from our study and the literature, it's noteworthy to mention the limited or absent observable influence of the histone demethylase domains in our fusion constructs. One would anticipate that coupling these demethylases with the transcription factors might enhance or modify the resulting phenotypes, owing to the potential chromatin accessibility changes brought about by demethylation. However, our observations suggest either that the histone demethylase domains might not have been functionally active in our constructs, or that the targets of the transcription factors were already accessible, rendering the demethylase action redundant. Further investigations could be directed towards validating the functionality of the demethylase domains in these constructs and understanding the intricacies of chromatin dynamics in this context.

5.4 Summary

Somatic embryogenesis is an attractive tool for plant biotechnology due to its high efficiency, rapid clonal multiplication and potential to produce genetically uniform populations. In our study, ectopic expression of somatic embryogenesis-related transcription factors fused to H3K27me3 histone demethylases in *Arabidopsis* had not lead to tissue dedifferentation but had had distinct effects on root length, possibly due to their roles in regulating gene expression. LEC2 and RKD4 transcription factors in combination with H3K27me3 histone demethylases, can influence root development through their roles in embryogenesis and plant growth regulation.

6 General Discussion

In this thesis, we have investigated the roles of histone demethylases RELATIVE OF EARLY FLOWERING 6 (REF6) and EARLY FLOWERING 6 (ELF6) in plant development and their contribution to the regulation of gene expression through chromatin modifications. Histone demethylases are enzymes that remove methyl groups from histone proteins, thereby modulating chromatin structure and influencing gene expression. REF6 and ELF6 are two Jumonji C (JmjC) domain-containing histone demethylases in plants which specifically target the repressive histone modification H3K27me3 and play crucial roles in diverse developmental processes. Although REF6 and ELF6 possess remarkably similar protein structures they display contrasting flowering phenotypes upon mutation (Noh et al., 2004). The primary aim of this thesis has been to conduct an in-depth comparative analysis of these two proteins, examining both their structural and functional attributes to elucidate the molecular basis for their phenotypic differences.

In our analysis we observed that the evolution of REF6 and ELF6 diverged at the emergence of flowering plants. It is possible that ELF6 has evolved to acquire a specific function in the regulation of flowering, a key aspect in the diversification and success of angiosperms (Irish, 2010; Theißen and Melzer, 2007). Histone modifications, including those mediated by histone demethylases like REF6 and ELF6, have been shown to play crucial roles in plant development and the regulation of gene expression (Berr et al., 2011; Zhang et al., 2015). This suggests that the diversification of histone modifiers could have contributed to the evolution of plants.

Differences in the zinc-finger domains of REF6 and ELF6 may affect their target recognition, leading to distinct genomic targets and regulatory roles (Cui et al., 2016a). Moreover, differences in the intrinsically disordered regions (IDRs) and their molecular features could impact protein-protein interactions, further contributing to functional diversification (Wright and Dyson, 2015; Katuwawala et al., 2019).

Our study also identified a shared conserved domain between REF6 and ELF6 which might play a fundamental role in their common functions. Additionally, we observed domain loss in both proteins which could have led to functional diversification during their evolution (Nasir et al., 2014). Future studies should validate the predicted protein interactions experimentally using techniques such as co-immunoprecipitation, yeast two-hybrid assays, or bimolecular fluorescence complementation (Fields and Song, 1989; Kerppola, 2006; Tang et al., 2010).

Furthermore, it would be interesting to investigate the roles of the conserved domains shared between REF6 and ELF6 and understand the functions of the lost domains. This information could provide valuable insights into the molecular mechanisms driving the diversification of these histone demethylases and their functional roles in plant development and evolution.

We also explored the functional roles of REF6 and ELF6 catalytic activity in our study. Current literature supports the notion that their enzymatic activities are similar as both proteins are known to demethylate H3K27me3 and H3K27me2 but not H3K27me1 (Lu et al., 2011a; Crevillén et al., 2014). However, our chromatin data suggested a possible divergence between ELF6 and REF6, with the former potentially capable of removing H3K27me1 as well. To investigate this hypothesis we designed chimeric proteins by shuffling the catalytic JmjC domain of REF6 and ELF6, aiming to reveal any differences in their activity within a plant system.

Despite the limitations encountered in our study, our results indicated notable differences in REF6 and ELF6 transgene expression. The expression level of the REF6 transgene was found to be 10 to 20 times higher than that of the ELF6 chimeric protein. This observation suggests that ELF6 overexpression in REF6 target genes could be detrimental to plants resulting in the survival of only those transgenic plants with low transgene expression. This could be due to ELF6 triggering a more robust gene expression compared to REF6.

Correspondingly, in the flowering time experiments, the REF6 construct showed a stronger complementation effect than ELF6 (Figure 4.12). Specifically, two REF6 lines presented statistically significant differences

from the negative control *ref6-5*, whereas only one ELF6 line showed such a deviation. Additionally, RNA-seq analysis revealed that the gene expression profile in the REF6 lines more closely resembled that of the WT (Figures 4.19, 4.20, and 4.21). This evidence suggests that the higher expression levels of REF6 may have contributed to its superior complementation efficacy compared to ELF6.

Another challenge we faced was the silencing of transgenes, potentially due to excessive transgene expression which might be related to the use of the Actin2 promoter. Future studies could address this issue by utilising alternative promoters that are either weaker or feature inducible expression (Guihur et al., 2023) or by employing silencing-free lines as backgrounds, such as *rdr6-11* (Schwach et al., 2005). The *rdr6-11* *Arabidopsis* mutant exhibits a deficiency in post-transcriptional gene silencing of sense transgenes. This means that in these mutants, the mechanism that typically "silences" or reduces the expression of certain genes is impaired.

Our findings also shed light on the potential interaction between plant demethylases and human cells. We demonstrated that the overexpression of exogenous plant demethylases in human cells can influence their gene expression levels, with distinct plant demethylases exhibiting unique effects on human cells. Notably, ELF6 appeared to have a significant impact on gene down-regulation. This observation raises intriguing questions about the evolutionary conservation of histone demethylase functions across species and the potential for cross-species applications in future research.

Overall, this study challenges the current understanding of the roles of histone catalytic domains in REF6 and ELF6. However, further investigation is warranted to elucidate the precise enzymatic activity of these demethylases using techniques such as in vitro histone peptide demethylation assays to assess enzyme activity in living cells.

Additionally, our findings open up new avenues for exploring the potential applications of plant demethylases in human cells which may contribute to advancements in both plant biology and human health research.

We also aimed to investigate the potential impact of expressing fusion proteins of H3K27me3 demethylases and transcription factors related to somatic embryogenesis, such as LEC2 and RKD4. Several constructs featuring the demethylase domain fused to transcription factors exhibited a greater number of lines with notable differences compared to the wild type than lines expressing only the transcription factor. This observation was especially notable in constructs like ELF6-LEC2 and REF6-RKD4, which showed a difference in lines diverging from the TF-only negative control by 5% and 30%, respectively. This suggests that the presence of the demethylase domain in these fusion proteins may enhance or alter their regulatory functions, leading to more pronounced phenotypic differences compared to the transcription factors alone. One limitation of this study could be the use of the ß-estradiol-inducible promoter as it may have affected the dedifferentiation of cells expressing fusion proteins or transcription factors. Future research could explore the use of lower estradiol concentrations or alternative promoters to improve efficacy. Further studies are needed to better understand the molecular mechanisms behind these observations and to optimise the expression of these fusion proteins for somatic embryogenesis applications. Overexpressing these demethylases in combination with embryogenesis-related transcription factors could result in a synergistic effect on target gene activation, enhancing the efficiency of somatic embryogenesis induction. This could lead to more efficient methods for plant propagation, genetic transformation, and germplasm conservation, especially for species like cereals and grapevine that are currently recalcitrant to somatic embryogenesis.

In summary, our thesis has shed light on the complexity of histone demethylase functions and their roles in plant development, evolution, and somatic embryogenesis. We hope that our work will serve as a foundation for future research in this area and inspire new strategies for advancing plant biotechnology and agricultural research.

References

Agger, K., Cloos, P. A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A. E., and Helin, K. (2007). Utx and jmjd3 are histone h3k27 demethylases involved in hox gene regulation and development. *Nature*, 449(7163):731–734.

Albert, V. A., Barbazuk, W. B., dePamphilis, C. W., Der, J. P., Leebens-Mack, J., Ma, H., Palmer, J. D., Rounsley, S., Sankoff, D., et al. (2013). The amborella genome and the evolution of flowering plants. *Science*, 342(6165):1241089.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2014). *Molecular Biology of the Cell*. Garland Science, 6 edition.

Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van't Veer, L. J., Berns, A., and van Lohuizen, M. (1997). Identification of bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes & development*, 11(2):226–240.

Allis, C. D. and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, 17(8):487–500.

Allshire, R. C. and Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology*, 19(4):229–244.

Almouzni, G. and Cedar, H. (2016). Maintenance of epigenetic information. *Cold Spring Harbor Perspectives in Biology*, 8(5):a019372.

Antunez-Sanchez, J., Naish, M., Ramirez-Prado, J. S., Ohno, S., Huang, Y., Dawson, A., Opassathian, K., Manza-Mianza, D., Ariel, F., Raynaud, C., et al. (2020). A new role for histone demethylases in the maintenance of plant genome integrity. *Elife*, 9:e58533.

Bah, A. and Forman-Kay, J. D. (2016). Modulation of intrinsically disordered protein function by post-translational modifications. *Journal of Biological Chemistry*, 291(13):6696–6705.

Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, 21(3):381–395.

Benveniste, D., Sonntag, H.-J., Sanguinetti, G., and Sproul, D. (2014). Transcription factor binding predicts histone modifications in human cell lines. *Proceedings of the National Academy of Sciences*, 111(37):13367–13372.

Bergink, S., Salomons, F. A., Hoogstraten, D., Groothuis, T. A., de Waard, H., Wu, J., Yuan, L., Citterio, E., Houtsmuller, A. B., Neefjes, J., et al. (2006). Dna damage triggers nucleotide excision repair-dependent monoubiquitylation of histone h2a. *Genes & development*, 20(10):1343–1352.

Bergmüller, E., Gehrig, P. M., and Gruissem, W. (2007). Characterization of post-translational modifications of histone h2b-variants isolated from arabidopsis thaliana. *Journal of proteome research*, 6(9):3655–3668.

Bernatavichute, Y. V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S. E. (2008). Genome-wide association of histone h3 lysine nine methylation with chg dna methylation in arabidopsis thaliana. *PLoS One*, 3(9):e3156.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2):315–326.

Berr, A., McCallum, E. J., Alioua, A., Heintz, D., Heitz, T., and Shen, W.-H. (2009). Arabidopsis histone methyltransferase set domain group8 mediates induction of the jasmonate/ethylene pathway genes in plant defense response to necrotrophic fungi. *Plant Physiology*, 151(3):1070–1084.

Berr, A., Shafiq, S., and Shen, W.-H. (2011). Histone modifications in transcriptional activation during plant development. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1809(10):567–576.

Berry, W. L. and Janknecht, R. (2013). Kdm4/jmjd2 histone demethylases: Epigenetic regulators in cancer cellsrole of kdm4/jmjd2 proteins in cancer. *Cancer research*, 73(10):2936–2942.

Bittrich, S., Burley, S. K., and Rose, A. S. (2020). Real-time structural motif searching in proteins using an inverted index strategy. *PLoS computational biology*, 16(12):e1008502.

Black, J. C., Van Rechem, C., and Whetstine, J. R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Molecular cell*, 48(4):491–507.

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant prc1 complex-dependent h2a ubiquitylation drives prc2 recruitment and polycomb domain formation. *Cell*, 157(6):1445–1459.

Blasi, T., Feller, C., Feigelman, J., Hasenauer, J., Imhof, A., Theis, F. J., Becker, P. B., and Marr, C. (2016). Combinatorial histone acetylation patterns are generated by motif-specific reactions. *Cell systems*, 2(1):49–58.

Bonnet, J., Boichenko, I., Kalb, R., Le Jeune, M., Maltseva, S., Pieropan, M., Finkl, K., Fierz, B., and Müller, J. (2022). Pr-dub preserves polycomb repression by preventing excessive accumulation of h2aub1, an antagonist of chromatin compaction. *Genes & Development*, 36(19-20):1046–1061.

Borg, M., Jacob, Y., Susaki, D., LeBlanc, C., Buendía, D., Axelsson, E., Kawashima, T., Voigt, P., Boavida, L. C., Becker, J. D., et al. (2020). Targeted reprogramming of h3k27me3 resets epigenetic memory in plant paternal chromatin. *Nature cell biology*, 22(6):621–629.

Borg, M., Papareddy, R. K., Dombey, R., Axelsson, E., Nodine, M. D., Twell, D., and Berger, F. (2021). Epigenetic reprogramming rewires transcription during the alternation of generations in arabidopsis. *Elife*, 10:e61894.

Boulard, C., Thévenin, J., Tranquet, O., Laporte, V., Lepiniec, L., and Dubreucq, B. (2018). Lec1 (nf-yb9) directly interacts with lec2 to control gene expression in seed. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1861(5):443–450.

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., Goodrich, J., Renou, J.-P., Grini, P. E., Colot, V., et al. (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS genetics*, 7(3):e1002014.

Bratzel, F., López-Torrejón, G., Koch, M., Del Pozo, J. C., and Calonje, M. (2010). Keeping cell identity in arabidopsis requires prc1 ring-finger homologs that catalyze h2a monoubiquitination. *Current biology*, 20(20):1853–1859.

Braybrook, S. A. and Harada, J. J. (2008). Lecs go crazy in embryo development. *Trends in plant science*, 13(12):624–630.

Butenko, Y. and Ohad, N. (2011). Polycomb-group mediated epigenetic mechanisms through plant evolution. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1809(8):395–406.

Byvoet, P., Shepherd, G., Hardin, J., and Noland, B. (1972). The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Archives of biochemistry and biophysics*, 148(2):558–567.

Calonje, M., Sanchez, R., Chen, L., and Sung, Z. R. (2008). Embryonic flower1 participates in polycomb group–mediated ag gene silencing in arabidopsis. *The Plant Cell*, 20(2):277–291.

Cardoso, C., Mignon, C., Hetet, G., Grandchamps, B., Fontes, M., and Colleaux, L. (2000). The human ezh2 gene: genomic organisation and revised mapping in 7q35 within the critical region for malignant myeloid disorders. *European Journal of Human Genetics*, 8(3):174–180.

Casas-Delucchi, C. S., van Bemmel, J. G., Haase, S., Herce, H. D., Nowak, D., Meilinger, D., Stear, J. H., Leonhardt, H., and Cardoso, M. C. (2012).

Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. *Nucleic acids research*, 40(1):159–169.

Cavalli, G. and Misteli, T. (2013). Functional implications of genome topology. *Nature Structural & Molecular Biology*, 20(3):290–299.

Chang, S., Yim, S., and Park, H. (2019). The cancer driver genes idh1/2, jarid1c/kdm5c, and utx/kdm6a: crosstalk between histone demethylation and hypoxic reprogramming in cancer metabolism. *Experimental & molecular medicine*, 51(6):1–17.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z. R., and Goodrich, J. (2004). Interaction of polycomb-group proteins controlling flowering inarabidopsis. *Development*, 131(21):5263–5276.

Chen, D., Molitor, A., Liu, C., and Shen, W.-H. (2010). The arabidopsis prc1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. *Cell research*, 20(12):1332–1344.

Chen, L.-J., Xu, X.-Y., Zhong, X.-D., Liu, Y.-J., Zhu, M.-H., Tao, F., Li, C.-Y., She, Q.-S., Yang, G.-J., and Chen, J. (2023). The role of lysine-specific demethylase 6a (kdm6a) in tumorigenesis and its therapeutic potentials in cancer therapy. *Bioorganic Chemistry*, page 106409.

Chhun, T., Chong, S. Y., Park, B. S., Wong, E. C. C., Yin, J.-L., Kim, M., and Chua, N.-H. (2016). Hsi2 repressor recruits med13 and hda6 to down-regulate seed maturation gene expression directly during arabidopsis early seedling growth. *Plant and Cell Physiology*, 57(8):1689–1706.

Choi, S. H., Ahn, W. S., Lee, M. H., Jin, D. M., Lee, A., Jie, E. Y., Ju, S. J., Ahn, S. J., and Kim, S. W. (2023). Effects of tsa, nab, aza in lactuca sativa l. protoplasts and effect of tsa in nicotiana benthamiana protoplasts on cell division and callus formation. *Plos one*, 18(2):e0279627.

Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., and Mann, M. (2014). The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature reviews Molecular cell biology*, 15(8):536–550.

Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., and Schofield, C. J. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *Journal of inorganic biochemistry*, 100(4):644–669.

Costa, S. and Dean, C. (2019). Storing memories: the distinct phases of polycomb-mediated silencing of arabidopsis flc. *Biochemical Society Transactions*, 47(4):1187–1196.

Couture, J.-F., Collazo, E., Ortiz-Tello, P. A., Brunzelle, J. S., and Trievel, R. C. (2007). Specificity and mechanism of jmjd2a, a trimethyllysine-specific histone demethylase. *Nature structural & molecular biology*, 14(8):689–695.

Crevillén, P., Yang, H., Cui, X., Greeff, C., Trick, M., Qiu, Q., Cao, X., and Dean, C. (2014). Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. *Nature*, 515(7528):587–590.

Cui, D., Zhao, J., Jing, Y., Fan, M., Liu, J., Wang, Z., Xin, W., and Hu, Y. (2016a). The arabidopsis idd14, idd15, and idd16 cooperatively regulate lateral organ morphogenesis and gravitropism by promoting auxin biosynthesis and transport. *PLoS Genetics*, 12(9):e1006383.

Cui, X., Lu, F., Qiu, Q., Zhou, B., Gu, L., Zhang, S., Kang, Y., Cui, X., Ma, X., Yao, Q., et al. (2016b). Ref6 recognizes a specific dna sequence to demethylate h3k27me3 and regulate organ boundary formation in arabidopsis. *Nature genetics*, 48(6):694–699.

De Santa, F., Totaro, M. G., Prosperini, E., Notarbartolo, S., Testa, G., and Natoli, G. (2007). The histone h3 lysine-27 demethylase jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *cell*, 130(6):1083–1094.

Disfani, F. M., Hsu, W.-L., Mizianty, M. J., Oldfield, C. J., Xue, B., Dunker, A. K., Uversky, V. N., and Kurgan, L. (2012). Morfpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics*, 28(12):i75–i83.

Durr, J., Papareddy, R., Nakajima, K., and Gutierrez-Marcos, J. (2018). Highly efficient heritable targeted deletions of gene clusters and non-coding regulatory regions in arabidopsis using crispr/cas9. *Scientific reports*, 8(1):4443.

Egertsdotter, U., Ahmad, I., and Clapham, D. (2019). Automation and scale up of somatic embryogenesis for commercial plant production, with emphasis on conifers. *Frontiers in Plant Science*, 10.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PloS one*, 3(11):e3647.

Farrona, S., Hurtado, L., Bowman, J. L., and Reyes, J. C. (2004). The arabidopsis thaliana snf2 homolog atbrm controls shoot development and flowering. *Development (Cambridge, England)*, 131(20):4965–75.

Fehér, A. (2015). Somatic embryogenesis—stress-induced remodeling of plant cell fate. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1849(4):385–402.

Feinberg, A. P., Koldobskiy, M. A., and Göndör, A. (2016). Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nature Reviews Genetics*, 17(5):284–299.

Felsenfeld, G. (2014). A brief history of epigenetics. *Cold Spring Harbor perspectives in biology*, 6(1):a018200.

Fields, S. and Song, O.-k. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246.

Franz, H., Mosch, K., Soeroes, S., Urlaub, H., and Fischle, W. (2009). Multimerization and h3k9me3 binding are required for cdyl1b heterochromatin association. *Journal of Biological Chemistry*, 284(50):35049–35059.

Fujisawa, T. and Filippakopoulos, P. (2017). Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nature reviews Molecular cell biology*, 18(4):246–262.

Gaj, M. D. (2004). Factors influencing somatic embryogenesis induction and plant regeneration with particular reference to arabidopsis thaliana (l.) heynh. *Plant Growth Regulation*, 43:27–47.

Gan, E.-S., Xu, Y., and Ito, T. (2015). Dynamics of h3k27me3 methylation and demethylation in plant development. *Plant signaling & behavior*, 10(9):e1027851.

Gan, E.-S., Xu, Y., Wong, J.-Y., Geraldine Goh, J., Sun, B., Wee, W.-Y., Huang, J., and Ito, T. (2014). Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of flc in arabidopsis. *Nature communications*, 5(1):5098.

Garton, M., Najafabadi, H. S., Schmitges, F. W., Radovani, E., Hughes, T. R., and Kim, P. M. (2015). A structural approach reveals how neighbouring c2h2 zinc fingers influence dna binding specificity. *Nucleic acids research*, 43(19):9147–9157.

Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E. M., and Coupland, G. (1997). A polycomb-group gene regulates homeotic gene expression in arabidopsis. *Nature*, 386(6620):44–51.

Greer, E. L. and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, 13(5):343–357.

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M. A., and Gagliano, W. B. (1998). Maternal control of embryogenesis by medea, a polycomb group gene in arabidopsis. *Science*, 280(5362):446–450.

Guan, P., Ripoll, J.-J., Wang, R., Vuong, L., Bailey-Steinitz, L. J., Ye, D., and Crawford, N. M. (2017). Interacting tcp and nlp transcription factors control plant responses to nitrate availability. *Proceedings of the National Academy of Sciences*, 114(9):2419–2424.

Guihur, A., Bourgine, B., Rebeaud, M. E., and Goloubinoff, P. (2023). Design of an arabidopsis thaliana reporter line to detect heat-sensing and signaling mutants. *bioRxiv*, pages 2023–03.

Guo, F., Liu, C., Xia, H., Bi, Y., Zhao, C., Zhao, S., Hou, L., Li, F., and Wang, X. (2013). Induced expression of atlec1 and atlec2 differentially promotes somatic embryogenesis in transgenic tobacco plants. *PLoS ONE*, 8(8):e71714.

Happel, N. and Doenecke, D. (2009). Histone h1 and its isoforms: contribution to chromatin structure and function. *Gene*, 431(1-2):1–12.

He, C., Chen, X., Huang, H., and Xu, L. (2012). Reprogramming of h3k27me3 is critical for acquisition of pluripotency from cultured arabidopsis tissues. *PLoS genetics*, 8(8):e1002911.

Hennig, L., Taranto, P., Walser, M., Schönrock, N., and Gruissem, W. (2003). Arabidopsis msi1 is required for epigenetic maintenance of reproductive development. *Development (Cambridge, England)*, 130(12):2555–65.

Hisanaga, T., Yamaoka, S., Kawashima, T., Higo, A., Nakajima, K., Araki, T., Kohchi, T., and Berger, F. (2019). Building new insights in plant gametogenesis from an evolutionary perspective. *Nature Plants*, 5(7):663–669.

Honda, B., Candido, P., and Dixon, G. (1975). Histone methylation. its occurrence in different cell types and relation to histone h4 metabolism in developing trout testis. *Journal of Biological Chemistry*, 250(22):8686–8689.

Hopkinson, R. J. (2013). Mechanistic studies on histone demethylases and related enzymes.

Horstman, A., Bemer, M., and Boutilier, K. (2017). A transcriptional view on somatic embryogenesis. *Regeneration*, 4(4):201–216.

Horton, J. R., Upadhyay, A. K., Qi, H. H., Zhang, X., Shi, Y., and Cheng, X. (2010). Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nature structural & molecular biology*, 17(1):38–43.

Hou, X., Zhou, J., Liu, C., Liu, L., Shen, L., and Yu, H. (2014). Nuclear factor y-mediated h3k27me3 demethylation of the soc1 locus orchestrates flowering responses of arabidopsis. *Nature communications*, 5(1):4601.

Houben, A., Kumke, K., Nagaki, K., and Hause, G. (2011). Cenh3 distribution and differential chromatin modifications during pollen development in rye (secale cereale l.). *Chromosome research*, 19:471–480.

Hu, Q., Jin, Y., Shi, H., and Yang, W. (2014). Gmfld, a soybean homolog of the autonomous pathway gene flowering locus d, promotes flowering in arabidopsis thaliana. *BMC Plant Biology*, 14:1–12.

Huang, C., Xiang, Y., Wang, Y., Li, X., Xu, L., Zhu, Z., Zhang, T., Zhu, Q., Zhang, K., Jing, N., et al. (2010). Dual-specificity histone demethylase kiaa1718 (kdm7a) regulates neural differentiation through fgf4. *Cell research*, 20(2):154–165.

Hyun, Y., Richter, R., Vincent, C., Martinez-Gallegos, R., Porri, A., and Coupland, G. (2016). Multi-layered regulation of spl15 and cooperation with soc1 integrate endogenous flowering pathways at the arabidopsis shoot meristem. *Developmental cell*, 37(3):254–266.

Ikeda-Iwai, M., Umehara, M., Satoh, S., and Kamada, H. (2003). Stress-induced somatic embryogenesis in vegetative tissues of arabidopsis thaliana. *The Plant Journal*, 34(1):107–114.

Ikeuchi, M., Iwase, A., Rymen, B., Harashima, H., Shibata, M., Ohnuma, M., Breuer, C., Morao, A. K., de Lucas, M., De Veylder, L., et al. (2015). Prc2 represses dedifferentiation of mature somatic cells in arabidopsis. *Nature Plants*, 1(7):1–7.

Irish, V. F. (2010). The flowering of arabidopsis flower development. *The Plant Journal*, 61(6):1014–1028.

Itoh, J.-I., Kitano, H., Matsuoka, M., and Nagato, Y. (2000). Shoot organization genes regulate shoot apical meristem organization and the pattern of leaf primordium initiation in rice. *The Plant Cell*, 12(11):2161.

Jenuwein, T. and Allis, C. D. (2001). Translating the histone code. *Science*, 293(5532):1074–1080.

Jeong, J.-H., Song, H.-R., Ko, J.-H., Jeong, Y.-M., Kwon, Y. E., Seol, J. H., Amasino, R. M., Noh, B., and Noh, Y.-S. (2009). Repression of flowering locus t chromatin by functionally redundant histone h3 lysine 4 demethylases in arabidopsis. *PLoS one*, 4(11):e8033.

Jiang, D., Kong, N. C., Gu, X., Li, Z., and He, Y. (2011). Arabidopsis compass-like complexes mediate histone h3 lysine-4 trimethylation to control floral transition and plant development. *PLoS genetics*, 7(3):e1001330.

Jiang, D., Yang, W., He, Y., and Amasino, R. M. (2007). Arabidopsis relatives of the human lysine-specific demethylase1 repress the expression of fwa and flowering locus c and thus promote the floral transition. *The Plant Cell*, 19(10):2975–2987.

Joseph, F. (2004). *Inferring phylogenies*, volume 2. Sinauer associates Sunderland, MA.

Karytinos, A., Forneris, F., Profumo, A., Ciossani, G., Battaglioli, E., Binda, C., and Mattevi, A. (2009). A novel mammalian flavin-dependent histone demethylase. *Journal of Biological Chemistry*, 284(26):17775–17782.

Katuwawala, A., Peng, Z., Yang, J., and Kurgan, L. (2019). Computational prediction of morfs, short disorder-to-order transitioning protein binding regions. *Computational and Structural Biotechnology Journal*, 17:454–462.

Kerppola, T. K. (2006). Design and implementation of bimolecular fluorescence complementation (bifc) assays for the visualization of protein interactions in living cells. *Nature Protocols*, 1(3):1278–1286.

Klein, J. S., Jiang, S., Galimidi, R. P., Keeffe, J. R., and Bjorkman, P. J. (2014). Design and characterization of structured protein linkers with differing flexibilities. *Protein Engineering, Design & Selection*, 27(10):325–330.

Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P., and Weisshaar, B. (2012). Gabi-kat simplesearch: new features of the arabidopsis thaliana t-dna mutant database. *Nucleic acids research*, 40(D1):D1211–D1215.

Klose, R. J., Kallin, E. M., and Zhang, Y. (2006). Jmjc-domain-containing proteins and histone demethylation. *Nature reviews genetics*, 7(9):715–727.

Ko, J.-H., Mitina, I., Tamada, Y., Hyun, Y., Choi, Y., Amasino, R. M., Noh, B., and Noh, Y.-S. (2010). Growth habit determination by the balance of histone methylation activities in arabidopsis. *The EMBO Journal*, 29(18):3208–3215.

Koi, S., Hisanaga, T., Sato, K., Shimamura, M., Yamato, K. T., Ishizaki, K., Kohchi, T., and Nakajima, K. (2016). An evolutionarily conserved plant rkd factor controls germ cell differentiation. *Current Biology*, 26(13):1775–1781.

Kooistra, S. M. and Helin, K. (2012). Molecular mechanisms and potential functions of histone demethylases. *Nature reviews Molecular cell biology*, 13(5):297–311.

Koonin, E. V., Zhou, S., and Lucchesi, J. C. (1995). The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. *Nucleic acids research*, 23(21):4229–4233.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4):693–705.

Kroj, T., Savino, G., Valon, C., Giraudat, J., and Parcy, F. (2003). Regulation of storage protein gene expression in arabidopsis. *Development*, 130(24):6065–6073.

Lan, F., Nottke, A. C., and Shi, Y. (2008). Mechanisms involved in the regulation of histone lysine demethylases. *Current opinion in cell biology*, 20(3):316–325.

Lan, F. and Shi, Y. (2009). Epigenetic regulation: methylation of histone and non-histone proteins. *Science in China Series C: Life Sciences*, 52:311–322.

Lau, O. S., Davies, K. A., Chang, J., Adrian, J., Rowe, M. H., Ballenger, C. E., and Bergmann, D. C. (2014). Direct roles of speechless in the specification of stomatal self-renewing cells. *Science*, 345(6204):1605–1609.

Lee, K., Park, O.-S., and Seo, P. J. (2018). Jmj30-mediated demethylation of h3k9me3 drives tissue identity changes to promote callus formation in arabidopsis. *The Plant Journal*, 95(6):961–975.

Lee, M. G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007). Demethylation of h3k27 regulates polycomb recruitment and h2a ubiquitination. *Science*, 318(5849):447–450.

Lelu-Walter, M.-A., Thompson, D., Harvengt, L., Sanchez, L., Toribio, M., and Pâques, L. E. (2013). Somatic embryogenesis in forestry with a focus on europe: state-of-the-art, benefits, challenges and future direction. *Tree Genetics Genomes*, 9(4):883–899.

Lewis, E. B. (1978). A gene complex controlling segmentation in drosophila. *Nature*, 276(5688):565–570.

Li, C., Chen, C., Chen, H., Wang, S., Chen, X., and Cui, Y. (2018a). Verification of dna motifs in arabidopsis using crispr/cas9-mediated mutagenesis. *Plant biotechnology journal*, 16(8):1446–1451.

Li, G., Huang, Z., Zhang, C., Dong, B.-J., Guo, R.-H., Yue, H.-W., Yan, L.-T., and Xing, X.-H. (2016). Construction of a linker library with widely controllable flexibility for fusion protein design. *Applied microbiology and biotechnology*, 100:215–225.

Li, Q.-F., Lu, J., Yu, J.-W., Zhang, C.-Q., He, J.-X., and Liu, Q.-Q. (2018b). The brassinosteroid-regulated transcription factors bzr1/bes1 function as a coordinator in multisignal-regulated plant growth. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1861(6):561–571.

Li, X., Wang, X., He, K., Ma, Y., Su, N., He, H., Stolc, V., Tongprasit, W., Jin, W., Jiang, J., et al. (2008). High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between dna methylation, histone methylation, and gene expression. *The Plant Cell*, 20(2):259–276.

Liang, Z., Yuan, L., Xiong, X., Hao, Y., Song, X., Zhu, T., Yu, Y., Fu, W., Lei, Y., Xu, J., et al. (2022). The transcriptional repressors val1 and val2 mediate genome-wide recruitment of the chd3 chromatin remodeler pickle in arabidopsis. *The Plant Cell*, 34(10):3915–3935.

Liu, C., Lu, F., Cui, X., and Cao, X. (2010). Histone methylation in higher plants. *Annual review of plant biology*, 61:395–420.

Liu, J., Feng, L., Gu, X., Deng, X., Qiu, Q., Li, Q., Zhang, Y., Wang, M., Deng, Y., Wang, E., et al. (2019). An h3k27me3 demethylase-hsfa2 regulatory loop orchestrates transgenerational thermomemory in arabidopsis. *Cell Research*, 29(5):379–390.

Lopez-Bigas, N., De, S., and Teichmann, S. A. (2008). Functional protein divergence in the evolution of homo sapiens. *Genome biology*, 9(2):1–14.

Lotan, T., Ohto, M.-a., Yee, K. M., West, M. A., Lo, R., Kwong, R. W., Yamagishi, K., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (1998). Arabidopsis leafy cotyledon1 is sufficient to induce embryo development in vegetative cells. *Cell*, 93(7):1195–1205.

Lu, F., Cui, X., Zhang, S., Jenuwein, T., and Cao, X. (2011a). Arabidopsis ref6 is a histone h3 lysine 27 demethylase. *Nature genetics*, 43(7):715–719.

Lu, F., Li, G., Cui, X., Liu, C., Wang, X.-J., and Cao, X. (2008). Comparative analysis of jmjc domain-containing proteins reveals the potential histone demethylases in arabidopsis and rice. *Journal of Integrative Plant Biology*, 50(7):886–896.

Lu, S. X., Knowles, S. M., Webb, C. J., Celaya, R. B., Cha, C., Siu, J. P., and Tobin, E. M. (2011b). The jumonji c domain-containing protein

jmj30 regulates period length in the arabidopsis circadian clock. *Plant physiology*, 155(2):906–915.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648):251–260.

Ma, S., Zhang, Z., Long, Y., Huo, W., Zhang, Y., Yang, X., Zhang, J., Li, X., Du, Q., Liu, W., et al. (2022). Evolutionary history and functional diversification of the jmjc domain-containing histone demethylase gene family in plants. *Plants*, 11(8):1041.

Margueron, R. and Reinberg, D. (2011). The polycomb complex prc2 and its mark in life. *Nature*, 469(7330):343–349.

Marmorstein, R. (2004). Structural and chemical basis of histone acetylation. In *Reversible Protein Acetylation: Novartis Foundation Symposium 259*, volume 259, pages 78–101. Wiley Online Library.

McDonough, M. A., Loenarz, C., Chowdhury, R., Clifton, I. J., and Schofield, C. J. (2010). Structural studies on human 2-oxoglutarate dependent oxygenases. *Current opinion in structural biology*, 20(6):659–672.

Mendoza, M. S., Dubreucq, B., Miquel, M., Caboche, M., and Lepiniec, L. (2005). Leafy cotyledon 2 activation is sufficient to trigger the accumulation of oil and seed specific mrnas in arabidopsis leaves. *FEBS letters*, 579(21):4666–4670.

Merini, W. and Calonje, M. (2015). Prc 1 is taking the lead in p c g repression. *The Plant Journal*, 83(1):110–120.

Merini, W., Romero-Campero, F. J., Gomez-Zambrano, A., Zhou, Y., Turck, F., and Calonje, M. (2017). The arabidopsis polycomb repressive complex 1 (prc1) components atbmi1a, b, and c impact gene networks throughout all stages of plant development. *Plant physiology*, 173(1):627–641.

Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. *Cell*, 128(4):787–800.

Miyajima, Y., Noguchi, S., Tanaka, Y., Li, J.-R., Nishimura, H., Kishima, M., Lim, J., Furuhata, E., Suzuki, T., Kasukawa, T., et al. (2022). Prediction of transcription factors associated with dna demethylation during human cellular development. *Chromosome Research*, 30(1):109–121.

Mosammaparast, N. and Shi, Y. (2010). Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annual review of biochemistry*, 79:155–179.

Mozgová, I., Muñoz-Viana, R., and Hennig, L. (2017). Prc2 represses hormone-induced somatic embryogenesis in vegetative tissue of arabidopsis thaliana. *PLOS Genetics*, 13(1):e1006562.

Nasir, A., Kim, K. M., and Caetano-Anollés, G. (2014). Global patterns of protein domain gain and loss in superkingdoms. *PLoS computational biology*, 10(1):e1003452.

Noble, D. (2015). Conrad waddington and the origin of epigenetics. *The Journal of experimental biology*, 218(6):816–818.

Noh, B., Lee, S.-H., Kim, H.-J., Yi, G., Shin, E.-A., Lee, M., Jung, K.-J., Doyle, M. R., Amasino, R. M., and Noh, Y.-S. (2004). Divergent roles of a pair of homologous jumonji/zinc-finger–class transcription factor proteins in the regulation of arabidopsis flowering time. *The Plant Cell*, 16(10):2601–2613.

Noh, Y.-S. and Amasino, R. M. (2003). Pie1, an iswi family gene, is required for flc activation and floral repression in arabidopsis. *The Plant Cell*, 15(7):1671–1682.

Ohira, M. J., Hendrickson, D. G., Scott McIsaac, R., and Rhind, N. (2017). An estradiol-inducible promoter enables fast, graduated control of gene expression in fission yeast. *Yeast*, 34(8):323–334.

Pais, M. S. (2019). Somatic embryogenesis induction in woody species: The future after omics data assessment. *Frontiers in Plant Science*, 10:240.

Pan, J., Zhang, H., Zhan, Z., Zhao, T., and Jiang, D. (2021). The ref6-dependent h3k27 demethylation establishes transcriptional competence to promote germination in arabidopsis. *bioRxiv*, pages 2021–06.

Pazos, F., Pietrosemoli, N., García-Martín, J. A., and Solano, R. (2013). Protein intrinsic disorder in plants. *Frontiers in plant science*, 4:363.

Pu, L., Liu, M.-S., Kim, S. Y., Chen, L.-F. O., Fletcher, J. C., and Sung, Z. R. (2013). Embryonic flower1 and ultrapetala1 act antagonistically on arabidopsis development and stress response. *Plant Physiology*, 162(2):812–830.

Quiroz, S., Yustis, J. C., Chávez-Hernández, E. C., Martínez, T., Sanchez, M. d. l. P., Garay-Arroyo, A., Álvarez-Buylla, E. R., and García-Ponce, B. (2021). Beyond the genetic pathways, flowering regulation complexity in arabidopsis thaliana. *International Journal of Molecular Sciences*, 22(11):5716.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., et al. (2000). Regulation of chromatin structure by site-specific histone h3 methyltransferases. *Nature*, 406(6796):593–599.

Richter, R., Kinoshita, A., Vincent, C., Martinez-Gallegos, R., Gao, H., van Driel, A. D., Hyun, Y., Mateos, J. L., and Coupland, G. (2019). Floral regulators flc and soc1 directly regulate expression of the b3-type transcription factor target of flc and svp 1 at the arabidopsis shoot apex via antagonistic chromatin modifications. *PLoS genetics*, 15(4):e1008065.

Riggs, A. D. and Porter, T. N. (1996). Overview of epigenetic mechanisms. *Cold Spring Harbor monograph series*, 32:29–46.

Rossetto, D., Avvakumov, N., and Côté, J. (2012). Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics*, 7(10):1098–1108.

Sakamoto, Y., Kawamura, A., Suzuki, T., Segami, S., Maeshima, M., Polyn, S., De Veylder, L., and Sugimoto, K. (2022). Transcriptional

activation of auxin biosynthesis drives developmental reprogramming of differentiated cells. *The Plant Cell*, 34(11):4348–4365.

Sani, E., Herzyk, P., Perrella, G., Colot, V., and Amtmann, A. (2013). Hyperosmotic priming of arabidopsis seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome biology*, 14(6):1–24.

Schwach, F., Vaistij, F. E., Jones, L., and Baulcombe, D. C. (2005). Mechanisms that can promote silencing of gene expression are active in arabidopsis. *The Plant Journal*, 41(5):745–762.

Seto, E. and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harbor perspectives in biology*, 6(4):a018713.

Sheldon, C. C., Hills, M. J., Lister, C., Dean, C., Dennis, E. S., and Peacock, W. J. (2008). Resetting of flowering locus c expression after epigenetic repression by vernalization. *Proceedings of the National Academy of Sciences*, 105(6):2214–2219.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog lsd1. *Cell*, 119(7):941–953.

Sims III, R. J., Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. *TRENDS in Genetics*, 19(11):629–639.

Sinha, K., Bilokapic, S., Du, Y., Malik, D., and Halic, M. (2023). Histone modifications regulate pioneer transcription factor binding and cooperativity. *bioRxiv*.

Stavropoulos, P., Blobel, G., and Hoelz, A. (2006). Crystal structure and mechanism of human lysine-specific demethylase-1. *Nature structural & molecular biology*, 13(7):626–632.

Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (2001). Leafy cotyledon2 en-

codes a b3 domain transcription factor that induces embryo development. *Proceedings of the National Academy of Sciences*, 98(20):11806–11811.

Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765):41–45.

Sun, X., Malhis, N., Zhao, B., Xue, B., Gsponer, J., and Rikkerink, E. H. (2019). Computational disorder analysis in ethylene response factors uncovers binding motifs critical to their diverse functions. *International Journal of Molecular Sciences*, 21(1):74.

Suzuki, M., Oda, M., Ramos, M.-P., Pascual, M., Lau, K., Stasiek, E., Agyiri, F., Thompson, R. F., Glass, J. L., Jing, Q., et al. (2011). Late-replicating heterochromatin is characterized by decreased cytosine methylation in the human genome. *Genome research*, 21(11):1833–1840.

Suzuki, M., Wang, H. H.-Y., and McCarty, D. R. (2007). Repression of the leafy cotyledon 1/b3 regulatory network in plant embryo development by vp1/abscisic acid insensitive 3-like b3 genes. *Plant physiology*, 143(2):902–911.

Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y. (2007). The histone h3k4 demethylase smcx links rest target genes to x-linked mental retardation. *Nature*, 447(7144):601–605.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., and Kennison, J. A. (1992). brahma: A regulator of drosophila homeotic genes structurally related to the yeast transcriptional activator snf2swi2. *Cell*, 68(3):561–572.

Tanaka, M., Kikuchi, A., and Kamada, H. (2008). The arabidopsis histone deacetylases hda6 and hda19 contribute to the repression of embryonic properties after germination. *Plant physiology*, 146(1):149–161.

Tang, L., Li, Z., Li, X., and Yang, W. (2010). Co-immunoprecipitation methods and protocols. *Methods in molecular biology*, 408:309–320.

Tang, L. P., Zhou, C., Wang, S. S., Yuan, J., Zhang, X. S., and Su, Y. H. (2017). Fusca 3 interacting with leafy cotyledon 2 controls lateral root formation through regulating yucca 4 gene expression in arabidopsis thaliana. *New Phytologist*, 213(4):1740–1754.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). Rybp-prc1 complexes mediate h2a ubiquitylation at polycomb target sites independently of prc2 and h3k27me3. *Cell*, 148(4):664–678.

Tessarz, P. and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nature reviews Molecular cell biology*, 15(11):703–708.

Theißen, G. and Melzer, R. (2007). The abc of flower development. *Flowering and its manipulation*, pages 1–27.

Thomas, G., Lange, H., and Hempel, K. (1972). Relative stability of lysine-bound methyl groups in arginie-rich histones and their subfrations in ehrlich ascites tumor cells in vitro. *Hoppe-Seyler's Zeitschrift fur physiologische Chemie*, 353(9):1423–1428.

Tian, Z., Li, X., Li, M., Wu, W., Zhang, M., Tang, C., Li, Z., Liu, Y., Chen, Z., Yang, M., et al. (2020). Crystal structures of ref6 and its complex with dna reveal diverse recognition mechanisms. *Cell discovery*, 6(1):17.

Torok, M. S. and Grant, P. A. (2004). Histone acetyltransferase proteins contribute to transcriptional processes at multiple levels. *Advances in protein chemistry*, 67:181–199.

Tsai, C.-L., Shi, Y., and Tainer, J. A. (2014). How substrate specificity is imposed on a histone demethylase—lessons from kdm2a. *Genes & development*, 28(16):1735–1738.

Tsukada, Y.-i., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of jmjc domain-containing proteins. *Nature*, 439(7078):811–816.

van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L., Söding, J., and Steinegger, M. (2023). Fast and accurate protein structure search with foldseek. *Nature Biotechnology*, pages 1–4.

Waddington, C. H. (1953). Epigenetics and evolution. In *Symp. Soc. Exp. Biol*, volume 7, pages 186–199.

Waki, T., Hiki, T., Watanabe, R., Hashimoto, T., and Nakajima, K. (2011). The arabidopsis rwp-rk protein rkd4 triggers gene expression and pattern formation in early embryogenesis. *Current Biology*, 21(15):1277–1281.

Walport, L. J. and Schofield, C. J. (2018). Adventures in defining roles of oxygenases in the regulation of protein biosynthesis. *The Chemical Record*, 18(12):1760–1781.

Wang, D., Liu, D., Yuchi, J., He, F., Jiang, Y., Cai, S., Li, J., and Xu, D. (2020a). Musitedeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Research*, 48(W1):W140–W146.

Wang, F.-X., Shang, G.-D., Wu, L.-Y., Xu, Z.-G., Zhao, X.-Y., and Wang, J.-W. (2020b). Chromatin accessibility dynamics and a hierarchical transcriptional regulatory network structure for plant somatic embryogenesis. *Developmental Cell*, 54(6):742–757.e8.

Wang, H., Liu, C., Cheng, J., Liu, J., Zhang, L., He, C., Shen, W.-H., Jin, H., Xu, L., and Zhang, Y. (2016a). Arabidopsis flower and embryo developmental genes are repressed in seedlings by different combinations of polycomb group proteins in association with distinct sets of cis-regulatory elements. *PLoS genetics*, 12(1):e1005771.

Wang, H., Liu, C., Cheng, J., Liu, J., Zhang, L., He, C., Shen, W.-H., Jin, H., Xu, L., and Zhang, Y. (2016b). Arabidopsis flower and embryo developmental genes are repressed in seedlings by different combinations of polycomb group proteins in association with distinct sets of cis-regulatory elements. *PLOS Genetics*, 12(1):e1005771.

Wang, X., Gao, J., Gao, S., Li, Z., Kuai, B., and Ren, G. (2019a). Ref6 promotes lateral root formation through de-repression of pin1/3/7 genes. *Journal of Integrative Plant Biology*, 61(4):383–387.

Wang, X., Gao, J., Gao, S., Song, Y., Yang, Z., and Kuai, B. (2019b). The h3k27me3 demethylase ref6 promotes leaf senescence through directly activating major senescence regulatory and functional genes in arabidopsis. *PLoS Genetics*, 15(4):e1008068.

Wang, Y., Gu, D., Deng, L., He, C., Zheng, F., and Liu, X. (2023). The histone h3k27 demethylase ref6 is a positive regulator of light-initiated seed germination in arabidopsis. *Cells*, 12(2):295.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.-H., Perlin, J. R., Leonelli, L., Sonbuchner, L. S., McDonald, C. H., Cook, R. G., Dou, Y., et al. (2004). Human pad4 regulates histone arginine methylation levels via demethylimination. *Science*, 306(5694):279–283.

Wang, Y., Yu, Y., Pang, Y., Yu, H., Zhang, W., Zhao, X., and Yu, J. (2021). The distinct roles of zinc finger cchc-type (zcchc) superfamily proteins in the regulation of rna metabolism. *RNA biology*, 18(12):2107–2126.

Weake, V. M. and Workman, J. L. (2008). Histone ubiquitination: triggering gene activity. *Molecular cell*, 29(6):653–663.

Whetstine, J. R., Nottke, A., Lan, F., Huarte, M., Smolikov, S., Chen, Z., Spooner, E., Li, E., Zhang, G., Colaiacovo, M., et al. (2006). Reversal of histone lysine trimethylation by the jmjd2 family of histone demethylases. *Cell*, 125(3):467–481.

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M. S., Burleigh, J. G., Gitzendanner, M. A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, 111(45):E4859–E4868.

Woodson, J. D., Joens, M. S., Sinson, A. B., Gilkerson, J., Salomé, P. A., Weigel, D., Fitzpatrick, J. A., and Chory, J. (2015). Ubiquitin facilitates

a quality-control pathway that removes damaged chloroplasts. *Science*, 350(6259):450–454.

Wright, P. E. and Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews Molecular cell biology*, 16(1):18–29.

Xiao, J. and Wagner, D. (2015). Polycomb repression in the regulation of growth and development in arabidopsis. *Current opinion in plant biology*, 23:15–24.

Yamaguchi, N. (2021). Removal of h3k27me3 by jmj proteins controls plant development and environmental responses in arabidopsis. *Frontiers in Plant Science*, 12:687416.

Yan, W., Chen, D., Smaczniak, C., Engelhorn, J., Liu, H., Yang, W., Graf, A., Carles, C. C., Zhou, D.-X., and Kaufmann, K. (2018). Dynamic and spatial restriction of polycomb activity by plant histone demethylases. *Nature plants*, 4(9):681–689.

Yan, Y., Kluz, T., Zhang, P., Chen, H.-b., and Costa, M. (2003). Analysis of specific lysine histone h3 and h4 acetylation and methylation status in clones of cells with a gene silenced by nickel exposure. *Toxicology and applied pharmacology*, 190(3):272–277.

Yan, Y., Shen, L., Chen, Y., Bao, S., Thong, Z., and Yu, H. (2014). A myb-domain protein efm mediates flowering responses to environmental cues in arabidopsis. *Developmental cell*, 30(4):437–448.

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). Val-and atbmi1-mediated h2aub initiate the switch from embryonic to postgerminative growth in arabidopsis. *Current Biology*, 23(14):1324–1329.

Yang, C.-H. and Chou, M.-L. (1999). Fld interacts with co to affect both flowering time and floral initiation in arabidopsis thaliana. *Plant and cell physiology*, 40(6):647–650.

Yang, H., Howard, M., and Dean, C. (2016). Physical coupling of activation and derepression activities to maintain an active transcriptional state at flc. *Proceedings of the National Academy of Sciences*, 113(33):9369–9374.

Yang, X. and Zhang, X. (2010). Regulation of somatic embryogenesis in higher plants. *Critical Reviews in Plant Science*, 29(1):36–57.

Yang, Z., Qiu, Q., Chen, W., Jia, B., Chen, X., Hu, H., He, K., Deng, X., Li, S., Tao, W. A., et al. (2018). Structure of the arabidopsis jmj14-h3k4me3 complex provides insight into the substrate specificity of kdm5 subfamily histone demethylases. *The Plant Cell*, 30(1):167–177.

Yoo, S.-D., Cho, Y.-H., and Sheen, J. (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature protocols*, 2(7):1565–1572.

Yoshida, N., Yanai, Y., Chen, L., Kato, Y., Hiratsuka, J., Miwa, T., Sung, Z. R., and Takahashi, S. (2001). Embryonic flower2, a novel polycomb group protein homolog, mediates shoot development and flowering in arabidopsis. *The Plant Cell*, 13(11):2471–2481.

Yu, X., Li, L., Li, L., Guo, M., Chory, J., and Yin, Y. (2008). Modulation of brassinosteroid-regulated gene expression by jumonji domain-containing proteins elf6 and ref6 in arabidopsis. *Proceedings of the National Academy of Sciences*, 105(21):7618–7623.

Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M., and Jacobsen, S. E. (2015). The histone modification landscape of plants. *Plant Physiology*, 139(1):1–9.

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H. (2006). Agrobacterium-mediated transformation of arabidopsis thaliana using the floral dip method. *Nature protocols*, 1(2):641–646.

Zhao, J., Herrera-Diaz, J., and Gross, D. S. (2005). Domain-wide displacement of histones by activated heat shock factor occurs independently of swi/snf and is not correlated with rna polymerase ii density. *Molecular and cellular biology*, 25(20):8985–8999.

Zhao, S. and Fernald, R. D. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of computational biology*, 12(8):1047–1064.

Zhao, T., Zhan, Z., and Jiang, D. (2019). Histone modifications and their regulatory roles in plant development and environmental memory. *Journal of genetics and genomics*, 46(10):467–476.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., et al. (2016). Resetting epigenetic memory by reprogramming of histone modifications in mammals. *Molecular cell*, 63(6):1066–1079.

Zheng, S., Hu, H., Ren, H., Yang, Z., Qiu, Q., Qi, W., Liu, X., Chen, X., Cui, X., Li, S., et al. (2019). The arabidopsis h3k27me3 demethylase jumonji 13 is a temperature and photoperiod dependent flowering repressor. *Nature communications*, 10(1):1303.

Zhou, Y., Romero-Campero, F. J., Gómez-Zambrano, Á., Turck, F., and Calonje, M. (2017). H2a monoubiquitination in arabidopsis thaliana is generally independent of lhp1 and prc2 activity. *Genome biology*, 18:1–13.

Zhou, Y., Tan, B., Luo, M., Li, Y., Liu, C., Chen, C., Yu, C.-W., Yang, S., Dong, S., Ruan, J., et al. (2013). Histone deacetylase19 interacts with hsl1 and participates in the repression of seed maturation genes in arabidopsis seedlings. *The Plant Cell*, 25(1):134–148.

Zhou, Y., Yan, A., Han, H., Li, T., Geng, Y., Liu, X., and Meyerowitz, E. M. (2018). Hairy meristem with wuschel confines clavata3 expression to the outer apical meristem layers. *Science*, 361(6401):502–506.

Zhu, D., Wen, Y., Yao, W., Zheng, H., Zhou, S., Zhang, Q., Qu, L.-J., Chen, X., and Wu, Z. (2023). Distinct chromatin signatures in the arabidopsis male gametophyte. *Nature Genetics*, pages 1–15.

Zuin, A., Isasa, M., and Crosas, B. (2014). Ubiquitin signaling: extreme conservation as a source of diversity. *Cells*, 3(3):690–701.

Zuo, J., Niu, Q.-W., and Chua, N.-H. (2000). An estrogen receptor-based transactivator xve mediates highly inducible gene expression in transgenic plants. *The Plant Journal*, 24(2):265–273.

Appendix

AtREF6 homologs	REF6_[Quercus_lobata]
REF6_[Arabidopsis_thaliana]	REF6_[Cannabis_sativa]
REF6_[Oryza_sativa_Japonica_Group]	REF6_[Rosa_chinensis]
REF6_[Zea_mays]	REF6_[Prunus_persica]
REF6_[Lactuca_sativa]	REF6-like_[Malus_domestica]
REF6_[Helianthus_annuus]	REF6_[Citrus_clementina]
REF6_[Erigeron_canadensis]	REF6_[Manihot_esculenta]
REF6_isoform_X2_[Arachis_hypogaea]	REF6_[Populus_trichocarpa]
REF6_[Glycine_max]	REF6_[Vitis_vinifera]
REF6_[Trifolium_pratense]	REF6_[Carica_papaya]
REF6_isoform_X1_[Medicago_truncatula]	REF6-like_[Gossypium_hirsutum]
REF6-like_[Cicer_arietinum]	REF6-like_[Gossypium_raimondii]
REF6_[Amborella_trichopoda]	REF6-like_isoform_X1_[Durio_zibethinus]
REF6_[Tarenaya_hassleriana]	REF6_putative_isoform_2_[Theobroma_cacao]
REF6_[Raphanus_sativus]	REF6_[Erythranthe_guttata]
REF6-like_[Brassica_napus]	REF6_[Striga_asiatica]
REF6_isoform_X1_[Brassica_rapa]	REF6-like_[Chenopodium_quinoa]
REF6_isoform_X2_[Brassica_rapa]	REF6-like_isoform_X1_[Nicotiana_tabacum]
REF6_[Eutrema_salsugineum]	REF6_[Capsicum_annuum]
REF6_[Capsella_rubella]	REF6_[Solanum_tuberosum]
REF6-like_[Papaver_somniferum]	REF6_[Solanum_pennellii]
REF6_[Cucumis_melo]	REF6_[Solanum_lycopersicum]
REF6-like_[Tripterygium_wilfordii]	REF6-like_[Camellia_sinensis]

Figure 6.1: List of AtREF6 homologs obtained from BLAST. The selected sequences have an e-value lower than e-50 and a bit score higher than 40.

AtELF6 homologs
ELF6_[Arabidopsis_thaliana]
ELF6_[Selaginella_moellendorffii]
putative_ELF6_[Zea_mays]
ELF6_isoform_X2_[Eutrema_salsugineum]
ELF6_isoform_X1_[Eutrema_salsugineum]
ELF6_[Capsella_rubella]
ELF6_[Brassica_rapa]
ELF6_[Brassica_napus]
ELF6_[Trifolium_pratense]
ELF6_isoform_X1_[Cicer_arietinum]
ELF6_isoform_X2_[Cicer_arietinum]
ELF6_[Pisum_sativum]
ELF6_[Medicago_truncatula]
ELF6_[Arachis_hypogaea]
ELF6_[Glycine_max]
ELF6_[Erythranthe_guttata]
ELF6_isoform_X2_[Chenopodium_quinoa]
ELF6_isoform_X1_[Chenopodium_quinoa]
ELF6_[Erigeron_canadensis]
ELF6_isoform_X1_[Lactuca_sativa]
ELF6_isoform_X2_[Lactuca_sativa]
ELF6_isoform_X5_[Amborella_trichopoda]
ELF6_isoform_X2_[Amborella_trichopoda]
ELF6_[Citrus_clementina]
ELF6_[Manihot_esculenta]
ELF6_isoform_X2_[Populus_trichocarpa]
ELF6_isoform_X1_[Populus_trichocarpa]
ELF6_[Papaver_somniferum]
ELF6_[Helianthus_annuus]
ELF6_[Gossypium_hirsutum]
ELF6_isoform_X2_[Durio_zibethinus]
probable_ELF6_[Eucalyptus_grandis]
ELF6_[Eucalyptus_grandis]
ELF6_isoform_X3_[Carica_papaya]
ELF6_isoform_X2_[Carica_papaya]
ELF6_isoform_X1_[Carica_papaya]
ELF6_[Cucumis_melo]
ELF6_isoform_X1_[Cucumis_sativus]
ELF6_isoform_X2_[Cucumis_sativus]
ELF6_[Punica_granatum]
ELF6_[Solanum_lycopersicum]
ELF6_isoform_X2_[Solanum_pennellii]
ELF6_isoform_X1_[Solanum_pennellii]
ELF6_isoform_X1_[Capsicum_annuum]
ELF6_isoform_X1_[Rosa_chinensis]
ELF6_isoform_X2_[Rosa_chinensis]
ELF6_[Cannabis_sativa]
ELF6_[Prunus_persica]
ELF6_[Malus_domestica]
ELF6[Camellia_sinensis]
ELF6_isoform_X2_[Tripterygium_wilfordii]

Figure 6.2: List of AtELF6 homologs obtained from BLAST. The selected sequences have an e-value lower than e-50 and a bit score higher than 40.

Uncategorised REF6/ELF6 homologs

JMJ705_[Selaginella_moellendorffii]

LOC112288597[Physcomitrium_patens]

SE14-like_isoform_X2_[Oryza_sativa_Japonica_Group]

SE14-like_isoform_X1_[Oryza_sativa_Japonica_Group]

MARPO_0006s0237_[Marchantia_polymorpha]

unnamed_protein_[Brassica_oleracea]

putative_[histone_H3]

zinc_finger_protein_[Striga_asiatica]

DKX38_012242_[Salix_brachista]

hypothetical_protein_[Gossypium_raimondii]

jumonji_family_protein_[Theobroma_cacao]

CDL15_Pgr026055_[Punica_granatum]

SE14_[Nicotiana_tabacum]

unnamed_protein_[Coffea_canephora]

JMJ705_[Zea_mays]

JMJ705-like_[Pisum_sativum]

PHAVU_003G207700g_[Phaseolus_vulgaris]

H3K36_demethylase_[Lupinus_albus]

HPP92_013548_[Vanilla_planifolia]

Bca52824_000810_[Brassica_carinata]

JMJ705_[Cucumis_sativus]

FH972_018901_[Carpinus_fangiana]

I3842_12G033300_[Carya_illinoinensis]

JMJ705-like_[Nelumbo_nucifera]

CCACVL1_02664_[Corchorus_capsularis]

unnamed_protein_[Coffea_canephora]

RHSIM_Rhsim12G0171300_[Rhododendron_simsii]

Figure 6.3: List of homologs of REF6 and ELF6 from basal plants obtained from BLAST. The selected sequences have an e-value lower than e-50 and a bit score higher than 40.

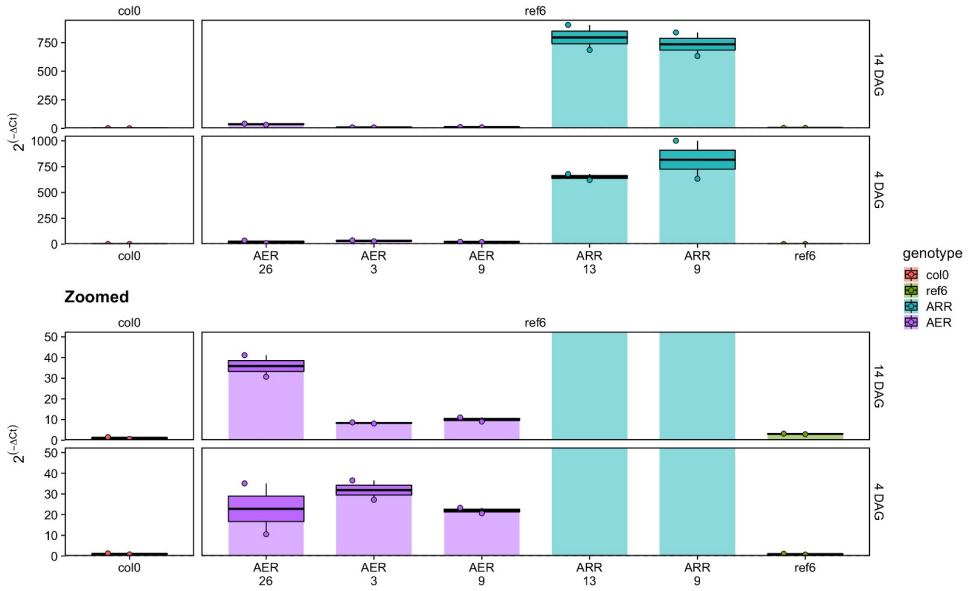


Figure 6.4: GFP expression normalised to GAPDH in selected lines with primer couple 2. Three independent lines per genotype (ARR and AER) were screened for GFP expression by qPCR analysis. *ref6* and col-0 were used as negative control for GFP expression. The lines were tested at four days after germination (DAG) and fourteen DAG. The results were normalised using GAPDH as housekeeping gene. The figure on top shows the GFP expression relative to the ARR lines whereas the figure at the bottom shows the GFP expression relative to the AER lines

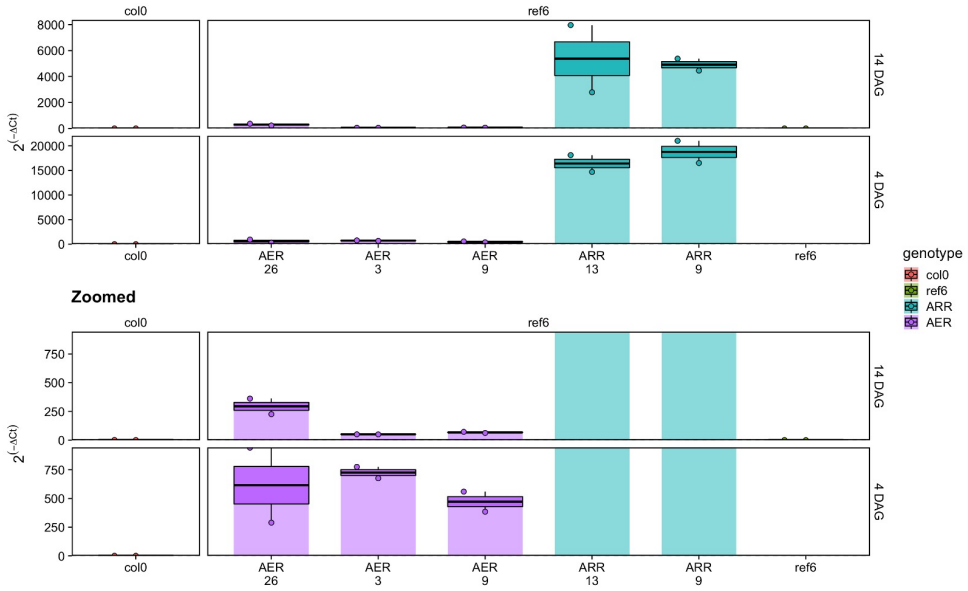


Figure 6.5: GFP expression normalised to GAPDH in selected lines with primer couple 3. Three independent lines per genotype (ARR and AER) were screened for GFP expression by qPCR analysis. *ref6* and col-0 were used as negative control for GFP expression. The lines were tested at four days after germination (DAG) and fourteen DAG. The results were normalised using GAPDH as housekeeping gene. The figure on top shows the GFP expression relative to the ARR lines whereas the figure at the bottom shows the GFP expression relative to the AER lines

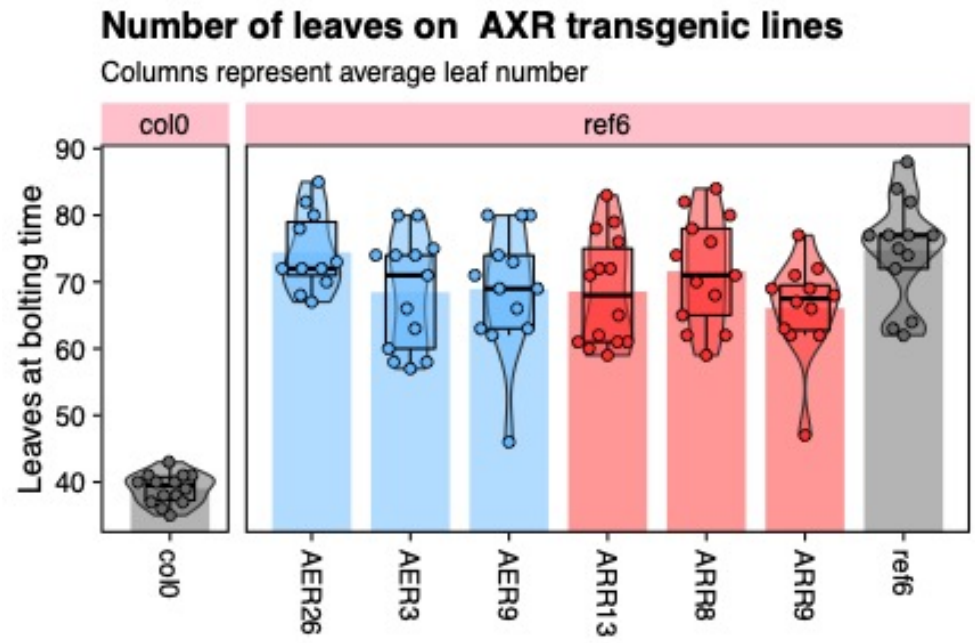


Figure 6.6: Number of leaves at bolting in ARR;*ref6-5* and AER;*ref6-5* under short day conditions. Box plots of col0, *ref6-5*, ARR;*ref6-5* and AER;*ref6-5* lines grown in soil until bolting.

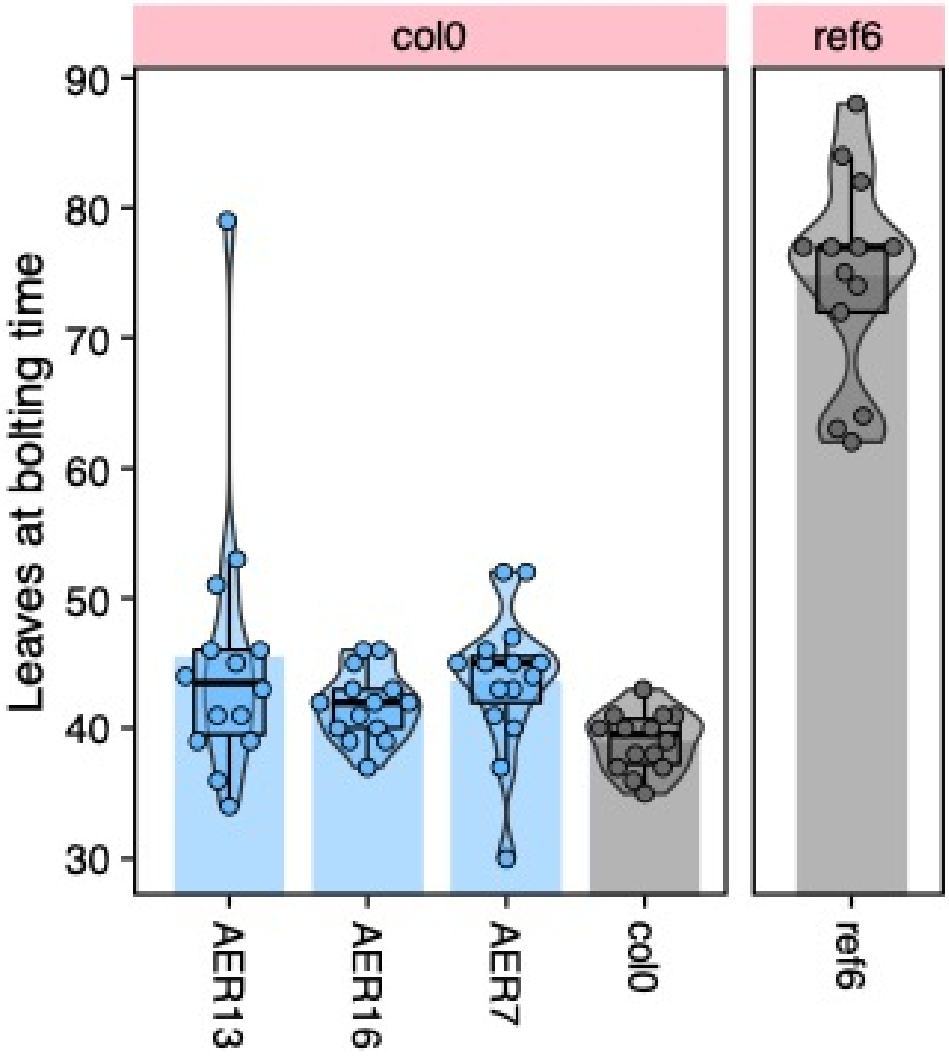


Figure 6.7: Number of leaves at bolting in AER;col0 under short day conditions. Box plots of col0, *ref6-5*, AER;col0 lines grown in soil until bolting.

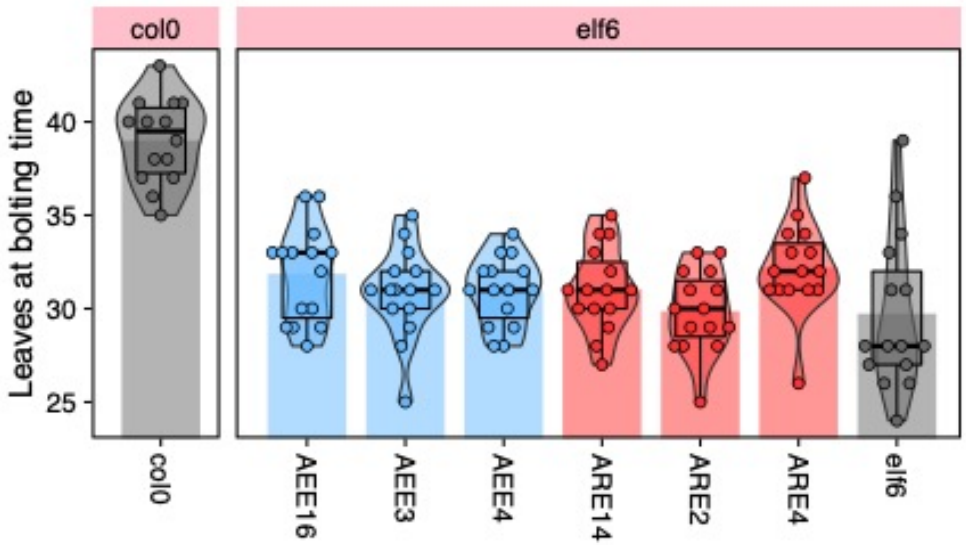


Figure 6.8: Number of leaves at bolting in AARE;*elf6-C* and AEE;*elf6-C* under short day conditions. Box plots of col0, *elf6-C*, ARE;*elf6-C* and AEE;*elf6-C* lines grown in soil until bolting.

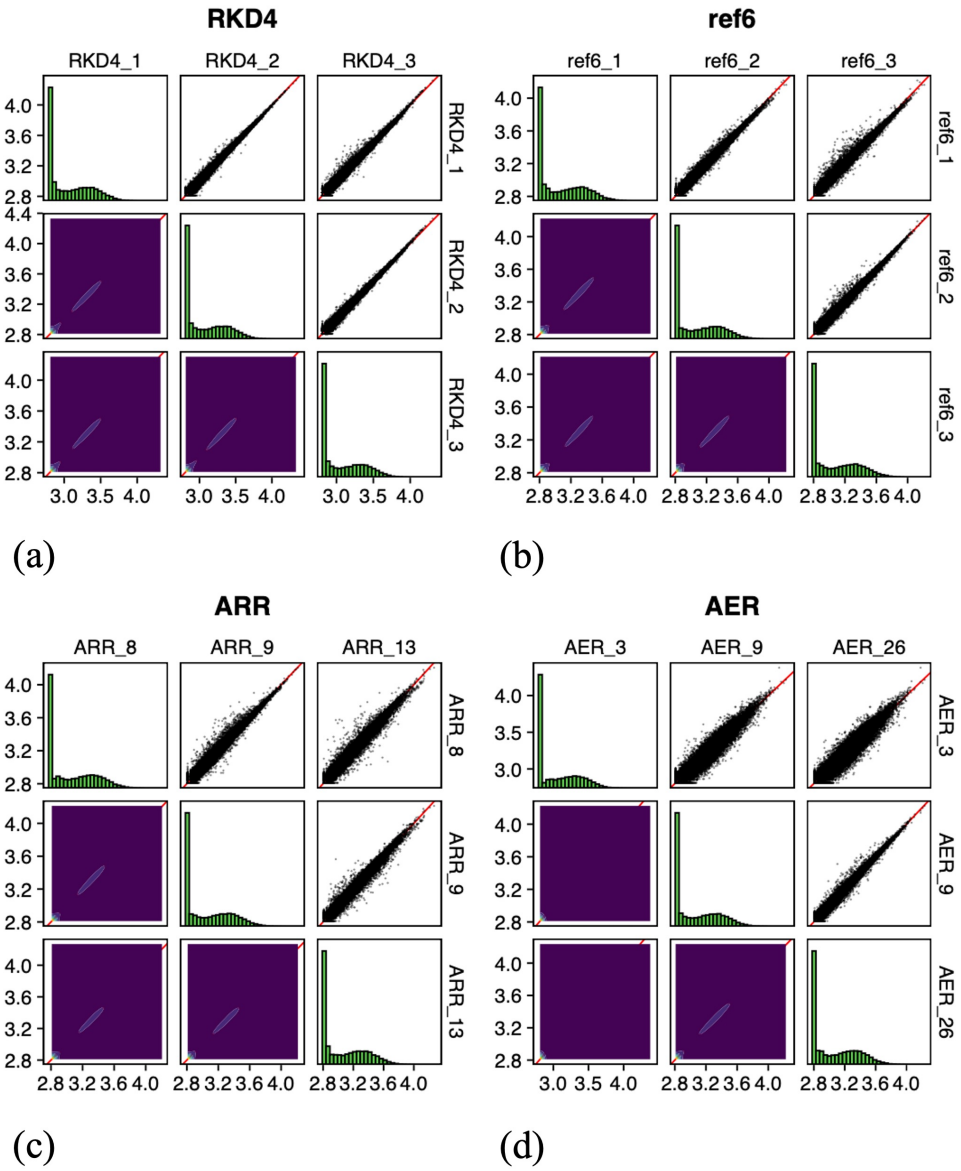


Figure 6.9: Scatterplot of RNA-seq gene expression data. Scatter plots of transcript expression data for replicates Arabidopsis RKD4 wildtype (a), mutant *ref6-5* (b), ARR;*ref6-5* (c), and AER;*ref6-5* (d) showing a high degree of correlation. Transcripts with $\geq$ 1 FPKM and minimum of 10 mapped reads were used. The plot is on a log-transformed scale.

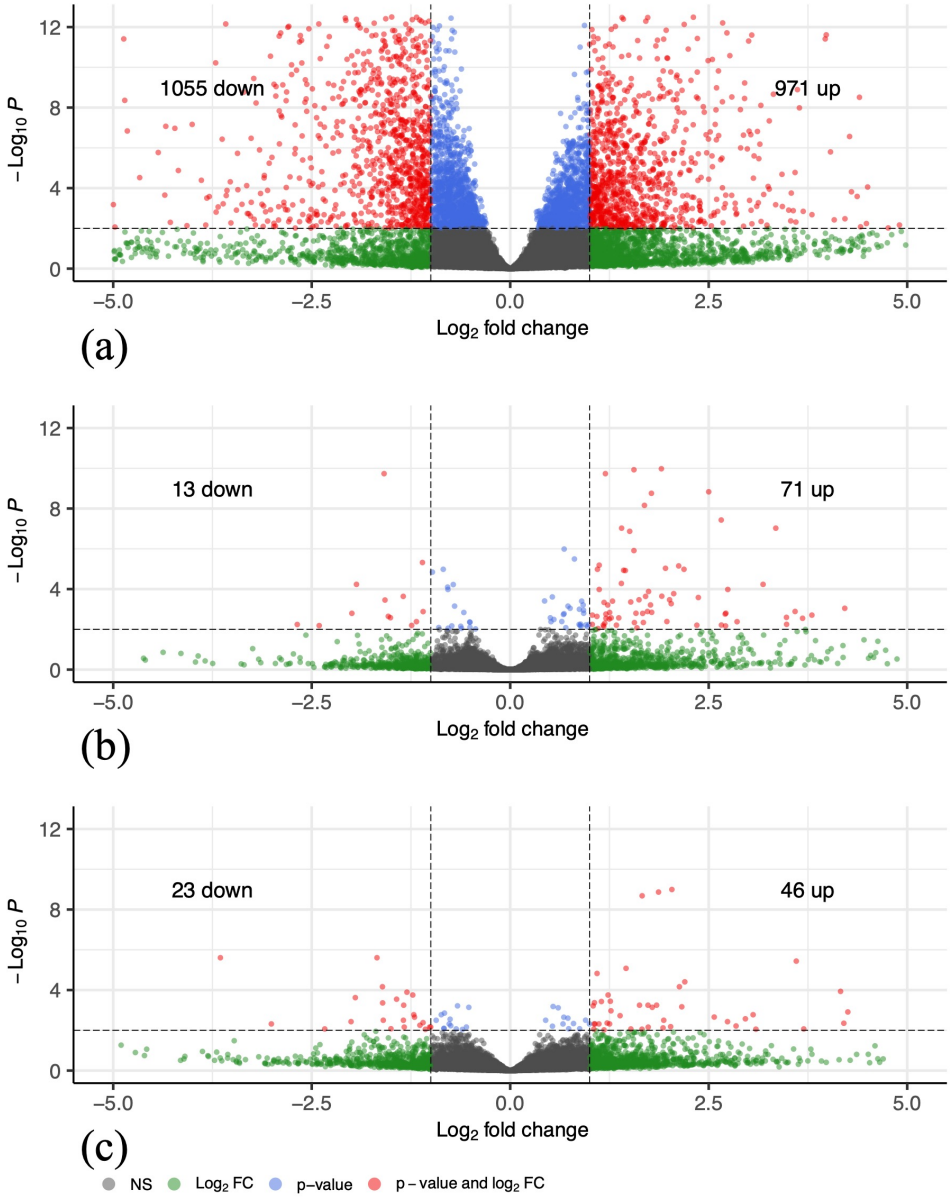


Figure 6.10: Volcano plots of RNA-seq differential gene expression analysis for the comparisons between (a) *ref6-5* vs RKD4, (b) ARR;*ref6-5* vs *ref6-5* and (c) AER;*ref6-5* vs *ref6-5*. The volcano plots illustrate the relationship between the log2 fold change (x-axis) and the -log10 adjusted p-value (y-axis) for each gene. Each point represents a single gene, with upregulated and and downregulated genes shown in red. Differentially expressed genes (DEGs) were determined using a log2 fold change threshold of > 1 and an adjusted p-value (FDR) < 0.05.

Genes downregulated between ref6 and WT			
AT2G06845	AT3G26170	AT4G28800	AT1G64100
AT5G56810	AT3G02115	AT1G34460	AT4G33370
AT4G22170	AT3G14710	AT4G23200	AT3G60640
AT1G21220	AT3G18320	AT1G43605	AT1G29600
AT3G14700	AT4G10265	AT4G37370	AT1G47370
AT2G33280	AT5G39850	AT1G73700	AT5G03390
AT2G05015	AT3G56780	AT2G25600	AT2G32780
AT3G53690	AT1G20680	AT5G59310	AT3G60960
AT5G27940	AT1G09773	AT1G64795	AT2G27402
AT5G44980	AT1G50830	AT4G29770	AT3G23240
AT3G41979	AT3G22700	AT4G16745	AT3G02790
AT3G21080	AT3G21330	AT1G80160	AT1G18830
AT2G35550	AT5G54550	AT1G04133	AT3G16650
AT5G57567	AT3G22860	AT5G64060	AT5G56380
AT1G01670	AT5G37750	AT5G17780	AT1G06753
AT5G57640	AT4G01820	AT5G41820	AT1G22980
AT5G53700	AT1G22240	AT3G01015	AT3G59200
AT1G21870	AT4G34400	AT2G24800	
AT4G05760	AT3G44120	AT2G38465	
AT5G57570	AT3G21000	AT5G12020	
AT3G26165	AT5G03965	AT1G33770	
AT5G55690	AT5G45990	AT4G36450	
AT3G59260	AT5G03435	AT2G08340	
AT4G19950	AT2G37770	AT5G60140	
AT5G09730	AT4G38870	AT3G01630	
AT3G44790	AT5G57650	AT3G03480	
AT5G14470	AT4G37860	AT5G52930	
AT2G09525	AT5G53870	AT1G33970	
AT1G07450	AT1G20350	AT5G63750	
AT2G21910	AT3G22750	AT5G41030	
AT5G60142	AT1G51670	AT5G46370	
AT5G37440	AT3G44805	AT2G32550	
AT3G02810	AT4G19570	AT3G29767	
AT1G21260	AT1G07160	AT4G10010	
AT1G14520	AT3G61120	AT1G64105	
AT4G35570	AT3G27620	AT5G42120	
AT1G50770	AT1G18197	AT1G69720	
AT5G13600	AT1G65370	AT1G64220	
AT1G30920	AT3G26220	AT3G57390	
	AT3G15170	AT5G54190	
	AT1G22130	AT2G42480	
	AT3G59190	AT1G21320	
	AT1G48400	AT3G60980	

Figure 6.11: Set of downregulated genes between ref6 and WT and filtered based on H3K27me3 gain in ref6.

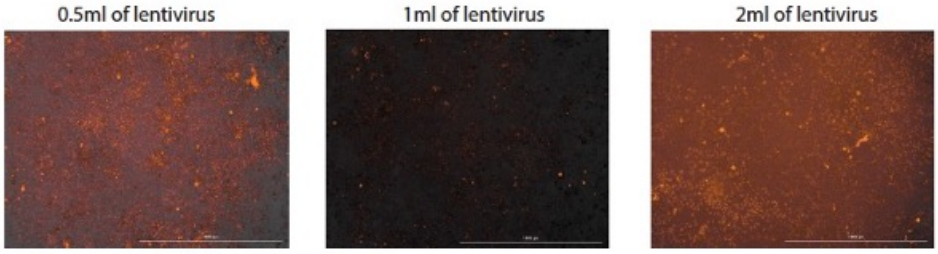


Figure 6.12: mCherry expression in HEK293 cells transfected with REF6-KDM6 48h after DOX induction. The figures show the expression at different lentivirus concentrations: 0.5ml, 1ml and 2ml.

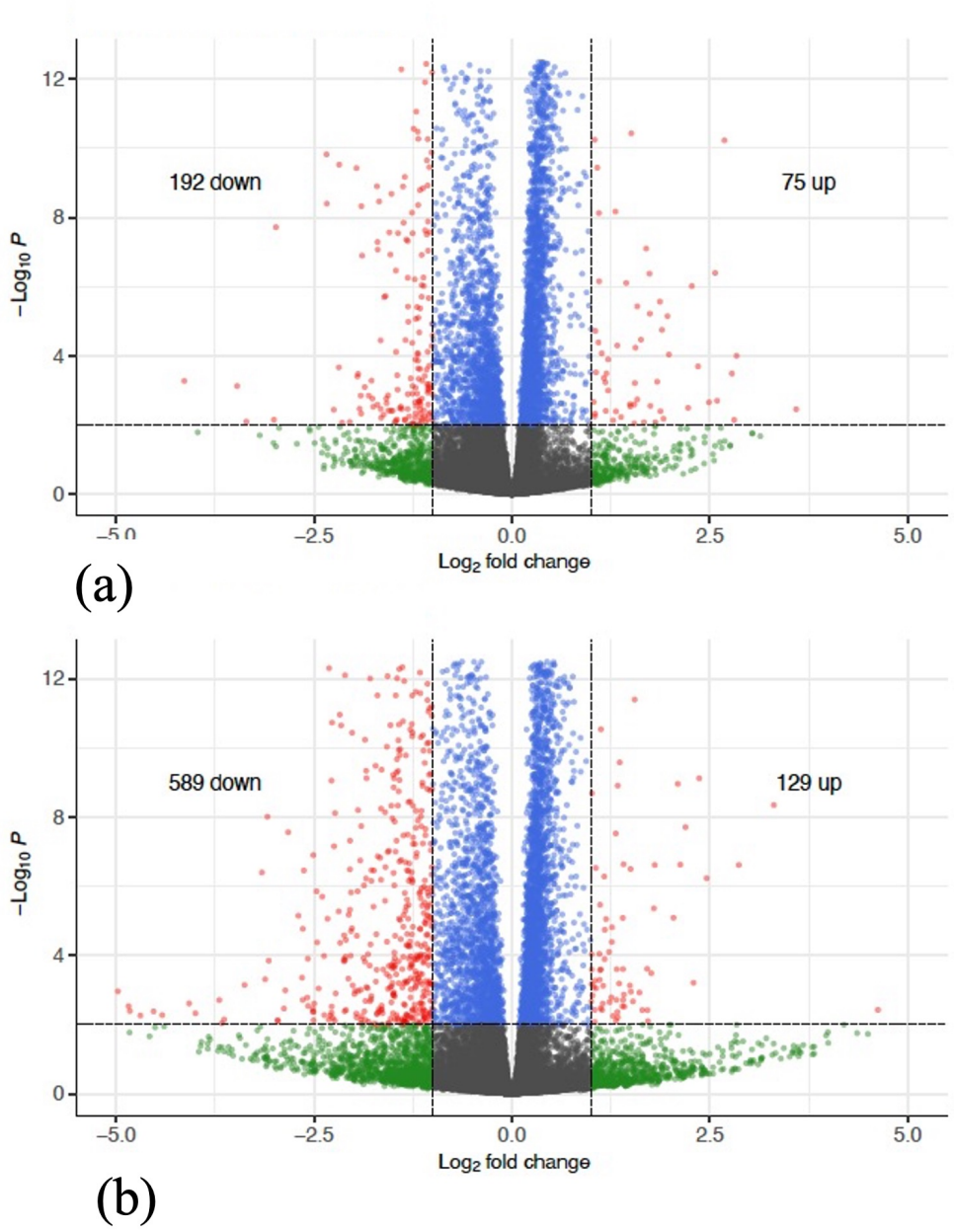


Figure 6.13: Volcano plots of RNA-seq differential gene expression analysis for the comparisons between JMJ13-KDM6 at (a) 3h and (b) 6h. The volcano plots illustrate the relationship between the log2 fold change (x-axis) and the -log10 adjusted p-value (y-axis) for each gene. Each point represents a single gene, with upregulated and and downregulated genes shown in red. Differentially expressed genes (DEGs) were determined using a log2 fold change threshold of > 1 and an adjusted p-value (FDR) < 0.05. A total of 75 upregulated and 192 downregulated genes were identified as significant DEGs between samples at 3h and a total of 129 upregulated and 589 downregulated genes at 6h.

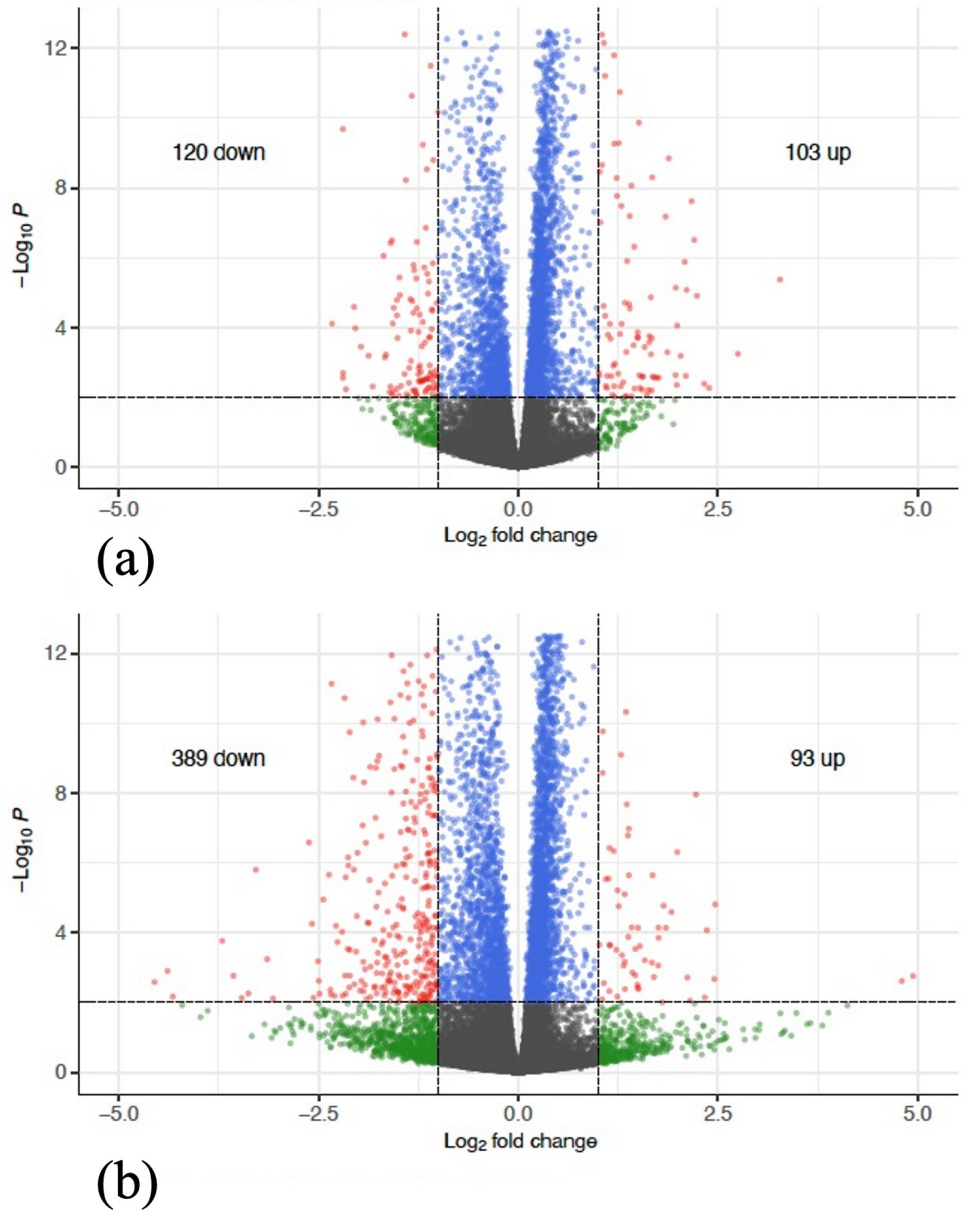


Figure 6.14: Volcano plots of RNA-seq differential gene expression analysis for the comparisons between REF6-KDM6 at (a) 3h and (b) 6h. The volcano plots illustrate the relationship between the log2 fold change (x-axis) and the -log10 adjusted p-value (y-axis) for each gene. Each point represents a single gene, with upregulated and and downregulated genes shown in red. Differentially expressed genes (DEGs) were determined using a log2 fold change threshold of > 1 and an adjusted p-value (FDR) < 0.05. A total of 103 upregulated and 120 downregulated genes were identified as significant DEGs between samples at 3h and a total of 93 upregulated and 389 downregulated genes at 6h.

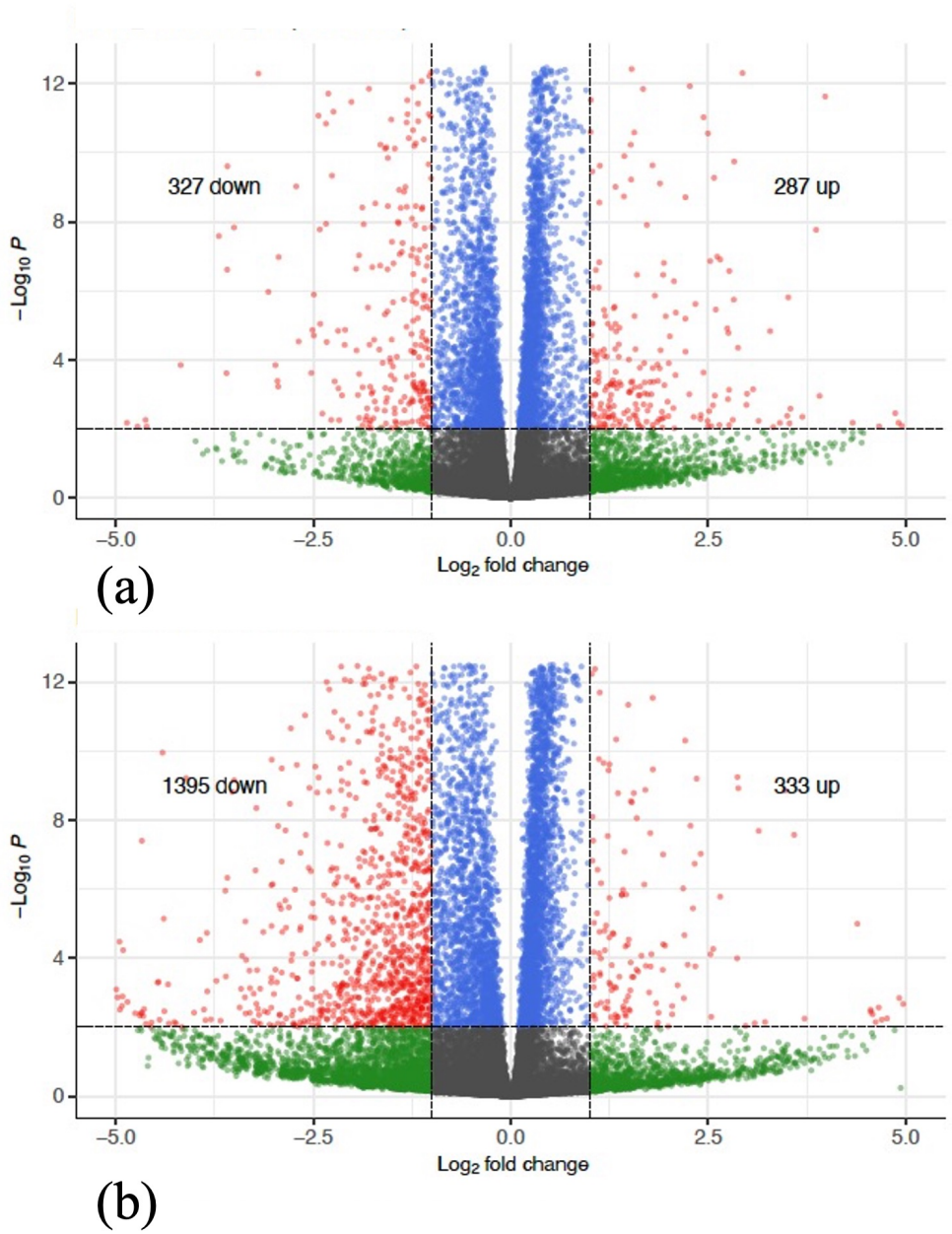


Figure 6.15: Volcano plots of RNA-seq differential gene expression analysis for the comparisons between ELF6-KDM6 at (a) 3h and (b) 6h. The volcano plots illustrate the relationship between the log2 fold change (x-axis) and the -log10 adjusted p-value (y-axis) for each gene. Each point represents a single gene, with upregulated and and downregulated genes shown in red. Differentially expressed genes (DEGs) were determined using a log2 fold change threshold of > 1 and an adjusted p-value (FDR) < 0.05. A total of 287 upregulated and 327 downregulated genes were identified as significant DEGs between samples at 3h and a total of 333 upregulated and 1395 downregulated genes at 6h.

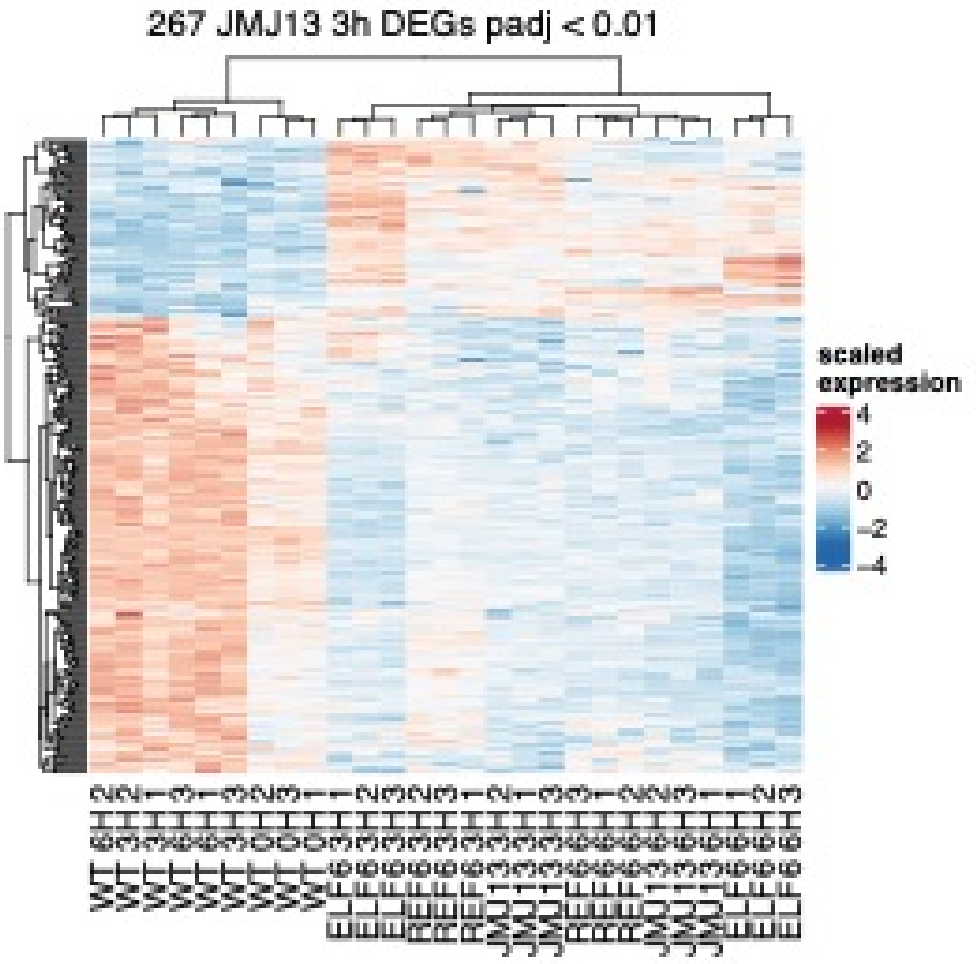


Figure 6.16: Heatmap of differentially expressed genes (DEGs) from RNA-seq data for JMJ13-KDM6. The heatmap displays the hierarchical clustering of DEGs (rows) and samples (columns) based on their normalized, log2-transformed gene expression levels. Each cell represents the expression level of DEGS between JMJ13-KDM6 and WT at 3h from induction, with colors indicating relative expression: high expression (red), average expression (white), and low expression (blue). DEGs were identified using a log2 fold change threshold of >1 and an adjusted p-value (FDR) < 0.05

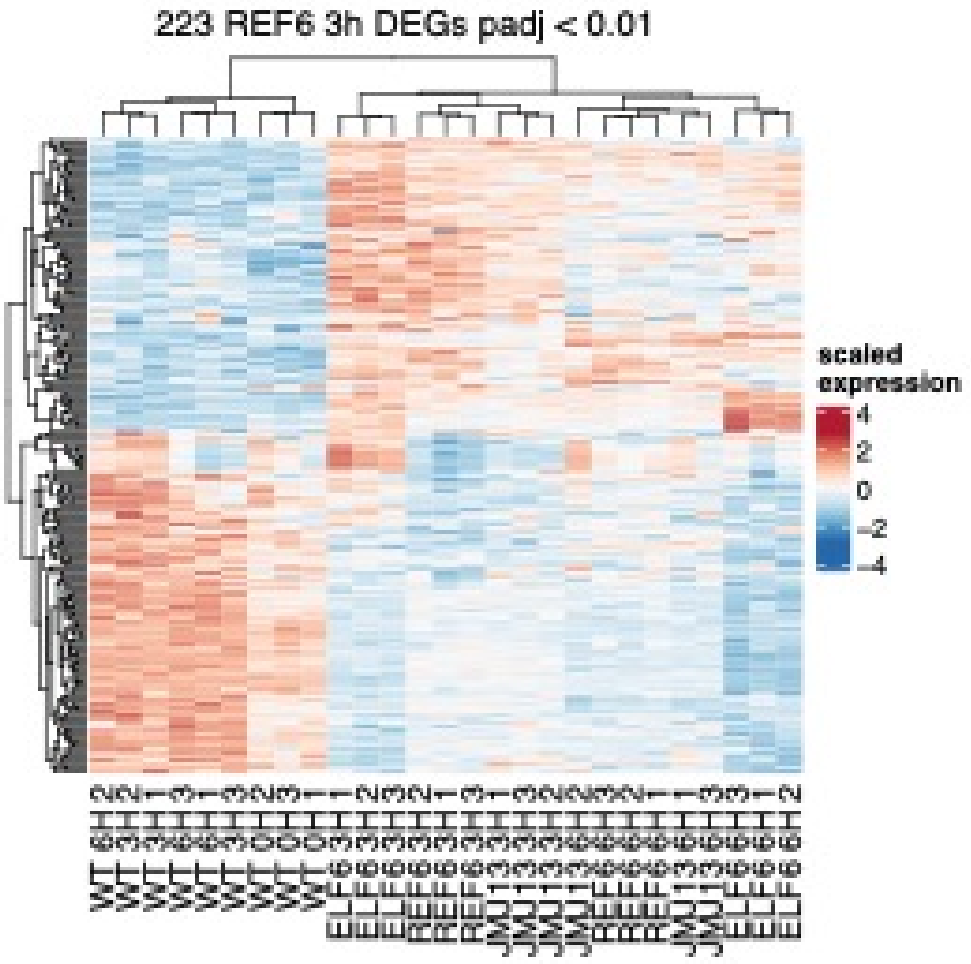


Figure 6.17: Heatmap of differentially expressed genes (DEGs) from RNA-seq data for REF6-KDM6. The heatmap displays the hierarchical clustering of DEGs (rows) and samples (columns) based on their normalized, log2-transformed gene expression levels. Each cell represents the expression level of DEGS between REF6-KDM6 and WT at 3h from induction, with colors indicating relative expression: high expression (red), average expression (white), and low expression (blue). DEGs were identified using a log2 fold change threshold of >1 and an adjusted p-value (FDR) < 0.05

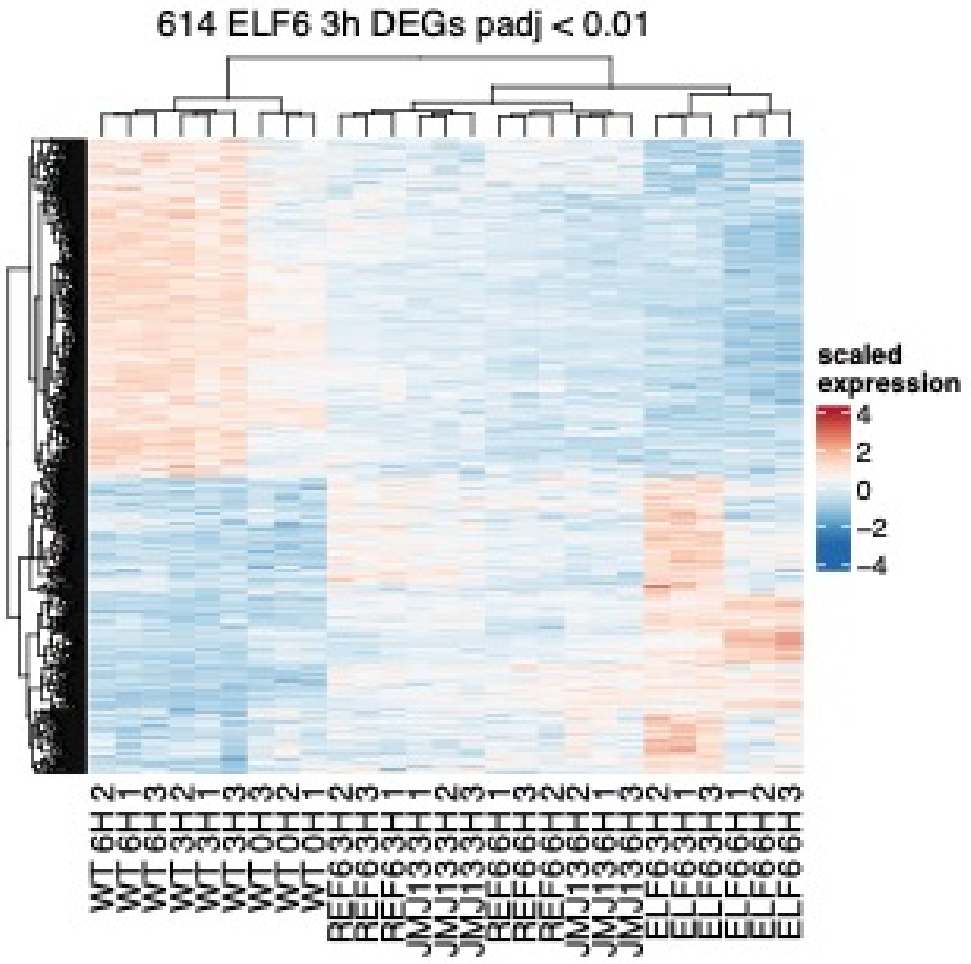


Figure 6.18: Heatmap of differentially expressed genes (DEGs) from RNA-seq data for ELF6-KDM6. The heatmap displays the hierarchical clustering of DEGs (rows) and samples (columns) based on their normalized, log2-transformed gene expression levels. Each cell represents the expression level of DEGS between ELF6-KDM6 and WT at 3h from induction, with colors indicating relative expression: high expression (red), average expression (white), and low expression (blue). DEGs were identified using a log2 fold change threshold of >1 and an adjusted p-value (FDR) < 0.05

Background	Construct	Line number																			
col0	RKD4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	LEC2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	JMJ13-RKD4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	JMJ13-LEC2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	REF6-RKD4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	REF6-LEC2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	ELF6-RKD4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	ELF6-LEC2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Figure 6.19: Table of T2 homozygous lines generated in Arabidopsis col0 background: RKD4, JMJ13-RKD4, REF6-RKD4, ELF6-RKD4, LEC2, JMJ13-LEC2, REF6-LEC2 and ELF6-LEC2.

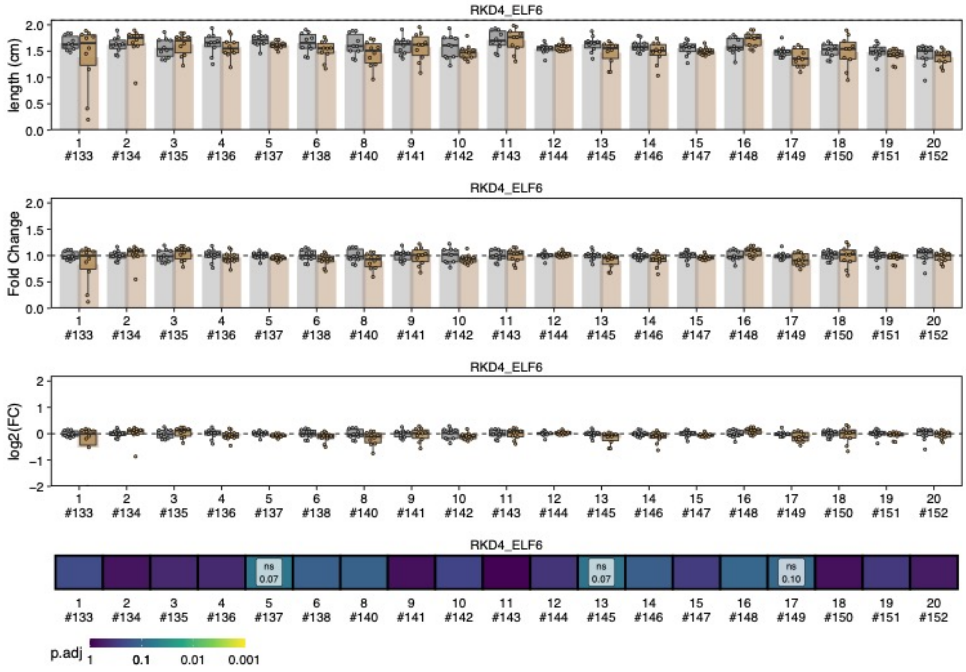


Figure 6.20: Barplot representing the comparison between the root length at day three for RKD4 lines and the control col0. Below the barplot are shown the lines which have significative difference from the control from T-test; p-value <0.05.

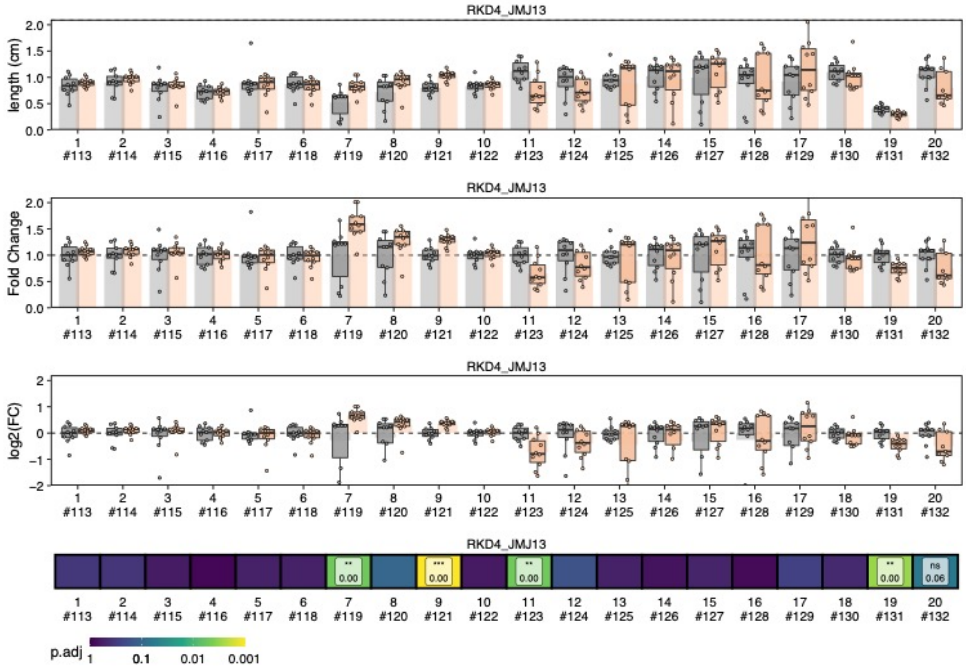


Figure 6.21: Barplot representing the comparison between the root length at day three for JMJ13-RKD4 lines and the control col0. Below the barplot are shown the lines which have significative difference from the control from T-test; p-value <0.05.

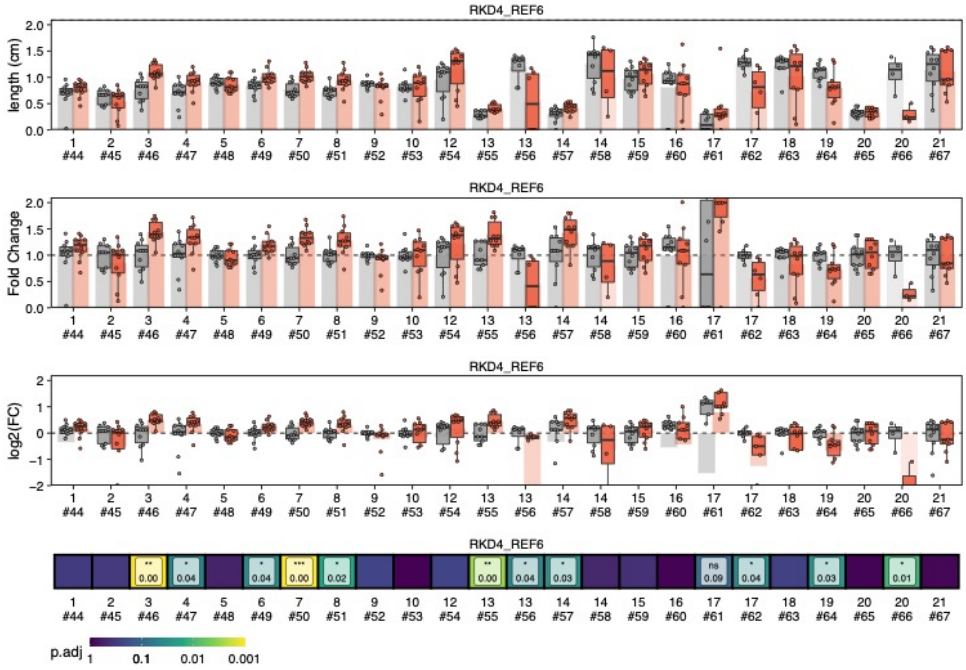


Figure 6.22: Barplot representing the comparison between the root length at day three for REF6-RKD4 lines and the control col0. Below the barplot are shown the lines which have significative difference from the control from T-test; p-value <0.05.

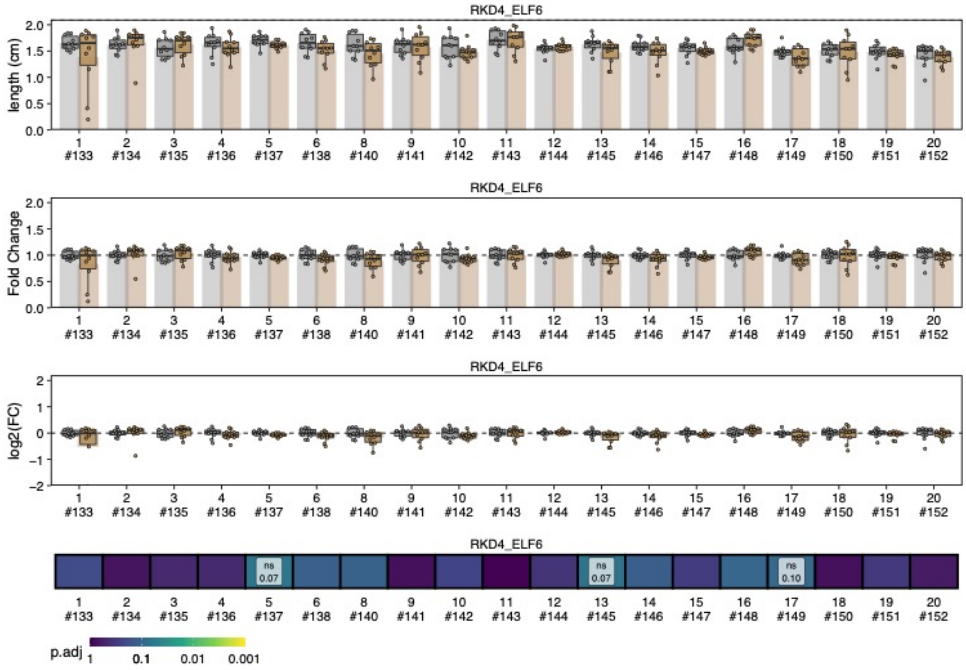


Figure 6.23: Barplot representing the comparison between the root length at day three for ELF6-RKD4 lines and the control col0. Below the barplot are shown the lines which have significative difference from the control from T-test; p-value <0.05.

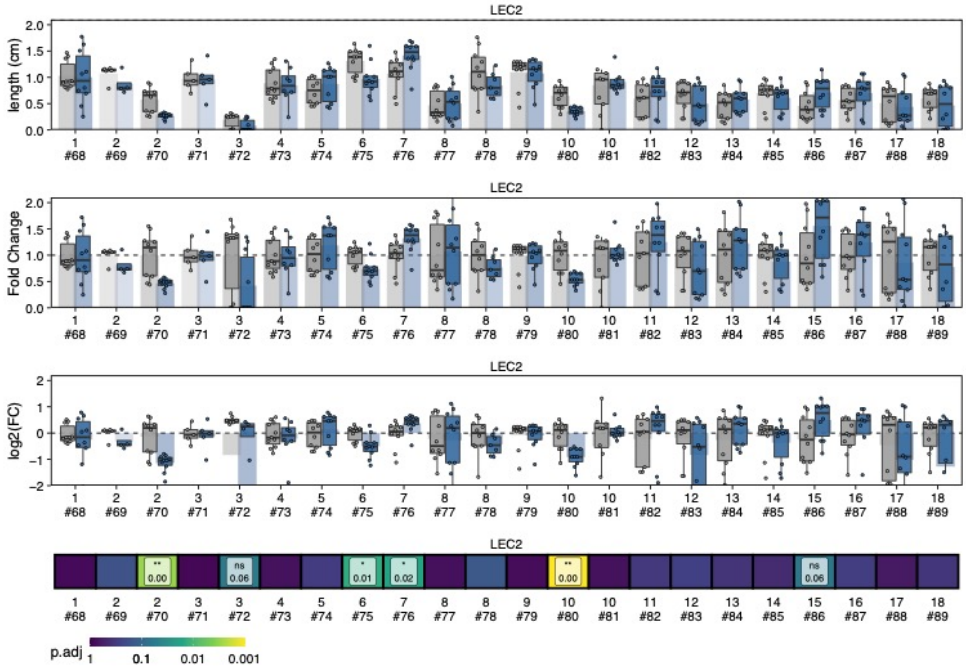


Figure 6.24: Barplot representing the comparison between the root length at day three for LEC2 lines and the control col0. Below the barplot are shown the lines which have significative difference from the control from T-test; p-value <0.05.

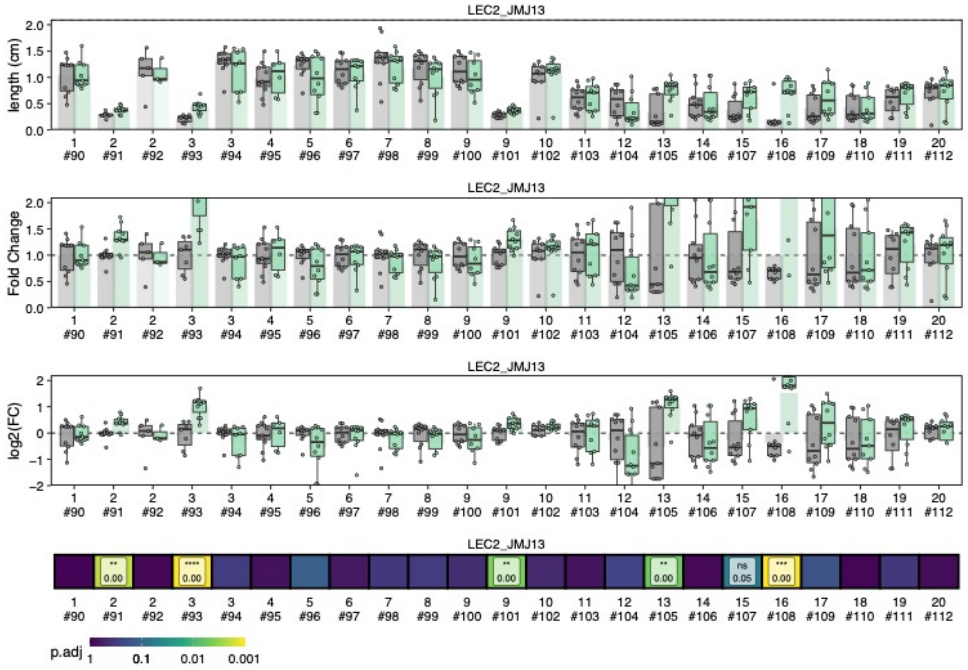


Figure 6.25: Barplot representing the comparison between the root length at day three for JMJ13-LEC2 lines and the control col0. Below the barplot are shown the lines which have significative difference from the control from T-test; p-value <0.05.

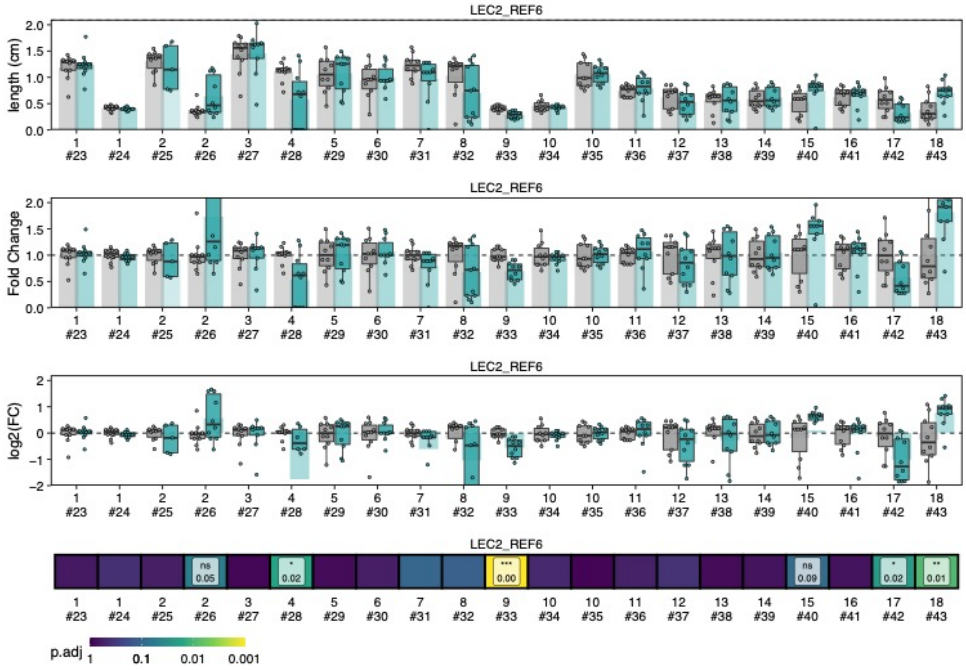


Figure 6.26: Barplot representing the comparison between the root length at day three for REF6-LEC2 lines and the control col0. Below the barplot are shown the lines which have significative difference from the control from T-test; p-value <0.05.

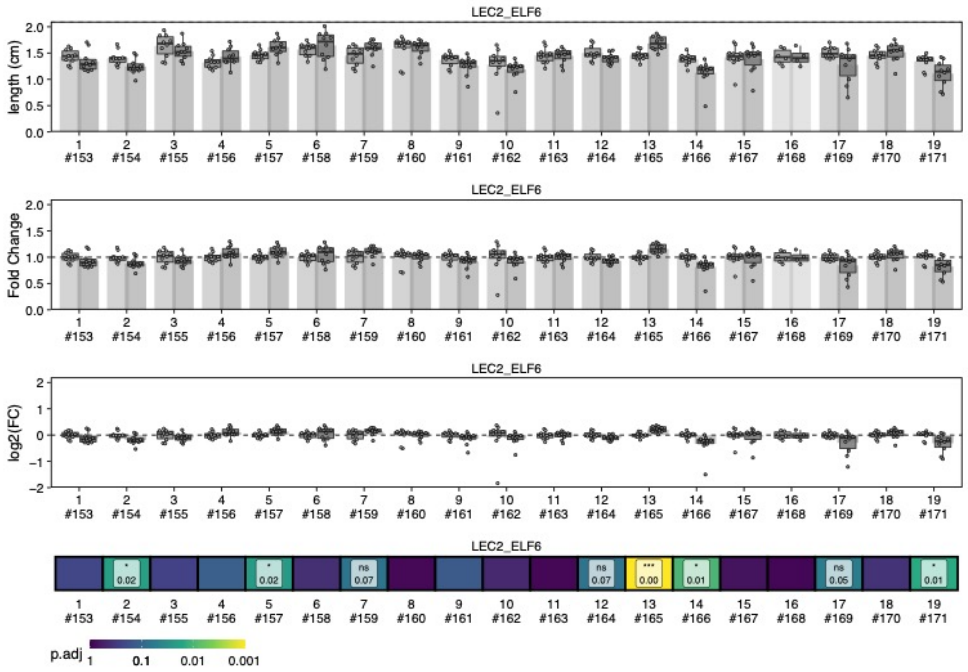


Figure 6.27: Barplot representing the comparison between the root length at day three for ELF6-LEC2 lines and the control col0. Below the barplot are shown the lines which have significative difference from the control from T-test; p-value <0.05.