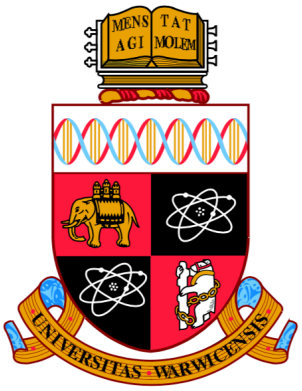

Long-range control of tissue specific gene expression in Arabidopsis

Nosheen Hussain

A thesis submitted in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy in Life Sciences

University of Warwick, Department of Life Sciences

September 2019

Abstract

The ability of plants to cope with a variety of environmental stresses depends on precise modulation of gene expression. Expression of stress related genes, usually, is under the control of long-range regulatory elements. In order to understand how gene regulation is modulated by these elements, it is important to identify distal regulatory regions and characterise them. One way to identify distal regulatory elements is to measure chromatin accessibility. We have devised a unique *in vivo* system (iNOMe-seq), to simultaneously measure chromatin accessibility, nucleosome occupancy and methylation profile. Our method outperforms current chromatin accessibility methods in identification of important genomic features with great accuracy in plants. Using our chromatin accessibility data, we identified potential distal regulatory regions (DREs) and a study was carried out to investigate their role in gene regulation by targeted epigenetic and genetic modifications at these regions. Our data shows that hypermethylation of distal regions leads to the down-regulation of gene expression and the phenotypic changes observed after DRE hypermethylation /CRISPR deletion were related to those seen in the mutants of their target genes. Our study concludes that the gene expression can be modulated by modifying their regulatory regions and this approach can be used to modify important crop traits to improve food security.

Contents

List of Figures

List of Tables

List of Abbreviations

bp	Base pair
cm	Centimeter
CREs	Cis regulatory elements
DREs	Distal regulatory elements
ddH2O	Double distilled water
dsRNA	Double stranded RNA
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
h	Hour
HAT	Histone acetyltransferase
kb	kilo bases
lincRNA	Long intergenic non-coding RNA
LB	Luria Bertani
mg	milligram
min	Minute
mL	Milliliter
mM	Milimolar
mRNA	Messenger RNA
Mb	Megabases
MS	Murashige and Skoog
OD	Optical density
PCR	Polymerase chain reaction
RdDM	RNA directed DNA methylation
RNA	Ribonucleic acid
rpm	Revolutions per minute
RO	Root origin
siRNA	Small interfering RNA

TE	Transposable element
TF	Transcription factor
µM	Micro molar

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. The work done in this thesis is original and has not been published or presented for any other qualification. The work presented is done by the author except the one described below.

Dr Julia Englehorn (Max Planck Institute / University of Duesseldorf)

Assisting in DMR selection (section 4.3.1, Table 4.1) and designing M.CViPI construct (section 3.2.2)

Dr Cora McAllister (University of Michigan, USA)

Phenotypic analysis of HRGP1-RNAi line (Figure 4.12B)

Javier Antunez Sanchez (University of Warwick)

Generation of Integrative genome viewer figures (Figure 4.3B, Figure 4.7B, Figure 4.13B)

Ryan Merritt (University of Warwick)

BSeq data Analysis (section 2.15.4, 2.16, 2.17, Figure 3.9- 3.19, Figure 7.6-7.16)

Acknowledgements

I am immensely grateful to my supervisor, Dr. Jose Gutierrez-Marcos for guiding me throughout the project. I also thank him for his valuable suggestions for research ideas. I am also thankful to my advisory panel members, Andre Pires da Silva and Graham Teakle for their valuable feedback during my advisory panel meetings. I am grateful to Julia Englehorn and Julius Durr for their much-needed guidance to learn important techniques. I also want to thank Gary Grant for his support in PBF and all C30 lab members for their help, support and nice company whenever I needed the most. Finally, my special thanks to my husband Hussain for his patience and morale-boosting encouragement and my lovely boys Abdul and Taimoor for being very understanding and source of immense love and support.

1. General Introduction

1.1. General overview

Plants are incredibly important organisms and account for 80% of human diet, thus, are essential for food security (United Nations 2017). The world population is increasing at an alarming rate and is likely to reach around 10 billion by 2050 (United Nations 2019). To feed an exponentially growing human population, a substantial increase in crop production is required. This puts pressure on an already squeezed agricultural capacity due to urbanisation of land and severe environmental changes such as heat and drought affecting crop production (Fahad et al. 2017). Extreme weather conditions negatively affect plant development and plants have to adapt to the changes in a physiologically challenging way, which results in lower crop yield (Dresselhaus and Hückelhoven 2018). Environmental stresses such as cold, heat, drought and salinity make plants more vulnerable to pathogens by altering plant physiology and pathogen defence mechanisms (Seherm and Coakley 2003). The ability of plants to cope with an array of environmental stresses depends on the precise modulation of gene expression (Asensi-Fabado et al. 2017). Gene regulation relies not only on regulatory sequences and sequence-specific transcription factors, but on a variety of epigenetic processes and chromatin organisation that interplay to modulate precise gene expression in response to environmental cues (Ariel et al. 2014).

In the introduction, basic gene regulation processes in eukaryotes will be described and different ways to study chromatin accessibility and identification of regulatory elements and their roles in regulating gene expression in plants will be discussed.

1.2. Regulation of gene expression

Quantitative and qualitative control of gene expression is crucial for all living organisms. In multicellular organisms, including plants, spatial and temporal regulation is critical to drive cellular differentiation and development, which leads to accurate morphogenesis (De Smet and Beeckman 2011). Control of gene expression can take place on different levels i.e. transcriptional, post-transcriptional, and translational stages (**Figure 1.1**) (Ohnishi 2012). However, transcriptional regulation is the first opportunity to control gene expression and therefore is the most important determinant for the expression of any gene (Halbeisen et al. 2007).

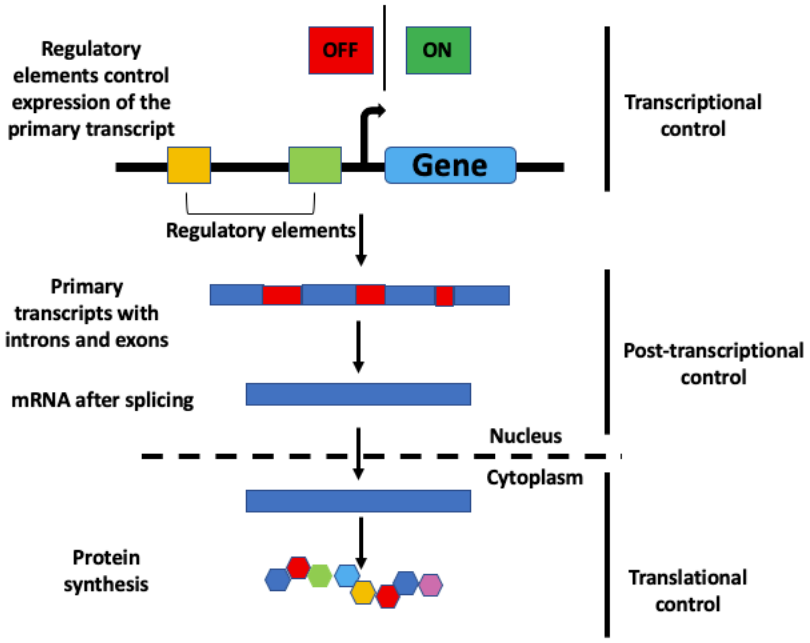

Figure 1.1: Overview of the gene expression regulation at various levels in a cell. Regulatory elements control "ON" or "OFF" state of a gene to control the expression of the primary transcript. RNA splicing leads to the formation of mRNA, which is later transported to the cytosol for protein synthesis.

1.3. Transcriptional control of gene expression

Eukaryotes display extensively elaborated transcriptional control mechanisms determined by a complex interplay among regulatory proteins, cis regulatory elements, DNA methylation and other epigenetic mechanisms (Alberts et al. 2002). In a given moment, in a typical human cell, 10,000-20,000 genes are expressed out of approximately 30,000 genes (Alberts et al. 2002). Also, the expression profile varies between different cell types and throughout the life cycle of the organism in a given cell (Wray et al. 2003). Transcriptional regulation determinants including regulatory proteins, cis regulatory elements, epigenetic modification and chromatin organisation are discussed as follows.

1.4. Transcription factors and regulatory proteins

Various regulatory proteins are involved in regulating chromatin interactions through their sequence specific DNA binding, including regulatory proteins forming transcription initiation complexes (Lee and Young 2000). Transcription factors are key regulatory proteins that work independently or make complexes with other proteins, which can either activate or repress the recruitment of RNA Polymerase to the promoter (Nikolov and Burley 1997). In eukaryotes, transcription initiation requires highly conserved Polymerase II (Pol II) and general transcription factors including TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Orphanides et al. 1996).

A variety of transcription factors (TFs) have been identified in eukaryotes. Based on the presence of conserved sequences of known DNA binding domains (DBDs), around 1500 genes encode TFs in the comparatively small genome of *A. thaliana* (Riechmann et al. 2000).

These transcription factors have been further grouped into more than 60 TF families in Arabidopsis consisting of more than 100 TFs with MYB, the largest family consisting of approximately 200 genes (Qu and Zhu 2006). MYB transcription factors have been implicated in many physiological processes, including development and morphogenesis, stress signalling and defence mechanisms (Yanhui et al. 2006). The MADs box family, which is not a plant-specific family includes genes responsible for flower development in Arabidopsis. for example, combinatorial action of floral homeotic genes, ABC, results in partial expression of these genes specified in discreet domains of the flower leading to accurate transition to flowering development (Irish 2017).

Many TFs carry DNA binding domains related to one another e.g. Zinc-finger domain which can fold into loop structures hence named 'fingers', were initially identified in Pol III TFs but are also been identified in Pol II TFs (Krishna et al. 2003). Zinc-finger TFs constitute half of the TFs identified in *D. melanogaster* and *C. elegans*, but only comprise 20% of all the TFs identified so far in Arabidopsis (Riechmann et al. 2000).

TF binding DNA sequences show various levels of binding specificity, ranging from single TF binding to one site, to many TFs recognising the same binding site (Wray et al. 2003). ChIP-seq (Chromatin Immuno Precipitation Sequencing) analysis in Arabidopsis has revealed that majority of genes (63%) are potentially regulated by more than one TFs which indicates the complexity of transcriptional gene regulation networks (Heyndrickx et al. 2014).

Nevertheless, technological advances have helped revolutionise the identification hundreds of putative TFs and their binding sites, but the accurate characterisation of TF binding sites is a complicated process. Precise characterisation requires input

from studying a variety of epigenetic mechanisms such as DNA methylation patterns, chromatin accessibility and 3D chromatin organisation studies to predict novel regulatory mechanisms in plants.

1.5. *cis*-regulatory elements

Every gene harbors flanking regulatory regions that, in conjunction with regulatory proteins, controls expression. These flanking regions are commonly known as cis-regulatory elements (CREs) (Adrian et al. 2010). CREs are stretches of non- coding DNA are normally present in the vicinity of protein-coding genes. Based on their position around the gene, CREs can be grouped as either proximal regulatory regions or distal regulatory elements (Wittkopp and Kalay 2012). Proximal promoter elements include a core promoter, a minimal section of promoter required for transcription initiation (Smale and Kadonaga 2003). Core promoters harbor a TSS (transcription start site), polymerase binding sites, and general TF binding sites such as a TATA box, although there is no such 'universal element' sequence that is conserved in all promoters (Juven-Gershon and Kadonaga 2010). Other proximal promoter elements found approximately 250 bp upstream of the TSS also contain primary regulatory elements and allow binding of specific TFs (Smale and Kadonaga 2003). Transcriptional regulation does not only depend on core promoter and proximal promoter elements but also depends on distal regulatory elements present far from their target genes. Based on their roles in gene expression, distal regulatory elements can be divided into enhancers, silencers or insulators (Heintzman and Ren 2009). Enhancers and promoters show similarities in possessing DNA biding motifs for tissue specific TFs and, contrary to CpG rich promoters, tissue specific promoters and enhancers carry DNA binding motifs that are CpG depleted (Taher et al. 2013).

Distal regulatory elements in coordination with certain chromatin states regulate temporal and tissue specific gene expression. In plants for example, *FLOWERING LOCUS T (FT)* is regulated via a 5kb upstream region by a complex interplay between a conserved distal DNA block and dynamic chromatin modifications (Adrian et al. 2010).

1.6. Epigenetic modifications and regulation of gene expression

Gene expression regulation is very dynamic and complex, not only due to the above-mentioned various elements and the interplay between their associated factors, but also due to the role of an array of epigenetic marks. Epigenetic marks are responsible for heritable changes in gene expression without involving any changes in DNA sequence. Epigenetic modifications can alter chromatin accessibility, which will allow modulation of transcription and thus may eventually cause phenotypic variation (Seymour and Becker 2017).

Unlike the genome, where the whole set of genetic information remains largely stable throughout the lifetime of an organism, the epigenome exhibits temporal and tissue specific variation (Gibney and Nolan 2010). In early stages of development, differentiating cells start accumulating epigenetic marks that are different from the undifferentiated cells, resulting in different lineages of cells carrying different epigenetic marks (Reik 2007). Epigenetic modifications also arise in mature organisms under the influence of environmental changes or after sensing developmental cues (Jaenisch and Bird 2003). Although the primary DNA sequence is not altered, these modifications are maintained through mitotic divisions and may remain heritable through many subsequent generations (Pikaard and Mittelsten

Scheid 2014). In the model plant *A. thaliana*, it has been demonstrated that epigenetic changes induced by exposure to osmotic stress evade epigenetic reprogramming in the subsequent generation and results in 'priming' of plants to the same type of stress (Wibowo et al. 2016). This indicates that epigenetic marks acquired in response to various environmental cues are potentially inherited and affect how plants adapt to changing environmental conditions.

Three distinct mechanisms are associated with deployment and maintenance of epigenetic modifications across the genome: DNA methylation, histone modification and RNA based epigenetic control (Henderson and Jacobsen 2007).

1.6.1. DNA methylation

DNA methylation is a covalent modification of DNA, where a methyl (CH3) group is added to the fifth carbon atom of the cytosine ring; a process carried out by DNA methyltransferases (Moore et al. 2013). It is a highly conserved epigenetic phenomenon in both animals and plants, it is vital for transposon silencing, stable repression of many genes and also critical for normal development (Chan et al. 2006).

In animals, DNA methylation is present only in a CG contexts throughout the genome, especially in short CG rich regions called CpG islands that are frequently found in gene promoter regions (He et al. 2011). A classic example of gene regulation by differential DNA methylation states is of *Agouti* gene expression in mice. The *Agouti* gene is flanked by a CG rich region at the 5'end, after birth, methylation of the CG rich region leads to a brown fur phenotype in developmentally normal mice while absence of methylation leads to developmental abnormalities such as yellow fur, obesity, and diabetes (**Figure 1.2**) (Dolinoy et al. 2006).

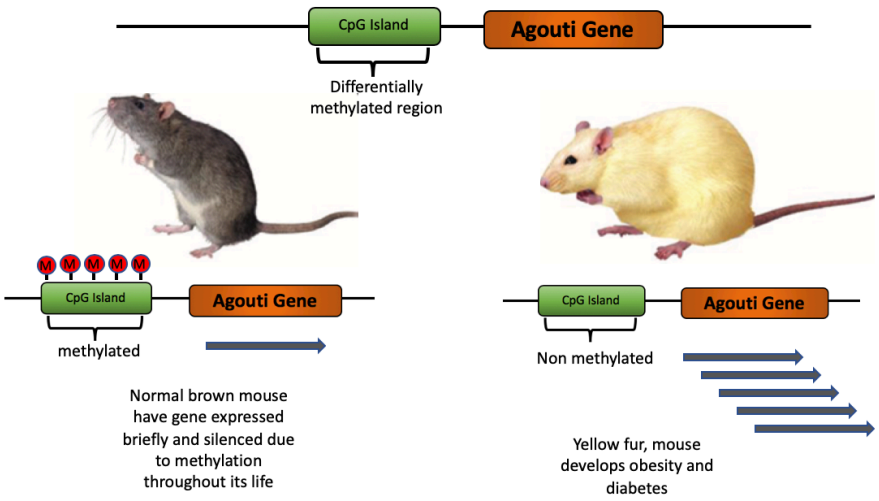

Figure 1.2**: DNA methylation dependent regulation of the Agouti gene in mice**. The *Agouti* gene is flanked by CG rich regions called CpG islands. Presence of an unmethylated CpG region can result in ectopic expression of *Agouti,* resulting in developmental abnormalities such as yellow fur, obesity and diabetes while a methylated state of CpG results in controlled expression/silencing of *Agouti,* leading to normal brown fur mice.

In plants, where highest levels of DNA methylation occurs i.e. 50% of cytosine methylation in some plant species, the phenomenon is observed in CG, CHG and CHH (H= A, C, T) contexts (Suzuki and Bird 2008).

De novo methylation in plants mainly at CG sites is carried out by a DNA METHYLTRANSFERASE 1 (DNMT1) homologue called DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), while the maintenance of methylation at CG and CHG is carried out by a DNMT1 homologue called MET1 (DNA METHYLTRANSFERASE 1) and a plant specific methyltransferase known as CMT3 (CHROMOMETHYLASE 3), respectively (Pikaard and Mittelsten Scheid 2014). CHH methylation is asymmetric in nature and is methylated by DRM2 in a *de novo* fashion (Goll and Bestor 2005). In both plants and animals, methylation is prevalent in centromeric, pericentromeric regions as well as in repeat elements such as transposons and is conventionally

associated with gene repression (Gibney and Nolan 2010). MET1 is highly conserved in plants and *met1* mutants in *A. thaliana* show a range of developmental abnormalities due to a lack of methylation at CG sites (Kankel et al. 2003).

Loss of methylation in plants can lead to epiallelic variation (heritable phenotypic changes brought about by epigenetic modification), that if present after meiotic transmission can lead to deleterious effects such as flowering time variation, leaf and floral defects and other developmental aberrations (Henderson and Jacobsen 2007). Hypomethylation also results in enhanced transposon activity, which leads to increased insertional mutations with potentially deleterious effects such as changes in meristem identity and female sterility (Ronemus 1996).

The impact of DNA methylation on gene expression in plants is however complex due to the fact that correlation of gene expression with DNA methylation differs between gene body methylation and methylation of TEs flanking genes. Gene body methylation generally results in upregulation of expression, in contrast to TE methylation being negatively correlated with gene expression, exceptions in both cases, however, are also present (Meng et al. 2016).

1.6.2. Histone modifications

Histones (H1, H2a, H2b, H3 and H4) are a crucial component of chromatin, where DNA is packaged around four histone pairs into a chromosomal structural unit called nucleosome (Dong and Weng 2013). Histones, mainly, H3 and H4, have long polypeptide tails protruding from the nucleosome structure, which are covalently modified post-translationally at different amino acid residues (Pfluger and Wagner 2007). These modifications, which include acetylation, methylation, phosphorylation

and ubiquitylation, can repress or activate gene expression by changing chromatin configuration (Berger 2007).

There are many ways in which histone modifications can alter gene expression, and predominantly three different phenomena are believed to be involved: i) by directly altering the chromatin structure, ii) by interfering with the binding of proteins that normally bind chromatin and iii) by facilitating binding of certain other proteins to the chromatin (Kouzarides and Berger 2007).

Acetylation of histones involves modifying the lysine residues of histone amino-terminal tails and is generally associated with up-regulation of transcription, mainly by inhibiting formation of secondary or tertiary chromatin structure. thus enabling chromatin to have a more open conformation (**Figure 1.3**) (Rice and Allis 2001). Although originally studied for its role in chromatin compaction during cell cycle, phosphorylation of H3 in eukaryotes, has been is found to be involved in chromatin remodeling, especially for the relaxation of chromatin to facilitate transcription (Wei et al. 1998). Phosphorylation of histone residues is positively correlated with gene transcription; for example, phosphorylation of serine 10 and 28 of H3 is involved in the regulation of epidermal growth factor (EGF) in mouse (Rossetto et al. 2012).

Histone methylation, opposite to phosphorylation, was initially thought to be associated specifically with gene repression. Although, over the last two decades, studies have revealed that histone methylation can have either positive or negative effect on gene regulation based on the position of the methylated amino acid residues (Zhang and Reinberg 2001). For instance, methylation of H3 at lysine 4, 36 and 79 is associated with gene activation whereas methylation of lysine residue 9

and 27 of H3 (Figure 1.3) are involved in gene repression (Gibney and Nolan 2010; Perrone et al. 2014).

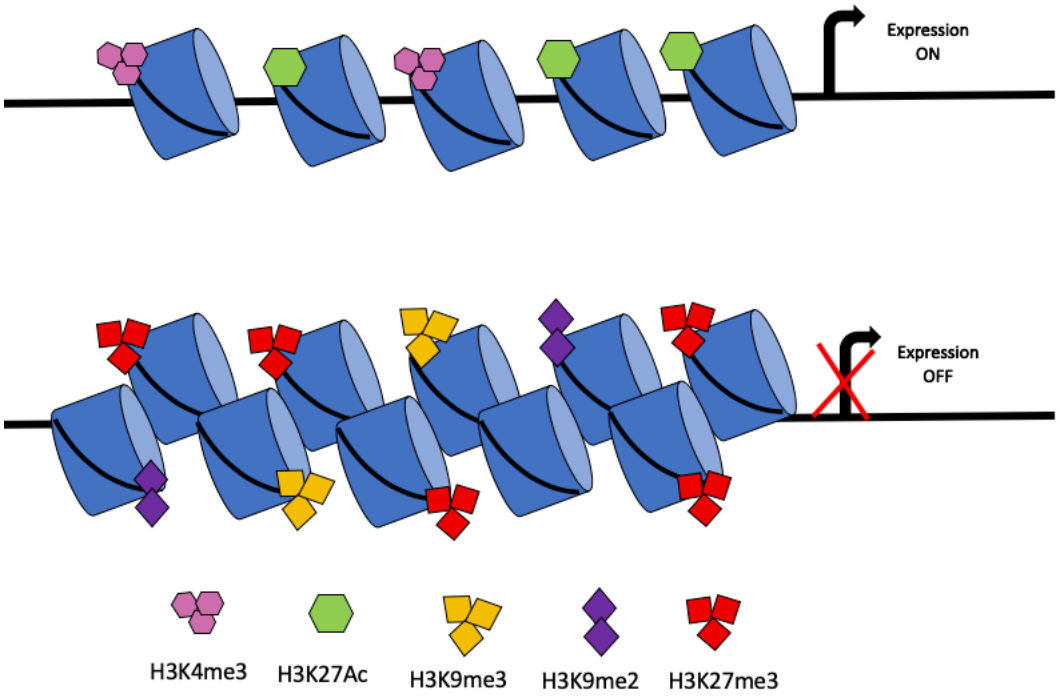

Figure 1.3: **Histone modifications leading to activation or repressive state of chromatin in eukaryotes**. H3K27Ac3 and H3K4m3 results in an open chromatin state leading to activation of genes whereas, H3K27me3, H3K9me2 and H3K9me3 lead to compact chromatin states and the repression of genes . Image adapted from (Boland Michael et al. 2014)

1.6.3. Interplay between DNA methylation and histone modification

There is evidence that the DNA methylation machinery and methyl CpG binding proteins (MBDs) recruit histone deacetylases (HDACs), reinforcing two layers of gene repression (Fuks 2005). However, it is not completely understood exactly how DNA methylation and histone acetylation/deacetylation are established during development. In plants, there is evidence that plant specific chromodomain methyl transferase (CMT3) is involved in gene repression by DNA methylation and by

facilitating heterochromatin formation (Liu et al. 2011). CMTs carry a chromo-domain which can bind to a K9 histone methyl transferase, SUVH4 (Liu et al. 2014; Nielsen et al. 2002), which indicates that interplay between histone marks and DNA methylation is a conserved phenomenon in eukaryotes.

According to the currently accepted model, histone acetylation facilitates an open chromatin structure thus allowing the transcriptional machinery to access the promoter. Removal of acetylation by HDAC (Histone deacetylase) can induce other epigenetic modifications e.g. DNA methylation, leading to gene silencing (Vaissiere et al. 2008) (**Figure 1.4A**). On the other hand, hemi-acetylated histones can recruit histone methyltransferase and DNA methyltransferases to enable histone methylation and DNA methylation respectively, which leads to compact chromatin formation and gene silencing (**Figure 1.4B**) (Vaissiere et al. 2008).

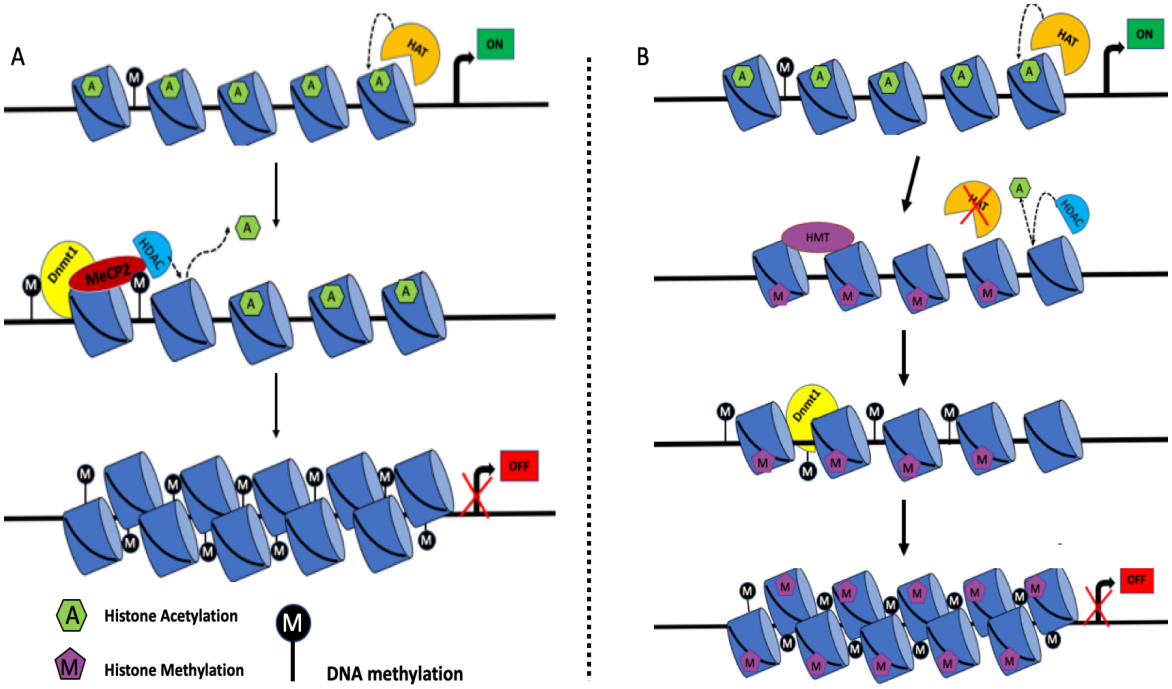

Figure 1.4. Models of histone modification and DNA methylation dynamics during gene silencing in eukaryotes (A) Partial methylation of DNA allows MePC2 (Methyl binding proteins) to bind to the methylated CpG region and help recruit histone deacetylases (HDACs) that deacetylase histones. In the presence of methyltransferase (Dnmt1) induced methylation, HDACs amplify gene silencing on the target region. **(B)** Deacetylated chromatin is recognised by Dnmt1, a *de novo* methyl transferase and histone methyltransferase HMT which leads to compact chromatin and gene silencing. (Image adapted from Vaissière et al. 2008)

1.6.4. RNA-based processes involved in epigenetic regulation

Non-coding RNAs (ncRNAs) are being recognised as important biological role players and are increasingly implicated in the epigenetic regulation of gene expression (Mercer et al. 2009). Long non-coding RNAs (lncRNAs) are RNAs longer than 200 nucleotides, stably repress certain genes by modulating DNA methylation of cytosine residues at CpG dinucleotides (Law and Jacobsen 2010).

LncRNAs are also implicated in the formation of chromatin modifying complexes by interacting with various proteins. In the HOXC (Homeobox C cluster of genes) cluster; *HOTAIR* (HOX transcript antisense RNA- a lncRNA) interacts with PRC2 (Polycomb repressive complex2) to repress *HOXD* genes in trans (Rinn et al. 2007).

There are also many different types of small non-coding RNAs generally derived from lncRNAs. These can be broadly categorised into the micro RNAs (miRNAs) and small interfering RNAs (siRNAs) in plants, though PIWI interacting RNAs (associated with proteins that harbour PIWI domain) and other lesser known types are found in animals. These small RNAs are typically 21-24 nucleotides in length and are actively involved in gene regulation at various levels (Gibney and Nolan 2010). In fission yeast, small RNAs have been found associated with various chromatin modifying complexes, facilitating the delivery of chromatin modifying enzymes to targeted regions (Verdel et al. 2004).

In the model organism, *Schizosaccharomyces pombe,* small RNAs have been found to be associated with chromatin modifying complexes and are implicated in the delivery of chromatin modifying enzymes to chromatin (Verdel et al. 2004). Silencing of transposons and repetitive sequences in eukaryotes is also mediated by different types of non-coding small RNAs, microRNAs and siRNAs that regulate gene

expression at different levels, by regulating, for instance, promoter activity, mRNA processing and translational interruption (Schwab et al. 2006). siRNAs especially are implicated in a number of epigenetic processes in plants i.e. especially their role in *de novo* DNA methylation via the RNA directed DNA methylation (RdDM) pathway (Zhang and Zhu 2011).

1.7. siRNAs directed DNA methylation in plants

siRNA directed DNA methylation was first discovered in tobacco plants infected with viruses (Wassenegger et al. 1994). This work revealed that *de novo* DNA methylation occurs frequently in plants and can be induced by the expression of double stranded RNA originating from inverted hairpins. The discovery of RNA interference allowed the engineering of plants lacking DNA methylation that could be propagated over multiple generations (Teixeira et al. 2009). It is now known that sequence specific siRNAs in plants can guide DNA methyltransferases to deposit new DNA methylation (Lu 2005; Zilberman et al. 2007) via RdDM. RdDM consists of three distinct components: (i) siRNA biogenesis (ii) production of scaffold RNA (iii) formation of guiding complex by the interplay of scaffold RNA and ARGONAUTE proteins (Kanhere et al. 2010). To achieve this, plants have evolved two atypical DNA polymerases, named Pol IV and Pol V, which are implicated in the formation of long non-coding RNAs to initiate siRNA biogenesis (Lahmy et al. 2010) and later direct the deposition of DNA methylation (Matzke and Mosher 2014).

The biogenesis of siRNA is initiated by a single stranded RNA (ssRNA) that is converted into double stranded RNA (dsRNA) by a RNA DEPENDENT DNA POLYMERASE 2 (RDR2) and later, in 24-nucleotide siRNAs by the activity of a DICER LIKE 3 (DCL3) protein (Law and Jacobsen 2010). These siRNAs are loaded onto

Argonaut 4 (AGO4) and bind to scaffold RNA (generated by the activity of Pol V), via sequence complementarity to form a guiding complex (**Figure 1.5**). The complex ultimately mediates *de novo* methylation by recruiting DNA methyltransferase to the target sequence (**Figure 1.5**) (Zhang and Zhu 2011).

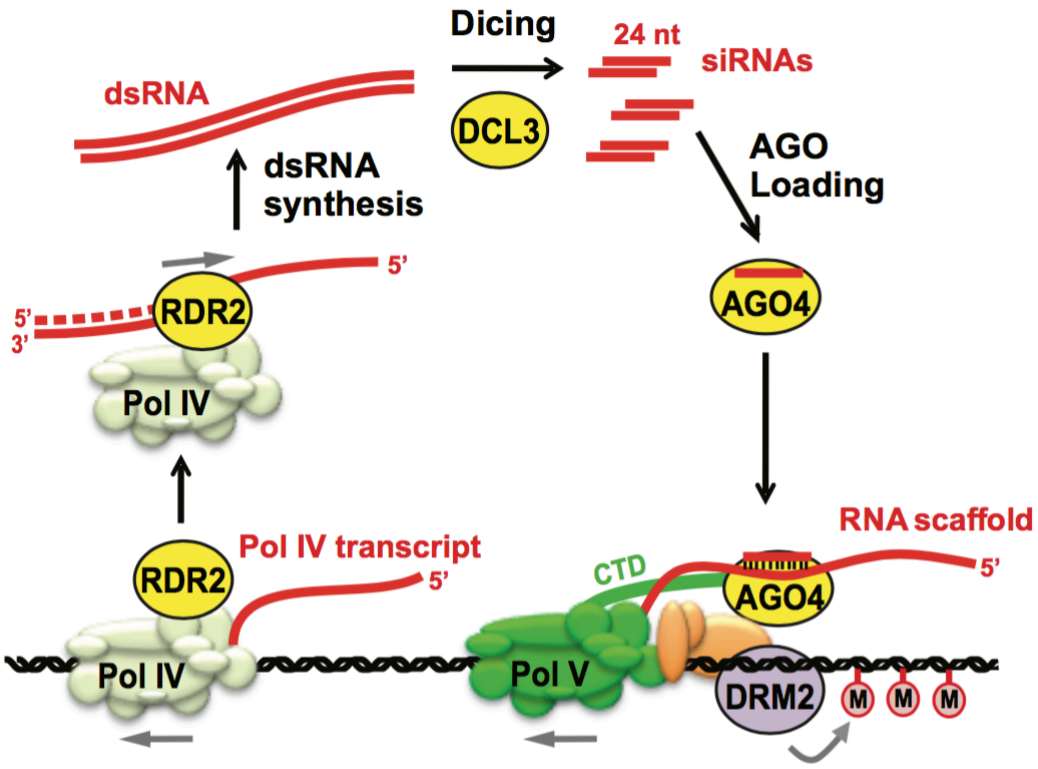

Figure 1.5. Overview of RNA directed DNA methylation (RdDM) pathway in plants. Pol IV generates single stranded RNA (ssRNA) and RDR2 converts ssRNA to double-stranded RNAs (dsRNA). These dsRNAs are diced by DCL3 into 24 nt siRNAs. Later siRNAs are loaded into AGO4 and are guided by PolV transcripts to the target DNA sequence where methyltransferases like DRM2 carries out methylation of Cytosine. (Image adapted from Blevins et al. 2015)

1.8. Chromatin accessibility and gene regulation

The conformation of chromatin is responsible for highly packaged DNA being able to fit into the nucleus of a microscopic cell. This high order chromatin organisation offers a physical barrier to DNA access by various transcription factors throughout the cell cycle (Biswas et al. 2011)

In-depth structural analysis has revealed that the nucleosome, which is the smallest subunit of chromatin, comprises 8 core histones (2 of each histone H2A, H2B, H3 and H4) wrapped with 147bp of DNA and with linker protein H1 sitting at the interface of the core proteins and the DNA entry and exit points (Van Holde et al. 1974). The core histones are globular in nature but have unstructured tails which are subjected to a variety of post translational modifications, mainly at their N terminal extensions (See 1.3.2) (Kouzarides 2007). These modifications include methylation, acetylation and phosphorylation, which play a key role in genome accessibility (Strahl and Allis 2000). The combined effect of these modifications along with DNA methylation, chromatin re-modelling and the presence of histone variants leads to more or less compact chromatin (**Figure 3.1**), resulting in differential gene expression (Clapier and Cairns 2009).

Less compact or open chromatin harbours actively transcribing genes which have distinct nucleosome depleted regions (NDRs) near transcription start sites, thus allowing various transcription factors to bind and initiate transcription (Yuan et al. 2005). Dynamic permissible chromatin is also characterised by the presence of regulatory regions, which carry binding sites for various regulatory proteins. Upon interaction with those regulatory proteins, chromatin is remodelled and displays

higher accessibility to various transcription factors and Pol II, resulting in up-regulation of certain genes (**Figure 1.6**) (Klemm et al. 2019).

The process that maintains dynamic chromatin is termed as 'nucleosome turnover' and this allows removal and the replacement of histones (Guillemette et al. 2005). Nucleosome turnover is rapid in nucleosome present closer to the promoter regions, in gene bodies and in regulatory regions (Deal et al. 2010).

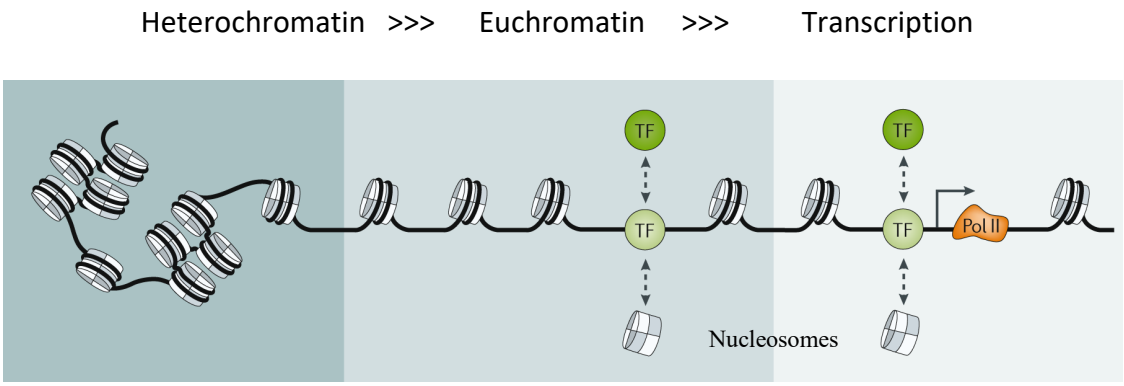

Figure 1.6: **Nucleosome turnover leads to the formation of highly dynamic chromatin.** Various chromatin remodelling factors such as INO80 and SWR1 complexes in plants (Wang et al. 2019), lead to decompaction of inaccessible chromatin through a process called nucleosome turnover. This leads to exclusion of older histones with newer ones, as a result, chromatin is remodelled and displays higher accessibility for Pol II and various transcription factors, resulting in up regulation of certain genes (Image adapted from Klemm et al. 2019)

1.9. Measuring nucleosome occupancy and regulatory landscape

Mapping of nucleosome occupancy or positioning indicates where nucleosomes are present in relation to the DNA sequence within the genome. The organisation of nucleosomes across the whole genome is not uniform and NDRs often harbour binding sites for transcription factors, RNA polymerases or other remodelling complexes, to regulate DNA access. Moreover, various studies have revealed that regulatory elements, such as enhancers, work independently of the distance of their target genes and their promoters. Therefore enhancers can be present several base pairs upstream or downstream, in intergenic regions, or even within the gene body (Zhu et al. 2015). Generally, the algorithms predict TF binding DNA motifs by searching shared cis-regulatory elements. These over-represented sequences are used to build position-specific weight matrix (PSWM) that specifies bases for each position in the matrix to identify conserved motifs by using statistical methods such as Gibbs sampling and expectation maximisation (Karnik and Beer 2015).

However, this makes it very challenging to identify non-conventional regulatory elements because these algorithms rely on specific DNA sequences (transcription binding sites) and distal regulatory elements may not fall within the parameters of identification algorithms (Maher et al. 2018). Therefore, it is important to couple nucleosome occupancy measurements with the available computational methods to make a robust mechanism of identification of regulatory landscape.

Currently, there are a number of methods available to measure chromatin accessibility. These include but are not limited to: DNase I Hypersensitive sites sequencing (DNase-seq), Micrococcal Nuclease sequencing (MNase-seq), Assay

for Transposase-Accessible Chromatin sequencing (ATAC-seq) and Nucleosome Occupancy and Methylome sequencing (NOMe-seq)

1.9.1. DNase I Hypersensitive sites Sequencing (DNase-seq)

Deoxyribonuclease I (DNase I) is an endonuclease encoded by human gene *DNASE1* that non-specifically cuts DNA on nucleosome depleted regions, also known as DNase 1 hypersensitive sites. Despite a lot of variation in in DNase-seq applications, this is a powerful method to study chromatin accessibility at the whole-genome scale. Current work flow which has been adapted from Boyles et al, (2008), involves DNase 1 treatment of genomic DNA followed by short read sequencing (**Figure 1.7A**). The genome wide nucleosome occupancy measurements have revealed that only a small minority of DHSs to be found in proximal promoter regions i.e the promoter itself and the transcription start sites (TSS). The majority (over 80%) of the accessible regions have been found in distal regions (Thurman et al. 2012). The method however, requires exceptionally large number of nuclei (1-50 million cells) and due to the requirement of multiple purification steps, DNA loss can occur, limiting its sensitivity (Lu et al. 2016a)

1.9.2. Micrococcal Nuclease Sequencing (MNase-seq)

Micrococcal nuclease (MNase) was the first DNA digestion enzyme used to study accessible chromatin and was initially isolated in the 1960s from *Staphylococcus aureus* (Heins et al. 1967). Nucleosome depleted DNA is sensitive to MNase digestion, in contrast to Nucleosome associated DNA, which makes it suitable to carry out genome wide nucleosome occupancy profiling (Pajoro A. et al. 2018). MNase acts as both exonuclease and endonuclease which has the ability to make a single stranded break followed by a double stranded break if the DNA is not

protected by the nucleosome or other DNA binding proteins (Heins et al. 1967). Like DNAse-seq, this method involves digestion of DNA with micrococcal nuclease coupled with high throughput sequencing, which can be analysed to reveal NDRs on a whole genome level (**Figure 1.7B**). MNase-seq generated reads have been found to be highly enriched at active regulatory regions i.e TSS, and proximal promoters but not at the 3' end of active genes or within gene bodies (Mieczkowski et al. 2016). This method also requires to have large number of cell nuclei (1-10 million) and its aggressive nuclease activity can lead to problems with the optimisation of digestion time and conditions (Chereji et al. 2019).

1.9.3. Assay for Transposase-Accessible Chromatin Sequencing (ATAC-seq)

MNase-seq and DNase-seq are time consuming, require sensitive enzymatic treatment, and pose a challenge of complicated sample preparations (Tsompana et al. 2104). Both methods also require millions of cells for nuclei isolation, which poses challenges such as heterogenicity of input material, meaning that some cell types which are rare will fail to meet the requirement of minimum input material for chromatin mapping.

A gold standard method to map genome accessibility, that addresses all of the above problems is an assay of transposase accessible chromatin (ATAC-seq) which employs an enzyme called hyperactive Tn5 transposase. Tn5 transposase is preloaded with sequencing adapters and can simultaneously cut DNA and ligate adapters to the fragmented DNA, a process called tagmentation, followed by next generation sequencing (**Figure 1.7C**) (Buenrostro et al. 2015). The method requires an average sample size of 50,000 cells in comparison to a million cells required for the other two

previously discussed methods. ATAC-seq works in different cell types and species and is also becoming popular in single cell analysis (Buenrostro et al. 2015).

1.9.4. Nucleosome Occupancy and Methylome Sequencing (NOMe-seq)

NOMe-seq is a unique assay which is different to the above-mentioned methods in that it does not involve any digestion or breakage of DNA fragments. This technique employs a viral methyltransferase M.CViPI that can methylate GpC motifs throughout the whole genome. MCViPI induces methylation, and when coupled with bisulfite sequencing, reveals the methylation status of the genome, which indicates the location of accessible sites as determined by the methylation status of GpC motifs (**Figure 1.7D**) (Kelly et al. 2012).

This method offers many advantages over other methods as NOMe-seq can identify nucleosome occupancy and the epigenetic state of the genome simultaneously. It requires fewer cells and is suitable for single cell chromatin analysis (Pott 2017), and is much less time consuming than other methods(Lay et al. 2018). However, the biggest disadvantage of NOMe-seq, which prevents its effective application in plants, is the requirement to isolate nuclei prior to enzymatic treatment. Due to the presence of cell wall, plant tissue poses a challenge for nuclei isolation. Therefore, it is important to devise new methods for chromatin accessibility in plant tissues.

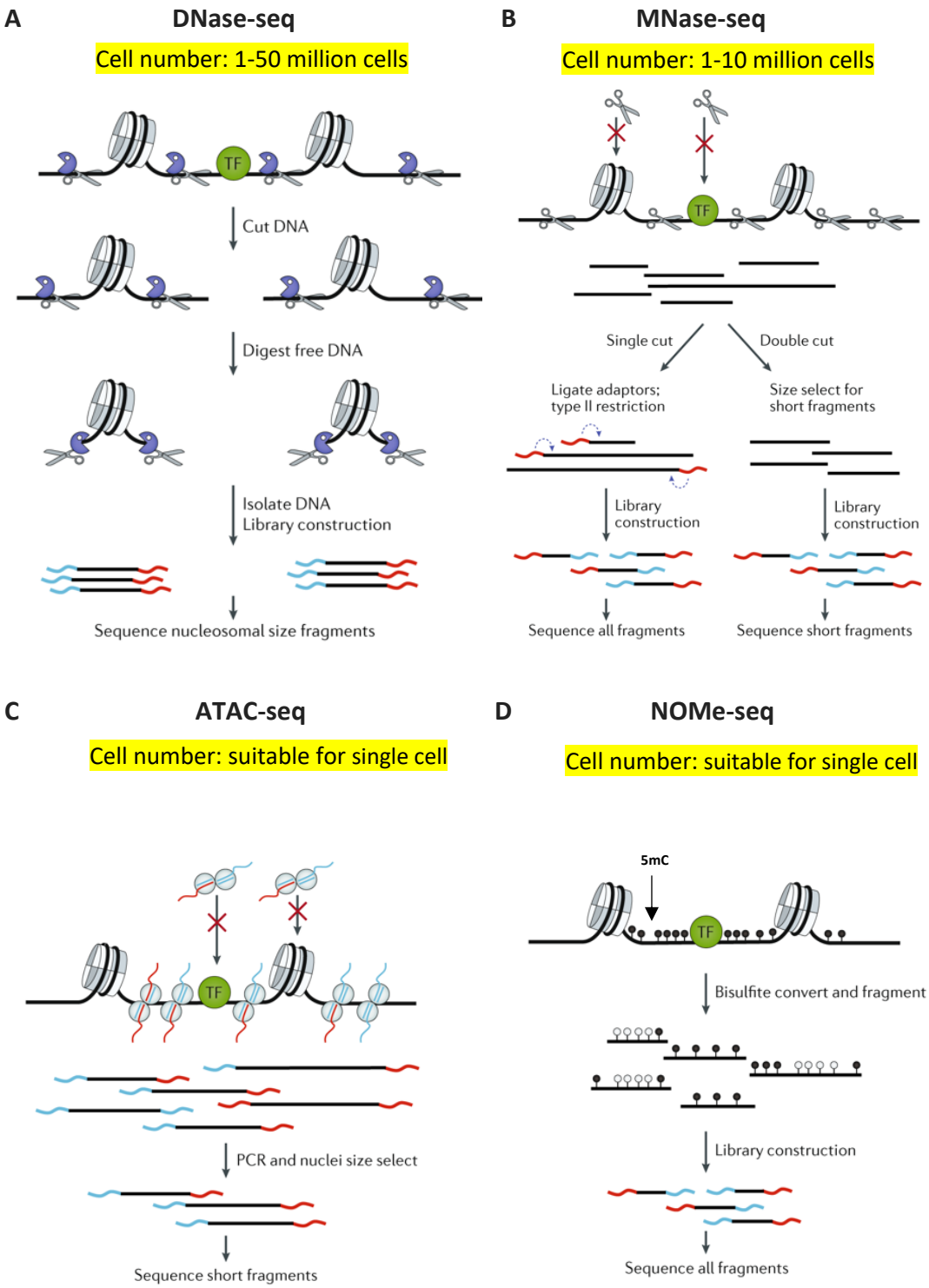

Figure 1.7: Various methods to study chromatin accessibility. (A) DNase-seq involves DNase 1 treatment of genomic DNA followed by short read sequencing (B) **MNase seq** requires digestion of DNA with micrococcal nuclease coupled with high throughput sequencing to obtain lots of short sequence reads which can be analysed to determine chromatin accessibility. (C) **ATAC-seq** employs a hyperactive Tn5 transposase that can simultaneously cut DNA and ligate adapters to the fragmented DNA- a process called tagmentation, which is followed by next generation sequencing. (D) **NOMe-seq** relies on a methyltransferase M.CViPI to methylate GpC motifs throughout the genome followed by bisulfite sequencing to reveal methylation profile and accessible chromatin simultaneously (Image adapted from Klemm et al. 2019)

Because of the current limitations of the available methodologies to study chromatin function, it can be hypothesised that there are many distal regulatory elements yet to be discovered and functionally characterised in plants.

1.10. Project aims

The aim of this thesis is to develop a method based on NOMe-seq to effectively map chromatin accessibility in plants. Furthermore, using this technique in *A. thaliana,* it would enable the identification and functional characterisation of known and novel long-range regulatory sequences, and to determine how these regions are influenced by epigenetic modification. Collectively, this work will shed light on the complex regulatory landscape of plant genomes

2. Materials and methods

2.1. Plant lines used in the study

For chromatin accessibility analysis, we generated ß-estradiol inducible transgenic lines for the controlled ectopic expression of MCViPI (M.CViPI-pMDC7-OLE1). To study role of the distal regulatory region in gene regulation, two hypermethylated lines, MeC_DRE-NAC82 (Me_DRE-NAC82-pJaw-OLE1) and MeC_DRE-HRGP1 (Me_DRE-HRGP1-pJaw-OLE1) were generated from root origin (RO) clonal plants (Wibowo et al. 2016). To carry out functional characterisation of the *HRGP1* gene, RNAi line, RNAi_HRGP1 (RNAi_HRGP1-pJaw-OLE1) was generated in the Col-0 background. To assess the function of regulatory elements, *apolo*^{CR}-1, *apolo*^{CR}-2 and DRE_HRGP1^{CR}-1 were generated using CRISPR/Cas9 system (Durr et al. 2018).

2.2. Growth conditions

A. thaliana plants were grown under long day conditions (16h/8h; day/night) with temperature set at 22^{o}C. Wild type Col-0 plants for protoplast generation were grown under short day conditions (12h/12h; day/night). For *in vitro* grown plants, seeds were sterilised in with 10% of sodium hypochlorite (V/V) for 5 minutes and washed in sterile water thoroughly 3 times. Plant growth media was prepared with half-strength MS (Murashige and Skoog (MS) salts) (Duchefa Biochemie), 1% sucrose (Sigma-Aldrich) and 0.8% Phyto agar (Duchefa Biochemie).

2.3. Bacterial strains and vectors

Escherichia coli (DH5α) was used for replicating different vectors. To stably transform *A. thaliana* lines, we used *A. tumefaciens* strain GV3103. For stable

transformation of plant lines with an inverted RNAi hairpin, we used *A. tumefaciens* strain GV3101:: pMP90RK . Vectors used in study are as follows;

Table 2.1 List of vectors used in the study

Vector	**Description**
pDONOR-207	A gateway entry vector (**Figure 2.1**) used for initial cloning of M.CViPI construct (**Table 7.1**) and hairpin sequences (**Table 7.2**)
pEN-chimera-2x	A donor vector for cloning two sgRNAs in one vector under the U6 promoter (**Figure 7.1**)
pUbi-cas9	A vector expressing Cas9 under the Ubiquitin promoter. Used for protoplast transfection to test the efficiency of synthetic sgRNAs. (**Figure 7.2**)
pJaw-OLE1	A binary vector, capable of expressing an inverted RNAi hairpin (**Figure 7.3**)
pMDC7-OLE1	A binary vector with an XVE based inducible expression system using ß-estradiol (**Figure 2.1B**)
pEC1-Cas9-OLE1	A binary vector expressing Cas9 under an egg cell specific promoter (EC1) (**Figure 7.4**)

2.4. Generation of *M.CViPI* construct

The *M.CViPI* sequence (**Figure7.1**) was codon optimised for *Arabidopsis* and chemically synthesized (Genesys). We added a nuclear localisation signal at the N-terminal end of the protein and a HA-tag was added at the C terminus. The construct was cloned into pDNOR207 and then cloned into binary vector pMDC7-OLE1 by gateway recombination (**Figure 2.1**).

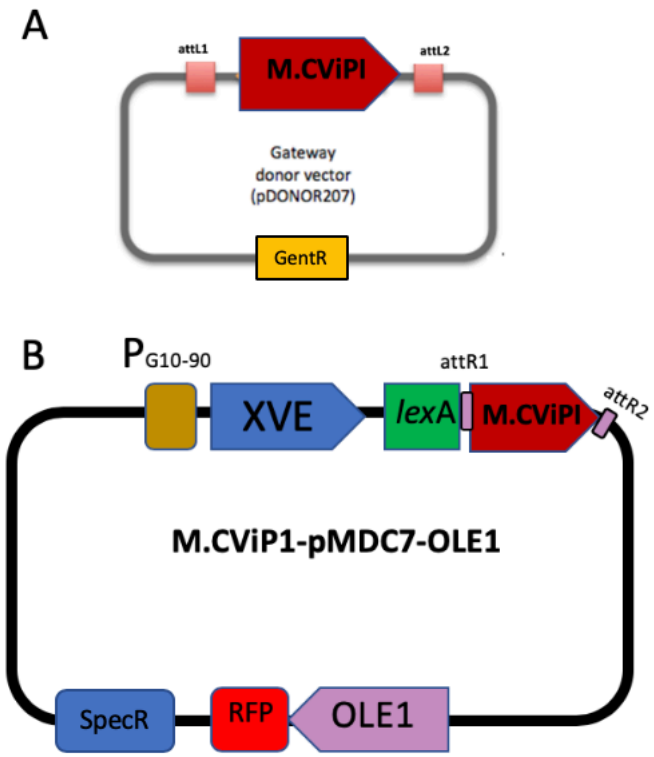

Figure 2.1 Cloning of *M.CViPI* sequence into gateway vectors. A) Synthetically designed sequence, NLS-*MCViP1*-3HA was engineered into an entry vector, pDONOR207. B) The *M.CViPI* construct from the entry vector was engineered into binary vector pMDC7-OE1 by LR cloning. A synthetic constitutive promoter G10-90 drives expression of a chimeric transcription factor, XVE, which consists of DNA-binding domain of the bacterial repressor LexA (X), a transactivating domain of VP16 (V) regulatory region of estrogen receptor (E). *XVE* controls activity of *lex*A promoter under influence of chemical induction by ß-estradiol.

2.5. Cloning reactions

BP and LR cloning was carried out according to the manufacturer's instructions (Invitrogen) and 5ul of reaction mixture was use to transform *ccdB* survival *E.coli* competent cells (TOP 10). 100-200ul of the transformed culture was plated on suitable selection media and incubated at 37^{o}C overnight. Cloning of sgRNAs for CRISPR/Cas9 system was carried out by restriction cloning. Initially, to clone sgRNA 'A', a gateway entry vector, pEN-2X-Chimera, was digested with restriction endonuclease *BpiI* (New England Biolabs) following manufacturer's instructions. After confirmation of accurate insertion of the first sgRNA, the plasmid carrying sgRNA 'A' was digested with *BsmBI* followed by cloning of sgRNA 'B'.

2.6. *E. coli* transformation

E. Coli strain DH5α was transformed by heat shock at 42^{o}C for 45 seconds and then plated on agar medium containing suitable selection antibiotic left overnight at 37^{o}C. The colonies growing on selection media were screened via colony PCR and sequencing.

2.7. Plasmid DNA extraction

Bacterial colonies were picked and inoculated into 5ml of LB overnight at 37 ^{o}C in suitable antibiotic selection. The bacterial cells were centrifuged at 5000g for 5 minutes and plasmid DNA extraction was carried out using Qiaprep Spin Miniprep Kit (Qiagen) following the manufacturer's manual. The plasmid DNA was stored at -20 ^{o}C for subsequent use.

2.8. PCR amplification of DNA

All PCR reactions were carried out in a MJ Research, PTC-225 Peltier thermal cycler. For genotyping of CRISPR deletions, 200ng of extracted DNA (in 2ul of TE buffer) was

used as a template. For colony PCR, multiple transformants from each construct were picked and re-suspended in 20ul of sterile water. 2ul from the mixture was used as DNA template in 25ul PCR mixture. PCR reaction mixture contained 0.15 units of Kapa Taq (KAPA biosystems), 2.5ul 10X PCR buffer, 0.3ul of 10mM dNTP mixture and 1ul of 5uM of each primer. The programme was set at 95^{o}C for 2 minutes initially for denaturation followed by 30 cycles at 95 ^{o}C for 30 seconds, 58^{o}C for 30 seconds and 72^{o}C with varied extension time depending on the length of target sequence (1kb per minute) followed by a final extension at 72^{o}C for 5 minutes. The amplified fragments were checked via gel electrophoresis.

2.9. Transformation of *A. tumefaciens*

Agrobacterium cells were transformed with the binary vector carrying DNA fragment of interest via electroporation. The procedure was performed in 0.1 cm cuvette at 2.2 kV with a Bio-Rad micropulser. 40ul of electro-competent cells of *A. tumefaciens* GV3131 were mixed with the binary vector by gently tapping the tube. After electroporation transformed competent cells were transferred into low-salt liquid LB medium and left for 1 hour at 28^{o}C. The cells were then spread on LB agar plates with suitable antibiotics and incubated at 28^{o}C for 48 hours.

2.10. Stable transformation of *A. thaliana*

A single colony from transformed *A. tumefaciens* strain that carries each DNA construct in a binary vector was inoculated in 5ml of low salt LB containing appropriate antibiotics. The overnight culture was used to inoculate 500ml low-salt liquid culture with appropriate antibiotics and grown overnight at 28^{o}C. Arabidopsis plants were transformed using the floral dip method described previously (Zhang et al. 2006). T0 seeds (RFP positive) were selected and grown to the T1 generation. The

subsequent lines that showed 3:1 RFP^+ to non-RFP seeds were selected as likely lines to carry a single insertion of the construct and they were grown to the T_2 generation. In the T_2 progeny, homozygous lines (lines with 100% RFP^+ seeds) were selected and propagated as stably expressing lines.

2.11. Heritable phenotypic analysis in M.CViP1-1a

Sterile M.CViPI-1a seeds were grown on ½ MS for 8 days. Seedlings were then transferred to ½ MS with 40uM ß-estradiol. The seedlings were grown for 4 days (termed as $\mathbf{I_1}$ seedlings) and transferred to soil. The seeds from $\mathbf{I_1}$ were grown to the $\mathbf{I_2}$ generation and phenotypically analysed. Seeds from the $\mathbf{I_2}$ progeny were grown to the $\mathbf{I_3}$ generation and phenotypically analysed. Selected plants from $\mathbf{I_2}$ were crossed with parental WT and the F1 generation was grown on soil for further phenotypic and molecular analysis.

2.12. Genomic DNA extraction

For genotyping of transgenic plants, high throughput DNA extraction was performed. Leaf samples of approximately $1cm^2$ were collected in a 96 well block with a 2mm diameter metal bead in each well. Samples were flash frozen and were ground at 18 revolutions per second for I minute. Then the block was centrifuged to allow tissue to settle in the bottom of each well. 300ul of DNA extraction buffer (100 mM Tris-Cl, pH 8.0; 50 mM EDTA, pH 8.0; 500 mM NaCl; 10 mM β-mercaptoethanol) was added in each well using a multi-channel pipette. After adding 40ul of 10% SDS, samples were mixed by vortexing and incubated at 65^{o}C for 20 minutes. Later, samples were then incubated on ice for 5 minutes before adding 100ul of 5M potassium acetate. Samples were further incubated for at least 20 minutes at ice. Samples were centrifuged at maximum speed for 20 minutes and 125 ul of supernatant was

transferred to a well in a fresh 96 well block. After adding 200ul of isopropanol, samples were incubated at -20 °C for at least 30 minutes. Samples were centrifuged at max speed for 15 minutes; isopropanol was removed and the pellet was washed with 70% ethanol. The DNA pellet was dried after removing 70% ethanol. Samples were resuspended in 100ul of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0 with 20ug/ul RNAse A (Invitrogen, UK).

For whole genome bisulfite sequencing, after ß-estradiol or mock treatment seedling root and leaf samples were collected from 20-30 seedlings at each treatment time. Root and leaf tissues were separated by cutting the seedlings in two at the lower part of the hypocotyl with a scalpel blade. The tissue samples were collected in 1.5ml Eppendorf tubes and are flash frozen in liquid nitrogen. DNA extraction was carried out using the Qiagen Plant DNeasy kit (Qiagen) and DNA was quantified using Qubit dsDNA HS assay kit (Invitrogen)

2.13. RNA extraction

Leaf tissue samples (50mg) were collected in 1.5ml Eppendorf tubes and flash frozen in liquid nitrogen and stored at -80°C. Frozen samples were ground with an electrical pulveriser. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) according to manufacturer's instructions. Quantification of total RNA was done using NanoDrop (Thermo Scientific). The quality of total RNA was analysed using gel electrophoresis.

2.14. qRT-PCR

Total RNA was treated a with TURBO DNA-free Kit (Invitrogen) according to manufacturer's instructions. 1ug of DNase treated RNA was used for cDNA synthesis.

cDNA was synthesised using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) according to manufacturer's instructions. For qRT-PCR, analysis was performed using a MyiQ System (BIO-RAD). qPCR primers were designed using the NCBI primer blast tool (**Table 7.2**) qPCR mixture was prepared to a final volume of 12.5µl containing 2.5µl of cDNA template, 0.2µM of each primer and 6.25µl of 2×MESA Blue qPCR MasterMix (Eurogentec Headquarters). All reactions were run in three technical replicates and a negative control was also run without DNA template. The thermal cycler programme included 95°C for 10 min then 40 cycles of 95°C for 10s, 60°C for 15s, and 72°C for 15s. A melting curve was calculated in the range of 60–95°C with a temperature increase of 0.01°C/s. Initial data BIORAD MyiQ System were exported in an excel file. Analysis of absolute expression of M.CViPI was performed according to according to the ΔC_T method while relative expression analysis of *NAC82* was performed with $\Delta\Delta C_T$ method (Schmittgen and Livak 2008) using PP2AA3 (At1g13320) as a house keeping gene for data normalisation.

2.15. Chromatin accessibility assay using iNOMe-seq

2.15.1. ß-estradiol induction of seedlings

Seedlings of MCViPI transgenic and control plants were subjected to ß-estradiol induction and mock treatment for 6, 12 and 24 hours, the stock of ß-estradiol was 10mM in DMSO, with mock being DMSO in water. 8 days old seedlings were transferred to a Whatman paper soaked wet with a film, so that the plants were in contact with the solution and to avoid drought stress. Approximately 30 seedlings were placed on one plate in two rows. Plates were wrapped with cling film before being moved to a growth cabinet.

2.15.2. BSeq library preparation

Approximately 100ng of genomic DNA in total 100ul volume was sonicated using a Bioruptor (Diagenode Inc.) with 30sec ON and 30 second OFF at a "low" settings to generate 350bp DNA fragments. The libraries were generated using the Illumina TruSeq Nano kit (Illumina, CA, U.S.A) according to the manufacturer's manual. Size-selected and adapter-ligated DNA was subjected to bisulfite treatment using an Epitect Plus DNA Bisulfite Conversion Kit (Qiagen, Hilden, Germany) according to the manufacturer's manual. PCR amplification was performed using Kapa Hifi Uracil+ DNA polymerase (Kapa Biosystem, MA, U.S.A) according to the manufacturer's instructions. To check size distribution of DNA fragments at each stage especially before library preparation of sonicated samples as well as after final library preparation, selected samples were run on bioanalyzer (Agilent 2100) for size distribution analysis after each stage.

2.15.3. Bisulfite sequencing

Indexed bisulfite treated libraries were sequenced with 2 × 150-bp paired-end reads on an Illumina HiSeq2000 instrument. Maximum of 16 different indexed libraries were pooled in a single lane run. The Illumina Real Time Analysis 1.13.48. software was used for image analysis and to extract sequence data.

2.15.4. Processing of iNOMe data and quality control

Before in depth analysis of regions identified as differentially methylated (DMRs), the quality of raw iNOMe data was analysed using FastQC (v0.11.5) (Andrews 2010). Raw reads were trimmed using Trimmomatic v0.36 (Bolger et al. 2014) to remove

adapter sequences and poor quality reads. Using Bismark v0.15.0 (Krueger and Andrews 2011) and bowtie2 v2.2.6 (Langmead and Salzberg 2012), trimmed reads were aligned to the reference genome with parameters '-N 1 -L 20 -X 1000 –score min L,0-0.6'.

2.16. iNOMe-seq VS ATAC-seq Analysis

A bins method was used to compute Differentially Methylated Regions (DMRs). Initially, DMRs from iNOMe-seq data were computed with varying methylation differences (10-100%) and bin sizes (50-5000 bp). After selecting optimal bin size and methylation difference, we computed DMRs based on the methylation difference between control and the test samples that is statistically significant with an adjusted P-value ≤ 0.05.

To compare these DMRs with ATAC-seq identified accessible regions, precision, recall and F score was calculated. Each set of the identified DMRs was compared to the ATAC-seq accessible regions and for each of them the proportion of true positives (TP - regions detected by both NOME- seq and ATAC-seq), false positives (FP - regions detected only by NOME-seq) and false negatives (FN - regions detected only by ATAC-seq) was calculated. We assumed that the regions detected by ATAC-seq were true positives. Using these calculated proportions, we computed the recall ($TP/(TP+FN)$), precision ($TP/(TP+FP)$) and F-score ($2TP/(2TP + FP + FN)$). Finally, we selected the methylation difference threshold and bin size that lead to the highest F-score. Based on calculation the precision, recall and F-score, the value of 30% for methylation difference and 700 bp were found to be the optimum bin size for the comparison of accessible chromatin regions by both methods. To compare

accessible regions between iNOMe-seq and ATAC-seq, Venn diagrams were then created using the draw.pairwise.venn function from the VennDiagram package v1.6.19 (Chen and and Boutros 2011).

Differentially methylated regions/Accessible regions identified by NOME-seq were compared to accessible open chromatin identified by ATAC-seq. Any difference in accessibility observed in control and induced samples around the targeted features was compared to accessibility of these features by ATAC-seq data. In order to carry out iNOMe signal detection around targeted features, the proportion of methylation percentage at each individual cytosine was used to compute a moving average with a window size of 500 bp for all samples. The moving averages were then used to determine the methylation difference between control and treatment samples. The difference was calculated by subtracting the moving average of control samples from the moving average of treatment samples.

2.17. Calculation of genome wide distribution of accessible chromatin regions

As iNOMe-seq detects a larger amount of accessible DNA than ATAC-seq, the distribution of iNOMe peaks throughout the genome was checked. To do this the R package ChIPpeakAnno (Zhu, 2013; Zhu et al., 2010) was used to annotate DMRs and to identify the closest gene associated with each differentially methylated region. This package also calculated the distance between genes and DMRs; these distances were used to determine the distribution of iNOMe peaks around genes. By comparing these data to various features, the iNOMe peaks were separated into different categories, these categories included genes, transposable elements, upstream, downstream and intergenic regions.

2.18. Hypermethylation of targeted genomic regions

For the targeted hypermethylation of genomic regions (**Table 7.2**), we used a RNAi haipin vector (**Figure 7.3**). This vector can be used to generate short inverted repeat hairpins directing *de novo* methylation to the targeted genomic sequences via the RdDM pathway (Law et al. 2013).

2.19. Designing guide RNAs for deletion of DMRs via CRISPR/Cas9

To design guide RNAs (protospacer), we used the online tool Breaking Cas (http://bioinfogp.cnb.csic.es/tools/breakingcas/. A 20nt protospacer sequence 5' of NGG was selected, on each side of the region targeted for the deletion (DMR), to generate a sequence specific sgRNA. A reverse complementary sequence for each of the protospacer was also designed to generate double stranded sgRNAs after annealing opposite strands. The sgRNAs were later cloned into pEN-2X-Chimera, to enable both sgRNAs to be expressed under individual U6 promoter (**Figure 7.1**). For each targeted region, two pairs of gRNAs were designed on each flanking side. The sgRNA on 5' of the target region was named as 'A' and the one on the 3' of the target region was named as 'B' sgRNA.

2.20. Protoplast isolation

Protoplast isolation was carried out according to protocol published by Yoo *et al.* (Yoo et al. 2007). *A. thaliana,* Col-0 plants were grown for four weeks at 20-22^{o}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, 10mMCaCl2 and 0.1% BSA) in a petri dish, with vacuum infiltration of leaf strips twice for 5 min. The leaf material was left at 25^{o}C

for 2.5 hours. Next, the tissue was filtered through a 70 µm nylon cell strainer and rinsed with the enzyme/protoplast solution with an equal volume of W5 solution (2mM MES PH5.7**,** 154mM NaCl**,** 125mM $CaCl_2$**,** 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 keep on ice.

2.21. Transfection of protoplasts and determination of the efficiency of sgRNAs

Approximately 80,000 cells were used for transfection with sgRNA carrying plasmid (pEN-2x-chimera) and a CRISPR/Cas9 (pUbi-Cas9) vector (**Figure 7.1, 7.2).** For 200ul of leaf protoplasts, 8ug of each of the plasmid was added. A total of 216ul of freshly prepared PEG solution (40% v/w PEG4000, 0.2M mannitol, 100mM $CaCl_2$) 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 min for transfection. 500ul of W5 was added to stop the transfection process. The samples were mixed gently and centrifuged at 100g for two min. The supernatant was removed and 500ul of WI solution (4mM MES PH 5.7**,** 0.5 M Mannitol, 20mM KCl) was added. Transfected protoplasts were left in the growth cabinet at 21 ^{o}C for 36-48 hours in long day conditions.

2.22. Selection of efficient sgRNAs and stable transformation of plants

Transfected protoplasts were used for genomic DNA extraction. Confirmation of the deletion was carried out by PCR using flanking oligos (**Table 7.3**). The sgRNA combinations, with higher efficiency were cloned into pEC1-Cas9-OLE1 by Gateway.

A tumefaciens mediated transformation of *A. thaliana* was carried out by floral dipping (Zhang et al. 2006).

2.23. Identification of heritable deletions

We selected T_0 seeds expressing RFP and screened individual plants by PCR to detect somatic deletions. To isolate heritable deletions, we grew T_2 plants that did not carry the Cas9 construct (RFP negative) and screened for deletions by PCR. Those plants were propagated to maturity and T_3 seeds were used to identify homozygous deletions (**Figure 7.5**).

.

3. *In vivo* nucleosome occupancy and methylome sequencing (iNOMe-seq)

3.1. Introduction

Eukaryotes have highly compact chromatin organisation, owing to the presence of highly conserved proteins called histones. DNA wrapped around 4 pairs of histones makes a basic structural unit of chromatin, called the nucleosome (Luger et al. 1997). Nucleosomes, which seem to be present unevenly throughout the genome are in fact, not located randomly, instead their position is determined by the DNA sequence and other chromatin modifications (Chereji and Clark 2018).

Accessibly of genomic DNA is altered due to its packaging into nucleosomes meaning that many DNA binding proteins are prevented from reaching targeted DNA sequences. On the other hand many distal sequences can interact together due to the optimal positioning of distal nucleosomes. (Anderson and Widom 2000).

The significance of nucleosome occupancy and positioning in eukaryotic gene regulation has been illustrated by many studies. It has been established that the chromatin at promoters and known transcription binding sites was more accessible with DNase treatment (Hsiung et al. 2015). In another study, in the model yeast *Saccharomyces cerevisiae*, it was demonstrated that active regulatory genomic regions are depleted of nucleosomes at the whole genome level and especially during mitotic growth of the organism. Moreover, nucleosome occupancy was greatly reduced at active gene promoters (Lee et al. 2004). Other findings have also confirmed that nucleosome can be moved, displaced or even lost during active gene expression (Guillemette et al. 2005; Schones et al. 2008a), indicating that

nucleosome occupancy is a dynamic feature and it goes hand in hand with other chromatin modifications.

To understand the relationship between the physical and functional genome, it is important to discover regulatory landscapes in relation to chromatin accessibility. Various regulatory elements such as proximal promoters, enhancers and other long distance regulatory elements possess binding sites for various transcription factors and other regulatory proteins (Bell et al. 2011). These binding sites are characterised by nucleosome-depleted regions flanked by a phased nucleosome on either side (Schones et al. 2008b).

Currently, there are various methods available to map nucleosome occupancy and subsequently identify accessible genomic regions. These methods include DNase-seq, MNase-seq, ATAC-seq and NOMe-seq, although other variants of some of these methods also exist. The first two methods involve intense enzyme titrations and require incredibly large amount nuclei (~ 1,000,000 cells) while ATAC-seq, which is currently a gold standard method requires a fewer number of cells (~ 50,000) to measure chromatin accessibility is also a most widely used method in eukaryotes (Tsompana et al. 2104). ATAC-seq employs a Tn5 a hyperactive transposase to cut and simultaneously ligate adaptors to accessible DNA regions.

The technique is relatively robust and offers many advantages over other techniques, however, there are certain limitations to the applicability of this technique to uncover regulatory regions in comparatively lower accessibility regions due to its inherent bias toward accessible active promotor sites such as TSS and other AT rich regions (Meyer and Liu 2014).

The Nucleosome Occupancy and Methylome sequencing (NOMe-seq) technique is a unique method to measure chromatin accessibility as well as its relationship with the epigenetic state of the genome. The technique employs M.CViPI, a methyltransferase from *Chlorella* virus that can methylate GpCs in accessible chromatin. When coupled with next generation sequencing, it can simultaneously reveal methylation profiles and chromatin accessibility (Kelly et al. 2012).

The M.CViPI methyltransferase encoded by *Paramecium bursaria Chlorella virus* (Xu et al. 1998) is a nucleic acid methyltransferases which use s-adenosyl methionine (SAM) as a methyl donor to methylate a substrate after transferring methyl group (CH_3) to the substrate (Martin and McMillan 2002). M.CViPI has the potential to methylate all available GpC dinucleotides (present in high abundance in the eukaryotic genome), creating an unbiased nucleosome occupancy footprint, which makes it a powerful, high-resolution nucleosome positioning identification method (Pott 2017).

The current method published by Kelly et al (2012) works by treating isolated nuclei with M.CViPI GpC methyltransferase to methylate GpCs present in nucleosome-free regions followed by bisulfite sequencing to reveal chromatin accessibility. Bisulfite sequencing involves treatment of DNA with bisulfite to convert all non-methylated cytosines to uracil. Uracil residues are then recognised as thymine during PCR amplification and are thus distinguished from methylated cytosines (Li and Tollefsbol 2011). M.CViPI induced methylation is distinguished from endogenous methylation profile when compared to non-treated samples (**Figure 3.1**).

.1

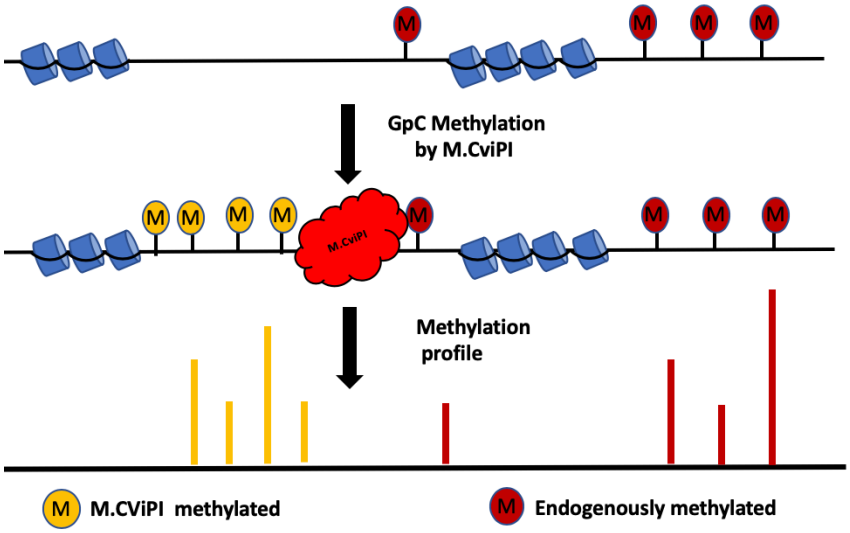

Figure 3.1: NOMe-seq can simultaneously reveal nucleosome occupancy and methylation profile. NOMe-seq coupled with bisulfite sequencing reveals whole genome methylation profile and nucleosome position, which can be confirmed by identifying M.CViPI inaccessible sites through their distinct methylation profiles.

Since the technique does not involve having to fragment DNA either by enzyme digestion or transposon insertion, there is less bias towards AT rich regions, and thus it can be relatively straightforward method to study chromatin accessibility.

NOMe-seq has been extensively used in mammalian cells and its employment has offered a valuable insight into complex tissues as well single cells chromatin profiling (Pott 2017).

In plants, however, there are certain barriers to the effective application of NOMe-seq. Plant tissues pose a huge challenge for effective M.CViP I treatment due to their rigid cell walls, making it problematic to carry out cell nuclei sorting without extensive cell preparation. In mammals, DNA methylation occurs only in CG dinucleotide contexts whereas in plans methylation occurs in three different

contexts i.e. CG, CHG, and CHH. Due to the presence of complex methylation contexts, ATAC-seq has been a preferred method by plant researchers to avoid confusion in plant methylome DNA analysis. ATAC-seq is, however, quite challenging to scale down sample input while single cell bisulfite sequencing is a widely used technique (Smallwood et al. 2014). It is imperative to devise an improved strategy for NOMe-seq specifically applied for studying chromatin accessibility in plants. Effective application of this method in plants will not only give information about the chromatin accessibility and methylation profile of the plant genome but also help reveal nucleosome position across the whole genome. Moreover, the method needs to be scalable and be effective in enabling of cell specific chromatin accessibility and methylation profile in different physical environmental conditions.

3.1.1. Chapter aims

This chapter described the development of a new methodology- *in vivo* NOMe-seq (iNOMe-seq) to investigate the regulatory DNA landscape in plants. This technique is based on the capacity of MCViPI to methylate open chromatin thus enabling the combined analysis of accessible chromatin and DNA methylation to reveal the regulatory regions of plant genomes.

3.2. Results

3.2.1. Experimental design

To study the chromatin regulatory landscape of different Arabidopsis tissues, we generated chemically inducible *M.CViPI* transgenic system. The controlled expression of methyltransferase M.CViPI enables the labelling of GpC motifs present in regions of the chromatin that are accessible and open to transcription. The genomic DNA was analysed by next generation bisulfite sequencing (Bseq) to identify the methylation profile compare this between induced and control samples. The genomic regions with open chromatin conformation were expected to show a distinct methylation profile when compared to control (**Figure 3.2**).

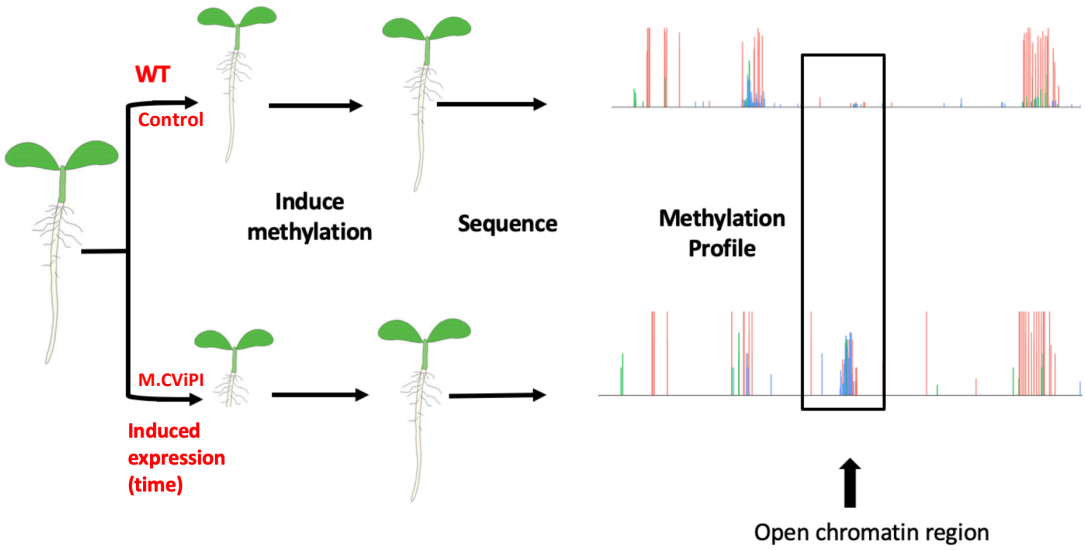

Figure 3.2: **Experimental design to identify open regions of chromatin in *A. thaliana* by iNOMe-seq.** The transgenic plants were subjected to MCViPI induction for certain length of time to mark open chromatin followed by BSeq to analysis.

3.2.2. Construction of inducible *M.CViPI* transgenic system

The *M.CViPI* transgene was cloned in a strictly regulated chemical inducible system that uses the chimeric transcription activator XVE (Zuo et al. 2000). *XVE* is a fusion gene that encodes a DNA binding domain of a bacterial repressor LexA (X) (Brent and & Ptashne 1980), the acidic transactivating domain VP16(V) (Cress et al. 1991) and the carboxyl region of human oestrogen receptor(E). Chemically inducible expression systems are commonly used to tightly regulate the expression of proteins that cause toxicity when expressed constitutively (Greene et al. 1986). *XVE* is normally expressed under a strong constitutive promoter (Ishige et al. 1999) and binds to *lex*A operator in presence of ß-estradiol thus resulting in transcriptional activation of the M.CViPI methyltransferase (**Figure 3.3**).

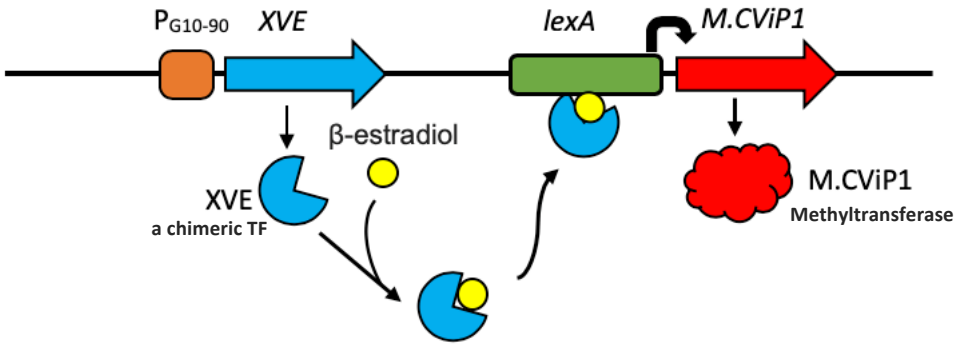

Figure 3.3: Controlled expression of M.CViPI using *XVE*: a chemically inducible system. A synthetic constitutive promoter G10-90 drives expression of a chimeric transcription factor, XVE (LexA-VP16-ER). *XVE* controls activity of the *lex*A promoter under influence of chemical induction by ß-estradiol. In absence of ß-estradiol, XVE cannot bind to *lex*A and M.CViP expression cannot be initiated. When ß-estradiol binds to XVE, it can activate *lex*A to promote *M.CViPI* expression.

3.2.3. Testing Expression of *M.CViPI* transgenic line

The inducible M.CViPI transgenic system (**Figure 3.3**) was engineered into a binary vector with seed specific RFP reporter (pMDC7-OLE1) and Arabidopsis transgenic lines were generated. To identify Arabidopsis *M.CViPI* transgenic lines suitable for analysis, we selected 20 T_0 transgenic seeds (RFP positive) and propagated these to the T_1 generation. Seeds from T_1 generation were screened for single copy Mendelian segregation ratio (3:1 RFP^+ to non-RFP) and 4 independent transgenic lines that were further propagated to generate homozygous lines (100% RFP seeds) (**Figure 3.4**).

To select a responsive and tightly controlled *M.CViPI* line, homozygous lines were subjected to prolonged ß-estradiol induction (two weeks) and monitored for growth and developmental differences to control plants. Plants were grown in a growth chamber for two weeks. WT lines were developmentally normal in presence or the absence of ß-estradiol (**Figure 3.5**). The *M.CViPI* transgenic lines that showed pleiotropic growth and developmental abnormalities (**Figure 3.5**) in the presence of ß-estradiol but were normal in absence of ß-estradiol, were selected for a molecular analysis.

To identify the most responsive *M.CViPI* line, we selected 4 independent homozygous lines (M.CViPI-1a, M.CViPI-2a, M.CViPI-5a and M.CViPI-10a) and analysed changes in DNA methylation after ß-estradiol induction. The selected lines were grown on media supplemented with 40uM ß-estradiol for 48 hours and BSeq analysis was performed. After initial methylation analysis, we found a line (M.CViPI-1a) that showed the highest level of CpG methylation (35.9%) compared to

endogenous CpG methylation (9.1%) observed in control plants. Other lines were also responsive to ß-estradiol treatment and showed high CpG methylation levels (25-30%) (**Figure 3.6**). Since, M.CViPI-1a exhibited maximum change in % CpG methylation, we selected this line for the expression analysis and optimisation of induction time.

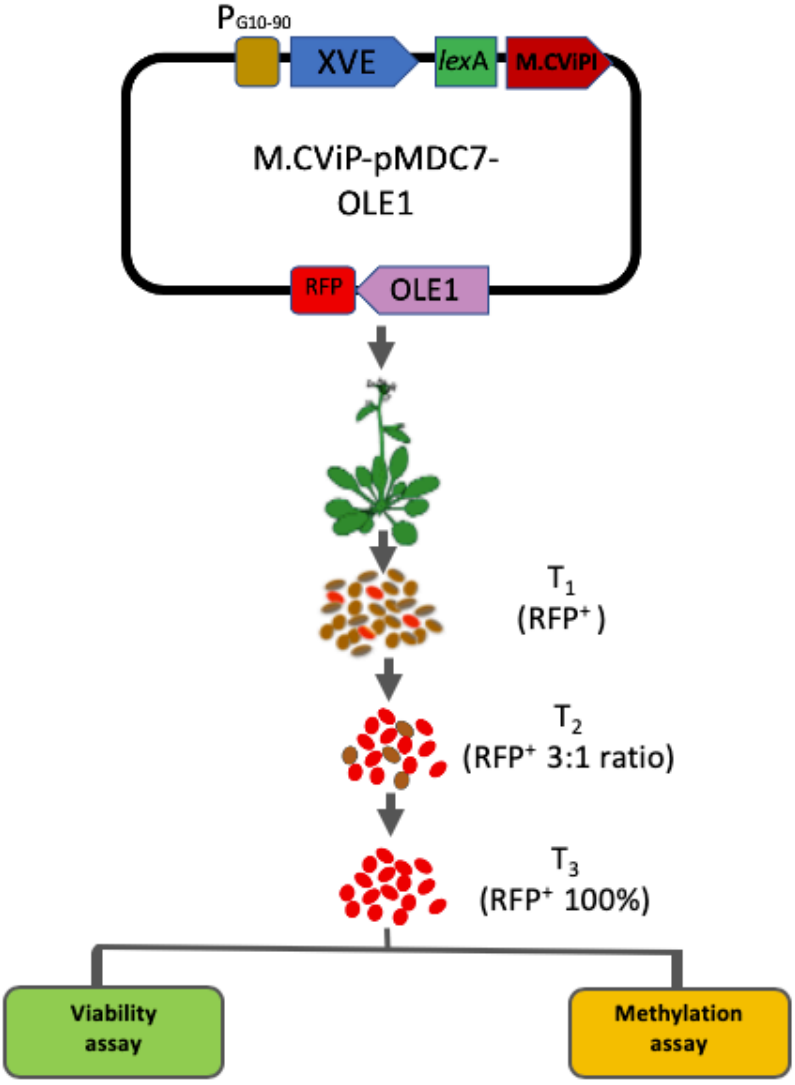

Figure 3.4: Strategy for the generation and selection of *M.CViPI* transgenic lines. *The M.CViPI* transgene was cloned under *XVE* inducible system. Transgenic plants were selected with RFP seed selection through T_1 and T_2 generation. Transgenic lines were tested for MCViPI expression by ß-estradiol induced phenotypic analysis

(viability assay) and methylation assay to select a tightly controlled and highly responsive T_3 line.

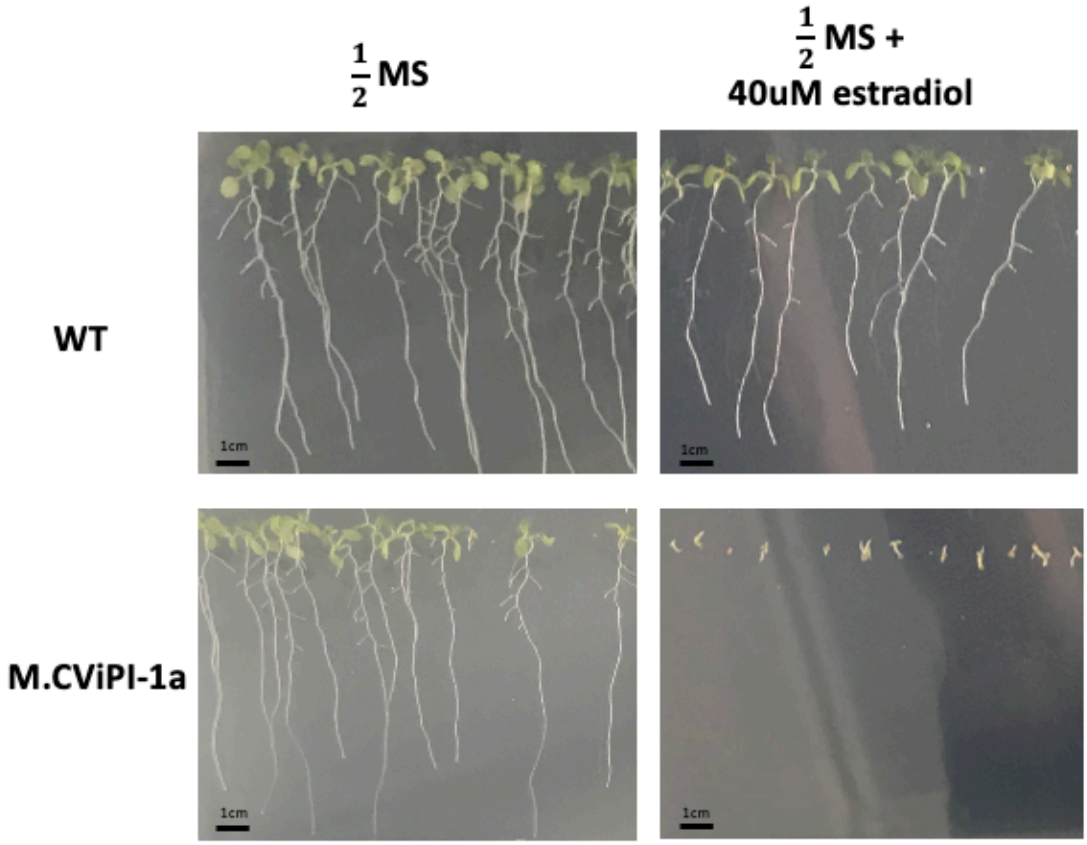

Figure 3.5 Identification of tightly controlled ß-estradiol responsive *M.CViPI* transgenic line through viability assay. Seeds from multiple homozygous lines were grown on media prepared with final concentration of 40uM ß-estradiol and ½ MS (only $\frac{1}{2}$MS for control). WT lines showed no growth or developmental abnormalities in presence or absence of ß-estradiol. M.CViPI transgenic lines, which exhibited severe growth abnormalities in presence of ß-estradiol (M.CViPI-1a only shown here) and showed normal development in $\frac{1}{2}$MS were selected for initial methylome analysis.

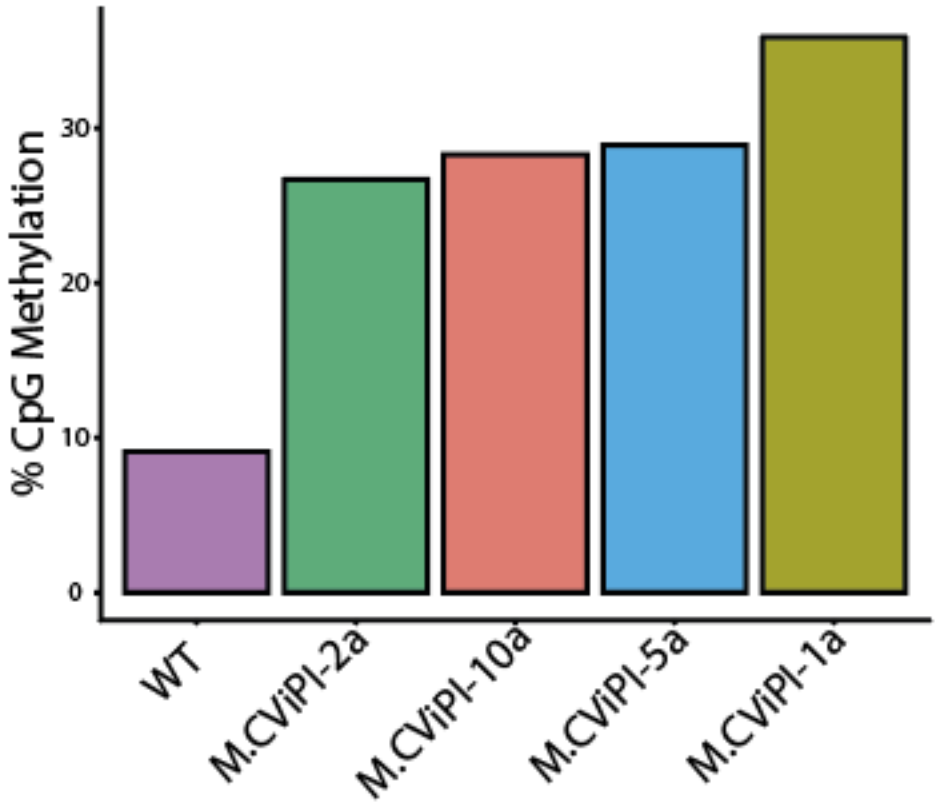

Figure 3.6: Percentage CpG methylation in M.CViPI lines after 48h ß-estradiol induction. 4 independent M.CViPI lines were subjected to methylome analysis to test %methylation after ß-estradiol induction for 48 hours. Root tissue was collected after 40um ß-estradiol treatment and a methylation assay was performed. MCViPI-1a showed highest % of methylation increase hence it was selected for iNOMe-seq analysis.

3.2.4. Expression Analysis of the M.CViPI-1a

After identification of MCViPI transgenic lines through phenotypic and methylation analysis, we selected M.CViPI-1a for iNOMe-seq analysis. To determine an optimal induction time, we carried out an expression analysis of MCViPI expression by qRT-PCR. We induced M.CViPI expression for various lengths of time (0h, 6h, 9h, 12h, 24h, 36h and 48h) and compared the absolute expression of M.CViPI to mock treated plants. A significant increase in expression was observed at 3h and 6h (p-value ≤ 0.05), however no substantial change in M.CViPI expression was observed between 6h and 12h. M.CViPI expression then gradually increased at 24h, 36h and 48h (**Figure 3.7**). M.CViPI expression in WT and MCViPI mock (0h) treatment remained undetectable (**Figure 3.7**). This indicated that M.CViPI expression in this line is tightly controlled (does not show any expression in the absence of ß-estradiol) and rapidly activated after chemical induction.

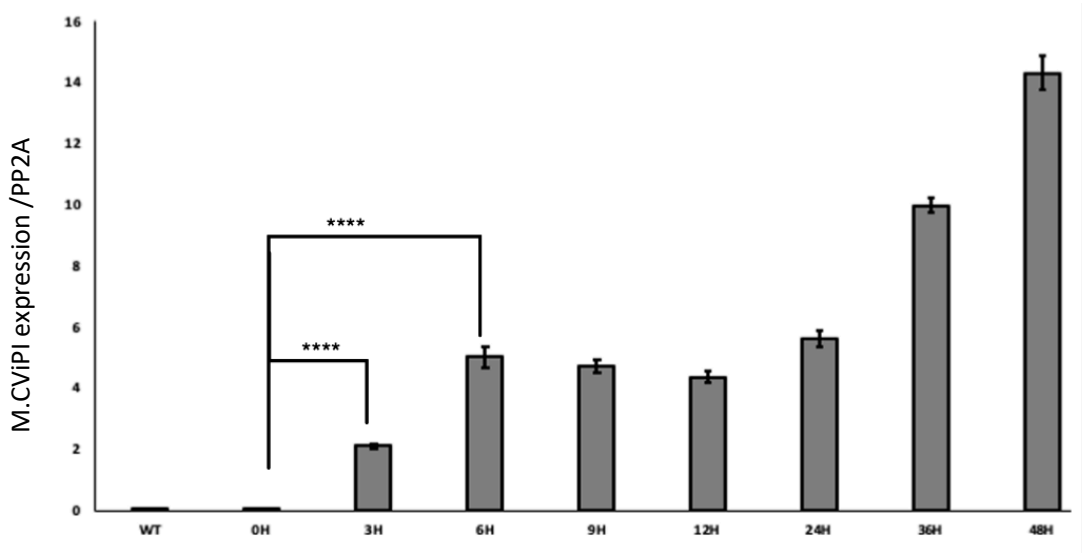

Figure 3.7: *M.CViPI* expression at different induction times. The qRT-PCR was done to quantify the absolute expression of *M.CViPI* transgene at various induction times in M.CViPI-1a (after addition of 40uM ß-estradiol) and WT control in 8 days old seedlings. The expression was normalised using *PP2A* (At1g13320) and expression data was analysed using ΔCT method (Schmittgen and Livak 2008). Difference in expression between 0H vs 3H/6H was statistically significant (3 technical replicates, t-test and p-value ≤ 0.05).

3.2.5. Heritable phenotypic changes induced by *M.CViPI* expression

During the screening process, we observed pleiotropic growth and development phenotypes with prolonged induction with ß-estradiol (**Figure 3.5**). Analysis of the methylation assay also revealed that the percentage of DNA methylation increased after M.CViPI induction (**Figure 3.6**). Previous studies have revealed that induced DNA methylation in Arabidopsis can be heritable over generations (Hofmeister et al. 2017). Stress induced DNA methylation has been shown to result in epigenetic inheritance, as indicated by the fact that progenies of stress induced plants show better stress tolerance (Paszkowski and Grossniklaus 2011).

We proposed that phenotypic changes resulting from M.CViPI induced methylation changes may be heritable, thus their progeny will inherit some of these phenotypes. We also hypothesised that M.CViPI induced phenotypes could be reversed by backcrossing with WT plants.

To test our first hypothesis, we treated 8 days old M.CViPI-1a and WT seedlings for 4 days with ß-estradiol (I_1)to ensure plants accumulate DNA methylation without affecting the plant survival. The plants did not seem to show growth or developmental defects in I1 generation (data not shown here). We transferred these plants to soil under normal growth conditions. We collected seeds that were grown to produce next generation (I_2 generation). These plants displayed variety of growth and development phenotypes compared to control plants (qualitative data only). The plants appeared to have disorganised leaf rosette and flower development delays (**Figure 3.8**). We grew plants from the I_3 generation and observed similar

developmental abnormalities (**Figure 3.8**). To test the second hypothesis, we crossed I_2 plants with WT and grew the F1 generation in soil under normal growth conditions. We found that these plants were similar to WT but they still displayed differences in growth (**Figure3.8**).

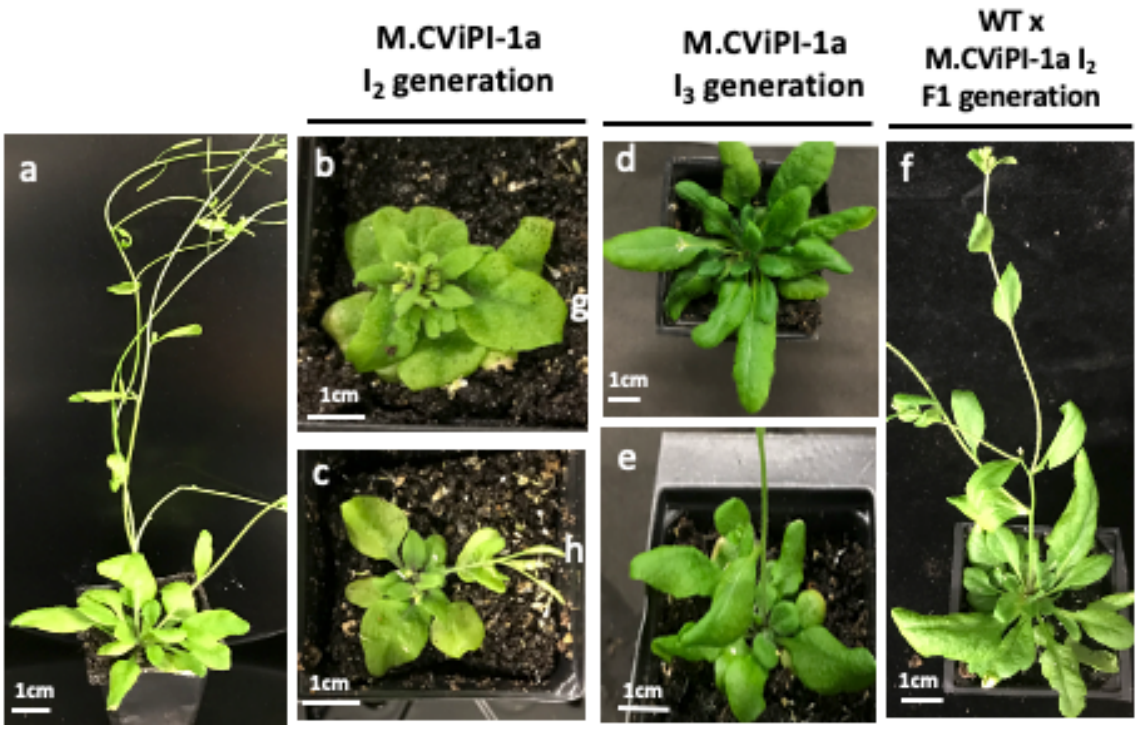

Figure 3.8. Heritable phenotypic changes induced by M.CViPI induction. 8 days old M.CViPI-1a seedlings were induced for 4 days with 40um ß-estradiol in ½ MS. Phenotypic changes in I_2 (second generation after induction) and I_3(third generation after induction) were observed. (a) WT plant. (b,c) MCViP1-I-1a I_2 plants showing disorganised leaf rosette and failing to develop flower. (d,e) MCViPI-1a I_3 plants seem to inherit asymmetrical rosette and flower development. (f) Phenotypic changes were reversed substantially in F_1 when M.CViPI-1a I_2 plants were crossed with WT.

3.2.6. iNOMe-seq successfully induced global methylation in plant genome

Since M.CViPI only methylates cytosines sites in a GpC context, we split GpC and non GpC methyaltion across the whole genome and only methylation in GpC context was computed for the analysis. In order to test if M.CViPI-1a has any effects on the DNA methylome due to leaky expression of M.CViPI, we compared Col-0 mock (Martin and McMillan) with MCViPI mock (MM) plants. We did not find any difference in DNA methylation between these samples (**Figure 7.6**). We, then treated WT plants with ß-estradiol (CE) to see any accumulation of DNA methylation changes and again, we did not observe any significant differences (**Figure 7.6**). We, then continued the analysis with MCViPI-1a plants by inducing with 20uM ß-estradiol and comparing to WT treated plants (CE).

Nucleosome distribution varies across the genome with regions of uniform distribution interrupted by regions of non-uniform distribution of nucleosome (Szerlong and Hansen 2010). Centromeric regions are distinctly marked by highly methylated repeat DNA sequences and display a highly compact chromatin and significantly less accessibility as compared to chromosomal arms (Koo et al. 2011). NOMe-seq has been used to reveal distinct nucleosome occupancy pattern and methylation profile simultaneously across whole chromosome length (Kelly et al. 2012). So, to visualize the distribution of iNOMe-seq (iNOMe) mediated DNA methylation across the whole genome, we analysed the profile on all chromosomes for leaf and root tissues. Overall, the data showed that DNA methylation increased in all non-centromeric regions across all chromosomes in both organs (**Figure 3.9, 3.10**, only chromosome 1 shown here). No significant changes in DNA methylation were

observed in centromeric regions of root or leaf tissues in GpCG and GpCHG contexts, while GpCHH methylation showed increase centromeric regions especially at 12-hour and 24-hour ß-estradiol induction time (**Figure 3.9, 3.10**, **Figure 7.7-7.12**).

Comparing methylation changes with various induction times between root and leaf tissue, we found that iNOMe mediated DNA methylation increase in root tissue was generally higher than that of leaf tissue (**Figure 3.9, 3.10**). For instance, there was no difference in methylation between mock and 6-hour induced samples in shoot tissues in GCG context (**Figure 3.9**) but on the other hand, approximately 20% increase in methylation (GCG context) was observed in root samples after 6-hours induction (**Figure 3.10**). Similarly after 24-hour induction, 20% methylation increase was observed in shoot tissues (**Figure 3.9**) in GCG context compared to 35% increase in root samples (Figure 3.10). The difference recorded in methylation increase may be due to differences in ß-estradiol transport in different plant organs.

The iNOMe mediated DNA methylation difference for leaves were visible only after 12h treatment and highest value was observed after 24 hours induction (Figure 3.9). In roots, iNOMe mediated DNA methylation was apparent 6h after induction reading the highest value 12h after treatment (**Figure 3.10**). Therefore, our analysis established optimal time points for the chromatin analysis using iNOMe in Arabidopsis transgenic line.

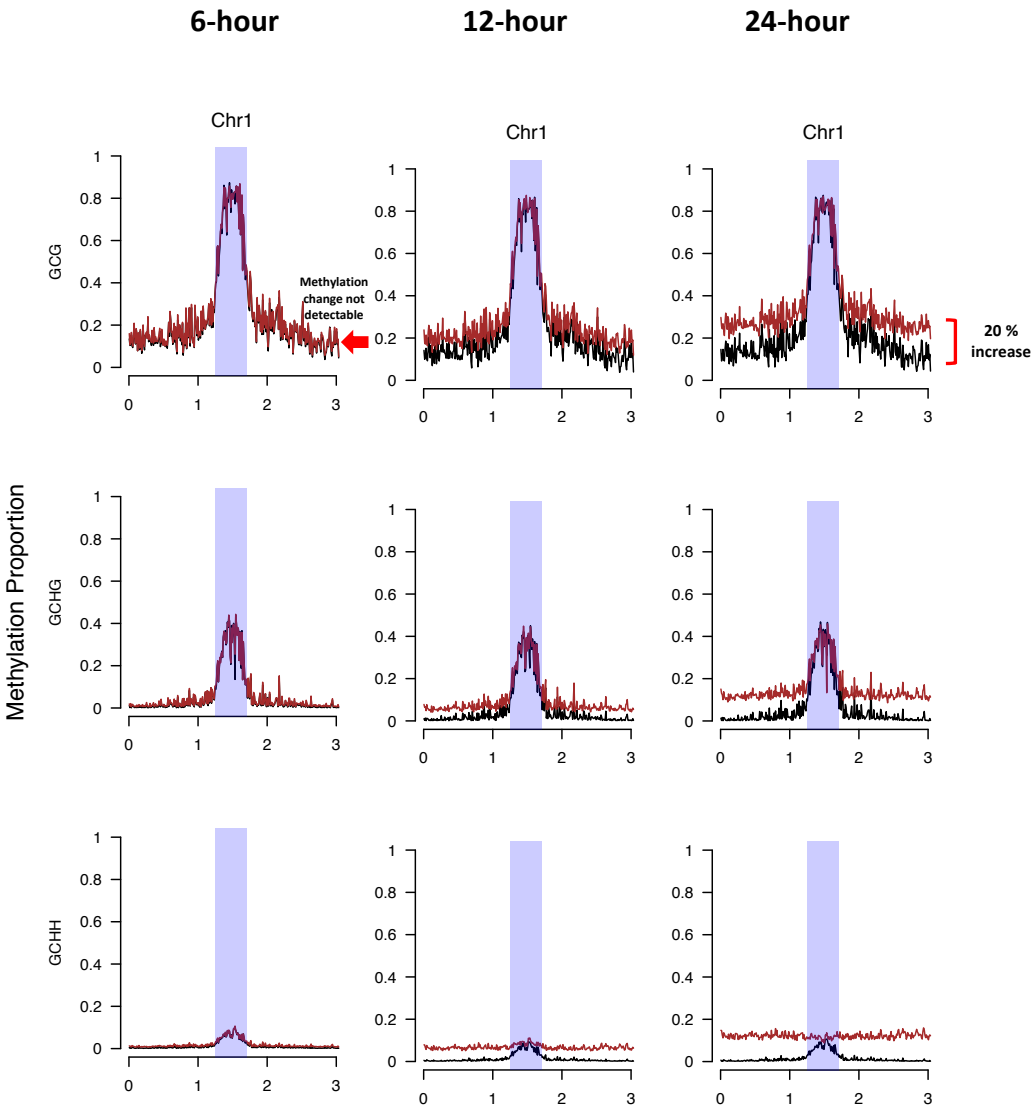

Figure 3.9: Methylation profile of *A. thaliana* chromosome 1 in leaf tissue in GpCG, GpCHG and GpCHH contexts with 6, 12 and 24-hour ß-estradiol induction. Cytosine methylation increased in all contexts in induced plants with increasing induction time compared to control plants. The black line indicates methylation profile of CE samples (Control) and red line represents methylation profile of ME (M.CViPI induced line).

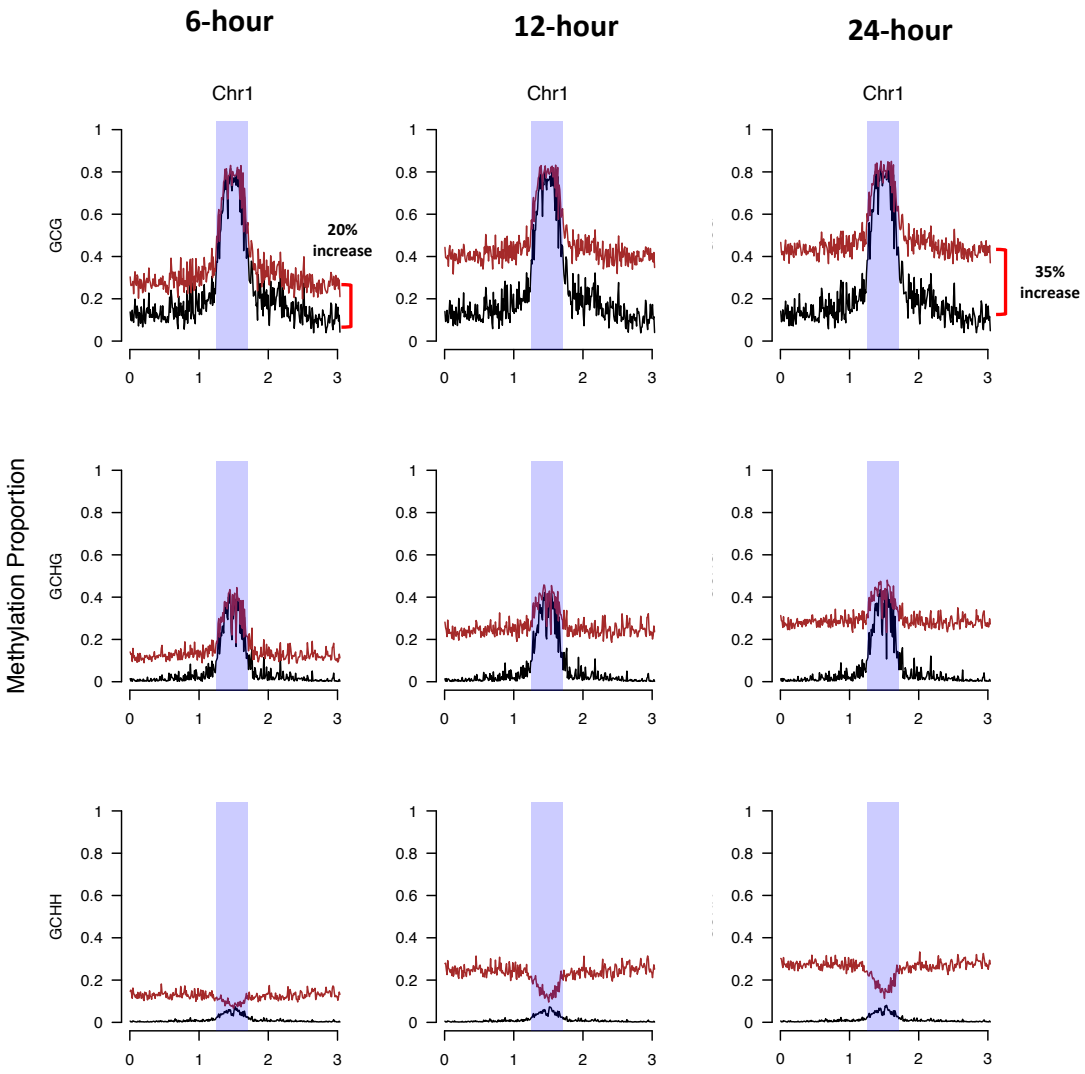

Figure 3.10: Methylation profile of *A. thaliana* chromosome 1 in root tissue in GpCG, GpCHG and GpCHH contexts with 6, 12 and 24-hour ß-estradiol induction. Cytosine methylation increased in all contexts in induced plants with increasing induction time compared to control plants. The black line indicates methylation profile of CE samples (Control) and red line represents methylation profile of ME (M.CViPI induced line).

3.2.7. Identification of accessible regions using iNOMe

Differentially methylated regions (M.CViPI induced methylation vs endogenous methylation of control) from iNOMe data were computed. To compare these DMRs with ATAC-seq identified accessible regions, precision, recall and F score was calculated (methods 2.16). In order to decide which time point was optimal to determine accessibility of the chromatin across the genome, we compared iNOMe accessible regions with genomic regions identified by ATAC-seq: a method to be the gold standard for chromatin accessibility (Buenrostro et al. 2015). ATAC-seq data analysis relies on peak callers where mapped reads are used to detect accessible chromatin shown as "peaks" (Chang et al. 2018). On the other hand, iNOMe data is generated by computing methylation difference for a bin (a fixed size window) averaged from quantitative readout for each cytosine ectopically methylated in the bin (Nordström et al. 2019).

The aggregate length of iNOMe regions at different time points in roots was compared to length of open chromatin regions detected by ATAC (methods 2.16). We found that the accessible regions found by iNOMe in GpCG context were higher in 6h treated roots vs shoots. However, the length of accessible regions in 12- 24h treated roots were similar in all DNA methylation contexts (**Figure 3.11**). Our data also show that the length of accessible chromatin regions identified by ATAC, (21.6 M bp) was similar to that found by iNOMe (23MMbp) (**Figure 3.11**). Only half of these regions were identified by both methods. Notably, iNOMe found additional 14 Mb in GpCG context, not detected by ATAC (**Figure 3.11**). In order to test if longer induction of M.CViPI would increase the overlap between both methods, we

compared iNOMe from 12 and 24h treated root samples. Our data show that iNOMe signal in 12 h treated roots was 75Mb covering the vast majority of regions identified by ATAC (**Figure 7.15**). Moreover, the overlap between both methods did not differ significantly in 24h treated roots (**Figure 7.16**). Our data shows that the chromatin regions marked by iNOMe in leaf were much shorter than those were found in roots. iNOMe identified 14.6Mb in 12h after induction in shoots and only 5.6M bp of DNA was found to be accessible by both methods (**Figure 3.12**).

Roots 6-hour

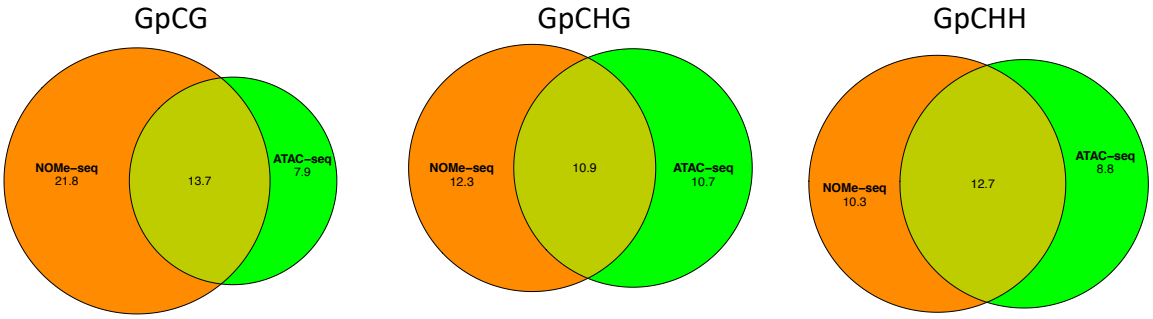

Figure 3.11: Accessible chromatin regions found by iNOMe compared to ATAC in root samples with 6-hour ß-estradiol induction. The orange circle indicates the length of accessible chromatin detected by iNOMe only (Mb) whereas the turquoise indicates accessible chromatin regions (Mb) identified by ATAC only. The overlap of both circles shows the length of accessible chromatin found by both methods.

Leaves 12-hour

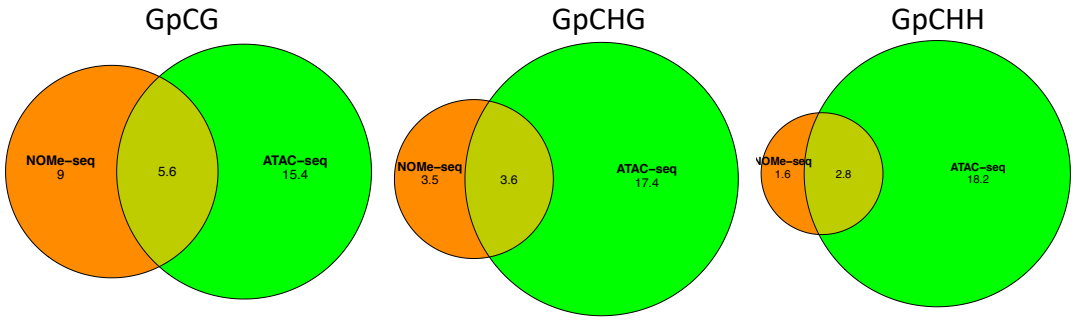

Figure 3.12: Accessible chromatin regions found by iNOMe compared to ATAC in leaf samples with 12 hour ß-estradiol induction. The orange circle indicates the length of accessible chromatin detected by iNOMe only (Mb) whereas the turquoise indicates accessible chromatin identified by ATAC only (Mb). The overlap of both circles shows the length of accessible chromatin found by both methods.

3.2.8. iNOMe signals around features detected by ATAC

To calculate the iNOMe signal we used a moving average of the difference in methylation between control ß-estradiol treated samples (methods 2.19). Since, iNOMe detects global DNA methylation values of the genome (i.e. methylation added by M.CViPI as well as endogenous methylation), we calculated the methylation differences between control and ß-estradiol samples to define the open chromatin region.

Once accessible chromatin regions were identified by iNOMe, we intersected these data with different genomic features (i.e genes, transposable elements, Transcription Start Sites (TSS) and Pol II occupancy). Ours data shows that iNOMe successfully detected signals around features detected by ATAC (**Figure 3.13, 3.14 and 3.15**)

Gene promoters and gene bodies show lower nucleosome occupancy and lower endogenous methylation (Thurman et al. 2012). Chromatin accessibility assays have confirmed that transcription levels are negatively correlated with the methylation of TF binding site (Stadler et al. 2011). We hypothesised that lower DNA methylation levels and higher chromatin accessibility at transcription binding sites would enhance the sensitivity of iNOMe and difference between endogenous methylation and M.CViPI-induced methylation is important indicator of chromatin accessibility with greater accuracy than other methods.

To test this hypothesis, we investigated the distribution of iNOMe signal (indicated as highlighted peaks see **Figure 3.16**) around genes and (upstream and downstream) flanking region. Signal for all annotated genes was distributed according their

expression. We found that as for ATAC, the signal detected by iNOMe was enriched near the transcription start sites of highly expressed genes (**Figure 3.13**).

The iNOMe signal was very similar to ATAC, increased at TSS and gradually decreased across the whole gene body and increased back again at TES (**Figure 3.13**). However, ATAC signal at TSS was more pronounced than that in iNOMe. These differences are likely due the selective amplification of sequence marked by ATAC (Meyer and Liu 2014).

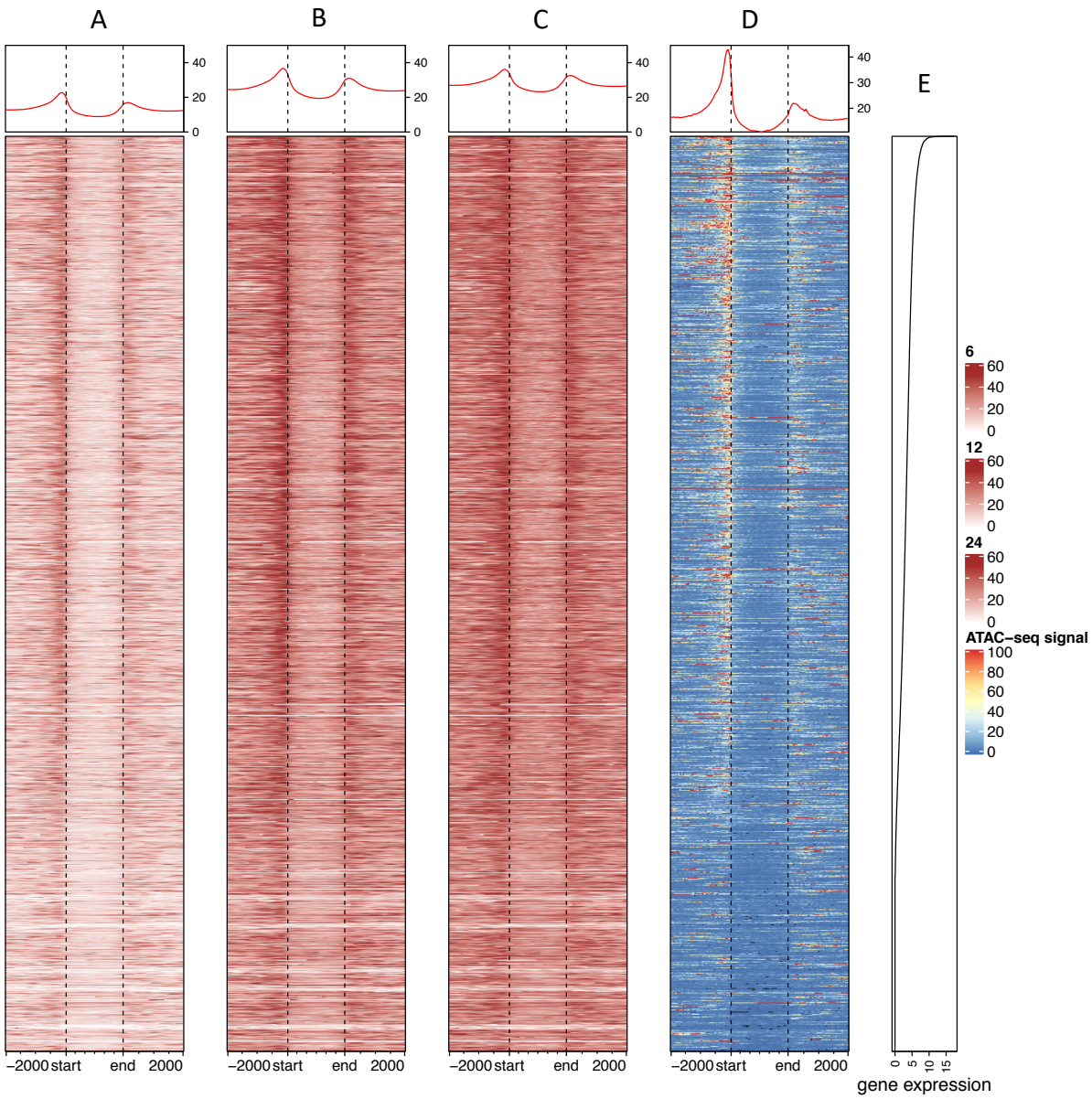

Figure 3.13: Whole genome heat map showing iNOMe and ATAC signal between 2kb upstream and downstream of genes in roots. (A) iNOMe-seq signal around genes after 6-hour induction compared to control (B) iNOMe-seq signal around genes after 12-hour induction compared to control (C) iNOMe-seq signal around genes afte 24-hour induction. (D) ATAC signal around genes. (E) The expression for each gene shown from the whole genome expression data. Each heatmap is ordered by gene expression with highly expressed genes at the top and low expressed genes at the bottom. The metaplot above each heatmap shows the average signal for each location around each gene.

Active transcription is dependent on the binding of Pol II to promoter sequence, however nucleosome hinder the accessibility of Pol II, therefore, nucleosome depleted regions (NDRs) are usually associated with RNA polymerase II binding (Maehara and Ohkawa 2016). An *in vitro* study has revealed that nucleosomes act as a physical barrier to Pol II binding in a variety of ways such as pausing, backtracking and arresting of Pol II (Kulaeva et al. 2013). We computed nucleosome occupancy at Pol II binding sites to compare their accessibility by ATAC-seq and iNOMe. Our data shows a periodic iNOMe signal downstream of the Pol II binding site, thus indicating that this new methodology could be also used to reveal nucleosome occupancy in plants (Figure 3.19). Our data also showed that methylation at Pol II peaks negatively correlated with endogenous methylation signal in control sample (Figure 3.14) supporting the view that DNA methylation imposes a repressive state in chromatin.

We, then analysed the iNOMe at transposable elements (TEs) because they constitute large portion of eukaryotic genomes, are generally transcriptionally inactive and heavily methylated (Lander 2001; Xie et al. 2013). TEs are abundant in heterochromatin regions, such as centromeres and have been implicated in tissue specific gene regulation as well in creating making physical boundaries to hinder spreading of gene silencing (Slotkin and Martienssen 2007). Since TEs are generally found in comparatively less open chromatin regions, we anticipated that iNOMe may detect more signal around TEs than other methods to profile chromatin accessibility. To test this hypothesis, we compared the iNOMe and ATAC signal around TEs. Our data shows that both methods detect open chromatin at TE flanking regions but almost none at the transposon body (**Figure 3.15**). Collectively, our analysis has

revealed that iNOMe detects genomic features such as genes, Pol II binding sites and TEs with great accuracy when compared to ATAC-seq.

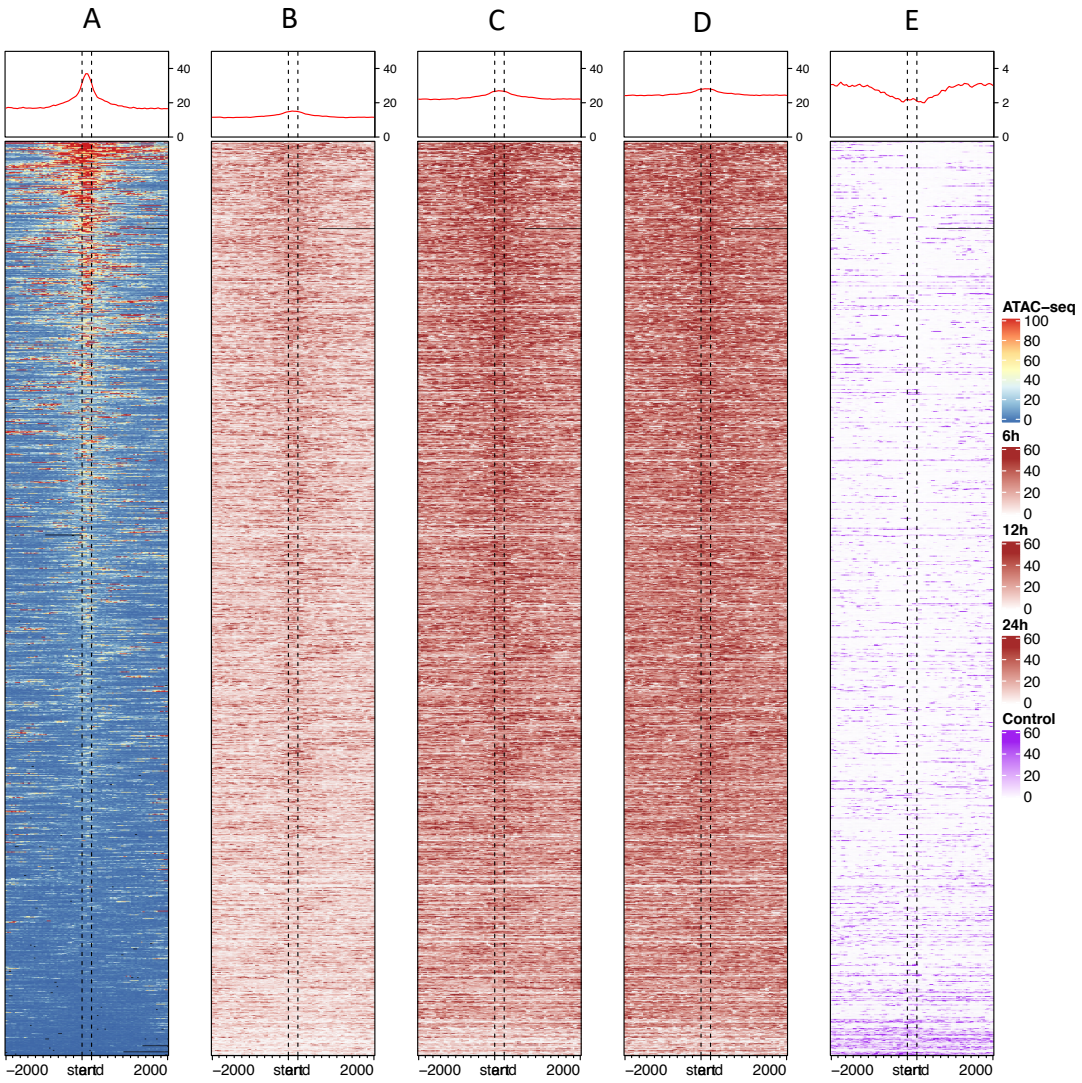

Figure 3.14: Whole genome heat map showing iNOMe and ATAC signal between 2kb upstream and downstream of pol II binding sites in roots. (A) Heatmap shows ATAC-seq signal around pol II peaks (B) iNOMe-seq signal around Pol II sites after 6-hour induction(C) iNOMe-seq signal around Pol II sites after 12-hour induction (D) iNOMe-seq signal around Pol II sites after 24-hour induction. (E) Heatmap shows methylation from a control sample around pol II sites. The metaplot above each heatmap shows the average signal for each location around each Pol II sites.

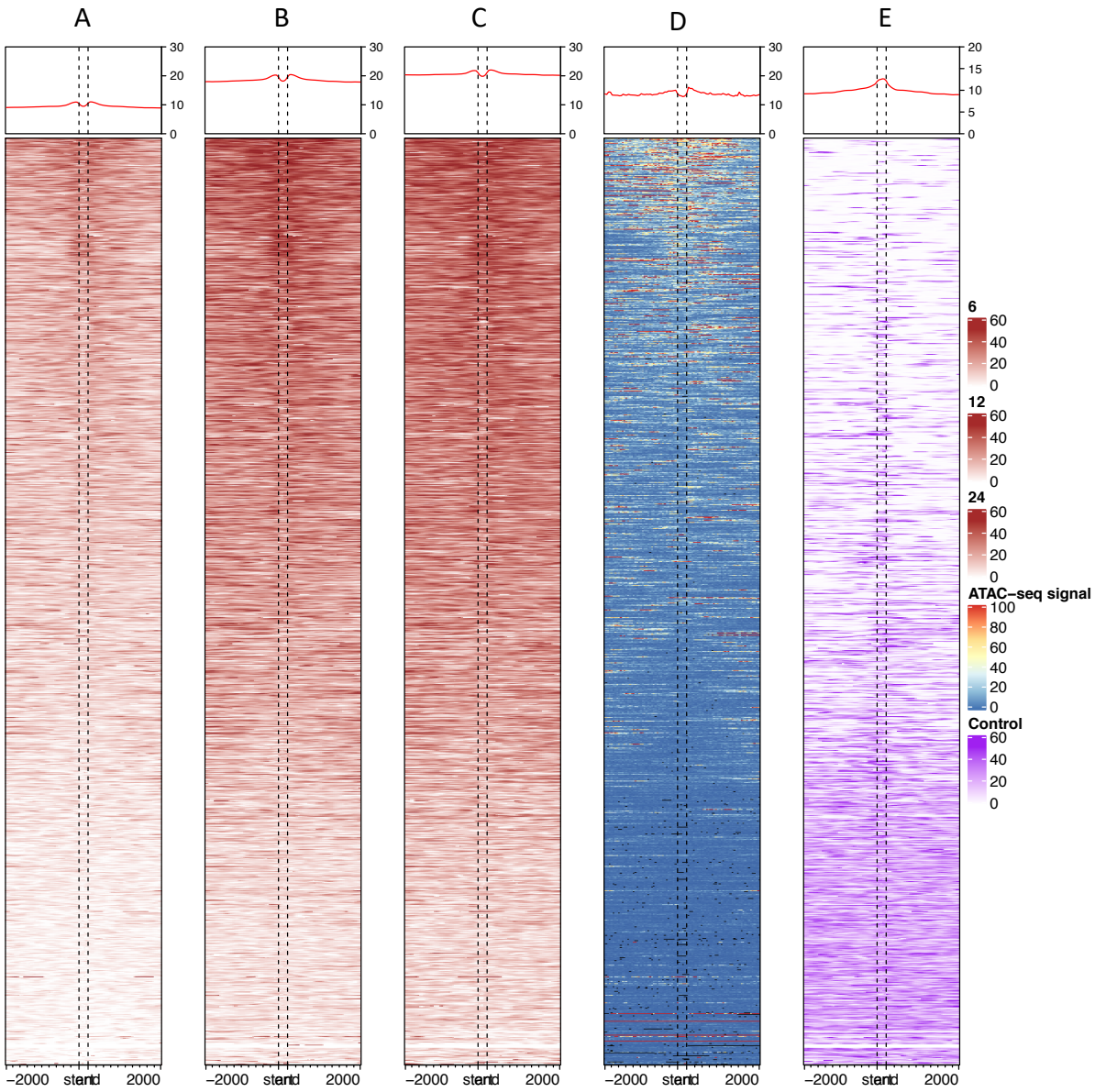

Figure 3.15: Whole genome heat map showing iNOMe and ATAC signal between 2kb upstream and downstream of transposable elements (TEs) in roots. (A) iNOMe-seq signal around TEs after 6-hour induction(C) iNOMe-seq signal TEs after 12-hour induction (D) iNOMe-seq signal around TEs after 24-hour induction. (D) ATAC signal around TEs. (E) heatmap shows methylation from a control sample around TEs. The metaplot above each heatmap shows the average signal for each location around each TE.

To determine the accuracy of iNOMe in chromatin analysis, we compared our data with other chromatin accessibility methods focusing on specific genomic regions. To accelerate this analysis we displayed all the genome wide data in an Integrative Genome Viewer (IGV) (Robinson et al. 2017) We found that iNOMe could detect all the genome regions marked by ATAC (**Figure 3.16A**). However, we also found out that some of the genomic regions were identified by our method only (**Figure 3.16B**) indicating that iNOMe can detect extra regions which other method fail to detect. This corroborates with the data previously discussed in section 3.2.7, where total number of regions identified by iNOMe were greater than ATAC (**Figure 3.11**).

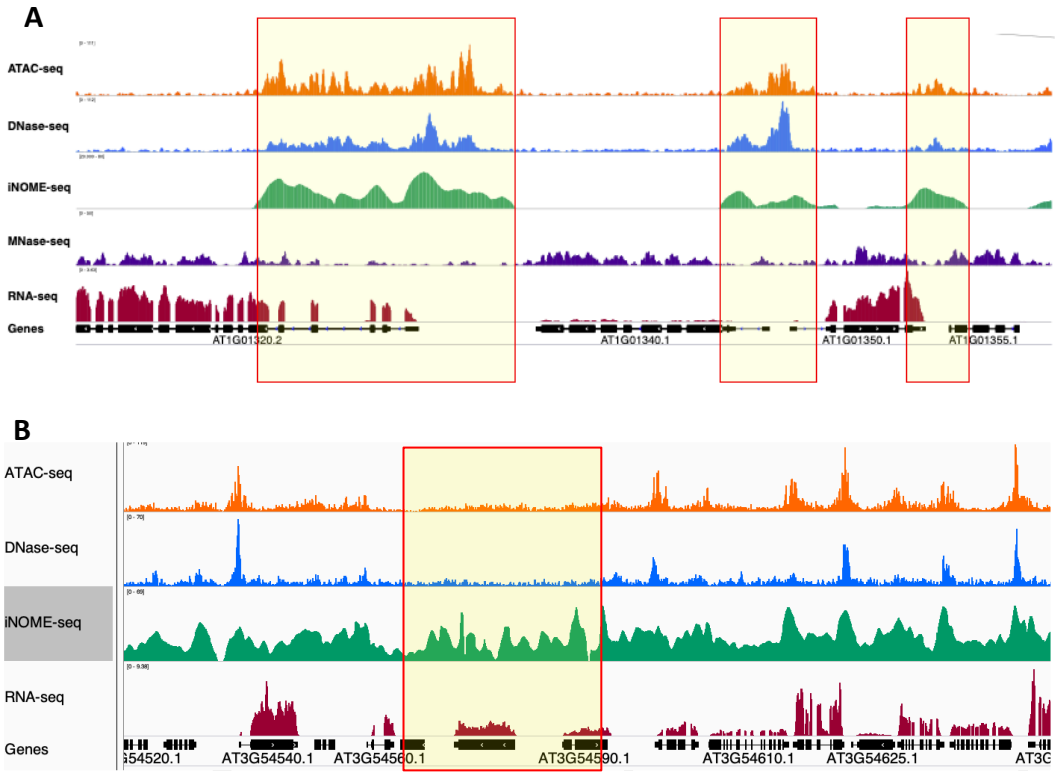

Figure 3.16: The iNOMe signal in a selected region in root tissue compared to other commonly used methods. (A) iNOMe accurately detects regions identified by other accessibility methods with great sensitivity. (B) iNOMe can detect some regions from the genome which other methods do not. The green tracks represent the iNOMe signal, the orange track shows the ATAC signal, the blue track represents the DNase-seq signal and the red track shows gene expression obtained using RNA-seq.

As iNOMe detects a larger fraction of accessible genomic regions than ATAC, we investigated the iNOMe signal throughout the genome. We separated iNOMe signal into different genomic features (genes, transposable elements, upstream, downstream and intergenic regions) and found that in all methylation contexts, iNOMe signal was detected primarily in genes but a significant fraction was also found associated with flanking intergenic regions (**Figure 3.17**).

When we plotted the distribution of iNOMe signal near genes, we found that most of the open chromatin regions detected were located within 2kb upstream regions. However, iNOMe also identified accessible regions located more than 2.5 kb upstream of expressed genes (**Figure 3.18**). This indicates that the accessible regions identified by iNOMe in intergenic regions present 2.5 kb upstream of the genes could potentially harbour distal regulatory elements.

Finally, to see if our method identifies accurate position of the nucleosome, we plotted iNOMe signal around genes. Nucleosome core particles wrap 147 bp of DNA fragments separated by varying lengths of linker DNA depending on the level of nucleosome compaction (Schones et al. 2008a). Measuring the precise distance between nucleosome positions indicates the level of chromatin compaction in certain genomic regions (Schones et al. 2008b). To determine whether iNOMe can measure the linker length accurately, we plotted iNOMe signal in a comparatively narrow window of 1000bp around genes that are being actively transcribed. We observed around 6 oscillatory peaks of iNOMe signal (**Figure 3.19**), which correspond to the predicted nucleosome positions around genes (i.e. ~200bp/oscillation). Our

data shows that iNOMe signal precisely detected nucleosomes at their estimated position around actively transcribing genes.

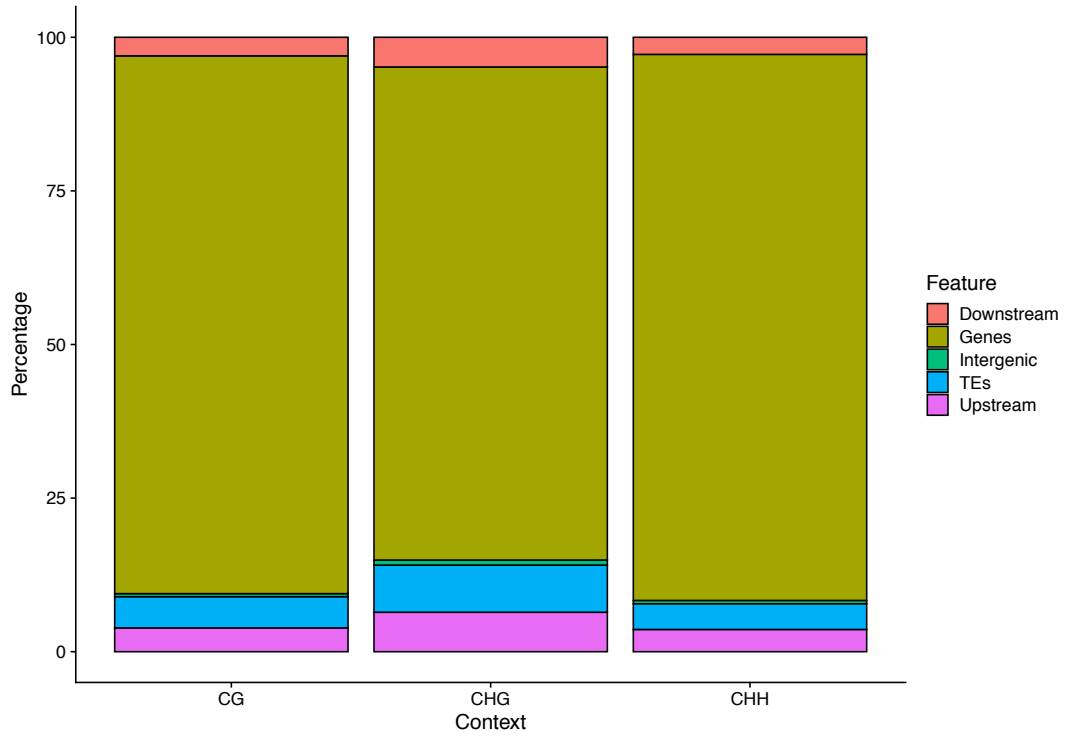

Figure 3.17: Distribution of iNOMe DMRs throughout the genome in relation to different features. By comparing DMRs to various features the DMRs were separated into different categories, these categories included genes, transposable elements, upstream, downstream and intergenic.

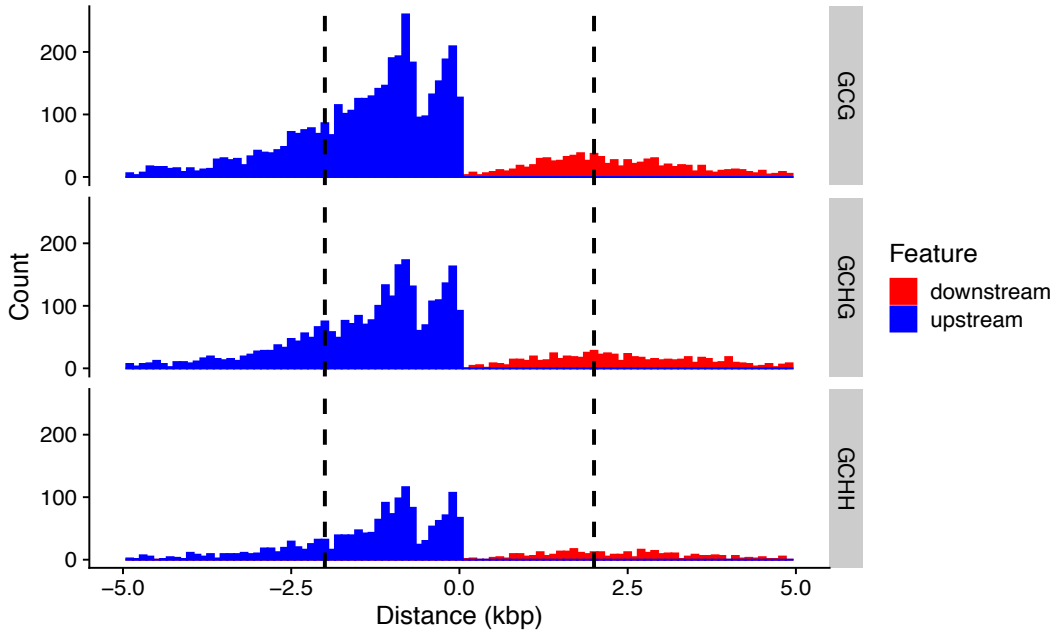

Figure 3.18: Distribution of DMRs upstream and downstream of genes. iNOMe signal near genes was plotted to see distribution of DMRs near genes. A large number of DMRs are distributed within 2.5 kb upstream of the genes. iNOMe also identified DMRs more than 2kb upstream and downstream of genes.

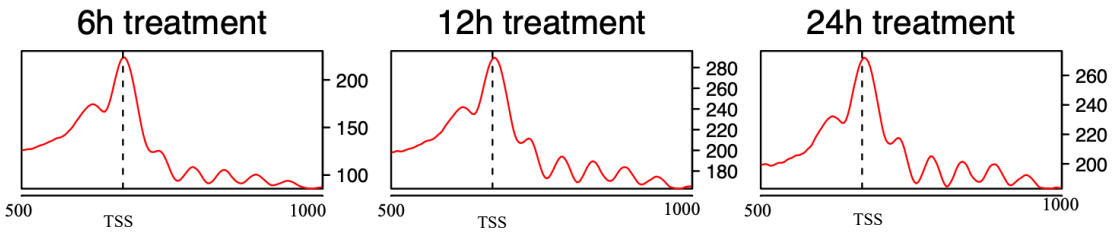

Figure 3.19 iNOMe detects nucleosome position with accuracy. iNOMe signal was plotted near genes to see nucleosome positions at 6h, 12h and 24h after treatment in root samples. The oscillations represent position of single nucleosome in 1000kb after TSS.

Taken together, our data shows that iNOMe is a simple and highly sensitive methodology for the study of chromatin accessibility, nucleosome positioning and DNA methylation in plants.

3.3. Discussion

Nucleosome occupancy is thought to be a principle determinant of genome accessibility and is dynamic in nature (Klemm et al. 2019). A small shift in nucleosome position can not only occlude regulatory DNA motifs but can potentially change chromatin structure (Brogaard et al. 2012). Measuring nucleosome occupancy allows researchers to identify regulatory elements, which cannot be characterised by conventional methods. A unique property of accessible chromatin is sensitivity of nucleosome depleted regions to DNA cleavage by nucleases such as micrococcal nuclease (MNase) and deoxyribonuclease I (DNase I) (Chereji et al. 2017; Boyle et al. 2008). MNase determines nucleosome occupancy while DNase primarily identifies TF binding sites situated in accessible chromatin regions (Cui and Zhao 2012; Hesselberth et al. 2009).

Another DNA cleavage-based method is ATAC-seq, which employs a hyperactive transposase to simultaneously cut and insert adapters preferentially at nucleosome depleted regions to create sequencing libraries (Buenrostro et al. 2015). ATAC-seq is considered the gold standard method for the identification of cis-regulatory elements in accessible chromatin regions in plants (Lu et al. 2016b), however, the efficiency of this method is limited due to its inherent bias towards AT-rich regions and the requirement for large amounts of purified nuclei (Meyer and Liu 2014; Lu et al. 2016b). All of the above-mentioned methods to map chromatin accessibility rely on DNA cleavage, which can potentially damage important functional DNA sequences and disrupt chromatin structure.

DNA methylation foot-printing is a better alternative to harsh enzymatic DNA treatments and a new strategy that employs a GpC methyltransferase added *in vitro*

to purified nuclei, has been developed to address these limitations and to uncover all accessible genome regions with accurate mapping of nucleosome occupancy (Lay et al. 2018). This methodology known as NOMe-seq has been recently used in mammalian single-cell epigenomic studies, however, it has not yet been implemented in other organisms. Here, we have devised an *in vivo* nucleosome positioning method (iNOMe-seq) based on the controlled expression of recombinant GpC methyltransferase, M.CViPI in *A. thaliana*.

The constitutive induction of M.CViPI resulted in plant lethality and controlled expression lead to heritable phenotypic defects, likely caused by the ectopic accumulation of epigenetic marks. Optimum methylation level is important for normal plant development and aberrations in methylation pattern can cause multiple growth and development changes (Kankel et al. 2003).

Our iNOMe data suggests that the accessible chromatin is located at transcribed genomic regions. Centromeric regions are known to be highly methylated and chromatin is highly compact, hindering access to chromatin modifying complexes and other regulatory proteins (Heslop-Harrison and Schwarzacher 2013). Overall, the regions marked by iNOMe in leaf are less abundant than those found in roots which is in line with the previous reports from ATAC-seq studies (Tannenbaum et al. 2018).

We found an increase in iNOMe signal near the transcription start sites of expressed genes correlated with nucleosome depletion, usually found in promoter regions (Yuan et al. 2005). The regions detected by iNOMe had a good overlap with RNA polymerase II binding sites, thus suggesting that iNOMe can uncover chromatin regions that are accessible and are in a potentially active transcriptional state.

However, not all genes marked by Pol II are actively transcribed (Levine 2011). In fact, a large proportion of genes both in plants and mammals, are associated with polymerase pausing (Gaertner and Zeitlinger 2014). Pol II pausing has been extensively studied in several higher organisms (Kujirai et al. 2018), and have been recently studied in Arabidopsis (Zhu et al. 2018).

When comparing iNOMe and ATAC in Arabidopsis we found that both methods efficiently discovered accessible chromatin regions. However, iNOMe was able to discover more regions than ATAC. The number of iNOMe regions increased significantly with an increase in induction of M.CViPI expression. This is likely due to the fact that iNOMe has the potential to mark regions of the genome that are dynamic in terms of their chromatin states over the time of M.CViPI induction. Live cell imaging techniques has previously revealed that many TFs and regulatory proteins transiently bind to regulatory regions, suggesting that chromatin organisation is not static. Its dynamic state allows access to various regulatory proteins to modulate gene expression (Voss and Hager 2008).

Nucleosomes organised into higher order structures of varying degree, which adds another level of complexity to gene regulatory mechanisms (Lorch et al. 1987). To detect nucleosome position, standard methods such as ATAC-seq and MNase-seq produce inaccurate data, which makes very challenging the analysis of nucleosome positioning in different samples (Chereji et al. 2017; Flores et al. 2014). Compared to these methods, iNOMe is able to accurately detect nucleosome positioning. This was confirmed by identifying iNOMe signal oscillations at estimated distance around highly expressed genes, which suggests that iNOMe can not only detect different

features detected by other chromatin accessibility methods accurately, but also nucleosome positioning with great accuracy.

3.4. Summary

Our study shows that iNOMe-seq is powerful technique for the study of chromatin accessibility in plants without involving lengthy and laborious procedures of cell/nuclei isolation or enzymatic optimisation that are required for current chromatin profiling techniques (Buenrostro et al. 2015; Cumbie et al. 2015; Mieczkowski et al. 2016). Our data shows that iNOMe-seq is more sensitive than ATAC-seq, thus could be used to identify accessible regions linked to long- range gene regulation. Taken together, we propose iNOMe-seq as a new gold standard methodology for chromatin accessibility and nucleosome occupancy profiling at cellular resolution in Arabidopsis and other plants amenable to genetic modifications.

4. Functional characterisation of distal regulatory elements

4.1. Introduction

Gene expression in eukaryotes is tightly regulated and selective gene expression is vital for accurate cell differentiation and development of an organism (Jaenisch and Bird 2003). Gene regulation can occur at various levels, however regulation at transcriptional level is of key importance and is dependent on the interaction of a variety of transacting proteins and genomic and epigenetic regulatory mechanisms (Mitchell and Tjian 1989).

Transcriptional regulatory elements are gene flanking DNA sequences that are involved in gene regulation and enable a complex interplay between regulatory proteins, genetic and epigenetic mechanisms. There are several kinds of regulatory elements that are recognized by regulatory proteins to either activate or repress gene expression (Narlikar and Ovcharenko 2009b). These include proximal promoter elements and distal regulatory elements named based on their position in relation to the target gene.

A typical CRE known as the promoter in eukaryotes is a collection of *cis*-regulatory elements, spanning the transcription start site (TSS) and extending a few hundred base pairs (bp) upstream and downstream of the TSS (Shahmuradov 2003; Jaenisch and Bird 2003). A eukaryotic promoter is further distinguished into a core promoter region and a proximal promoter region; both regions consist of various elements involved in transcriptional regulation (Smale and Kadonaga 2003). A core promoter is a minimal stretch of DNA required to drive the initiation of transcription by RNA polymerase and its associated machinery (Doerks et al. 2002). Core promoters

harbor various elements or functional motifs termed as ‘core promoter elements’

which add diversity to the functionality of the core promoter and contribute to the combinatorial regulation of expression (Smale and Kadonaga 2003). Some examples of the core promoter elements include: (i) The TATA box was the first ever core element identified and is generally present 25-30 bp upstream of the TSS (Breathnach and Chambon 1981) (**Figure 4.1**). The TATA box has not only been reported to be involved in RNA Pol II transcription but also in RNA Pol III related transcription (Wang et al. 1996) (ii) The Initiator (Inr) spans 6bp upstream and 11bp downstream of the TSS and like the TATA box, it facilitates binding of transcription factor IIB (TFIIB) (Xi et al. 2007) and can also trigger transcription initiation on its own or together with other core elements (O'Shea-Greenfield and Smale 1992). (iii) TFIIB recognition elements (BRE) are located upstream of the TATA box, where TFIIB binds in a sequence specific manner (Lagrange et al. 1998). (iv) CpG Islands are un-methylated CG rich regions and typically lack typical core promoter elements such as the TATA box (Brandeis et al. 1994). However, they do include downstream promoter elements (DPE) generally found in TATA-less promoters and located +28 to +32 to Inr (Burke and Kadonaga 1997). (v) Proximal promoter elements lie within ≈ 200bp upstream of TSS (**Figure 1**), do not always act like activators or repressors but rather facilitate contact between enhancers and the core promoter (Su et al. 1991).

Some *cis*-regulatory elements can also be present distal from the promoter i.e. a few thousand base pairs away from the transcription initiation site (Chandler 2001). These CREs are termed as distal regulatory elements (**Figure 4.1**) and are responsible

for transcriptional regulation after being recognised by different activators or repressors (Narlikar and Ovcharenko 2009a).

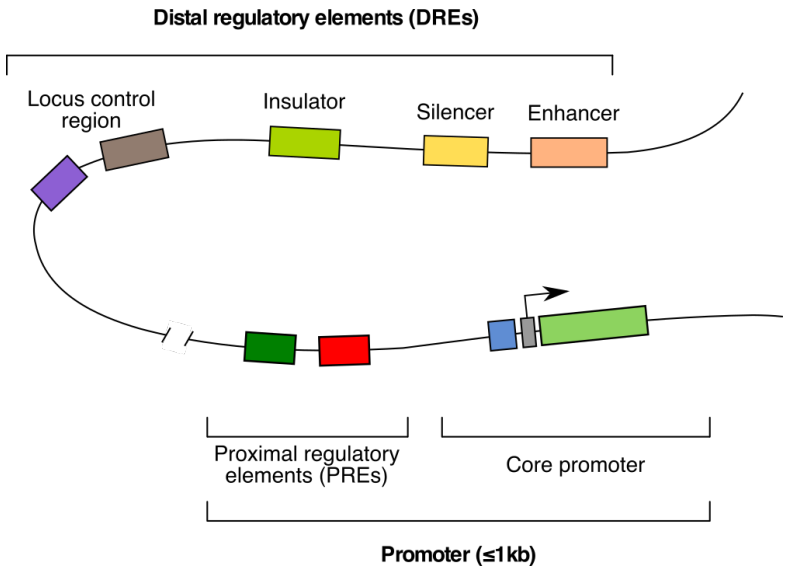

Figure 4.1: Schematic representation of eukaryotic *cis*-regulatory elements. Proximal promoter elements are present in close proximity to the core promoter, normally within 1 kb upstream of the gene whereas distal regulatory elements include enhancers, silencers, insulators and locus control regions are situated farther away from the core promoter (Maston et al. 2006).

Based on their specific role in transcriptional regulation, distal elements are generally classified as enhancers, silencers and insulators. Enhancers are DNA motifs located at a great distance either upstream, downstream or even inside the intron of the target gene (Levine 2010). They regulate cell functions such as stem cell multi-potency and chromosomal organisation as well as influence human evolution and disease (Anger et al. 2014). For instance, an enhancer region of *POU3F4* gene located on X chromosome is responsible for X-linked deafness in humans characterised by conductive hearing loss. Microdeletions have been found in a region present many hundred kb upstream of *POU3F4* gene, show similar hear loss phenotype suggesting

presence of enhancers in this region (Anger et al. 2014). Insulators on the other hand, act like 'barriers' between different chromosomal domains harboring protein-coding genes thus physically separating enhancers from the genes (Kellum and Schedl 1991). This suggests that any chromosomal deletions or inversions where these insulator elements reside can potentially lead to activation by an enhancer present in a separate domain resulting in altered expression of the targeted genes (Levine 2010). Silencers are short, specific stretches of DNA normally present upstream of the associated gene and are capable of repressing transcription (Ogbourne and Antalis 1998). Silencers are involved in many aspects of negative gene regulation such as affecting the activity of transcription factors, chromatin structure and 3' upstream un-translated region signal recognition (Clark and Docherty 1993). There are other distant elements other than enhancers, repressors or silencers, which can be involved in long distance regulation of the expression. B-globin genes in mammals, for example, are regulated by a set of elements located 20-30kb upstream of the gene cluster. This specific set of elements is termed as locus control region (LCR) and is significantly conserved in mammals (Kim and Dean 2012). There are only a few well characterised examples of DREs in plants. In maize, for instance, a quantitative trait locus, *vegetative to generative transition1* (*Vgt1*) acts as major flowering time locus and is located 70 kb upstream of a transcription factor encoding gene, *Ap2-like,* involved in flowering time regulation (Salvi et al. 2007). Another study on maize found that the B' region required for paramutation (interaction between alleles resulting in heritable changes in expression level) , is controlled by long-distance regulatory elements ~200 kb away (Stam 2002).

4.1.1. Mechanisms involved in long distance regulation

How do long distance regulatory elements physically interact with their target genes to regulate their expression? To answer this complex question, various models have been proposed to explain the physical interaction between distant genome regions. Due to the advancement of chromatin assay techniques, long range chromosomal interactions have been examined suggesting a loop configuration model (**Figure 4.2**), which is responsible for most of the distal physical interactions in the genome (Kadauke and Blobel 2009).

For instance, interplay among distant and proximal regulatory elements facilitated by nuclear transcription factor Y (NF-Y) and a floral regulator CONSTANTS (CO) results in regulation of *FLOWERING LOCUS T* (*FT*) expression in *A. thaliana* (Cao et al. 2014; Tiwari et al. 2010). Chromatin conformational capture (3C) assay has confirmed a physical interaction between *FT* promotor and far distant regulatory region called block C through chromatin loop formation (Cao et al. 2014).

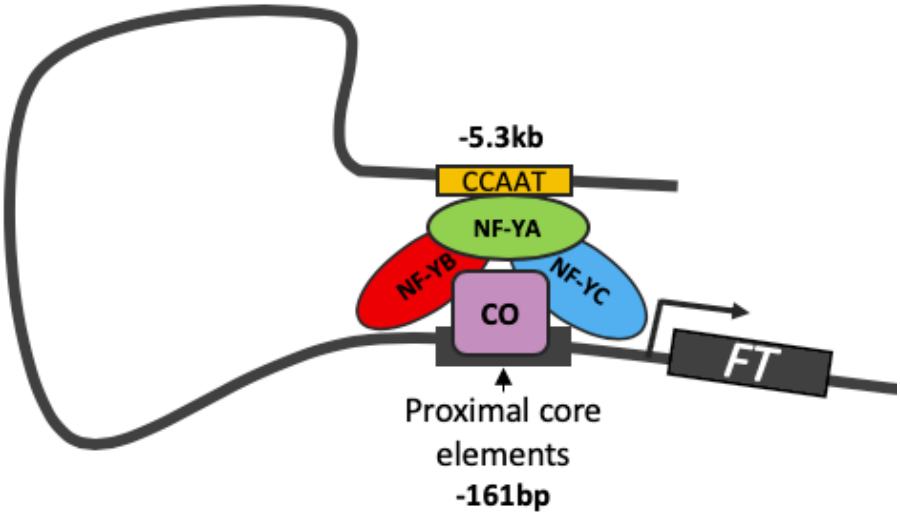

Figure 4.2: In *A. thaliana*, *FT* transcriptional activation by a chromatin looping by long distance interactions. A distal enhancer block C allows recruitment of a trimeric complex comprising NF-Y transcription factors (NF-Y A, B and C) through CCAAT boxes. This results in stable interaction of *CO* with its DNA binding site and bound NF-Y trimeric complex leading to *FT* activation in long day conditions (Cao et al. 2014).

Similarly, in mammals, a highly conserved TF, the CTCF (CCCTC-binding factor), has ability to make homodimers and facilitates chromatin looping (Filippova et al. 1996). CTCF has been has been found colocalised with cohesins (protein complexes that mediate chromosomal interactions) in chromatin immune-precipitation sequencing assay (ChIP-seq) indicating its ability to facilitate physical interactions with distal genomic regions (Lee and Iyer 2012).

4.1.2. Epigenetic modification of long-range regulatory regions

Recent genome-wide studies in plants have revealed that the expression of key developmental regulators is usually under the control of long-range regulatory sequences (Priest et al. 2009). Intriguingly, some of these regulatory sequences are modified epigenetically in response to an environmental stress, primarily by changes in DNA methylation. DNA methylation acquired during the stress response is "remembered" and transmitted to the next generation, which in most cases allow plants to respond efficiently to repeated exposure to the same stress, a process known as "priming" (Thiebaut et al. 2019; Crisp et al. 2016).

A study has revealed that RdDM might be involved in the alteration of chromosomal interactions between methylated sites and distal genes, thus altering the expression of associated genes (Rowley et al. 2017). *Ago4* mutants, which are defective in RdDM have higher rates of chromosomal looping, resulting in enhanced expression of certain genes (Rowley et al. 2017). DNA methylation of distant regions can contribute to dynamic reversible changes to chromatin topology, resulting in varied expression patterns (Ariel et al. 2014). Some of the known regulatory elements are long non coding RNAs (lncRNAs), which influence gene expression in a variety of ways, including direct interference with RNA polymerase binding and blocking transcription initiation (De Lucia and Dean 2011). LncRNAs can also recruit and guide chromatin modifying complexes to different chromosomal regions in *trans* (Rinn et al. 2007). In Arabidopsis, a long noncoding intergenic RNA coding region, *APOLO* (A*UXIN REGULATED* P*ROMOTER* LO*OP),* which is generally highly methylated via the RdDM pathway regulates the expression of a distal gene *PINOID,* a kinase involved in auxin response (Ariel et al. 2014).

By unravelling multiple layers of transcriptional regulation, gene function can be modulated ectopically without any genetic change. Although in plants, some of the regulatory mechanisms have been uncovered through genetic analyses, the function of far-distant controlling elements is still poorly understood. A significant factor is that the annotation of these *cis*-regulatory regions is extremely challenging, primarily due to their unpredictable position from their target promoters. In chapter 3, we devised a method for identification of regulatory elements by mapping chromatin accessibility and epigenetic modification. Identified accessible regions can be intersected with epigenomic data to predict the sequence of novel distal regulatory elements.

Since, the regulation of genes by these elements is also modulated by epigenetic marks, ectopic modification of these elements can be used to enable a direct functional characterisation of the associated genes. One way to modify chromatin in plants is to target *de novo* DNA methylation to the regulatory region using RdDM pathway, an endogenous *de novo* methylation pathway unique to plants (Li et al. 2015). Methylation can be targeted by generating an inverted hairpin to produce double stranded RNAs recognized by the RdDM pathway (**Figure 1.7**) (Dupre et al. 2015).

4.2. Chapter aims

The aim of this chapter is to identify and characterise the function of long-range regulatory sequences in plants. Using *A. thaliana*, work was carried out to investigate how these regions are influenced by epigenetic modifications and elucidate their function.

4.3. Results

4.3.1. Selection of tissue specific DREs

Multiple studies have confirmed that distal regulatory elements, such as enhancers and silencers, are predominantly found in accessible chromatin regions distal from the genes that they regulate (Shlyueva et al. 2014; Zhang et al. 2012; Pajoro et al. 2014). A recent study in maize has found 50% of the distal regulatory elements in intergenic regions are present in accessible regions of chromatin (Li et al. 2019).

To identify DREs in Arabidopsis, we intersected whole genome methylome and transcriptome data with our iNOMe data. We started this analysis by identifying 670 CG DMRs that were hypo-methylated in roots compared to shoots. First, we selected all genes which are located within 5000bp upstream or downstream of the DMRs (1566 genes). To analyse the expression pattern of these genes during developmental stages and tissues, the developmental series of "the AtGenExpress project" (Schmid et al. 2005) was employed. Data was available for 898/1566 genes, for which k-means clustering analysis was performed using a Pearson correlation coefficient as distance measurement (using the software Genesis, Sturn *et al*. 2003). We focused on genes especially expressed or repressed in root tissue. K=10 yielded sufficiently diverse clusters and we could identify one cluster with genes up-regulated in root tissues compared to other tissues (92 genes) and one with genes down-regulated in roots compared to other tissues (91 genes). These genes were then further characterised according to their known functions, DNA methylation pattern, expression values in wild type and methylation mutants (*met1* and *rdd*; information taken from http://neomorph.salk.edu/epigenome/epigenome.html) and the location of the DMR relative to the gene. We favoured genes that displayed

a defined methylation region and selected them according to their proximity (≤5kb upstream) to the transcription start site. Genes with a known function in development, response to hormones or stress were preferred. Using this strategy, we initially selected 7 candidates showing a good correlation between intergenic DMR, other epigenetic modifications and expression between tissues.

To narrow our list further, we compared the tissue specific DNA methylation created during cloning using different tissues (Wibowo et al. 2016). We selected (172) differentially methylated regions created during cloning and determined genes linked to them. We used K-means clustering of expression data (Schmid et al. 2005) We identified four distinct clusters. We selected gene clusters displaying higher expression in roots (54) and found three candidate intergenic regions that overlapped with the regions identified in the differential methylation analysis.

Overall, the analysis has identified 12 intergenic regions that could act as distal regulatory elements including a previously characterised long-distance regulatory lncRNA, *APOLO* (Ariel et al. 2014)(**Table 4.1)** . However, for the proof of concept, the study will focus on the analysis on the first three regions from identified list of DREs. These DREs will be epigenetically (by RNAi hypermethylation) and genetically (by CRISPR/Cas9) modified, followed by phenotypic/molecular characterisation of the modified lines.

Table 4.1: List of selected genes and their associated DMRs

No.	Gene	Gene description	Chromosome Coordinate
1	At5g09330	NAC domain containing protein 82 (NAC82)	Chr5:2,898,643-2,898,792
2	At3g54590	Proline rich extensin-like family protein (HRGP1)	Chr 3: 20209297-20209566
3 *(APOLO)*	AT2G34650	Auxin signalling (*PINOID*)	Chr2:14,596,565-14,597,543
4	At2g28720	Histone superfamily protein, DNA binding function	Chr2:12,324,429-12,324,628
5	At2g01620	RNI-like superfamily protein, protein binding function	Chr2:276,760-276,959
6	At5g02360	DC1 domain-containing protein	Chr5:503064-503316
7	At1g43040	SAUR-like auxin-responsive protein family	Chr1:16,184,300-16,184,599
8	At3g28220	TRAF-like family protein	Chr3:10,526,110-10,526,269
9	At3g14640	Putative cytochrome P450	Chr3:4918629-4918659
10	At2g04025	Root Meristem Growth factor 3	Chr2:1,279,755-1,279,994
11	At1g51190	Encodes a member of the AINTEGUMENTA-like (AIL) subclass of the AP2/EREBP family of transcription factors	Chr1:18,982,110-18,982,359
12	At4g12980	Auxin-responsive family protein	Chr4:7,592,720-7,592,969

4.3.2. DRE_NAC82 is a putative distal regulatory element

The first region identified with the potential of acting as a putative distal regulatory element, was a 150 bp long stretch of DNA present 3kb downstream of *NAC82* (DRE_NAC82). The accessibility of the region was confirmed by iNOMe assay, as well as with other chromatin accessibility datasets (**Figure 4.3**). The region seems to have high signal for both iNOMe and ATAC data sets but lower signal for DNase-seq data (**Figure 4.3**). This might be due to the sequence preference of each of the chromatin accessibility methods, for instance, DNase hypersensitive sites are mostly in proximal promoter regions and associated mainly with active gene expression (Pascoe et al. 2017) .

Tissue specific DNA methylation plays an essential regulatory role in plants, for instance, in the endosperm, DNA methylation regulates imprinted gene expression and crucial for normal embryonic development (Zhang et al. 2011). This indicates that the presence of tissue specific DNA methylation is an important indicator of the regulatory role of the DNA sequence. The DRE_NAC82 region displays tissue specific differences in DNA methylation in all methylation contexts (**Figure 4.3B**).

We also looked at presence of histone modification marks at this region to establish a link between chromatin accessibility, histone methylation and gene expression. H3K27me3 is required for the tight regulation of different transcriptional regulators expressed at different stages of development (Köhler and Makarevich 2006). In Arabidopsis, this modification is present at the 5’ end of the transcribed genes, indicating that regulatory regions are the main targets of H3K27me3 (Zhang et al. 2007). Post-translational modifications of histones also play an important role in chromatin organisation and methylation of H3K27me3 is well characterised histone

modification involved in regulation of gene expression by generating 'closed' chromatin (Li et al. 2008). The decrease in H3K27me3 methylation facilitates decompaction of the 'closed' chromatin and increase the accessibility of DNA sequence for DNA binding proteins (Pan et al. 2018). This genomic region lacks H3K27me3 but there is a strong enrichment of H3H27me1 in this region as well as flanking regions (**Figure 4.3A**). Differences in epigenetic modifications indicate that this region is dynamically accessible, thus could act putatively functional element. Tissue specific DRE_NAC82 is linked to NAC82, which is a transcriptional regulator implicated in stress response (Nuruzzaman et al. 2010). It is ubiquitously expressed in all tissue types but its expression is significantly high in inflorescence shoot apex and in tissues undergoing senescence (**Figure 4.4**) (Schmid et al. 2005). Since the expression of NAC82 is temporally and spatially regulated and associated with epigenetic variability of flanking sequences, we hypothesised that NAC82 may be regulated via distal regulatory sequence binding.

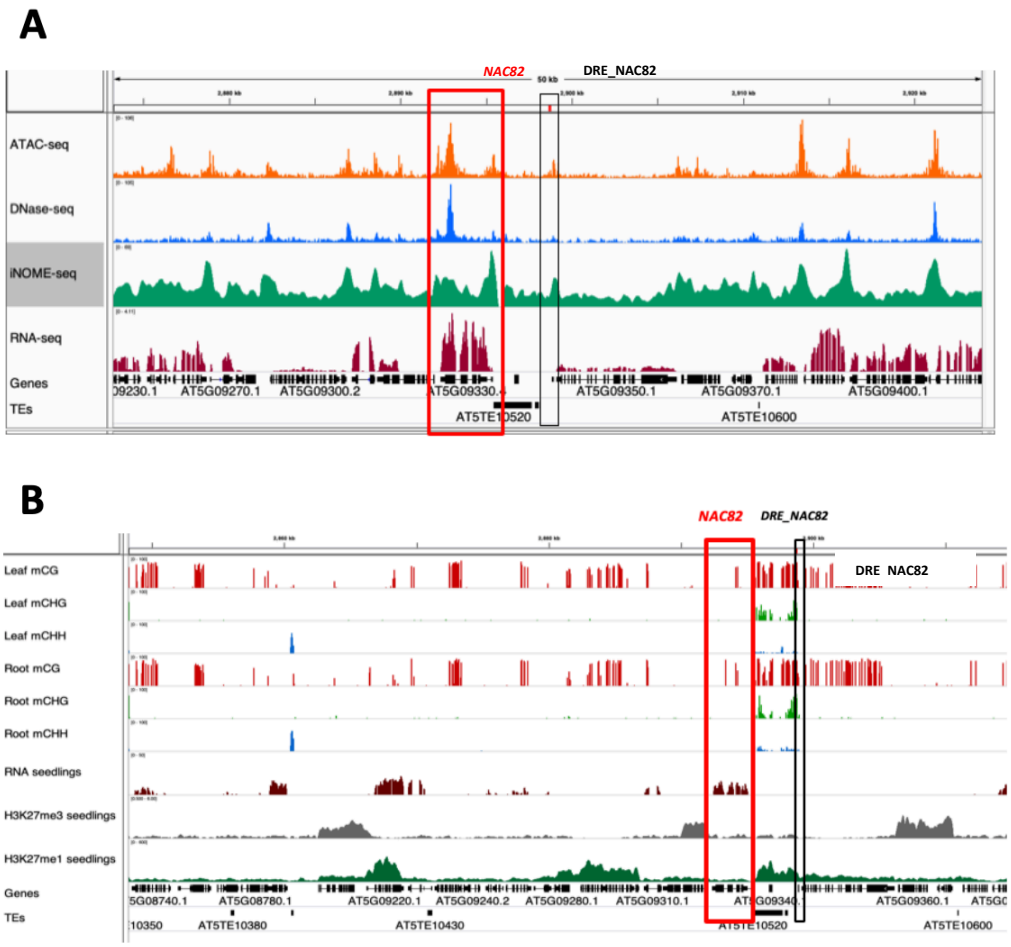

Figure 4.3: Chromatin accessibility and epigenomic profile of regions flanking *NAC82*. (A) iNOMe data compared with other accessibility methods shows that NAC_82 is present in an accessible region identified by different chromatin accessibility methods. The DMR is highlighted by a black rectangle and the associated gene is highlighted by a red rectangle. The orange track represents the ATAC signal, the blue track represents DNase signal, the green track represents iNOMe signal, while the gene expression is highlighted in the maroon peaks **(B)** Methylation status of DRE_NAC82 in different methylation contexts and tissue types (Wibowo et al. 2016) along epigenetic marks such as H3K27me3 and H3K27me1 (Data from Jose Marcos's lab, unpublished). The DMR shows higher methylation in roots than shoots (Differentially methylated) and is H3K27me3 depleted. The DMR is highlighted by a black rectangle and the associated gene is highlighted by a red rectangle. The red track indicates methylation in CG context, the green indicates CHG context and the blue track indicated CHH methylation. RNA expression, HeK27me3 and H3K27me1 are shown by the maroon, the grey and the dark green tracks respectively. Both images have been created by integrative genome viewer (IGV) (Robinson et al. 2011)

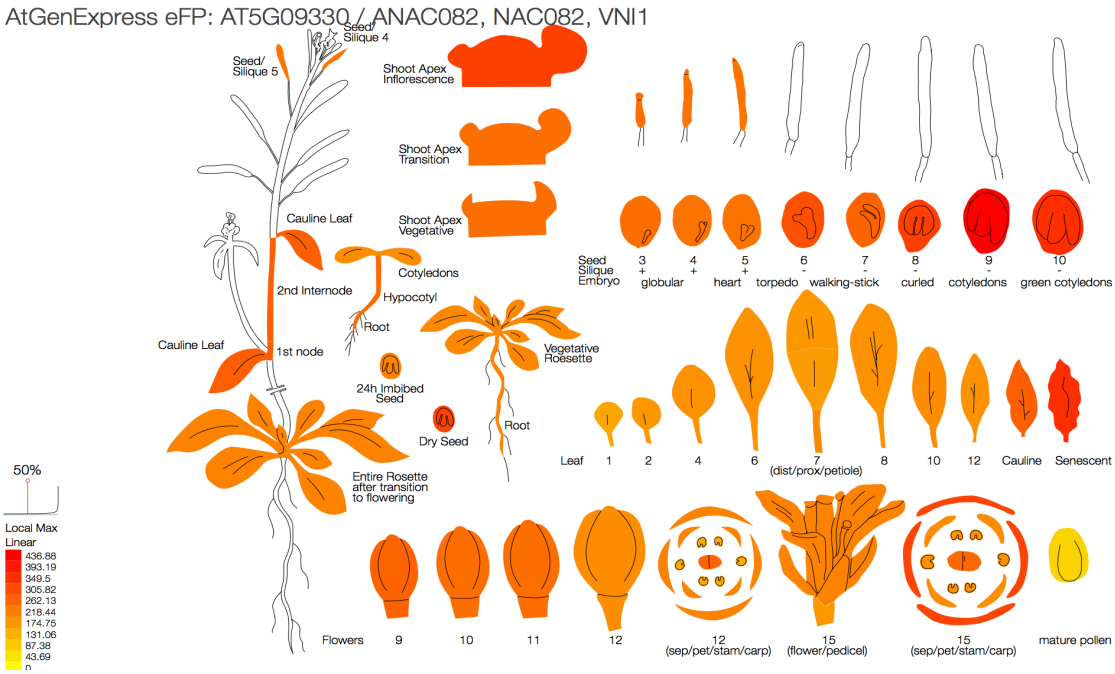

Figure 4.4: *NAC82* displays temporal spatial differences in gene expression. A snapshot of expression pattern of NAC domain containing protein 82 expression during various stages in *A. thaliana* published by AtGeneExpression (Schmid et al. 2005). NAC82 is ubiquitously expressed in all tissue types, however, the expression is very high in shoot apex inflorescence and at the leaf senescence stage. Image modified from http://bar.utoronto.ca/eplant/.

4.3.2.1. Hypermethylation of DRE-NAC82 leads to pleiotropic developmental phenotypes

RdDM mediated targeted methylation using dsRNA has already been used for transcriptional gene silencing in plants (Kasai and Kanazawa 2013). RdDM induced methylation can also be used modify the epigenetic state of the targeted genomic regions. Modified epigenetic state of the targeted region established by RdDM pathway is maintained in successive generations by endogenous methyltransferases (Chan et al. 2005). There is evidence that tissue specific differences in cytosine methylation result in phenotypic variation (Angers et al. 2010). This indicates that phenotypic traits can be controlled by altering the epigenetic state of the DNA. To test if epigenetic modification of a distal regulatory element may result in phenotypic or molecular changes, we hypermethylated DRE_NAC82 by using a RNAi hairpin (2.19). Analysis of the transgenic plants carrying a DRE_NAC82 inverted repeat hairpin revealed that the targeted methylation of this region resulted in phenotypic abnormalities (**Figure 4.5**). Hypermethylation of DRE-NAC82 resulted in significantly reduced rosette size and late flowering (n=30, t-test, p-value ≤ 0.05) (**Figure 4.5B, 4.5C**). The leaves appeared necrotic in 3/10 plants and asymmetrical rosette was observed in 6/10 hypermethylated plants (quantitative data not included) (**Figure 4.5A**).

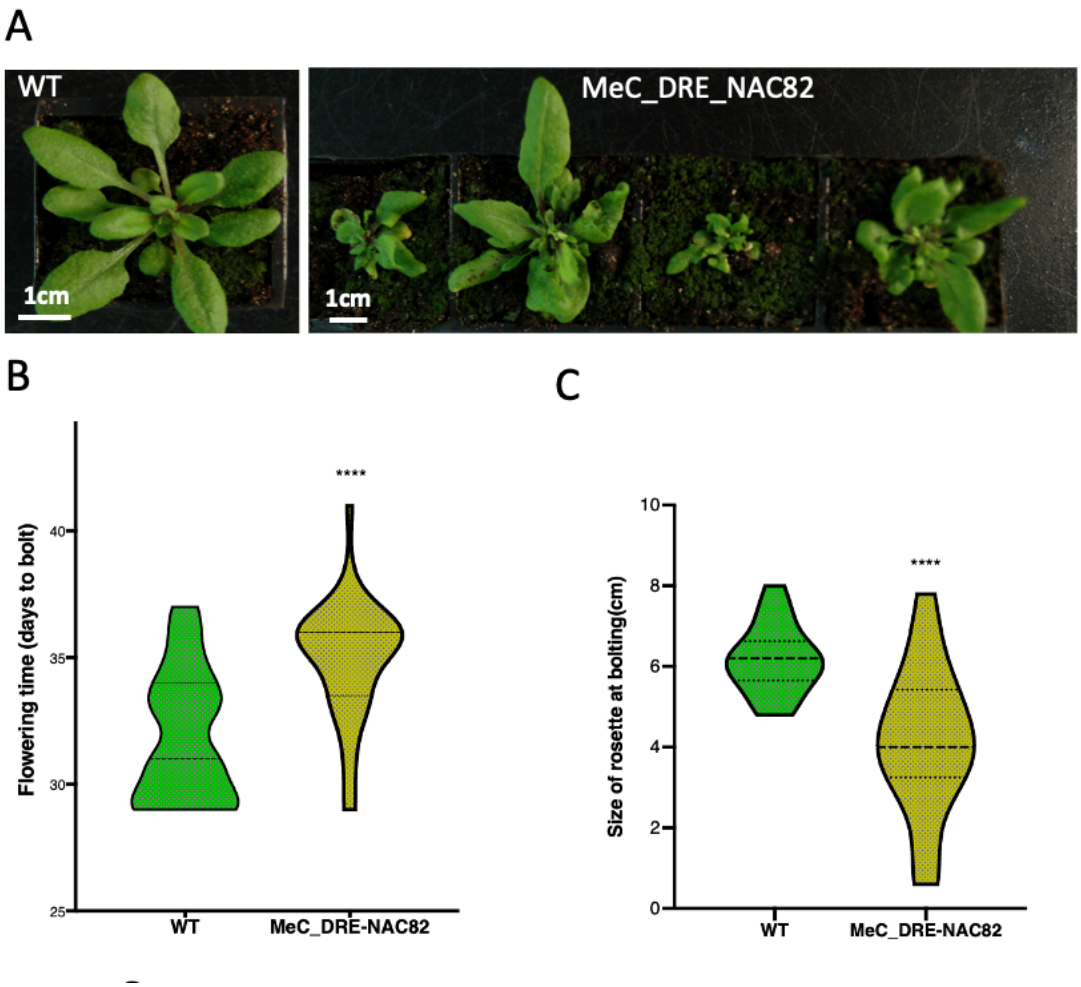

Figure 4.5: Hypermethylation of DRE_NAC82 leads to various growth and developmental phenotypes (A) Hypermethylation of DRE_NAC82 resulted in pleiotropic growth abnormalities (necrotic leaves, asymmetrical rosette appearance) in hypermethylated line (MeC_DRE-NAC82) vs control (WT). **(B)** Flowering is significantly delayed (n=30, t-test, p-value ≤ 0.05) in hypermethylated line (MeC_DRE-NAC82) compared to control (WT). **(C)** Rosette size is significantly smaller (n=30, t-test, p-value ≤ 0.05) in MeC_DRE-NAC82 compared with the control (WT).

4.3.2.2. Hypermethylation of DRE-NAC82 leads to downregulation of *NAC82*

The phenotypic abnormalities observed in DRE_NAC82 lines indicate underlying molecular changes. We hypothesise that these molecular changes are related to *NAC82* expression, the gene flanking DRE_NAC82.

NAC domain TFs are conserved in many plant species and have been implicated in a variety of abiotic stresses such as cold and drought (Nakashima et al., 2012, (Aida et al. 1997). A number of studies have revealed that the NAC proteins family are also involved in range of developmental stages such as apical meristem, seed and embryo development and a variety of other physiological processes (Duval et al. 2002; Kim et al. 2007; Kim et al. 2008). In Arabidopsis, *NAC82* has been found to mediate environmental stress response by reversing the proliferation defects caused by perturbation in ribosomal biogenesis (Ohbayashi et al. 2017).

Since the NAC82 is highly expressed in inflorescence shoot apex, the delayed flowering in hypermethylated DRE_NAC82 line, is likely to be caused by mis-regulation of NAC82. To test this hypothesis, DRE-NAC82 hypermethylated and WT lines were subjected to RT-qPCR analysis using two independent lines. The data supports the view that *NAC82* is downregulated in DRE_NAC82 hypermethylated lines (**Figure 4.6**). This supports our hypothesis that NAC82 expression is regulated by a distal regulatory element and that the expression of this gene is modulated epigenetically. To carry out functional characterisation of the region genetically, we attempted to generate a targeted deletion of DRE_NAC82 by CRISPR/Cas9. Unfortunately, due to the presence of tRNA in the close proximity of DRE-NAC82, we could not find highly efficient sgRNAs in the available intergenic region to generate the deletion lines.

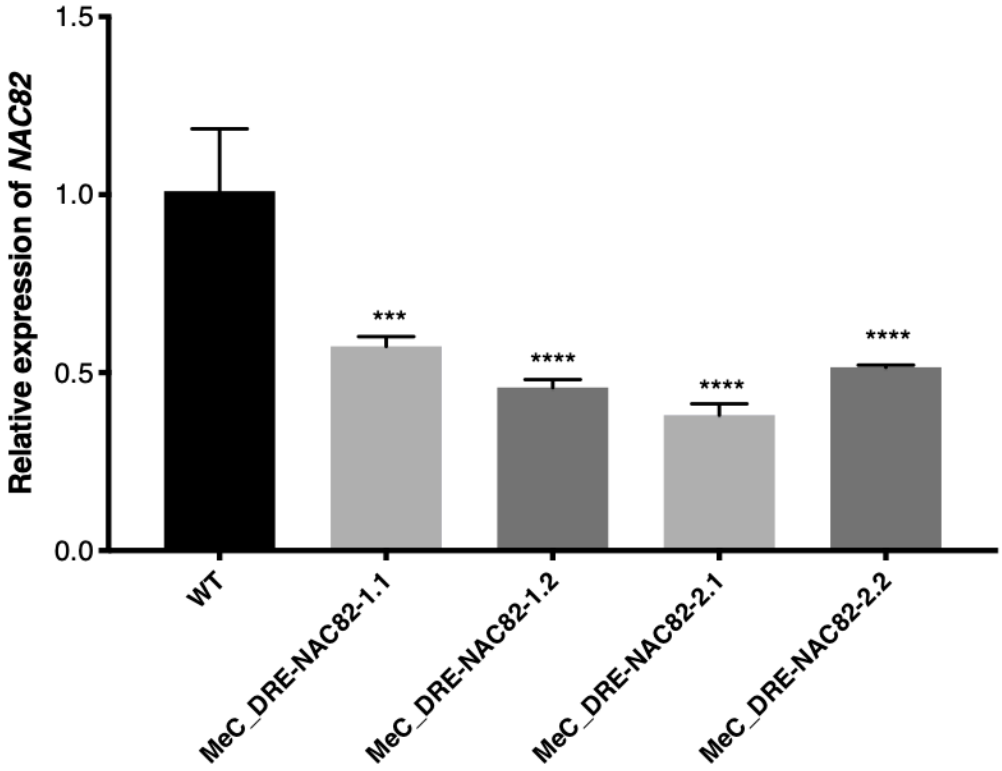

Figure 4.6: *NAC82* expression is regulated by hypermethylation of DRE_NAC82. Hypermethylation of DRE_NAC82 leads to significant downregulation (one-way ANOVA, p-value ≤ 0.05) of *NAC82*. Expression analysis of *NAC82* was carried out in 2 independent MeC_DRE_NAC82 lines (only MeC_DRE_NAC82-1 is shown in **Figure 4.5A)** compared to control (WT) expression.

4.3.3. DRE_HRGP1 is a putative distal regualtory element

A second potential regulatory region that could be epigentically controlled is a 270bp DNA sequence present approximately 1kb downstream of a proline-rich extensin-like gene *hydroxyproline-rich glycoprotein* (*HRGP1*). The accessibility of DRE_HRGP1 was confirmed by our iNOMe data but was not detected by other chromatin accessibilty datasets (**Figure 4.7A**). This region shows tissue specific diffrences in DNA methylation and also enriched for both H3K27me3 and H3K27me1 methylation (**Figure 4.7B**). The dynamic epigentic state of the region indicates that it might be involved in regualtion of tissue-specific gene expression. The gene flanking this region, *HRGP1* belongs to the extensin-like protein family, members of which are important components of the primary cell wall and crucial for cell wall strength and flexibility (Lamport and Northcote 1960). Gene expression data by Schmid et al (2005) has revealed that *HRGP1* is highly expressed in mature pollen and embryonic roots (**Figure 4.8**). Since the expression of *HRGP1* is temporally and spatially regulated and associated with epigenetic variability of flanking sequences, we hypothesised that it may be regulated by distal regulatory sequence.

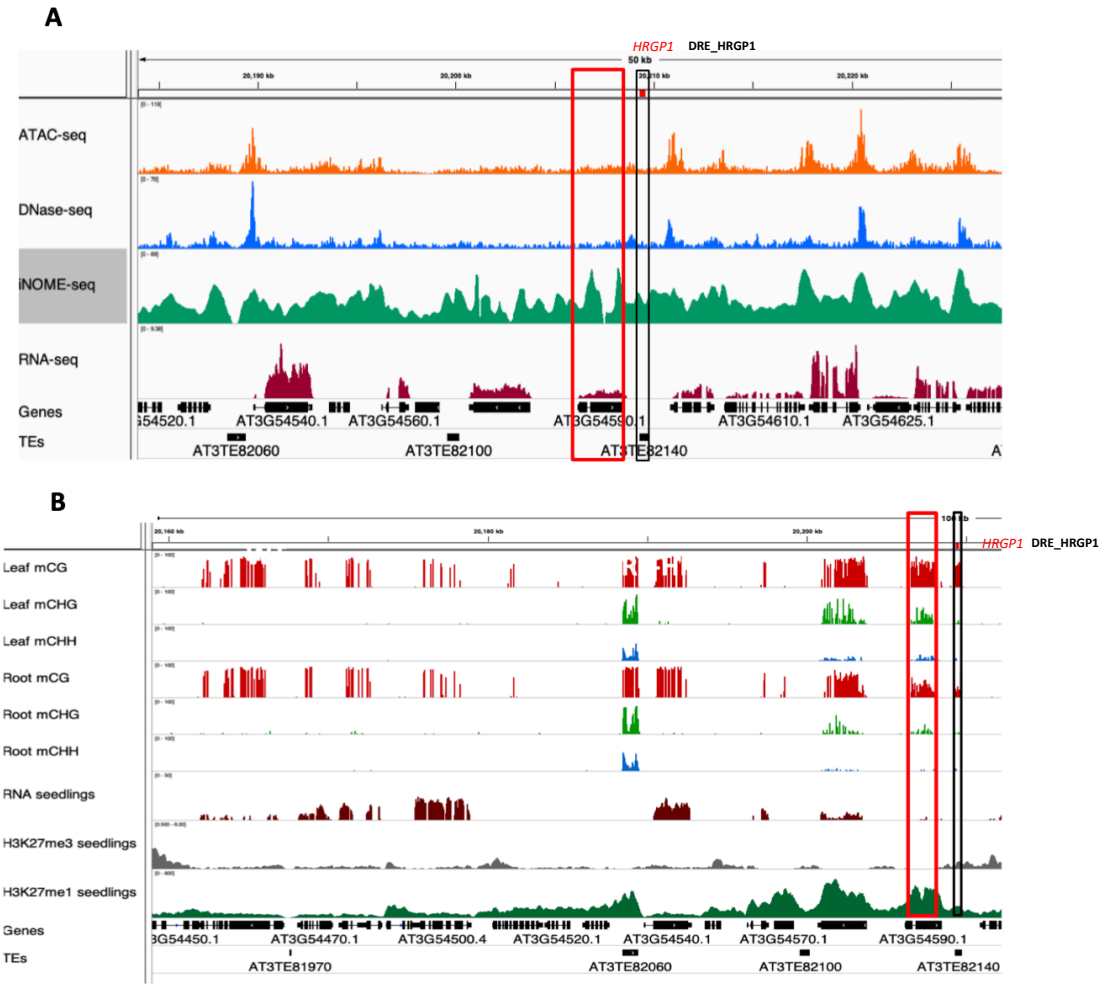

Figure 4.7: Chromatin accessibility and epigenomic profile of regions flanking DRE_HRGP1 region iNOMe data compared with other accessibility methods shows that DRE_HRGP1 is present in an accessible region identified by iNOMe only. The DMR is highlighted by a black rectangle and the associated gene is highlighted by a red rectangle. The orange track represents the ATAC signal, the blue track represents DNase signal, the green track represents iNOMe signal, while the gene expression is highlighted in the maroon peaks **(B)** Methylation status of DRE_HRGP1 in different methylation contexts and tissue types (Wibowo et al. 2016) along epigenetic marks such as H3K27me3 and H3K27me1 (Data from Jose Marcos's lab, unpublished). The DMR shows higher methylation in leaves than roots (Differentially methylated) and is moderately enriched for both H3K27me3 and H3K27me1. The DMR is highlighted by a black rectangle and the associated gene is highlighted by a red rectangle. The red track indicates methylation in CG context, the green indicates CHG context and the blue track indicated CHH methylation. RNA expression, HeK27me3 and H3K27me1 are shown by the maroon, the grey and the dark green tracks respectively. Both images have been created by integrative genome viewer (IGV) (Robinson et al. 2011)

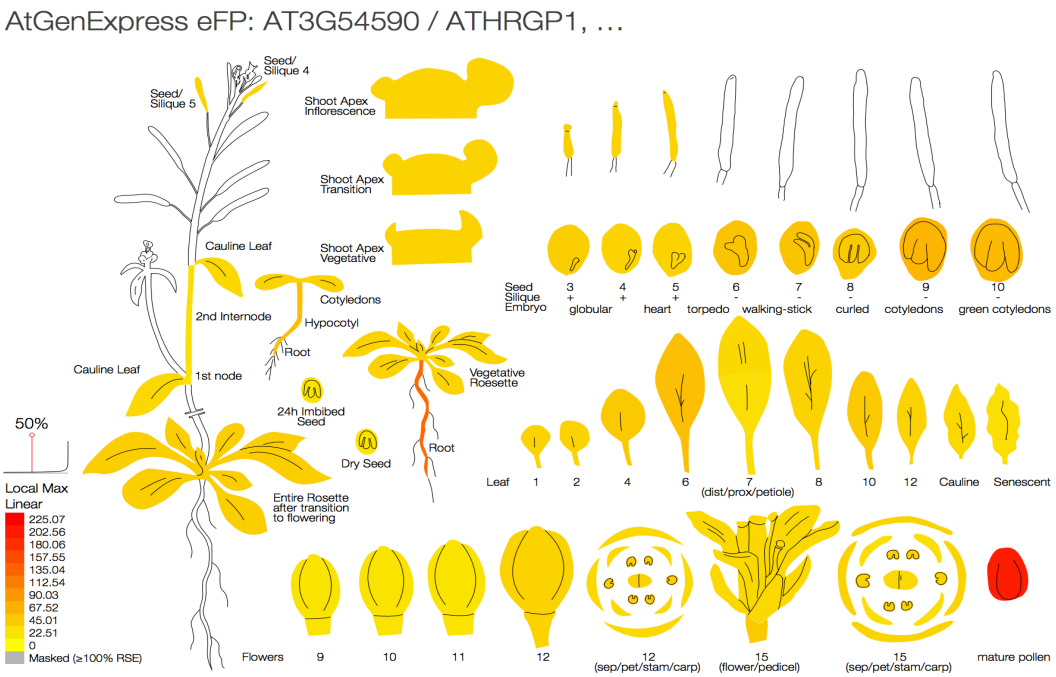

Figure 4.8: *HRGP1* is spatially expressed in roots and pollen. A snapshot of expression pattern proline-rich extension-like gene encoding hydroxyproline-rich glycoprotein (HRGP1) during various stages in *A. thaliana,* published by AtGeneExpression (Schmid et al. 2005). *HRGP1* is highly expressed in mature pollen and embryonic root. Image modified from http://bar.utoronto.ca/eplant/.

4.3.3.1. Hypermethylation of DRE_HRGP1 leads to growth and reproductive abnormalities

To assess if the DRE_HRGP1 region acts as a regualtory element, we genetrated a targeted hypermethylated line using a RNAi inverted repeat hairpin (2.19). We found that hypermethyalted DRE_HRGP1 lines displayed pleitropic developmental abnormalities, from plants failing to reach fertility that was associated with abnormal anther development and reduced silique size (**Figure 4.9A, 4.9B, 4.9C**).

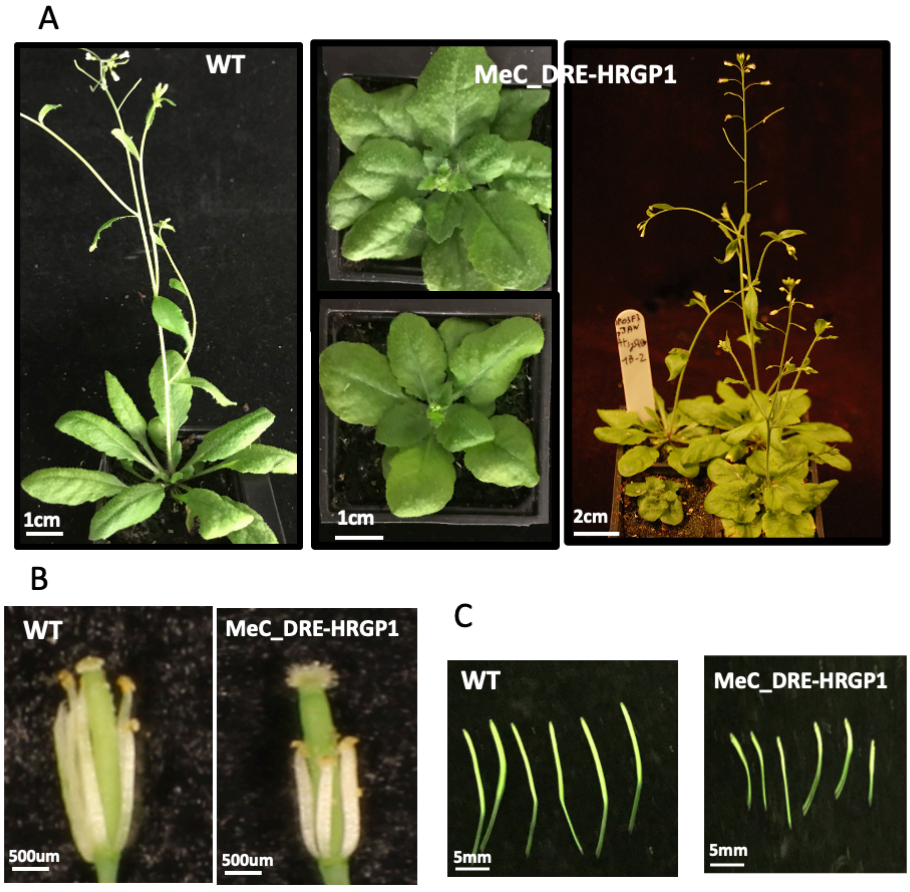

Figure 4.9: Targeted hypermethylation of DRE_HRGP1 leads to pleiotropic growth and reproductive defects. (A) DRE-HRGP1, the plants show rounded leaves and developmental delays as compared to WT. **(B)** DRE-HRGP1 **plants show** shorter anther size (qualitative data only included). **(C)** DRE-HRGP1 plants show shorter silique phenotype (qualitative data only included) compared to WT plants.

We, then decided to carry out targeted deletion of the DRE linked to HRGP1 to see if it affects regulation of its expression. We employed the CRISPR/Cas9 system for the targeted deletion of the DRE. Cas9 is RNA guided endonuclease associated with CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), involved in defense mechanism of *Streptococcus pyogenes* against certain viruses (Deltcheva et al. 2011). Due to its ability to cut DNA, it has been heavily used in genome modification in different organisms.

For the targeted deletion of HRGP1, sgRNAs were designed both upstream and downstream of DRE_HRGP1 flanking regions (**Figure 4.11A).** Targeted deletion was confirmed by PCR analysis of the region (**Figure 4.11B**). Sequencing of the DRE_HRGP1^{CR} lines confirmed a 420bp deletion (**Figure 4.11C**). Surprisingly, DRE-HRGP1^{CR} line did not exhibit any obvious developmental or reproductive abnormalities (**Figure 4.12 A, 4.12B, 4.12C**).

We, then predicted that the developmental abnormalities observed in hypermethylated DRE_HRGP1 may have been caused by mis-regulation of *HRGP1*. To test this hypothesis, DRE-HRGP1 hypermethylated and WT lines were subjected qRT-PCR analysis. However, due to the highly repetitive nature of the sequence of this gene, we could not design suitable assays.

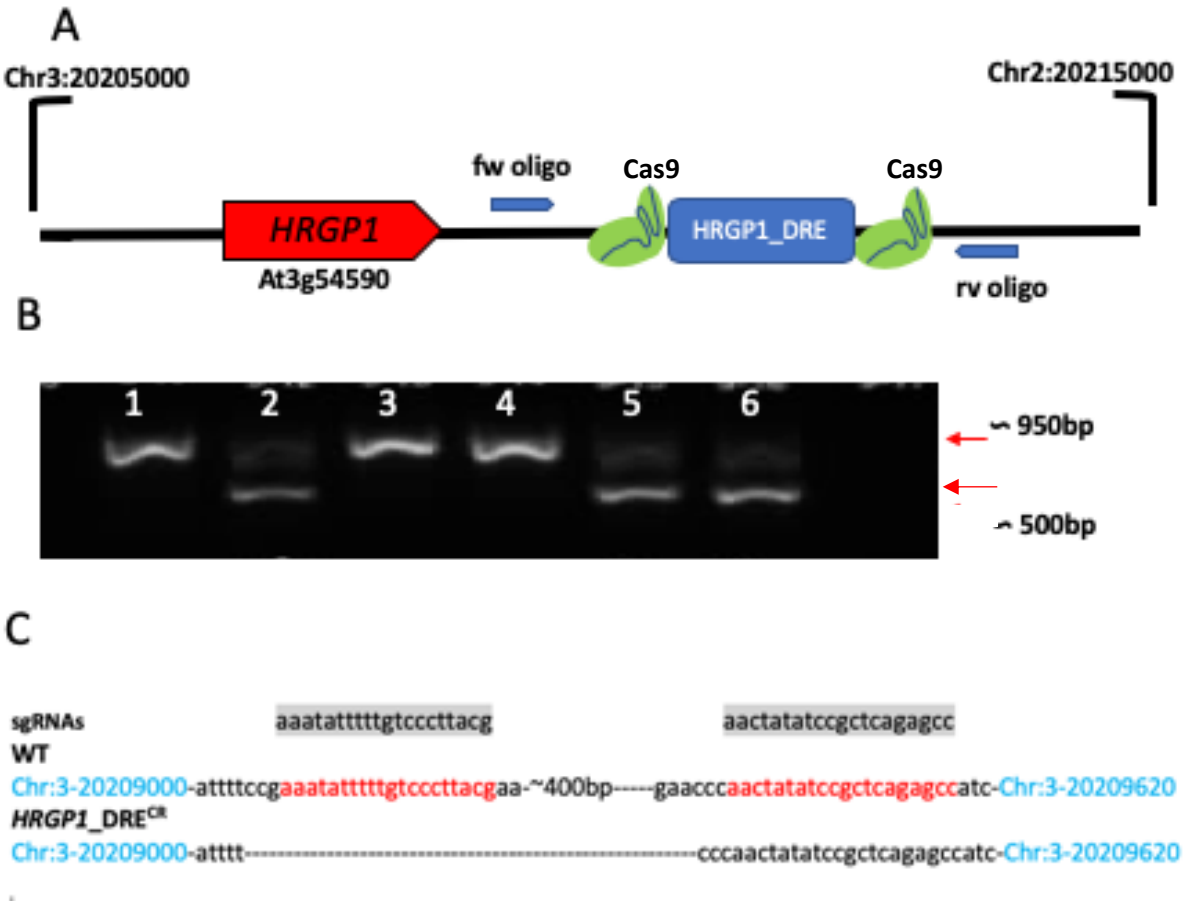

Figure 4.10: CRISPR/Cas9 targeted deletions for DRE_HRGP1 (A) Schematic representation of the **DRE_HRGP1** region: sgRNAs were designed on the flanking regions upstream and downstream of **DRE_HRGP1** (indicted by the position of Cas9). Genotyping oligos were designed outside of each flanking region (indicated by blue arrows). **(B)** *PCR* analysis to confirm deletion: expected size of PCR band is approximately 450 bp: WT control was run in lane 1 and in lane 2-6 test samples were run (positive deletion in lane 2,5 and 6) **(C)** Sequencing results for DRE_HRGP1 deletion**.** sgRNA sequence shown on the top highlighted in grey.

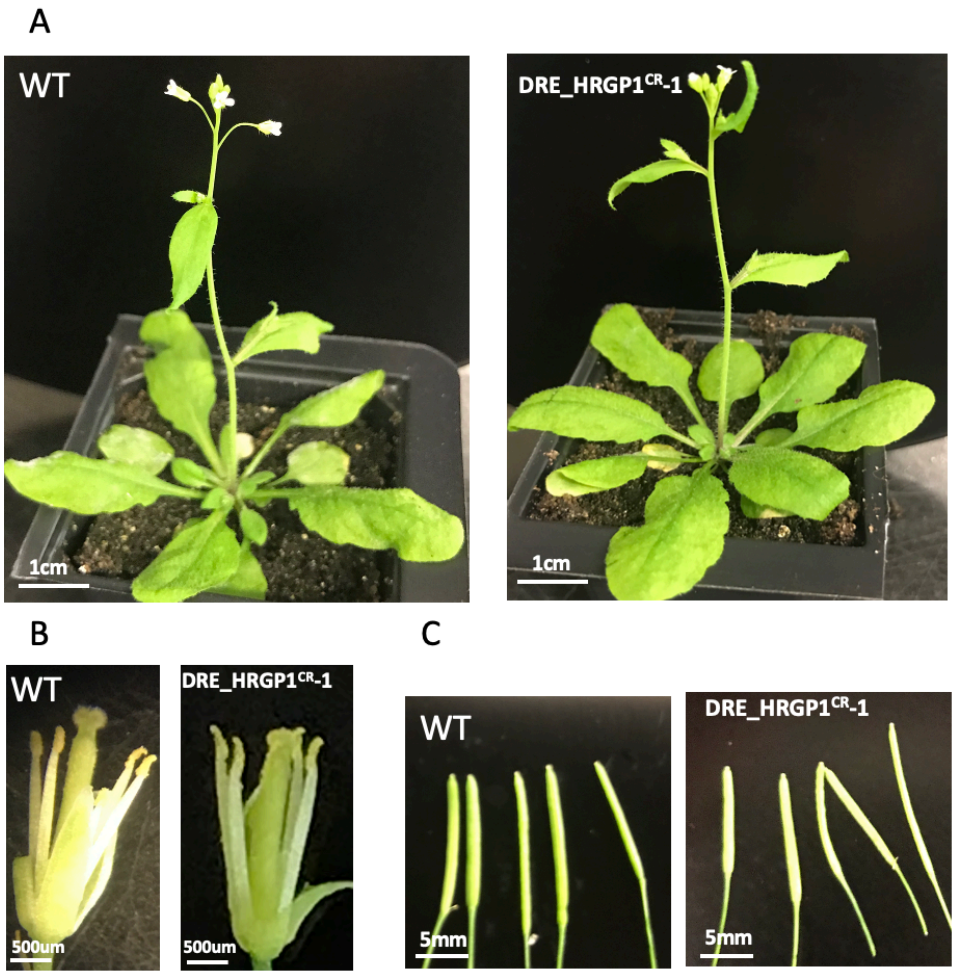

Figure 4.11: Phenotypic characterisation of DRE_HRGP1 deletion. Targeted deletion of DRE-HRGP1 was carried out by CRISPR cas9 and phenotypes were analysed (qualitatively). (A) Developmentally normal plants after targeted deletion of DRE-HRGP1. (B) No changes in anther development in DRE-HRGP1^{CR}-1 compared to WT plants (C) No differences in silique size in DRE_HRGP1^{CR}-1 compared to WT plants.

4.3.3.2. Functional characterisation of *HRGP1*

Although we were unable to carryout expression assay to establish that *HRGP1* is regulated by a distal regulatory element. Since this gene is highly expressed at pollen development stage, the fertility-related abnormalities observed in hypermethylated DRE_HRGP1 lines strongly suggested mis-regulation of HRGP1.

To test the hypothesis that *HRGP1* expression is regulated by a distal regulatory element, we generated RNAi lines by expressing inverted hairpins targeting the coding region of HRGP1 (2.19). Our data shows that knocking down *HRGP1* expression resulted in similar developmental abnormalities (shorter anther and silique size) to those observed in DRE_HRGP1 hypermethylated lines (**Figure 4.10 B, 4.10C**). HRGP1-RNAi line exhibited defects in anther development, reduced silique size and poor sed production (**Figure 4.10 B, 4.10C**). Our data (only qualitative data included here) suggests that HRGP1 is regulated by a distal regulatory element and is potentially involved in regualtion of develomental patways.

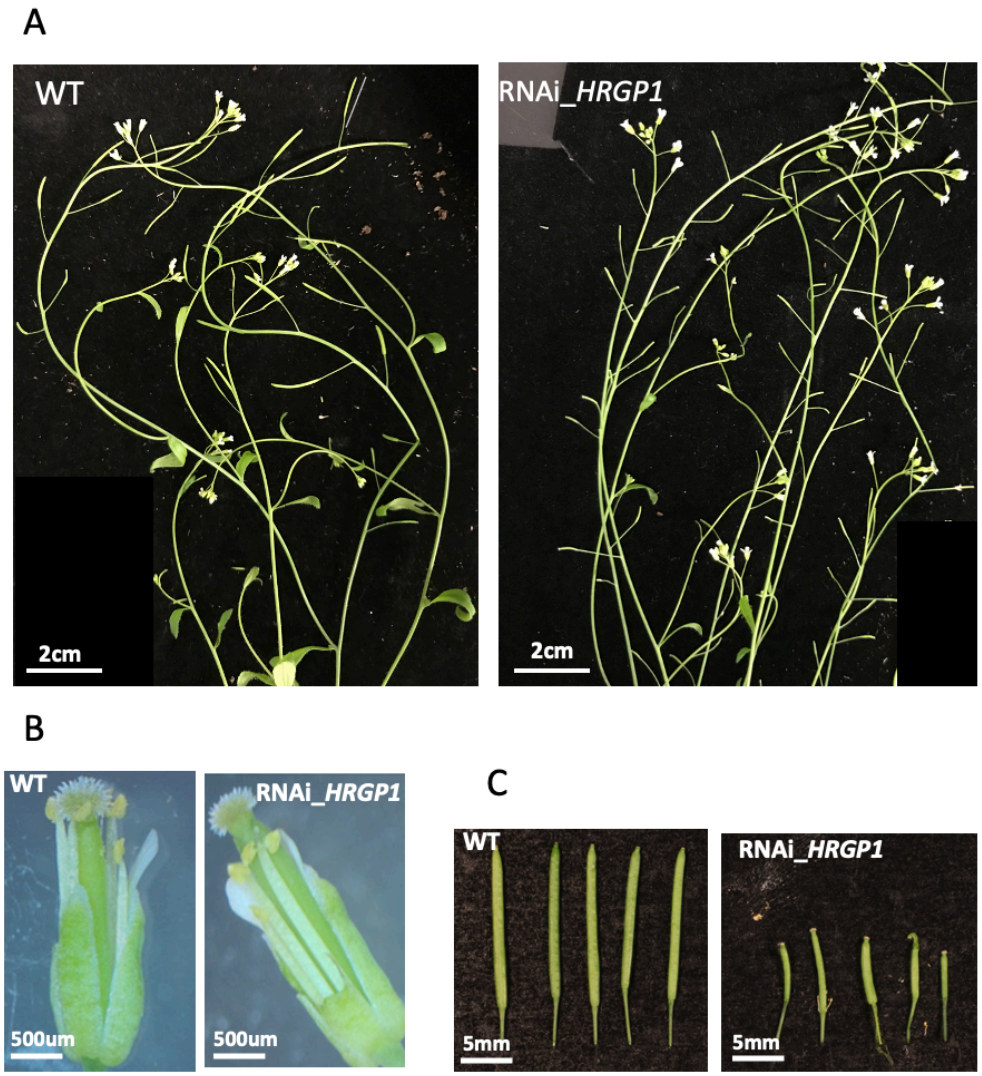

Figure 4.12 HRGP1_RNAi line show reproductive defects. (A) RNAi_*HRGP1* showing visibly shorter siliques compared to WT plant **(B)** RNAi_*HRGP1*) exhibit abnormal anther development (image kindly sent by our collaborator Cora MacAlister, University of Michigan **(C)** (RNAi_*HRGP1* plants show shorter silique size compared to WT plants. (Only qualitative data was collected for this experiment)

4.3.4. *APOLO* acts as distal regulator of *PID*

Our chromatin accessibility assay iNOMe identified *APOL*O, a lncRNA present 5kb upstream of *PINOID* (*PID*), encoding a kinase which regulates polar localisation of auxin in roots (Huang et al. 2010). Intriguingly, APOLO was identified as accessible by iNOMe, but not by other chromatin accessibility datasets (**Figure 4.13A**). This genomic region also shows tissue specific differences in DNA methylation in all contexts and is enriched for H3K27me1 marks (**Figure 4.13B**). Differences in epigenetic modifications also indicate that the region is dynamically accessible and a good candidate for long-distance regulation of the flanking genes.

APOLO expression is regulated in response to auxin, a phytochrome involved in primary axis formation, trophic movements of roots and shoots and root meristen patterning (Vanneste and Friml 2009). Expression data by Schmid et al (2005) showed that *APOLO* is highly expressed in inflorescence shoot apex (**Figure 4.14**). Impaired auxin transport in roots results in root hair abnormalities and gravitropic defects (Ganguly et al. 2010). A previous study has shown that hypermethylation of *APOLO*, dynamicallly controls *PID* and downregulation of *APOLO* results in altered root growth similar to *pid* mutants (Ariel et al., 2014). The study also revealed that the H3K27me3 in this region is dynamically regulatedted which results in dynamic regualtion of chromatin loop formation encompassing the *PID* promotor.

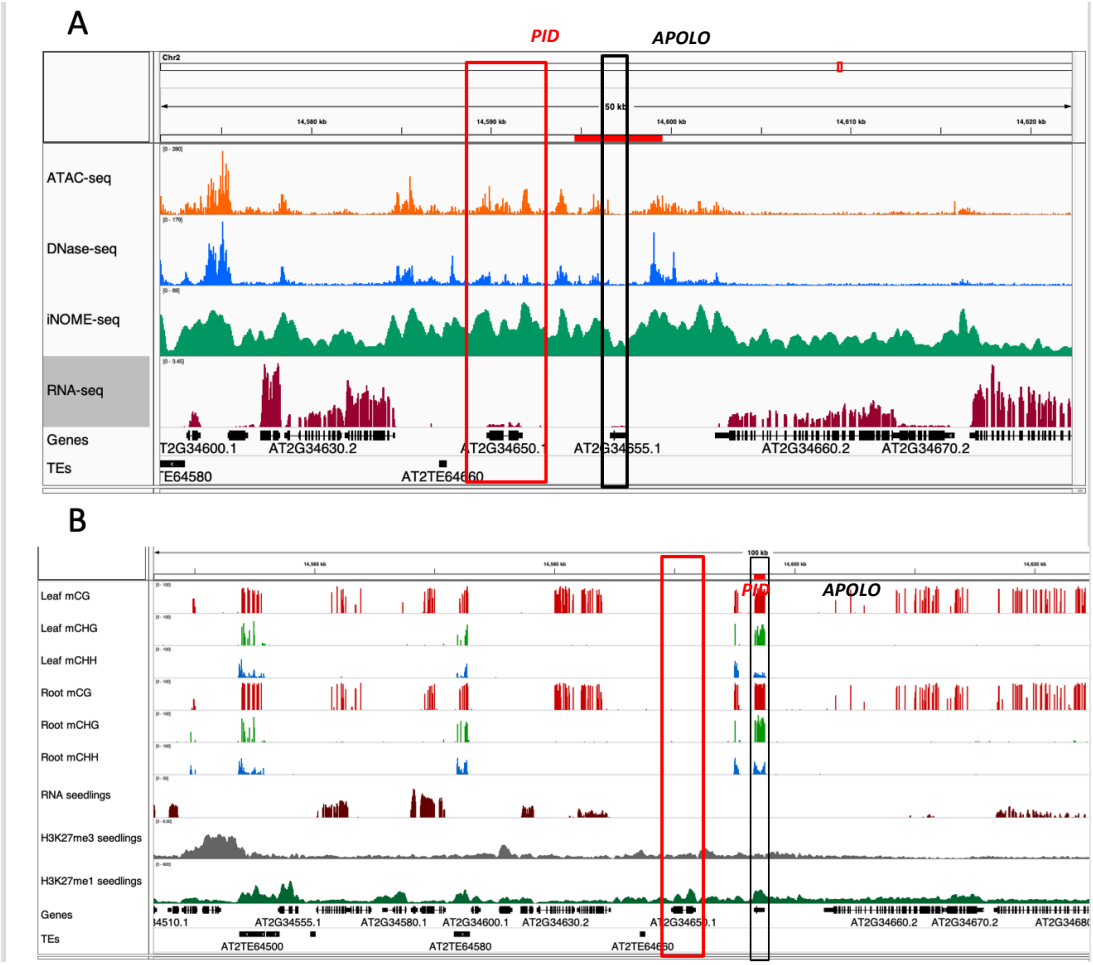

Figure 4.13. Chromatin accessibility and methylation profile at *APOLO* region (A) iNOMe data compared with other accessibility methods shows that *APOLO* is present in an accessible region identified by iNOMe only. The DMR is highlighted by a black rectangle and the associated gene is highlighted by a red rectangle. The orange track represents the ATAC signal, the blue track represents DNase signal, the green track represents iNOMe signal, while the gene expression is highlighted in the maroon peaks **(B)** Methylation status of *APOLO* in different methylation contexts (Wibowo et al. 2016) and tissue types along epigenetic marks such as H3K27me3 and H3K27me1 (Data from Jose Marcos's lab, unpublished). The DMR shows higher methylation in roots than leaves (Differentially methylated) and is enriched for both H3K27me3 and H3K27me1. The DMR is highlighted by a black rectangle and the associated gene is highlighted by a red rectangle. The red track indicates methylation in CG context, the green indicates CHG context and the blue track indicated CHH methylation. RNA expression, HeK27me3 and H3K27me1 are shown by the maroon, the grey and the dark green tracks respectively. Both images have been created by integrative genome viewer (IGV) (Robinson et al. 2011)

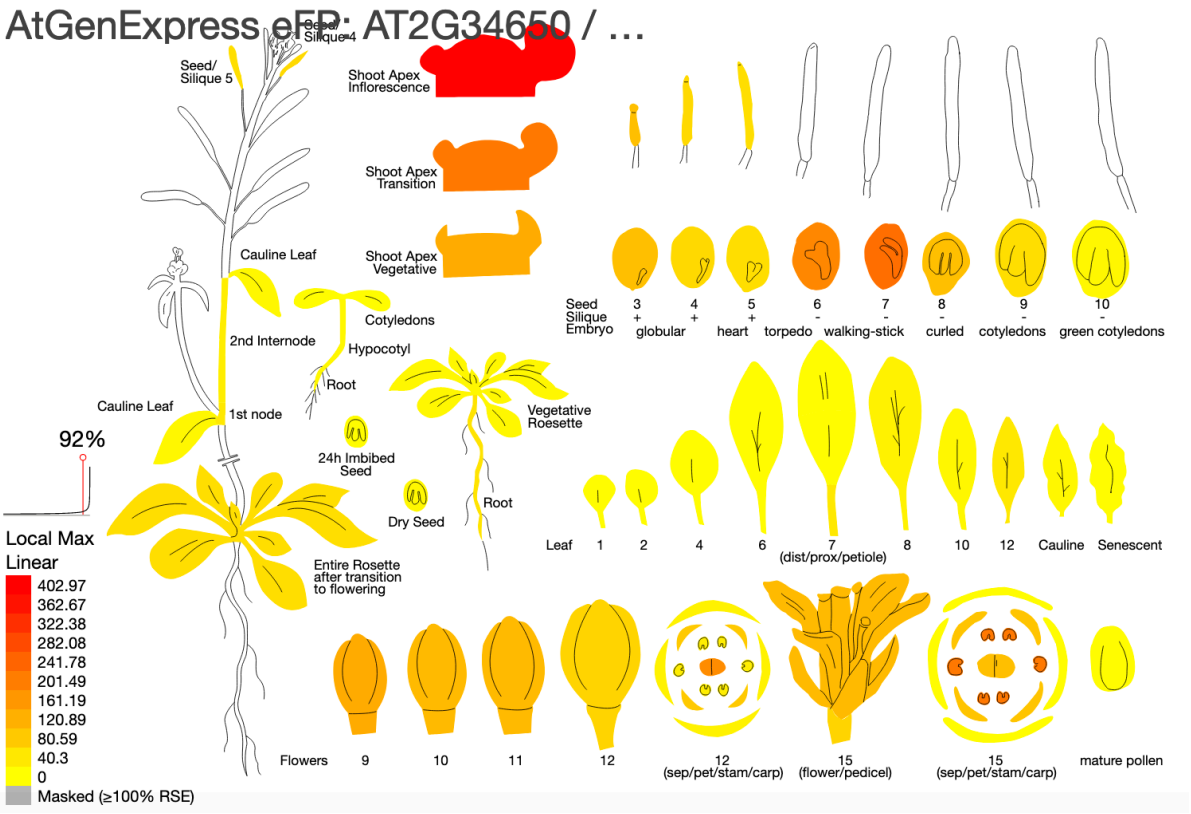

Figure 4.14: *PINOID* (*PID)* is differentially expressed spatially and temporally. A snapshot of expression pattern of PID (At2g34650) during various stages in *A. thaliana* published by AtGeneExpression (Schmid et al. 2005). The data shows that *APOLO* is highly expressed in inflorescence shoot apex. Image modified from http://bar.utoronto.ca/eplant/.

4.3.4.1. Targeted deletion of APOLO leads to changes in root hair growth

Since the the fuction of APOLO has already been established by RNAi hypermethylation, we focused our efforts on generating genetic lesion at *APOLO*. To do this, we carried out targeted deletions of APOLO using CRISPR/cas9. sgRNAs were designed both upstream and downstream flanking regions (**Figure 4.15A**). PCR analysis confirmed that *apolo*^{CR} lines had heritable deletions (**Figure 4.15B**) and sequencing validated our results by confirming a ~450bp deletion of the *APOLO* locus (**Figure 4.15C**). Once *APOLO* deletion was confirmed, we carried out a phenotypic analysis. Because *APOLO* regulates *PID* expression, we focused our analysis on root growth phenotypes. Our data shows that compared to WT, two independent *apolo*^{CR} lines displayed longer hair root hair length. Notably, *PID* overexpression results in suppressed root-hair growth (Lee and Cho 2006) and therefore, this data was in line with the known rule. This indicated that the *APOLO* region plays a crucial role in regulating root hair growth.

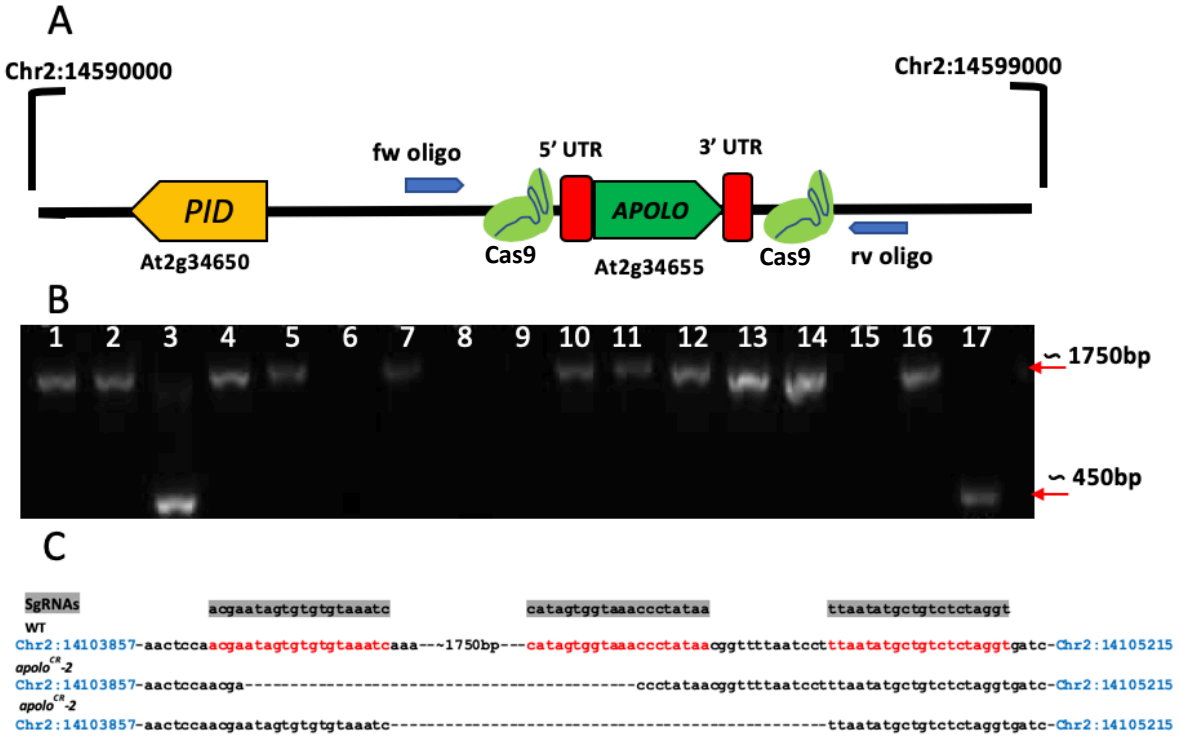

Figure 4.15: CRISPR/Cas9 targeted deletions for *APOLO*. (A) Schematic representation of *APOLO* region. sgRNAs were designed on the flanking regions upstream and downstream of *APOLO* (indicated by the position of Cas9). Genotyping oligos were designed outside of each flanking region (indicated by blue arrows). **(B)** PCR analysis to confirm deletion: expected size of PCR band is approximately 450 bp. WT band is shown in lane 1 *apolo*^{CR}*-1* deletion lines are shown in band 2-8 and *apolo*^{CR}*-2* lines are shown in lane 9-16. **(C)** Sequencing results for two independent deletion lines, both lines share same sgRNA sequence on the upstream ('A' sgRNA) of the APOLO region but downstream ('B' sgRNA) is different for each line.

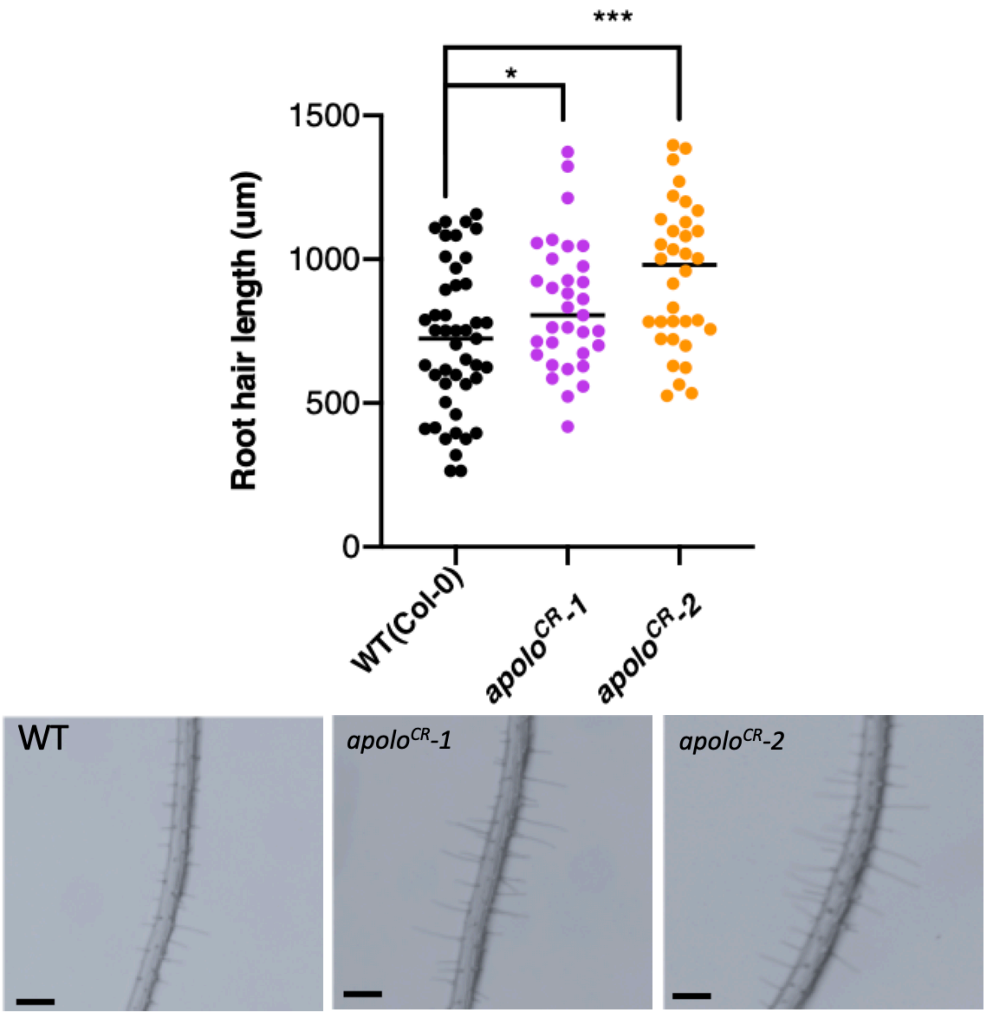

Figure 4.16 Longer root hair length in APOLO CRISPR deletion lines (A) Root hair length of Two APOLO mutants, *apolo*^{CR}*-1* and *apolo*^{CR}*-2* was compared with WT Col-0 plants. Both CRISPR deletion lines show significantly longer root hair length compared to WT (n=40, t-test and p-value ≤ 0.05) (B) *apolo*^{CR}*-1*, *apolo*^{CR}*-2* and WT root images showing longer root hair length (scale bar is set at 1000um).

4.4. Discussion

Long range regulation gene expression is very complex and it is dependent on many factors including transacting proteins, epigenetic mechanisms and spatial organisation of chromatin (Narlikar and Ovcharenko 2009a). Long-range regulatory elements, especially enhancers have been studied to some extent in animals, and they have been found to regulate a variety of functions (Kleinjan and van Heyningen 2005). However, long-range regulatory elements in plants are poorly understood as spatial organisation of functional elements remains unclear due to presence of an extra layer of epigenetic regulation through the RdDM pathway (Matzke et al. 2015). Chromatin accessibility is an important genomic feature and has been used to characterise the regulatory landscape in a variety of organisms (Bell et al. 2011).

Intergenic sequences generally have stable DNA methylation, but some regulatory regions show cell-specific differences in methylation, suggesting that cell specific gene expression can be regulated by the epigenetic regulation of these regions (Weber and Schübeler 2007). DNA methylation at intergenic regions is found to be dynamically regulated by other epigenetic marks and plays an important role in gene expression regulation (Wu et al. 2010). Taking the above facts into consideration, we have intersected our iNOMe chromatin accessibility data with whole-genome -omics data to identify putative intergenic functional elements. Through genetic and epigenetic manipulation of the identified genomic regions, we corroborated a direct regulatory relationship between genes and putative regulatory elements present in the flanking regions.

Hypermethylation of a distal regulatory element of *NAC82*, resulted in pleiotropic growth and developmental defects. A previous study has revealed that *NAC82* is

involved in mitigating ribosomal stress and cell proliferation defects in plants. (Ohbayashi et al. 2017). The epigenetic modification of the identified regulatory region resulted in downregulation of *NAC82*. These findings demonstrated the functional role of this DRE in *NAC82*-related regulation.

Another putative regulatory element was found in the regions flanking a proline-rich extension-like gene encoding hydroxyproline-rich glycoprotein (HRGP1). Epigenetic modification of this region resulted in reproductive organ defects and fertility. We observed similar phenotypic defects in knockdown plant lines of *HRGP1*. Previous studies have found extensin-like proteins to be associated with the regulation of reproductive organ development (Goldman et al. 1992). Thus, our data shows that the regulatory region flanking *HRGP1* is has a functional significance in reproductive development regulatory pathways. What mechanisms underpins crosstalk between these regions is not yet known. One possible explanation is that methylation of DRE can allow the recruitment of histone methyltransferases. Histone methyl transferases such as KRYTONITE(KYP) can epigenetically modify histones, which results in chromatin compaction, making it inaccessible for TF recruitment (Du et al. 2015). Another study also confirmed the formation of heterochromatin due to hypermethylation of the distal enhance region of *FLOWERING LOCUS T (FT)*, which lead to significant delayed flowering (Zicola et al. 2019b).

Regulatory regions generally harbour binding sites for multiple transcription factors and enhancers have been shown to interact with their target genes by making physical contacts (Krivega and Dean 2012). Deletion of the regulatory region can directly affect gene expression (Adrian et al. 2010). Targeted deletion allows functional characterisation of distal regulatory elements either by removing of TFs

binding sites or by disrupting 3D chromatin loop structures. Along these lines, we carried out targeted deletion of a regulatory lncRNA implicated in the auxin transport regulatory pathway. Our data shows that deletion of this regulatory region lead to phenotypic changes related to those seen with mis-regulation of auxin transport. *APOLO* influences 3D chromatin topology of the region by forming a loop by RdDM induced methylation resulting in dynamic control of auxin regulator, PID (Ariel et al., 2014). Targeted deletion of the region may have disrupted its loop structure, resulting in mis-regulation of the flanking gene *PID*. Studying 3D chromatin organisation is gaining more focus to help explain the role of long-distance interactions. A recent study employing this technique in maize has revealed that thousands of genes interact with intergenic regions and that some regulatory regions are gene-specific while others can regulate multiple targets (Peng et al. 2019). This indicates that distal regulation is a more prevalent in plant genomes than previously thought and characterising the role of distal regulatory is vital to help explain complex expression patterns in plants.

4.5. Conclusion

Our study concludes that gene expression can be modulated through modification of distal regulatory elements. Targeted methylation and genetic modification are efficient tools to characterise functional elements. However, to precisely elucidate interplay between the gene and its distal element, input from high resolution chromatin visualisation is vital. We anticipate that this approach could be used to modify traits related to stress tolerance in crops to improve food security.

5. General Discussion

5.1. Mapping chromatin accessibility in Arabidopsis using iNOMe-seq

In this study we have demonstrated that iNOMe is a powerful technique to simultaneously map chromatin accessibility, nucleosome occupancy and DNA methylation in plants (**Figure 5.1**) without relying on lengthy cell/nuclei isolation or enzymatic optimization processes (Chereji et al. 2017; Buenrostro et al. 2015). Like other chromatin accessibility methods previously developed, iNOMe is suited to identify accessible regions of chromatin where non canonical regulatory regions can be searched for. These regulatory regions are critical to understand the dynamics of transcriptional control that underline biological processes such as differentiation, growth, development and response to environmental signals (Adrian et al. 2010; Gibney and Nolan 2010). Regulatory sequences can be accessible or hindered based on the temporal or spatial organisation of the chromatin state of the cell (Allis and Jenuwein 2016). Mapping accessibility can be very powerful technique to identify distal regulatory elements that do not necessarily display any constraints in their direction, position or sequence and that are very challenging otherwise to identify based on comparative genomic analysis (Dickmeis and Müller 2005).

Current methods employed for chromatin accessibility rely on DNA fragmentation or transposase insertion, which requires optimisation and poses bias towards specific sequences. On the other hand iNOMe relies only on the presence of GpC dinucleotides, thus is technically simpler and does not show bias towards specific sequences (Rhie et al. 2018).

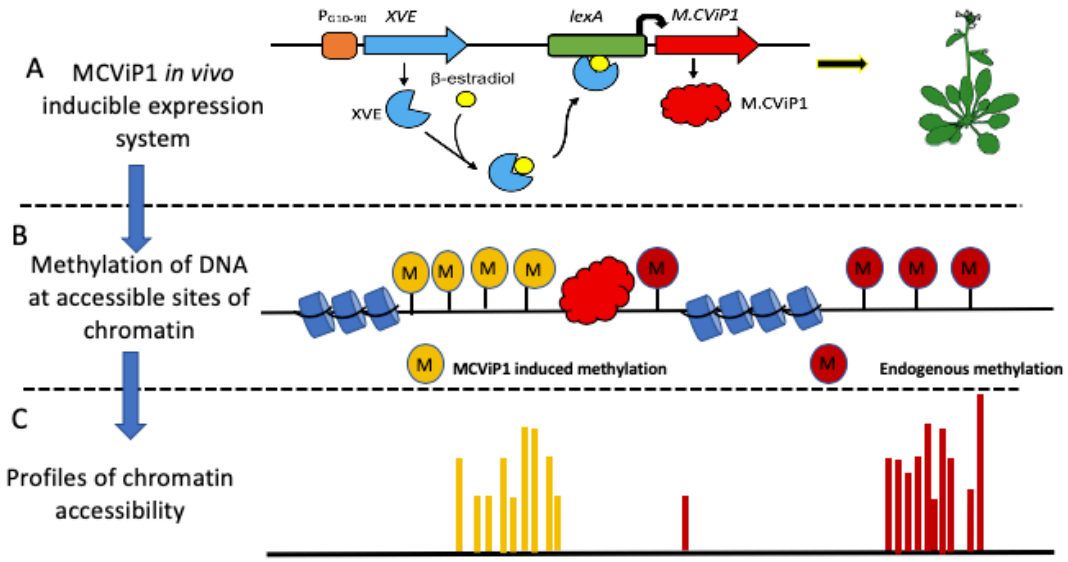

Figure 5.1 Schematic illustration of iNOMe-seq (A) *MCViPI* was cloned into a ß-estradiol based inducible system to generate *in vivo* expression line in *A. thaliana* (B) GC motifs in the regions of accessible chromatin are methylated, whilst MCViPI access in areas of condensed chromatin is restricted hence no methylation in those regions. (C) Methylated DNA is detected to generate chromatin accessibility profiles.

5.1.1. iNOMe-seq vs other methodologies

When we compared iNOMe-seq to other methodologies used to map chromatin accessibility, we found that iNOMe outperformed all the other because of its sensitivity but also because it can be used to detect chromatin, nucleosome occupancy and methylome in a single step. iNOMe efficiently discovered accessible regions that overlap with genes, TSS and Pol II binding sites. Since these regions are marked by nucleosome depletion, the identification is more straightforward than using other chromatin accessibility methods.

iNOMe is also able to estimate nucleosome positioning and determine nucleosome phasing, which adds another layer of information about the level of accessibility at particular regulatory regions (Kelly et al. 2012). Nucleosome positioning has a strong correlation with chromatin folding, thus iNOMe data could also aid in the

identification of differences in DNA topology for different chromatin compartments in the genome (Blank and Becker 1996). How the basic nucleosome structure is organised in a higher order structure remains unknown, however, differences in nucleosome positioning can be compared between different cell types to reveal the transition between active and silent chromatin (Iyer 2012). For example, chromatin accessibility at the regulatory lincRNA *APOLO*, was only detectable by iNOMe but not with other chromatin accessibility methods.The genomic region where *APOLO* is located has been shown to have dynamic accessibility due to the formation of chromatin loop after deposition of repressive histone marks (Ariel et al. 2014). It is thus possible that iNOMe allows to mark chromatin over a period of time, this will enable to detect highly dynamic changes in chromatin which contrast with other methodologies used for chromatin profiling.

5.1.2. Single cell analyses

Plant tissues, like in other eukaryotes, are complex in nature and consist of numerous cell types. Each cell type has a distinct epigenetic and transcriptional state owing to the specialized role of that cell in particular tissue. Therefore, single cell genomic analysis are critical to fully understand the molecular complexity present in different tissues (Schwartzman and Tanay 2015). DNA methylome analyses have been recently used for single cell epigenomic analyses in isolated mouse and human cells (Guo et al. 2015; Schwartzman and Tanay 2015). Single cell iNOMe analysis could reveal the relationship between DNA methylation and chromatin accessibility, but more importantly the link between epigenetic and transcriptional dynamics in response to developmental and environmental cues. Recent single-cell studies in mammals combining chromatin accessibility, DNA methylation and transcriptional

data have revealed novel associations between different layers of gene regulation (Clark et al. 2018). Therefore, the use of iNOMe-seq in plant cells for the combined analysis of chromatin accessibility and epigenetic analysis should be straightforward.

5.1.3. Application to other ecotypes and species

Plants grow in a variety of habitats and has gone through evolutionary process to adapt to a variety of environmental conditions over many generations (Bouchabke et al. 2008). Arabidopsis genotypes collected from different habitats show remarkable genetic and phenotypic variation (Koornneef et al. 2004). iNOMe could be used for comparative epigenomic analyses, using different Arabidopsis genotypes, to decipher mechanisms underpinning adaptive responses under different environments. Plants are sessile organisms, thus their response to environmental fluctuations is underpinned by dynamic changes in chromatin (Lämke and Bäurle 2017). Chromatin regulation relies on dynamic changes in nucleosome positioning, either through displacement or removal, as a result of the accumulation of epigenetic marks in response to environmental signals (Crisp et al. 2016; Zentner and Henikoff 2013). Therefore, a methodology such as iNOMe is capable of simultaneously quantifying different components of chromatin regulation is fundamental to understand the role of chromatin dynamics in plant adaptation to environmental stress. This methodology could be translated to economically valuable crops to identify regulatory elements implicated in valuable traits, thus enabling genomic assisted selection or to modify these genomic regions by targeted genetic/epigenetic manipulation.

5.2. Functional characterisation of distal regulatory elements

Eukaryotic genomes exhibit highly complex and tightly regulated gene expression, which largely depends on the effective interaction between regulatory elements and associated transcription factors (McBryant et al. 2006). Generally, proximal promoter elements are sufficient for the binding of transcriptional machinery and to initiate gene expression, however the involvement of distal regulatory elements can enhance or suppress gene expression through unknown mechanisms (Marand et al. 2017).

There has been growing interest in recent years to understand the mechanism(s) implicated in long-range gene regulation due to the rapid development in high throughput genome sequencing, genome and epigenome editing techniques and a growing realisation of the crucial role played by three-dimensional chromatin organisation (Long et al. 2016). Distal regulatory elements, especially enhancers, tend to exhibit sequence conservation and have been primarily identified following the study of TF binding sites (He et al. 2009). However, this approach is not sufficient to predict all enhancers and/or other regulatory elements such as suppressors and insulators. Several studies have identified a variety of regulatory elements in plants, most of them implicated in the regulation of development and stress responses (Adrian et al. 2010). Distal regulatory elements in plants have been predominantly found in open chromatin regions sensitive to DNase I treatment (Shlyueva et al. 2014). Another important genomic feature that offers functional contribution to regulatory activity is the presence of DNA methylation and/or post-translational histone modifications at regulatory sequences (Taberlay et al. 2014, Kouzarides,

2007). Although enhancers are generally present in nucleosome depleted regions (NDRs), nucleosomes in regions flanking NDRs are post-translationally modified by methyl or acetyl transferases (Kouzarides and Berger 2007). Recent genome-wide analyses using these defined chromatin features have identified thousands of long-distance functional regions in mammal, however their functional characterisation remains challenging (Hwang et al. 2013; Eveland et al. 2014). It is now well-acknowledged that the dynamic relationship between epigenetic modifications and distal regulatory elements is critical for fine transcriptional regulation (Rowley et al. 2017).

In this study we have identified putative distal regulatory elements by intersecting chromatin accessibility, whole genome DNA methylation and expression data. We have employed targeted genetic and epigenetic modifications to define the function of the identified regulatory elements. We have taken advantage of the RNA directed DNA methylation (RdDM) pathway that exists in Arabidopsis (Chan 2008) using RNA hairpins (RNAi) for the targeted hypermethylation of genome target sequences. RNAi-induced methylation has been similarly used in Arabidopsis to elucidate the role of distal regulatory elements (Zicola et al. 2019a). Study of the transcriptional profile of genes linked to newly identified DREs revealed that they are involved in long-distance modulation of gene expression. We also found that the genetic and/or epigenetic modification of these DREs resulted in developmental and reproductive differences, thus suggesting that targeted modification of these genomic regions could be an efficient tool to generate phenotypic variability in plants. Although our study strongly supports that gene expression in plants is influenced by the genetic and epigenetic status of DREs, the precise mechanisms underlying this regulation

remains unknown. Recent insights in to three dimensional chromatin organization in eukaryotes has revealed that chromosomes are present in dynamic topological domains where distal interactions take place between enhancers and their target genes (Li et al. 2018). By integrating chromosome capture data with our iNOMe data, it could be possible to enhance our understanding of regulatory mechanisms. Recent work in plants has shown that single distal regulatory element can regulate multiple target genes (Zhu et al. 2015) and that plants with complex genomes and that have undergone human domestication are rich in DREs (Li et al. 2019) The regulatory function of the candidate DREs identified in plant could be easily tested using reporter assays similar to ß-glucuronidase (GUS) assay, used to validate the enhancer function (Yan et al. 2019).

Distal regulatory elements are also present as redundant elements, acting as buffers to lethal effects of genetic variation (Wittkopp and Kalay 2012). Functional characterisation of long-distance regulatory elements could unravel the evolutionary forces that drives differences between plant genotypes. We foresee that the targeted genetic/epigenetic modification of conserved distal regulatory sequences will be fundamental to understand and manipulate these processes.

5.3. Future Work

Future work could be directed to applying iNOMe-seq in different Arabidopsis ecotypes and hybrids, which may pave the way to translating this methodology to other plant species. Using iNOMe-seq to investigate the dynamics in chromatin organisation between ecotypes or species would allow the identification of regulatory regions that are evolutionary conserved.

In addition, developing high throughput methodologies for the targeted genetic and epigenetic manipulation of potential regulatory elements will be critical for the functional characterisation of these genomic regions and to ascertain their role in plant development.

5.4. Conclusion

In this study we have developed in Arabidopsis, a novel method (iNOMe) to simultaneously map chromatin accessibility, DNA methylation and nucleosome positioning. iNOMe is a versatile technique, which can be used to gain a better understanding of chromatin accessibility in variety of cell/specie types and aid with the identification of genomic regions involved in development and responses to environmental conditions. Moreover, our iNOMe method could be used to reveal the dynamics of chromatin organisation during the developmental transition from undifferentiated to highly specialised cells (**Figure 5.2**).

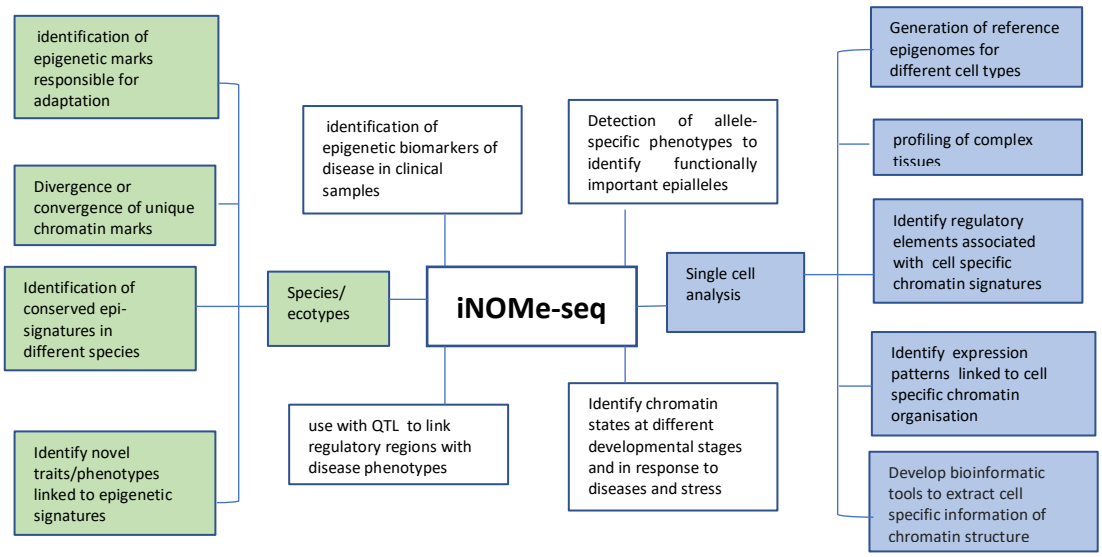

Figure 5.2: Applications of iNOMe-seq. iNOMEe-seq allows us to examine chromatin in number of cell/tissue types and species. Boxes highlighted in blue are related iNOMe-seq application to single-cell based analyses, the green boxes highlight iNOMe applications in different species and ecotypes and the white boxes highlight application in different settings/environments.

Our study also confirmed that the distal regulatory regions can be modified genetically and epigenetically to finely tune gene expression that might result in phenotypic variability. Identifying functional elements and their associated epigenetic signatures then characterising them will shed light on the regulatory components underpinning developmental decisions. The ability to epigenetically modify accessible chromatin identified with iNOMe could aid in the generation of novel epimutations that could be stably inherited over generations. These novel epimutations could be linked to beneficial phenotypes and therefore selected to enhance valuable traits in economically crop plants.

6. References

Adrian J, Farrona S, Reimer JJ, Albani MC, Coupland G, Turck F (2010) Cis-regulatory elements and chromatin state coordinately control temporal and spatial expression of *FLOWERING LOCUS T* in Arabidopsis. Plant Cell 22 (5):1425-1440. doi:10.1105/tpc.110.074682

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M (1997) Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. The Plant Cell 9 (6):841-857. doi:10.1105/tpc.9.6.841

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, and Walter P (2002) Molecular Biology of the Cell. Garland Science. doi:https://www.ncbi.nlm.nih.gov/books/NBK21054/

Allis CD, Jenuwein T (2016) The molecular hallmarks of epigenetic control. Nature Reviews Genetics 17 (8):487

Anderson JD, Widom J (2000) Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. Journal of Molecular Biology 296 (4):979-987. doi:10.1006/jmbi.2000.3531

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.

Anger GJ, Crocker S, McKenzie K, Brown KK, Morton CC, Harrison K, MacKenzie JJ (2014) X-Linked Deafness-2 (DFNX2) Phenotype associated with a paracentric inversion upstream of POU3F4. American Journal of Audiology 23 (1):1. doi:10.1044/1059-0889(2013/13-0018)

Angers B, Castonguay E, Massicotte R (2010) Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. Molecular Ecology 19 (7):1283-1295. doi:10.1111/j.1365-294X.2010.04580.x

Ariel F, Jegu T, Latrasse D, Romero-Barrios N, Christ A, Benhamed M, Crespi M (2014) Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. Molecular cell 55 (3):383-396

Asensi-Fabado MA, Amtmann A, Perrella G (2017) Plant responses to abiotic stress: The chromatin context of transcriptional regulation. Biochimica et Biophysica Acta Gene Regulatory Mechanisms 1860 (1):106-122. doi:10.1016/j.bbagrm.2016.07.015

Bell O, Tiwari VK, Thoma NH, Schubeler D (2011) Determinants and dynamics of genome accessibility. Nature Reviews Genetics 12 (8):554-564. doi:10.1038/nrg3017

Berger SL (2007) The complex language of chromatin regulation during transcription. Nature 447 (7143):407

Biswas M, Voltz K, Smith JC, Langowski J (2011) Role of histone tails in structural stability of the nucleosome. PLoS Computational Biology 7(12):e1002279. doi:10.1371/journal.pcbi.1002279

Blank TA, Becker PB (1996) The effect of nucleosome phasing sequences and DNA topology on nucleosome spacing. Journal of Molecular Biology 260 (1):1-8. doi:10.1006/jmbi.1996.0377

Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, Tang H, Pikaard CS (2015) Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. Elife 4 (4)

Boland Michael J, Nazor Kristopher L, Loring Jeanne F (2014) Epigenetic regulation of pluripotency and differentiation. Circulation Research 115 (2):311-324. doi:10.1161/CIRCRESAHA.115.301517

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15):2114-2120. doi:10.1093/bioinformatics/btu170

Bouchabke O, Chang F, Simon M, Voisin R, Pelletier G, Durand-Tardif M (2008) Natural variation in Arabidopsis thaliana as a tool for highlighting differential drought responses. PLoS One 3 (2):e1705

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132 (2):311-322. doi:10.1016/j.cell.2007.12.014

Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Temper V, Razin A, Cedar H (1994) Spl elements protect a CpG island from de novo methylation. Nature 371 (6496):435-438

Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. Annual Review of Biochemistry 50 (1):349-383

Brent R, & Ptashne M (1980) The lexA gene product represses its own promoter. Proceedings of the National Academy of Sciences of the United States of America 77 (4):1932–1936. doi:doi:10.1073/pnas.77.4.1932

Brogaard K, Xi L, Wang JP, Widom J (2012) A map of nucleosome positions in yeast at base-pair resolution. Nature 486 (7404):496-501. doi:10.1038/nature11142

Buenrostro JD, Wu B, Chang HY, and Greenleaf WJ (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. Current Protocols in Molecular Biology 109 (1):21-29.

Burke TW, Kadonaga JT (1997) The downstream core promoter element, DPE, is conserved fromDrosophila to humans and is recognized by TAFII60 of Drosophila. Genes & Development 11 (22):3020-3031

Cao S, Kumimoto RW, Gnesutta N, Calogero AM, Mantovani R, Holt BF (2014) A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the *FLOWERING LOCUS T* promoter and regulates the timing of flowering in Arabidopsis. The Plant Cell 26 (3):1009-1017

Chan SW (2008) Inputs and outputs for chromatin-targeted RNAi. Trends in Plant Science 13 (7):383-389

Chan SW, Henderson IR, Zhang X, Shah G, Chien JS, Jacobsen SE (2006) RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in arabidopsis. PLoS Genetics 2 (6):e83. doi:10.1371/journal.pgen.0020083

Chan SWL, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. Nature Reviews Genetics 6 (5):351-360. doi:10.1038/nrg1601

Chandler VLaV, H. (2001) Gene activation and gene silencing. Plant Physiology 125 (1):145-148

Chang P, Gohain M, Yen M-R, Chen P-Y (2018) Computational methods for assessing chromatin hierarchy. Computational and Structural Biotechnology Journal 16:43-53. doi:https://doi.org/10.1016/j.csbj.2018.02.003

Chen H, and Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC bioinformatics 12 (1):35

Chereji RV, Bryson TD, Henikoff S (2019) Quantitative MNase-seq accurately maps nucleosome occupancy levels. Genome Biology 20 (1):198. doi:10.1186/s13059-019-1815-z

Chereji RV, Clark DJ (2018) Major determinants of nucleosome positioning. Biophysical Journal 114 (10):2279-2289. doi:10.1016/j.bpj.2018.03.015

Chereji RV, Ocampo J, Clark DJ (2017) MNase-sensitive complexes in yeast: nucleosomes and non-histone barriers. Molecular cell 65 (3):565-577. e563

Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. Annual Review of Biochemistry 78 (1):273-304. doi:10.1146/annurev.biochem.77.062706.153223

Clark AR, Docherty K (1993) Negative regulation of transcription in eukaryotes. Biochemical Journal 296 (Pt 3):521

Clark SJ, Argelaguet R, Kapourani C-A, Stubbs TM, Lee HJ, Alda-Catalinas C, Krueger F, Sanguinetti G, Kelsey G, Marioni JC, Stegle O, Reik W (2018) scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nature Communications 9 (1):781. doi:10.1038/s41467-018-03149-4

Cress WD, and Triezenberg SJ, (1991) Critical structural elements of the VP16 transcriptional activation domain. Science 251 (4989):pp.87-90.

Crisp PA, Ganguly D, Eichten SR, Borevitz JO, Pogson BJ (2016) Reconsidering plant memory: intersections between stress recovery, RNA turnover, and epigenetics. Science Advances 2 (2):e1501340

Cui K, Zhao K (2012) Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. Methods in Molecular Biology 833:413-419. doi: 10.1007/978-1-61779-477-3_24.

Cumbie JS, Filichkin SA, Megraw M (2015) Improved DNase-seq protocol facilitates high resolution mapping of DNase I hypersensitive sites in roots in Arabidopsis thaliana. Plant Methods 11:42. doi:10.1186/s13007-015-0087-1

De Lucia F, Dean C (2011) Long non-coding RNAs and chromatin regulation. Current Opinion in Plant Biology 14 (2):168-173. doi:10.1016/j.pbi.2010.11.006

De Smet I, Beeckman T (2011) Asymmetric cell division in land plants and algae: the driving force for differentiation. Nature Revies Molecular Cell Biology 12 (3):177-188. doi:10.1038/nrm3064

Deal RB, Henikoff JG, Henikoff S (2010) Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science (New York, NY) 328 (5982):1161-1164. doi:10.1126/science.1186777

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471 (7340):602-607. doi:10.1038/nature09886

Dickmeis T, Müller F (2005) The identification and functional characterisation of conserved regulatory elements in developmental genes. Briefings in Functional Genomics 3 (4):332-350

Doerks T, Copley RR, Schultz J, Ponting CP, Bork P (2002) Systematic identification of novel protein domain families associated with nuclear functions. Genome Research 12 (1):47-56. doi:10.1101/

Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL (2006) Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. Environmental Health Perspect 114 (4):567-572. doi:10.1289/ehp.8700

Dong X, Weng Z (2013) The correlation between histone modifications and gene expression. Epigenomics 5 (2):113-116. doi:10.2217/epi.13.13

Dresselhaus T, Hückelhoven R (2018) Biotic and abiotic stress responses in crop plants. Agronomy 8 (11). doi:10.3390/agronomy8110267

Du J, Johnson LM, Jacobsen SE, Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. Nature Reviews Molecular Cell Biology 16 (9):519-532. doi:10.1038/nrm4043

Dupre E, Herrou J, Lensink MF, Wintjens R, Vagin A, Lebedev A, Crosson S, Villeret V, Locht C, Antoine R, Jacob-Dubuisson F (2015) Virulence regulation with Venus flytrap domains: structure and function of the periplasmic moiety of the sensor-kinase BvgS. PLoS Pathogens 11 (3):e1004700. doi:10.1371/journal.ppat.1004700

Durr J, Papareddy R, Nakajima K, 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. doi:10.1038/s41598-018-22667-1

Duval M, Hsieh TF, Kim SY, Thomas TL (2002) Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. Plant Molecular Biology 50 (2):237-248. doi:10.1023/a:1016028530943

Eveland AL, Goldshmidt A, Pautler M, Morohashi K, Liseron-Monfils C, Lewis MW, Kumari S, Hiraga S, Yang F, Unger-Wallace E (2014) Regulatory modules controlling maize inflorescence architecture. Genome Research 24(3):431-443

Fahad S, Bajwa AA, Nazir U, Anjum SA, Farooq A, Zohaib A, Sadia S, Nasim W, Adkins S, Saud S, Ihsan MZ, Alharby H, Wu C, Wang D, Huang J (2017) Crop production under drought and heat Ssress: Plant Responses and Management Options. Frontiers in Plant Science 8 (1147). doi:10.3389/fpls.2017.01147

Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV (1996) An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Molecular Cell Biology 16 (6):2802-2813. doi:10.1128/mcb.16.6.2802

Flores O, Deniz Ö, Soler-Lopez M, Orozco M (2014) Fuzziness and noise in nucleosomal architecture. Nucleic acids research 42 (8):4934-4946

Fuks F (2005) DNA methylation and histone modifications: teaming up to silence genes. Current Opinion in Genetics & Development 15 (5):490-495. doi:https://doi.org/10.1016/j.gde.2005.08.002

Gaertner B, Zeitlinger J (2014) RNA polymerase II pausing during development. Development 141 (6):1179-1183. doi:10.1242/dev.088492

Ganguly A, Lee SH, Cho M, Lee OR, Yoo H, Cho H-T (2010) Differential auxin-transporting activities of PIN-FORMED proteins in Arabidopsis root hair cells. Plant Physiology 153 (3):1046-1061. doi:10.1104/pp.110.156505
Gibney ER, Nolan CM (2010) Epigenetics and gene expression. Heredity (Edinb) 105 (1):4-13. doi:10.1038/hdy.2010.54
Goldman MH, Pezzotti M, Seurinck J, Mariani C (1992) Developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins. The Plant Cell 4 (9):1041-1051. doi:10.1105/tpc.4.9.1041
Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annual Review of Biochemistry 74:481-514. doi:10.1146/annurev.biochem.74.010904.153721
Greene GL, Gilna P, Waterfield, M. B, A., , Hort YaS, J., (1986) Sequence and expression of human estrogen receptor complementary DNA. Science 231 (4742):1150-1154.
Guillemette B, Bataille AR, Gevry N, Adam M, Blanchette M, Robert F, Gaudreau L (2005) Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biology 3 (12):e384. doi:10.1371/journal.pbio.0030384
Guo H, Zhu P, Guo F, Li X, Wu X, Fan X, Wen L, Tang F (2015) Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. Nature Protocols 10:645. doi:10.1038/nprot.2015.039
Halbeisen RE, Galgano A, Scherrer T, Gerber AP (2007) Post-transcriptional gene regulation: From genome-wide studies to principles. Cellular and Molecular Life Sciences 65 (5):798. doi:10.1007/s00018-007-7447-6
He X, Ling X, Sinha S (2009) Alignment and prediction of cis-regulatory modules based on a probabilistic model of evolution. PLoS Computational Biology 5 (3):e1000299. doi:10.1371/journal.pcbi.1000299
He X-J, Chen T, Zhu J-K (2011) Regulation and function of DNA methylation in plants and animals. Cell Research 21 (3):442
Heins JN, Suriano JR, Taniuchi H, and Anfinsen CB (1967) Characterization of a nuclease produced by Staphylococcus aureus. Journal of Biological Chemistry 242 (5):1016-1020
Heintzman ND, Ren B (2009) Finding distal regulatory elements in the human genome. Current opinion in genetics & development 19 (6):541-549
Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. Nature 447 (7143):418-424. doi:10.1038/nature05917
Heslop-Harrison JSP, Schwarzacher T (2013) Nucleosomes and centromeric DNA packaging. Proceedings of the National Academy of Sciences of the United States of America 110 (50):19974-19975. doi:10.1073/pnas.1319945110
Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nature Methods 6 (4):283
Heyndrickx KS, Van de Velde J, Wang C, Weigel D, Vandepoele K (2014) A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. The Plant cell 26 (10):3894-3910. doi:10.1105/tpc.114.130591

Hofmeister BT, Lee K, Rohr NA, Hall DW, Schmitz RJ (2017) Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation. Genome Biology 18 (1):155. doi:10.1186/s13059-017-1288-x

Hsiung CCS, Morrissey CS, Udugama M, Frank CL, Keller CA, Baek S, Giardine B, Crawford GE, Sung M-H, Hardison RC, Blobel GA (2015) Genome accessibility is widely preserved and locally modulated during mitosis. Genome Research 25 (2):213-225. doi:10.1101/gr.180646.114

Huang F, Zago MK, Abas L, van Marion A, Galván-Ampudia CS, Offringa R (2010) Phosphorylation of conserved PIN motifs directs Arabidopsis PIN1 polarity and auxin transport. The Plant Cell 22 (4):1129-1142

Hwang Y-C, Zheng Q, Gregory BD, Wang L-S (2013) High-throughput identification of long-range regulatory elements and their target promoters in the human genome. Nucleic Acids Research 41 (9):4835-4846. doi:10.1093/nar/gkt188

Irish V (2017) The ABC model of floral development. Current Biology 27 (17):R887-R890. doi:https://doi.org/10.1016/j.cub.2017.03.045

Ishige F, Takaichi M, Foster R, Chua NHaO, K., (1999).AG-box motif (GCCACGTGCC) tetramer confers high-level constitutive expression in dicot and monocot plants. . The Plant Journal 18 (4):443-448

Iyer VR (2012) Nucleosome positioning: bringing order to the eukaryotic genome. Trends in Cell Biology 22 (5):250-256. doi:10.1016/j.tcb.2012.02.004

Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature genetics 33 (3s):245

Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Develomental Biology 339 (2):225-229. doi:10.1016/j.ydbio.2009.08.009

Kadauke S, Blobel GA (2009) Chromatin loops in gene regulation. Biochimica et Biophysica Acta 1789 (1):17-25. doi:10.1016/j.bbagrm.2008.07.002

Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, Pombo A, Fisher AG, Young RA, Jenner RG (2010) Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Molecular Cell 38 (5):675-688. doi:https://doi.org/10.1016/j.molcel.2010.03.019

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) Arabidopsis cytosine methyltransferase Mutants. Genetics 163 (3):1109-1122

Karnik R, Beer MA (2015) Identification of predictive cis-Regulatory elements using a discriminative objective function and a dynamic search space. Plos One 10 (10):e0140557. doi:10.1371/journal.pone.0140557

Kasai M, Kanazawa A (2013) Induction of RNA-directed DNA methylation and heritable transcriptional gene silencing as a tool to engineer novel traits in plants. Plant Biotechnology 30 (3):233-241. doi:10.5511/plantbiotechnology.13.0319a

Kellum R, Schedl P (1991) A position-effect assay for boundaries of higher order chromosomal domains. Cell 64 (5):941-950

Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Research 22(12):2497-2506. doi:10.1101/gr.143008.112

Kim A, Dean A (2012) Chromatin loop formation in the β-globin locus and its role in globin gene transcription. Molecules and Cells 34 (1):1-5. doi:10.1007/s10059-012-0048-8

Kim S-G, Kim S-Y, Park C-M (2007) A membrane-associated NAC transcription factor regulates salt-responsive flowering via *FLOWERING LOCUS T* in Arabidopsis. Planta 226 (3):647-654

Kim SG, Lee AK, Yoon HK, Park CM (2008) A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt signaling in Arabidopsis seed germination. The Plant Journal 55 (1):77-88

Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. American Journal of Human Genetics 76 (1):8-32. doi:10.1086/426833

Klemm SL, Shipony Z, Greenleaf WJ (2019) Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics 20 (4):207-220. doi:10.1038/s41576-018-0089-8

Köhler C, Makarevich G (2006) Epigenetic mechanisms governing seed development in plants. EMBO Reports 7 (12):1223-1227

Koo D-H, Han F, Birchler JA, Jiang J (2011) Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. Genome Research 21 (6):908-914. doi:10.1101/gr.116202.110

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in Arabidopsis thaliana. Annual Review of Plant Biology 55:141-172

Kouzarides T (2007) Chromatin modifications and their function. Cell 128 (4):693-705. doi:10.1016/j.cell.2007.02.005

Kouzarides T, Berger SL (2007) Chromatin modifications and their mechanism of action. Epigenetics:191-209

Krishna SS, Majumdar I, Grishin NV (2003) Structural classification of zinc fingers: survey and summary. Nucleic Acids Research 31 (2):532-550. doi:10.1093/nar/gkg161

Krivega I, Dean A (2012) Enhancer and promoter interactions—long distance calls. Current Opinion in Genetics & Development 22 (2):79-85

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27 (11):1571-1572. doi:10.1093/bioinformatics/btr167

Kujirai T, Ehara H, Fujino Y, Shirouzu M, Sekine S-i, Kurumizaka H (2018) Structural basis of the nucleosome transition during RNA polymerase II passage. Science 362 (6414):595-598. doi:10.1126/science.aau9904

Kulaeva OI, Hsieh FK, Chang HW, Luse DS, Studitsky VM (2013) Mechanism of transcription through a nucleosome by RNA polymerase II. Biochimica et Biophysica Acta 1829 (1):76-83. doi:10.1016/j.bbagrm.2012.08.015

Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes & Development 12 (1):34-44
Lahmy S, Bies-Etheve N, Lagrange T (2010) Plant-specific multisubunit RNA polymerase in gene silencing. Epigenetics 5 (1):4-8. doi:10.4161/epi.5.1.10435
Lämke J, Bäurle I (2017) Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biology 18 (1):124. doi:10.1186/s13059-017-1263-6
Lamport DT, Northcote D (1960) Hydroxyproline in primary cell walls of higher plants. Nature 188 (4751):665
Lander ES (2001) Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. Nature 409:860-921
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9 (4):357-359. doi:10.1038/nmeth.1923
Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AM, Strahl BD, Patel DJ, Jacobsen SE (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498 (7454):385-389. doi:10.1038/nature12178
Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics 11 (3):204-220. doi:10.1038/nrg2719
Lay FD, Kelly TK, Jones PA (2018) Nucleosome occupancy and methylome sequencing (NOMe-seq). Methods in Molecular Biology 1708:267-284. doi:10.1007/978-1-4939-7481-8_14
Lee B-K, Iyer VR (2012) Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. The Journal of Biological Chemistry 287 (37):30906-30913. doi:10.1074/jbc.R111.324962
Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. Nature Genetics 36 (8):900-905. doi:10.1038/ng1400
Lee SH, Cho H-T (2006) PINOID positively regulates auxin efflux in Arabidopsis root hair cells and tobacco cells. The Plant Cell 18 (7):1604-1616. doi:10.1105/tpc.105.035972
Lee TI, Young RA (2000) transcription of protein-coding genes. Annual Review of Genetics 34 (1):77-137. doi:10.1146/annurev.genet.34.1.77
Levine M (2010) Transcriptional enhancers in animal development and evolution. Current Biology 20 (17):R754-763. doi:10.1016/j.cub.2010.06.070
Levine M (2011) Paused RNA polymerase II as a developmental checkpoint. Cell 145 (4):502-511. doi:10.1016/j.cell.2011.04.021
Li E, Liu H, Huang L, Zhang X, Dong X, Song W, Zhao H, Lai J (2019) Long-range interactions between proximal and distal regulatory regions in maize. Nature Communications 10 (1):2633. doi:10.1038/s41467-019-10603-4
Li Q, Gent JI, Zynda G, Song J, Makarevitch I, Hirsch CD, Hirsch CN, Dawe RK, Madzima TF, McGinnis KM, Lisch D, Schmitz RJ, Vaughn MW, Springer NM (2015) RNA-directed DNA methylation enforces boundaries between

heterochromatin and euchromatin in the maize genome. PNAS 112 (47):14728-14733. doi:10.1073/pnas.1514680112
Li X, Wang X, He K, Ma Y, Su N, He H, Stolc V, Tongprasit W, Jin W, Jiang J (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
Li Y, Hu M, Shen Y (2018) Gene regulation in the 3D genome. Human Molecular Genetics 27 (R2):R228-r233. doi:10.1093/hmg/ddy164
Li Y, Tollefsbol TO (2011) DNA methylation detection: bisulfite genomic sequencing analysis. Methods in Molecular Biology (Clifton, NJ) 791:11-21. doi:10.1007/978-1-61779-316-5_2
Liu J, Bai G, Zhang C, Chen W, Zhou J, Zhang S, Chen Q, Deng X, He XJ, Zhu JK (2011) An atypical component of RNA-directed DNA methylation machinery has both DNA methylation-dependent and -independent roles in locus-specific transcriptional gene silencing. Cell Res 21 (12):1691-1700. doi:10.1038/cr.2011.173
Liu ZW, Shao CR, Zhang CJ, Zhou JX, Zhang SW, Li L, Chen S, Huang HW, Cai T, He XJ (2014) The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genetics 10 (1):e1003948. doi:10.1371/journal.pgen.1003948
Long HK, Prescott SL, Wysocka J (2016) Ever-changing landscapes: Transcriptional enhancers in development and evolution. Cell 167 (5):1170-1187. doi:10.1016/j.cell.2016.09.018
Lorch Y, LaPointe JW, Kornberg RD (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell 49 (2):203-210
Lu C, Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C. and Green, P.J. (2005) Elucidation of the small RNA component of the transcriptome. Science 309 (5740):1567-1569
Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, Zhang Y (2016a) Establishing chromatin regulatory landscape during mouse preimplantation development. Cell 165 (6):1375-1388. doi:10.1016/j.cell.2016.05.050
Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ (2016b) Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. Nucleic Acids Research 45 (6):e41-e41
Luger K, Mäder AW, Richmond RK, Sargent DF, and Richmond TJ, 997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature, 389(6648), p.251. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. . Nature 389 (6648):251
Maehara K, Ohkawa Y (2016) Exploration of nucleosome positioning patterns in transcription factor function. Scientific Reports 6:19620-19620. doi:10.1038/srep19620
Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, Zumstein K, Woodhouse M, Bubb K, Dorrity MW, Queitsch C, Bailey-Serres J, Sinha N, Brady SM, Deal RB (2018) Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory

principles and new control modules. Plant Cell 30 (1):15-36. doi:10.1105/tpc.17.00581
Marand AP, Zhang T, Zhu B, Jiang J (2017) Towards genome-wide prediction and characterization of enhancers in plants. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1860 (1):131-139. doi:https://doi.org/10.1016/j.bbagrm.2016.06.006
Martin JL, McMillan FM (2002) SAM (dependent) IAM: theS-adenosylmethionine-dependentmethyl transferase fold. Current Opinion in Structural Biology 12:783–793
Maston GA, Evans SK, Green MR (2006) Transcriptional regulatory elements in the human genome. Annual Review in Genomics and Human Genetics 7:29-59. doi:10.1146/annurev.genom.7.080505.115623
Matzke MA, Kanno T, Matzke AJ (2015) RNA-Directed DNA methylation: The evolution of a complex epigenetic pathway in flowering plants. Annual Review of Plant Biology 66:243-267. doi:10.1146/annurev-arplant-043014-114633
Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nature Review Genetics 15 (6):394-408. doi:10.1038/nrg3683
McBryant SJ, Adams VH, Hansen JC (2006) Chromatin architectural proteins. Chromosome Research 14 (1):39-51
Meng D, Dubin M, Zhang P, Osborne EJ, Stegle O, Clark RM, Nordborg M (2016) limited contribution of DNA methylation variation to expression regulation in *Arabidopsis thaliana*. PLoS Genetics 12 (7):e1006141. doi:10.1371/journal.pgen.1006141
Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nature Reviews Genetics 10 (3):155
Meyer CA, Liu XS (2014) Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nature Reviews Genetics 15 (11):709
Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ, Kingston RE, Tolstorukov MY (2016) MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nature Communications 7:11485. doi:10.1038/ncomms11485
Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245 (4916):371-378
Moore LD, Le T, Fan G (2013) DNA methylation and its basic function. Neuropsychopharmacology 38 (1):23
Narlikar L, Ovcharenko I (2009a) Identifying regulatory elements in eukaryotic genomes. Briefings in Functional Genomics and Proteomics 8 (4):215-230. doi:10.1093/bfgp/elp014
Narlikar L, Ovcharenko I (2009b) Identifying regulatory elements in eukaryotic genomes. Briefings in Functional Genomics and Proteomics 8 (4):215-230. doi:10.1093/bfgp/elp014
Nielsen PR, Nietlispach D, Mott HR, Callaghan J, Bannister A, Kouzarides T, Murzin AG, Murzina NV, Laue ED (2002) Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature 416 (6876):103

Nikolov DB, Burley SK (1997) RNA polymerase II transcription initiation: a structural view. Proceedings of the National Academy of Sciences of the United States of America 94 (1):15-22. doi:10.1073/pnas.94.1.15

Nordström KJV, Schmidt F, Gasparoni N, Salhab A, Gasparoni G, Kattler K, Müller F, Ebert P, Costa IG, consortium D, Pfeifer N, Lengauer T, Schulz MH, Walter J (2019) Unique and assay specific features of NOMe-, ATAC- and DNase I-seq data. Nucleic Acids Research 47 (20):10580-10596. doi:10.1093/nar/gkz799

Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S (2010) Genome-wide analysis of NAC transcription factor family in rice. Gene 465 (1-2):30-44. doi:10.1016/j.gene.2010.06.008

O'Shea-Greenfield A, Smale ST (1992) Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. Journal of Biological Chemistry 267 (2):1391-1402

Ogbourne S, Antalis TM (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochemical Journal 331 (1):1-14. doi:10.1042/bj3310001

Ohbayashi I, Lin CY, Shinohara N, Matsumura Y, Machida Y, Horiguchi G, Tsukaya H, Sugiyama M (2017) Evidence for a role of ANAC082 as a ribosomal stress response mediator leading to growth defects and developmental alterations in Arabidopsis. Plant Cell 29 (10):2644-2660. doi:10.1105/tpc.17.00255

Ohnishi Y, Totoki, Y, Toyoda, A, Watanabe, T, Yamamoto, Y, Tokunaga, K, Sakaki, Y, Sasaki, H, & Hohjoh, H (2012) Active role of small non-coding RNAs derived from SINE/B1 retrotransposon during early mouse development. Molecular Biology Reports 39 (2):903-909. doi:10.1007/s11033-011-0815-1)

Orphanides G, Lagrange T, Reinberg D (1996) The general transcription factors of RNA polymerase II. Genes & Development 10 (21):2657-2683

Pajoro A, Madrigal P, Muiño JM, Matus JT, Jin J, Mecchia MA, Debernardi JM, Palatnik JF, Balazadeh S, Arif M (2014) Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biology 15 (3):R41

Pajoro A., Muiño J.M., Angenent G.C., K. K (2018) Profiling Nucleosome Occupancy by MNase-seq: Experimental protocol and computational analysis. In: Bemer M., Baroux C. (eds) Plant Chromatin Dynamics. Methods in Molecular Biology, vol 1675. Humana Press, New York, NY. Plant Chromatin Dynamics Methods in Molecular Biology 1675

Pan M-R, Hsu M-C, Chen L-T, Hung W-C (2018) Orchestration of H3K27 methylation: mechanisms and therapeutic implication. Cell and Molecular Life Science 75 (2):209-223. doi:10.1007/s00018-017-2596-8

Pascoe CD, Obeidat Me, Arsenault BA, Nie Y, Warner S, Stefanowicz D, Wadsworth SJ, Hirota JA, Yang SJ, Dorscheid DR (2017) Gene expression analysis in asthma using a targeted multiplex array. BMC Pulmonary Medicine 17 (1):189

Paszkowski J, Grossniklaus U (2011) Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Current Opinion in Plant Biology 14 (2):195-203. doi:10.1016/j.pbi.2011.01.002

Peng Y, Xiong D, Zhao L, Ouyang W, Wang S, Sun J, Zhang Q, Guan P, Xie L, Li W, Li G, Yan J, Li X (2019) Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. Nature Communications 10 (1):2632. doi:10.1038/s41467-019-10602-5

Perrone L, Matrone C, P. Sing L (2014) Erratum to "Epigenetic modifications and potential new treatment targets in diabetic retinopathy", vol 2014. doi:10.1155/2014/789120

Pfluger J, Wagner D (2007) Histone modifications and dynamic regulation of genome accessibility in plants. Current Opinion in Plant Biology 10 (6):645-652. doi:10.1016/j.pbi.2007.07.013

Pikaard CS, Mittelsten Scheid O (2014) Epigenetic regulation in plants. Cold Spring Harb Perspect Biol 6 (12):a019315. doi:10.1101/cshperspect.a019315

Pott S (2017) Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. Elife 6. doi:10.7554/eLife.23203

Priest HD, Filichkin SA, Mockler TC (2009) Cis-regulatory elements in plant cell signaling. Current Opinion in Plant Biology 12 (5):643-649. doi:10.1016/j.pbi.2009.07.016

Qu L-J, Zhu Y-X (2006) Transcription factor families in Arabidopsis: major progress and outstanding issues for future research. Current Opinion in Plant Biology 9 (5):544-549. doi:https://doi.org/10.1016/j.pbi.2006.07.005

Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447 (7143):425

Rhie SK, Schreiner S, Farnham PJ (2018) Defining regulatory elements in the human genome using nucleosome occupancy and methylome sequencing (NOMe-Seq). In: CpG Islands. Methods in Molecular Biology. pp 209-229. doi:10.1007/978-1-4939-7768-0_12

Rice JC, Allis CD (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation. Current Opinion in Cell Biology 13 (3):263-273. doi:https://doi.org/10.1016/S0955-0674(00)00208-8

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C-Z, Keddie J, Adam L, Pineda O, Ratcliffe O, Samaha R (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290 (5499):2105-2110

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough H, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by non-coding RNAs. Cell 129 (7):1311-1323. doi:10.1016/j.cell.2007.05.022

Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP (2017) Variant Review with the Integrative Genomics Viewer. Cancer Research 77 (21):e31-e34. doi:10.1158/0008-5472.CAN-17-0337

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nature Biotechnology 29 (1):24-26. doi:10.1038/nbt.1754

Ronemus MJ, Galbiati, M., Ticknor, C., Chen, J. and Dellaporta, S.L. (1996) Demethylation-induced developmental pleiotropy in Arabidopsis. Science 273 (5275):654

Rossetto D, Avvakumov N, Côté J (2012) Histone phosphorylation: A chromatin modification involved in diverse nuclear events. Epigenetics 7 (10):1098-1108. doi:10.4161/epi.21975

Rowley MJ, Rothi MH, Bohmdorfer G, Kucinski J, Wierzbicki AT (2017) Long-range control of gene expression via RNA-directed DNA methylation. PLoS Genetics 13 (5):e1006749. doi:10.1371/journal.pgen.1006749

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proceedings of the National Academy of Sciences 104 (27):11376-11381

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nature Genetics 37 (5):501-506. doi:10.1038/ng1543

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nature Protocols 3 (6):1101-1108

Schones DE, Cui K, Cuddapah S, Roh T-Y, Barski A, Wang Z, Wei G, Zhao K (2008a) Dynamic regulation of nucleosome positioning in the human genome. Cell 132 (5):887-898

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K (2008b) Dynamic regulation of nucleosome positioning in the human genome. Cell 132 (5):887-898. doi:10.1016/j.cell.2008.02.022

Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell 18 (5):1121-1133. doi:10.1105/tpc.105.039834

Schwartzman O, Tanay A (2015) Single-cell epigenomics: techniques and emerging applications. Nature Reviews Genetics 16:716. doi:10.1038/nrg3980

Seherm H, Coakley SM (2003) Plant pathogens in a changing world. Australasian Plant Pathology 32 (2):157-165

Seymour DK, Becker C (2017) The causes and consequences of DNA methylome variation in plants. Current Opinion in Plant Biology 36:56-63. doi:10.1016/j.pbi.2017.01.005

Shahmuradov IA (2003) PlantProm: a database of plant promoter sequences. Nucleic Acids Research 31 (1):114-117. doi:10.1093/nar/gkg041

Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. Nature Reviews Genetics 15 (4):272-286

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics 8:272. doi:10.1038/nrg2072

Smale ST, Kadonaga JT (2003) The RNA Polymerase II core promoter. Annual Review of Biochemistry 72 (1):449-479. doi:10.1146/annurev.biochem.72.121801.161520

Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nature Methods 11:817. doi:10.1038/nmeth.3035

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480 (7378):490

Stam M, Belele, C., Ramakrishna, W., Dorweiler, J.E., Bennetzen, J.L. and Chandler, V.L. (2002) The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162 (2):917-930

Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403 (6765):41-45. doi:10.1038/47412

Su W, Jackson S, Tjian R, Echols H (1991) DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes & Development 5 (5):820-826

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nature Reviews Genetics 9. doi:10.1038/nrg2341

Szerlong HJ, Hansen JC (2010) Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. Biochemistry and Cell Biology 89 (1):24-34

Taberlay PC, Statham AL, Kelly TK, Clark SJ, Jones PA (2014) Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Research 24 (9):1421-1432

Taher L, Smith RP, Kim MJ, Ahituv N, Ovcharenko I (2013) Sequence signatures extracted from proximal promoters can be used to predict distal enhancers. Genome Biology 14 (10):R117. doi:10.1186/gb-2013-14-10-r117

Tannenbaum M, Sarusi-Portuguez A, Krispil R, Schwartz M, Loza O, Benichou JIC, Mosquna A, Hakim O (2018) Regulatory chromatin landscape in Arabidopsis thaliana roots uncovered by coupling INTACT and ATAC-seq. Plant Methods 14 (1):113. doi:10.1186/s13007-018-0381-9

Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, Voinnet O, Wincker P, Esteller M, Colot V (2009) A role for RNAi in the selective correction of DNA methylation defects. Science 323 (5921):1600-1604. doi:10.1126/science.1165313

Thiebaut F, Hemerly AS, Ferreira PCG (2019) A Role for epigenetic regulation in the adaptation and stress responses of non-model plants. Frontiers in Plant Science 10:246-246. doi:10.3389/fpls.2019.00246

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA (2012) The accessible chromatin landscape of the human genome. Nature 489 (7414):75-82. doi:10.1038/nature11232

Tiwari SB, Shen Y, Chang H-C, Hou Y, Harris A, Ma SF, McPartland M, Hymus GJ, Adam L, Marion C, Belachew A, Repetti PP, Reuber TL, Ratcliffe OJ (2010) The flowering time regulator CONSTANS is recruited to the *FLOWERING LOCUS T* promoter via a unique cis-element. New Phytologist 187 (1):57-66. doi:10.1111/j.1469-8137.2010.03251.x

Tsompana M, and Buck MJ, . (2104) Chromatin accessibility: a window into the genome. Epigenetics Chromatin 7 (33)

United Nations DoEaSA, Population Division (2019) World Population Prospects 2019: Highlights.

United Nations F (2017) IPPC (International Plant Protection Convention) annual report, Food and Agriculture Organisation of the United Nations

Vaissiere T, Sawan C, Herceg Z (2008) Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutatation Research 659 (1-2):40-48. doi:10.1016/j.mrrev.2008.02.004

Vaissière T, Sawan C, Herceg Z (2008) Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation Research/Reviews in Mutation Research 659 (1):40-48. doi:https://doi.org/10.1016/j.mrrev.2008.02.004

Van Holde KE, Sahasrabuddhe CG, Shaw BR (1974) A model for particulate structure in chromatin. Nucleic Acids Research 1 (11):1579-1586. doi:10.1093/nar/1.11.1579

Vanneste S, Friml J (2009) Auxin: a trigger for change in plant development. Cell 136 (6):1005-1016

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303 (5658):672-676. doi:10.1126/science.1093686

Voss TC, Hager GL (2008) Visualizing chromatin dynamics in intact cells. Biochimica et biophysica acta 1783 (11):2044-2051. doi:10.1016/j.bbamcr.2008.06.022

Wang J, Gao S, Peng X, Wu K, Yang S (2019) Roles of the INO80 and SWR1 Chromatin Remodeling Complexes in Plants. International Journal of Molecular Sciences 20 (18):4591. doi:10.3390/ijms20184591

Wang Y, Jensen RC, Stumph WE (1996) Role of TATA box sequence and orientation in determining RNA polymerase II/III transcription specificity. Nucleic acids Research 24 (15):3100-3106. doi:10.1093/nar/24.15.3100

Wassenegger M, Heimes S, Riedel L, Sanger HL (1994) RNA-directed *de novo* methylation of genomic sequences in plants. Cell 76 (3):567-576

Weber M, Schübeler D (2007) Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Current Opinion in Cell Biology 19 (3):273-280. doi:https://doi.org/10.1016/j.ceb.2007.04.011

Wei Y, Mizzen CA, Cook RG, Gorovsky MA, Allis CD (1998) Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proceedings of the National Academy of Sciences 95 (13):7480-7484

Wibowo A, Becker C, Marconi G, Durr J, Price J, Hagmann J, Papareddy R, Putra H, Kageyama J, Becker J, Weigel D, Gutierrez-Marcos J (2016) Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile

sites in the genome and is restricted in the male germline by DNA glycosylase activity. Elife 5. doi:10.7554/eLife.13546
Wittkopp PJ, Kalay G (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics 13 (1):59-69. doi:10.1038/nrg3095
Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. Molecular Biology and Evolution 20 (9):1377-1419. doi:10.1093/molbev/msg140
Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, Li E, Zhang Y, Sun YE (2010) Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science 329 (5990):444-448
Xi H, Yu Y, Fu Y, Foley J, Halees A, Weng Z (2007) Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. Genome Research 17 (6):798-806
Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, Gascard P, Sigaroudinia M, Tlsty TD, Kadlecek T, Weiss A, O'Geen H, Farnham PJ, Madden PAF, Mungall AJ, Tam A, Kamoh B, Cho S, Moore R, Hirst M, Marra MA, Costello JF, Wang T (2013) DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nature Genetics 45 (7):836-841. doi:10.1038/ng.2649
Xu M, Kladde MP, Simpson RT, Van Etten JL (1998) Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. Nucleic Acids Research 26 (17):3961-3966. doi:10.1093/nar/26.17.3961
Yan W, Chen D, Schumacher J, Durantini D, Engelhorn J, Chen M, Carles CC, Kaufmann K (2019) Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. Nature Communications 10 (1):1705. doi:10.1038/s41467-019-09513-2
Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Molecular Biology 60 (1):107-124
Yoo SD, Cho YH, Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nature Protocols 2 (7):1565-1572. doi:10.1038/nprot.2007.199
Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ (2005) Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. Science 309 (5734):626-630. doi:10.1126/science.1112178
Zentner GE, Henikoff S (2013) Regulation of nucleosome dynamics by histone modifications. Nature Structural & Molecular Biology 20 (3):259
Zhang H, Zhu JK (2011) RNA-directed DNA methylation. Current Opinion in Plant Biology 14 (2):142-147. doi:10.1016/j.pbi.2011.02.003
Zhang M, Xu C, von Wettstein D, Liu B (2011) Tissue-specific differences in cytosine methylation and their association with differential gene expression in sorghum. Plant Physiology 156 (4):1955-1966. doi:10.1104/pp.111.176842

Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J (2012) High-resolution mapping of open chromatin in the rice genome. Genome Research 22 (1):151-162

Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J, Jacobsen SE (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biology 5 (5):e129

Zhang X, Henriques R, Lin SS, Niu QW, Chua NH (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nature Protocols 1 (2):641-646. doi:10.1038/nprot.2006.97

Zhang Y, Reinberg D (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes & Development 15 (18):2343-2360

Zhu B, Zhang W, Zhang T, Liu B, (2015) Genome-wide prediction and validation of intergenic enhancers in Arabidopsis using open chromatin signatures. The Plant Cell 27 (9):2415–2426. doi: doi:10.1105/tpc.15.00537

Zhu J, Liu M, Liu X, Dong Z (2018) RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis. Nature Plants 4 (12):1112

Zicola J, Liu L, Tanzler P, Turck F (2019a) Targeted DNA methylation represses two enhancers of *FLOWERING LOCUS T* in Arabidopsis thaliana. Nature Plants 5 (3):300-307. doi:10.1038/s41477-019-0375-2

Zicola J, Liu L, Tänzler P, Turck F (2019b) Targeted DNA methylation represses two enhancers of *FLOWERING LOCUS T* in Arabidopsis thaliana. Nature Plants 5 (3):300-307. doi:10.1038/s41477-019-0375-2

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics 39 (1):61-69. doi:10.1038/ng1929

Zuo J, Niu Q, Chua N (2000) An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. The Plant Journal 24:265-273. doi:doi:10.1046/j.1365-313x.2000.00868.x

7. Appendix

Appendix for chapter 2

Table 7.1 DNA and protein sequence for M.CViP1

DNA Sequence Nuclear localisation signal (NLS) M.CViPI sequence 3xHA tag	*ATGGCCCCCAAAAAGAAACGAAAAGTGATGACATTGAAAGCATTAGAACTTTTT* *GCTGGGATTGCAGGCATTACGCATGGCCTACGAGGCTTCGTAGAGCCCGTCGCG* *TTTGTTGAGATCAATAAGGATGCGCAGGAATTCCTTTCAACCAAATTTCCAGACA* *AGCCAGTCTTCGATGACGTAACTAAATTCAGTAAGCGTGACTTCGACGAGCCTA* *TTGATATGATCACAGGAGGGTTTCCGTGCACAGGTTTCTCAATCGCCGGGAAGA* *GGAACGGCTTTGAGCATGCCGAGAGTGGTCTATTTGGTGAGGTAGTAAGAATT* *ACCAAAGAATATATGCCCAAGATGGTCTTTCTAGAAAACTCTGGCATGCTCTCAC* *ATAAGTACAACCTGGACATCGTGATCAGGAGTATGGACAGCTTGGGTTACGATT* *GTCGATGGGTAACGTTACGTGCAACGGTCGTGGGGGCATTACATACCCGTCATA* *GGTGGTTCTGTCTGTGCACACGAAAAGATCACATACGTGAAACTCTAATCTGCG* *ACAGGGAAGTAACAAAATTCGACTGGGAGAATGACAGGCCTCCAATTCAAGTT* *GACAGCCGATCATACGAGAATAGTAGGTTGGTAAGATTTGCCGGTTACAGTGT* *GGTCCCAGATCAGATCAGATATGCATTCACCGGCCTATATACAGGAAATTTTTCA* *CCGAGTTTCAGTAAAACGCTTGTCCCTGGCAGTTTAGAGGGTAGTATATGCTTCA* *ACGAGGACAAAATTACAAATGGTTACTACAAAGATGGCGTGTATTACGAATTCG* *TCAGGACAGAAACGCACCGAGAACCCGTTAATATTTTGCTCACGCCGAGGGAGA* *TACCCAATAAGCATAACGGTAAGAAACTGCTGACTTTGCCGGTGACCAAGAGGT* *ATTGGTGCACGCCTTGTGCAAGCTATGGTAAAGGTACAGCAGGAGGGAGAGTG* *CTAACAGACCGAAGCAGTCACAGCTTGCCGACGCAGGTTAAATTTTCCCCTGAG* *GGCGAAGACGGTAAACACCTCTCTGGCAAATTCTGCGCTTGGTTAATGGGCTAC* *GATAAGGAGTATCTGGGAAACTTACTGGAGTATTACCCTTACGATGTGCCCGAC* *TACGCATATCCATATGACGTCCCAGATTACGCCTATCCTTATGACGTGCCAGACT* *ATGCCTAG*
Protein sequence Nuclear localisation signal (NLS) M.CViPI sequence 3xHA tag	**MAPKKKRKV**MTLKALELFAGIAGITHGLRGFVEPVAFVEINKDAQEFLSTKFPDKPV FDDVTKFSKRDFDEPIDMITGGFPCTGFSIAGKRNGFEHAESGLFGEVVRITKEYMP KMVFLENSGMLSHKYNLDIVIRSMDSLGYDCRWVTLRATVVGALHTRHRWFCLCT RKDHIRETLICDREVTKFDWENDRPPIQVDSRSYENSRLVRFAGYSVVPDQIRYAFT GLYTGNFSPSFSKTLVPGSLEGSICFNEDKITNGYYKDGVYYEFVRTETHREPVNILLT PREIPNKHNGKKLLTLPVTKRYWCTPCASYGKGTAGGRVLTDRSSHSLPTQVKFSPE GEDGKHLSGKFCAWLMGYDKEYLGNLLEYYPYDVPDYAYPYDVPDYAYPYDVPDY AStop

Table 7.2: Hairpin sequences for hypermethylated lines and RNAi lines

Plant line	Hairpin sequence
NAC82-DRE	CACTACTGGTACCCATTTTTTTTTGAAAATGCAAAACTCAAAAGTCCC TCGCTGGGTTAACGACGACGATTCAATTGCACGAATCCAATAAACTTT GATCCCGAAAGGAACATGATTCAAACATTAAGCAATCTTATCACAAA TTCAAAA
HRGP1-DRE	CAAAACTTTCGCGAGCCCGAACTTTCTTAGCTAAACAAAAAGAA ATCTAATATAAAAATAAGAAGAAGAAACAACCAAGGAACGATA AGAGGTAATTATGCTATAAAAAGCGGTCACTTGTCTTGGTTCTC GAGGATAGAAATCAAGTTTGCACAATCGGAAGCGAAAATACCA AACATCTTGCCGGTAGAGATCATGCATTAATTCCATTGCCCATA ACAGATAATCAAATTCTGCATGGAGCGACAATAAACTCCGACG GCTACTTCT
*HRGP1*_RNAi	CCTCTC**CG**CCAC**CG**TTGTATTCTTCAC**CG**CTTC**CTG**AAGTTGAATACA AGACTCCTCCACTACCATA**CG**T**CG**A**CAG**TTCCCCACCACCAACTTACA CAC**CAG**CTC**CTG**AAGTGGAATACAAGTCTCCACCACCACCATATGTCT ATAGTTCTCCACCCCCACCAACATA CTCCCCATCTCCAAAGGTTG

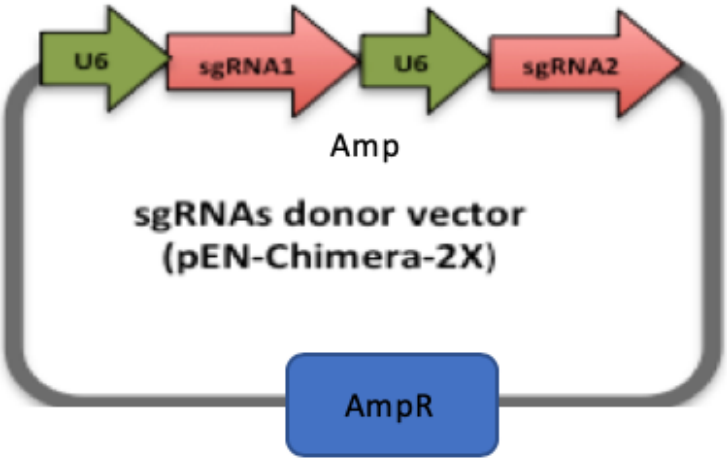

Figure 7.1: pEN-2x-Chimera to clone two sgRNAs in a single vector, each gRNA expressed under a separate U6 promoter.

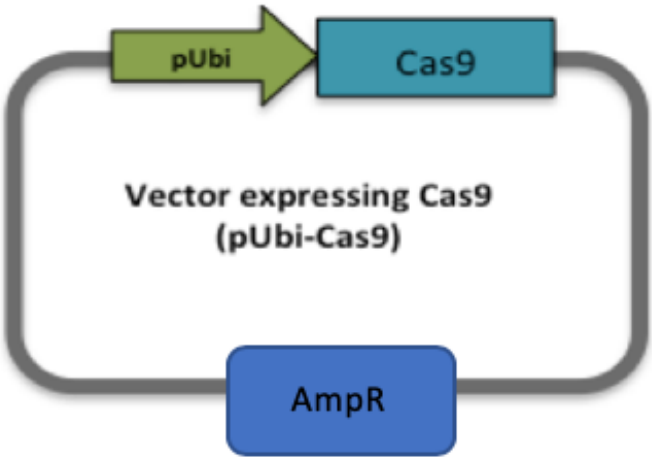

Figure 7.2: pUbi-cas9, A vector expressing Cas9 under Ubiquitin promoter. Used for protoplast transfection to test the efficiency of synthetic sgRNAs.

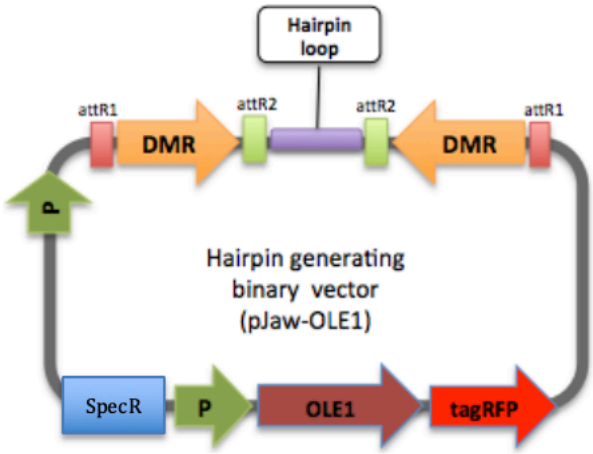

Figure 7.3: pJaw-OLE1, A binary vector, capable of generating an inverted RNAi hairpin

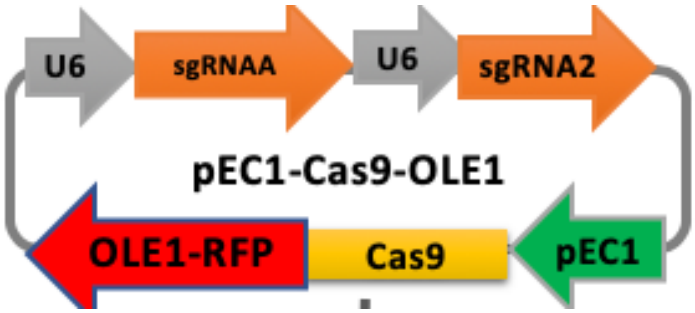

Figure 7.4: pEC1-Cas9, a binary vector expressing Cas9 under an egg cell specific promoter (EC1)

Table 7.3 List of oligos used in study

Oligo name	Sequence (5’-3’)	Description
MCViP1_RT_fw	ACAGACCGAAGCAGTCACAG	M.CViPI expression
MCViPI_RT_rev	TAGCCCATTAACCAAGCGCA	
AT5g09330-qPCR-F	GACGACAAAGTACTGACGCAG	NAC82 Expression
AT5g09330-qPCR-R	AGTCCTCTTCCTTAAACGGAGC	
PP2AA3.FP	TAACGTGGCCAAAATGATGC	qPCR Housekeeping
PP2AA3.RP	GTTCTCCACAACCGCTTGGT	
NH-AP-CR F	AGGTCCACATCATCGGCTAAA	*APOLO* genotyping
NH-AP-CR R	TGCAGATTGTATTGACTGCGT	
NH9-CR F	TATGAGACGACAATTCCCAAAGA	*DRE_HRGP1*
NH9-CR R	TGTAGTGCGCTTGCGACAT	
NH-AT2g34655-A1 F	ATTGGTCCACACGATCTCACATGT	*Sg*RNA-A1 for *APOLO* deletion
NH-AT2g34655-A1 R	AAACACATGTGAGATCGTGTGGAC	
NH-AT2g34655-A2 F	ATTGGATTTACACACACTATTCGT	*Sg*RNA-A2 for *APOLO* deletion
NH-AT2g34655-A2 R	AAACACGAATAGTGTGTGTAAATC	
NH-AT2g34655-B1 F	ATTGACCTAGAGACAGCATATTAA	*Sg*RNA-B1 for *APOLO* deletion
NH-AT2g34655-B1 R	AAACTTAATATGCTGTCTCTAGGT	
NH-AT2g34655-B2 F	ATTGCATAGTGGTAAACCCTATAA	*Sg*RNA-B2 for *APOLO* deletion
NH-AT2g34655-B2 R	AAACTTATAGGGTTTACCACTATG	
NH-At3g54590-A1 F	ATTGCGTAAGGGACAAAAATATTT	*Sg*RNA-A1 for *DRE_HRGP1*
NH-At3g54590-A1 R	AAATAAATATTTTTGTCCCTTACG	
NH-At3g54590-B1 F	ATTGGGCTCTGAGCGGATATAGTT	*S* RNA-B1 for *DRE_HRGP1*
NH-At3g54590-B1 R	AAACAACTATATCCGCTCAGAGCC	
NH-At3g54590-B2 F	ATTGGATCTCTGTTTAATGCGCCT	*Sg*RNA-B2 for *DRE_HRGP1*
NH-At3g54590-B2 R	AAACAGGCGCATTAAACAGAGATC	

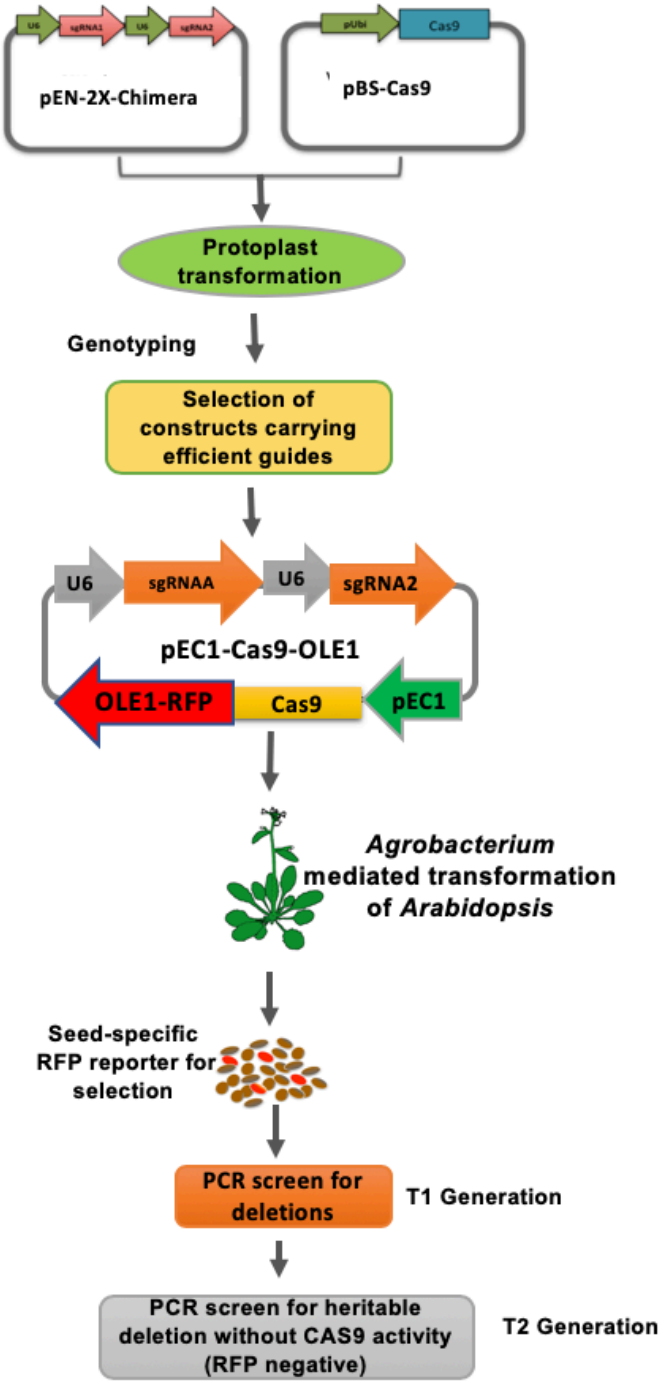

Figure 7.5: Basic strategy of gene editing in *Arabidopsis* via CRISPR/cas9 system. For each region to be deleted, sgRNAs were designed and cloned into a binary vector carrying CRISPR/cas9 system. The constructs were delivered into plants via Agrobacterium transformation after testing them in protoplast. T1 generation is screened for somatic deletions through PCR and the in the second generation the plants lacking Cas9 activity are screened for heritable changes and propagated to generate homozygous lines with stable deletions.

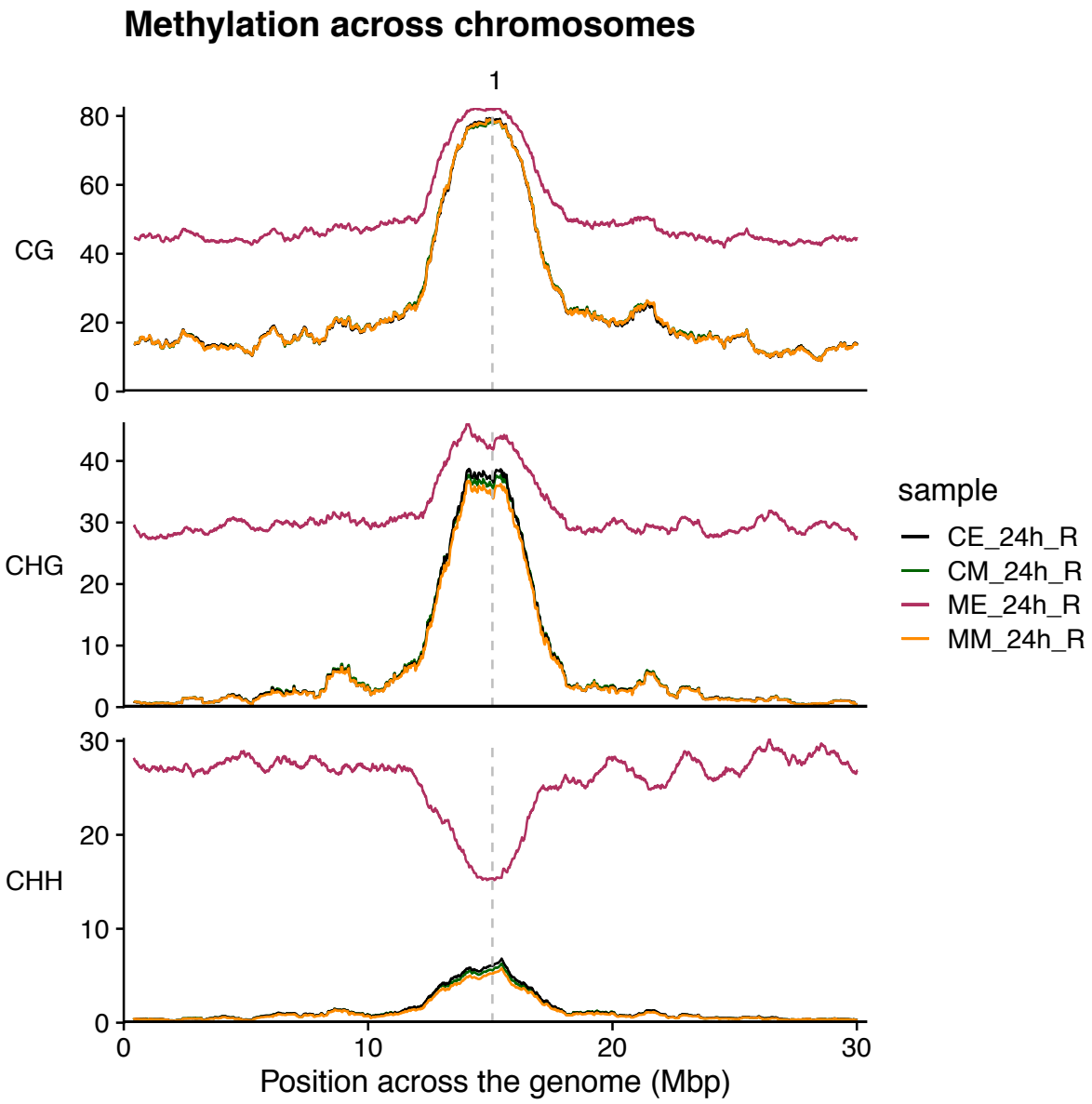

Figure 7.6 Methylation across chromosome 1 in root tissues in GCG, GCHG and GCHH contexts with 24-hour ß-estradiol induction. Black line represents Control plants with ß-estradiol treatments. Green line represents with mock treatment. Red line represents M.CViPI-1a plants with ß-estradiol treatment and orange line represents M.CViPI-1a plants with mock treatment.

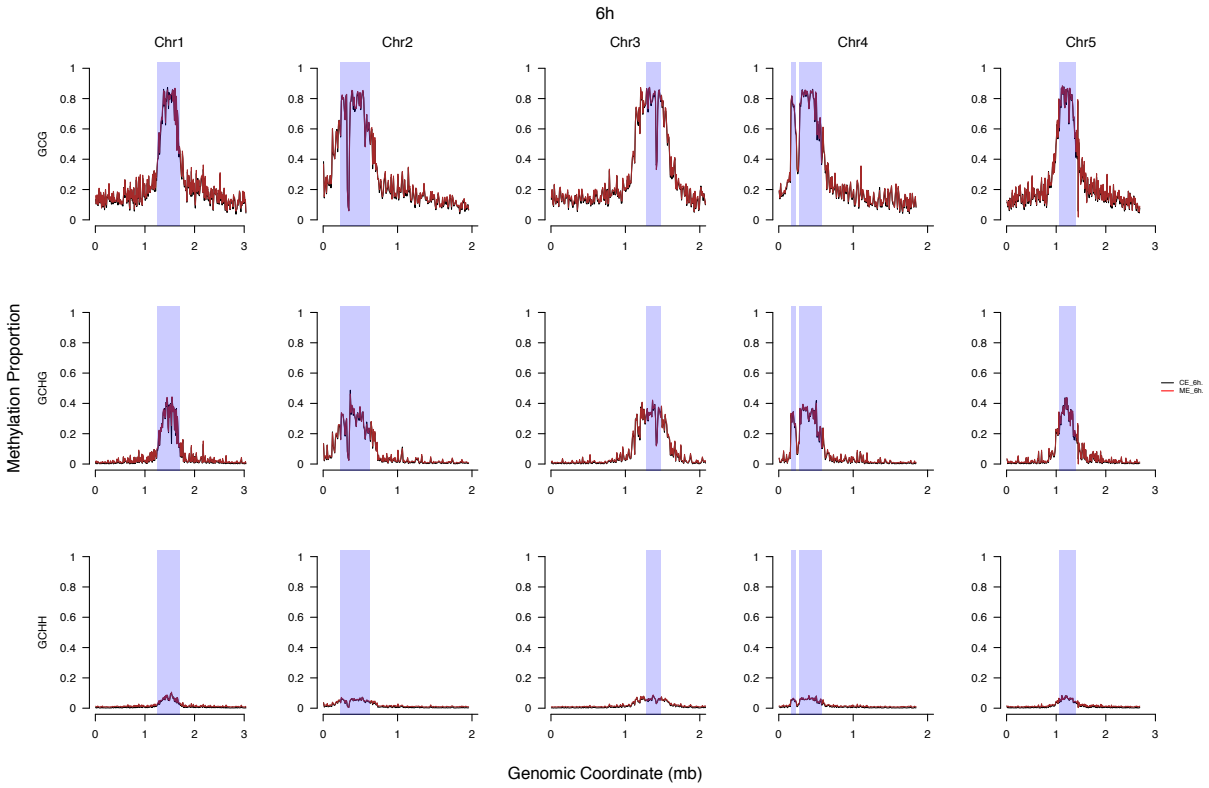

Figure 7.7.: Methylation profile of all chromosome in leaf tissue in GCG, GCHG and GCHH contexts with 6-hour ß-estradiol induction. No detectable differences in cytosine methylation in all contexts in induced plants at 6-hour induction time compared to control plants. The red line indicates methylation profile of CE samples (Control) and black line represents methylation profile of ß-estradiol induced line.

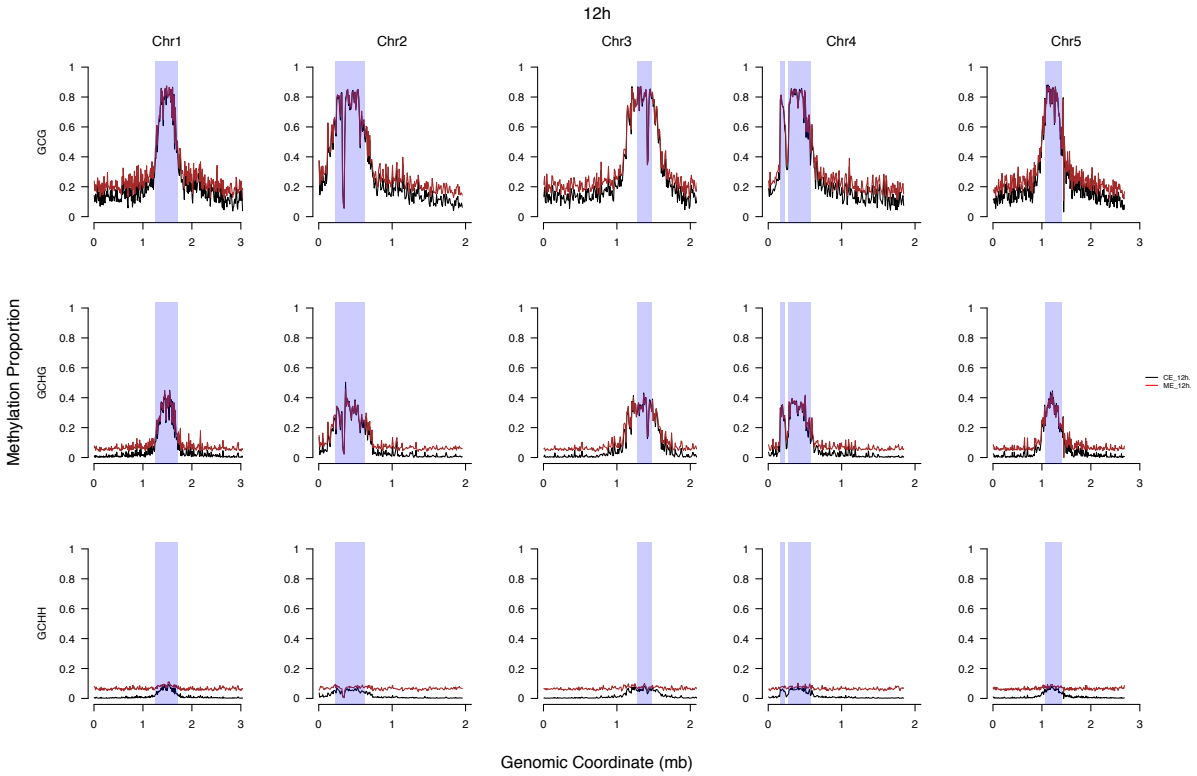

Figure 7.8: Methylation profile of all chromosome in leaf tissue in GCG, GCHG and GCHH contexts with 12-hour ß-estradiol induction. Cytosine methylation increased in all contexts at 12-hour induction time compared to control plants. The red line indicates methylation profile of CE samples (Control) and black line represents methylation profile of ß-estradiol induced line.

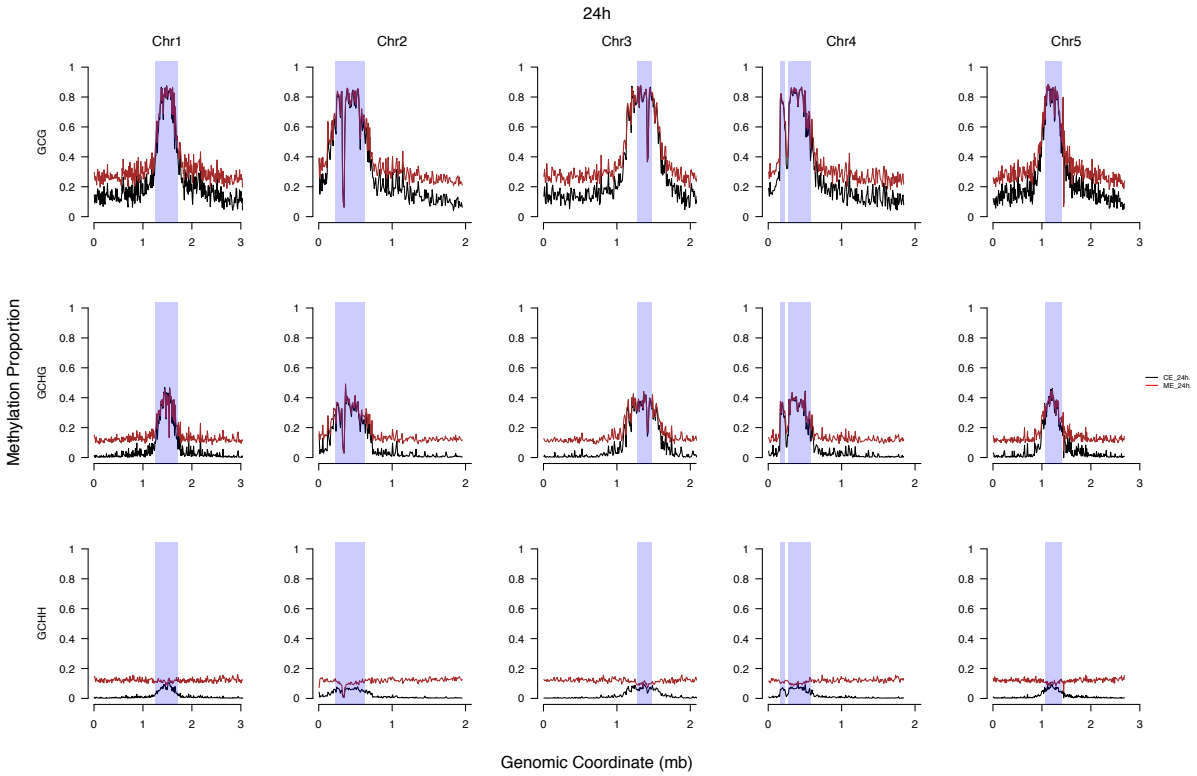

Figure 7.9: Methylation profile of all chromosome in leaf tissue in GCG, GCHG and GCHH contexts with 24-hour ß-estradiol induction. Cytosine methylation increased in all contexts at 24-hour induction time compared to control plants. The red line indicates methylation profile of CE samples (Control) and black line represents methylation profile of ß-estradiol induced line.

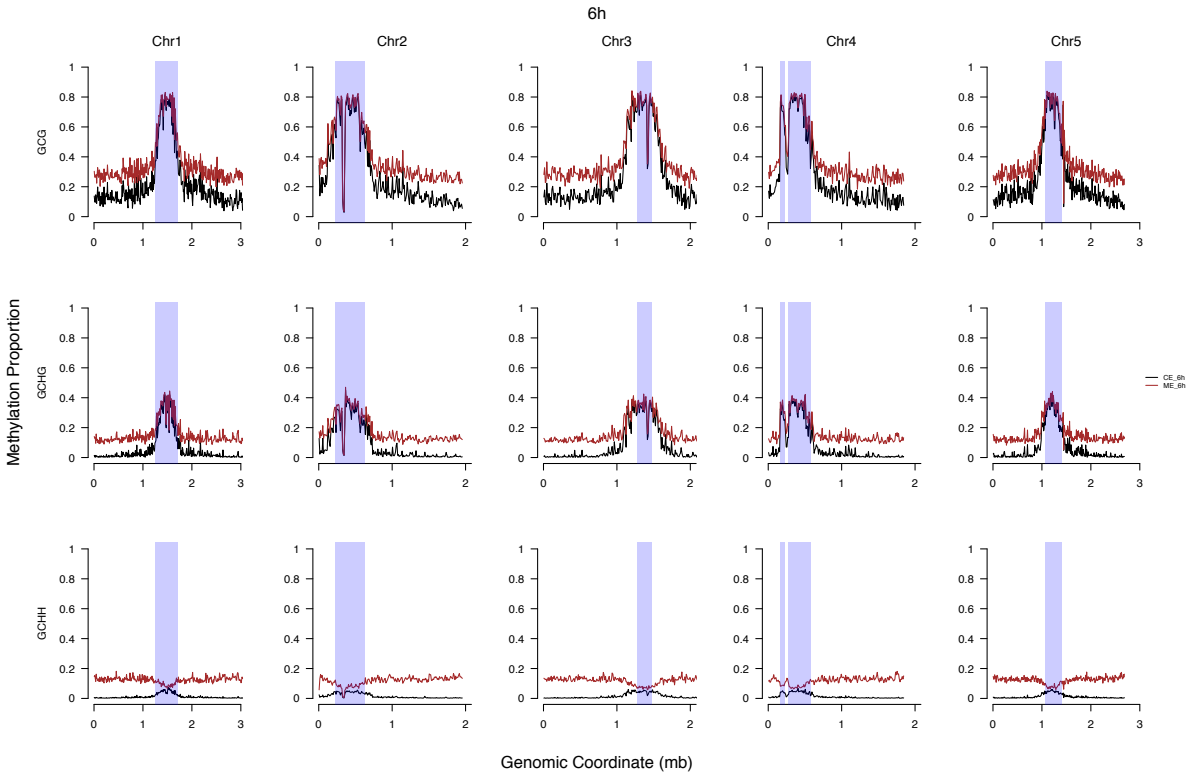

Figure 7.10: Methylation profile of all chromosome in root tissue in GCG, GCHG and GCHH contexts with 6-hour ß-estradiol induction. Cytosine methylation increased in all contexts at 6-hour induction time compared to control plants. The red line indicates methylation profile of CE samples (Control) and black line represents methylation profile of ß-estradiol induced line.

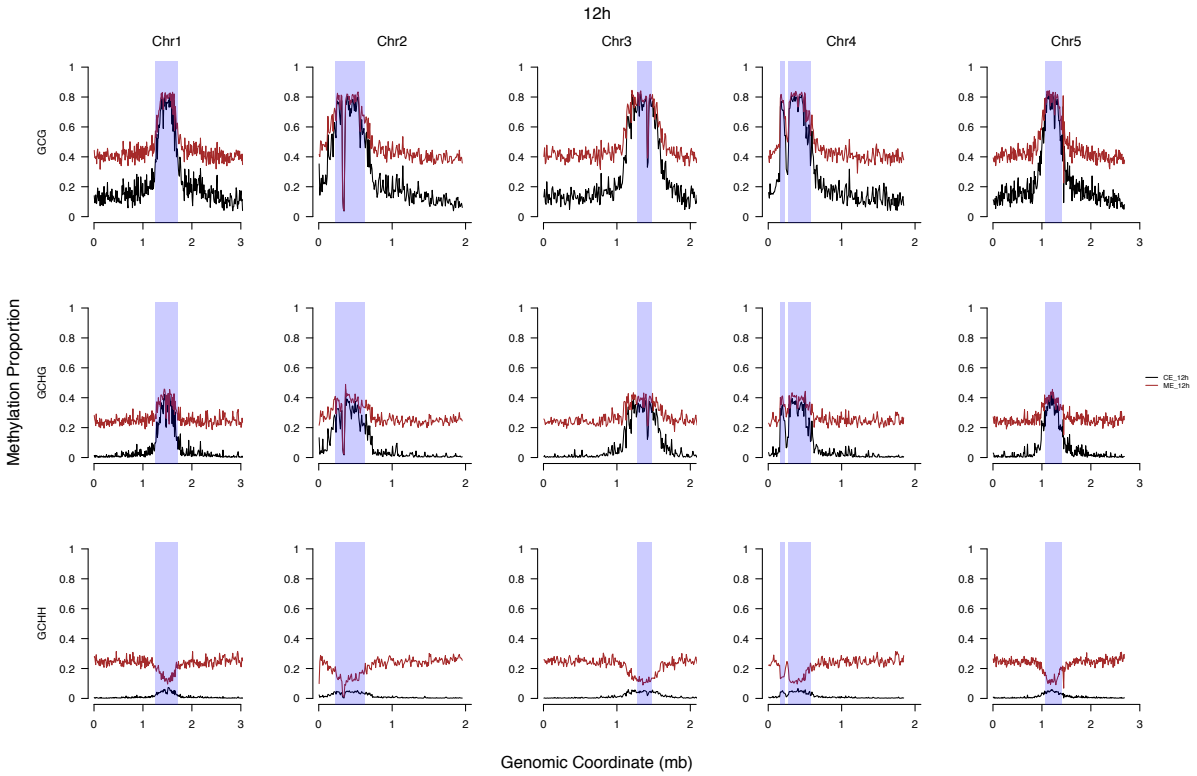

Figure 7.11: Methylation profile of all chromosome in root tissue in GCG, GCHG and GCHH contexts with 12-hour ß-estradiol induction. Cytosine methylation increased in all contexts at 12-hour induction time compared to control plants. The red line indicates methylation profile of CE samples (Control) and black line represents methylation profile of ß-estradiol induced line.

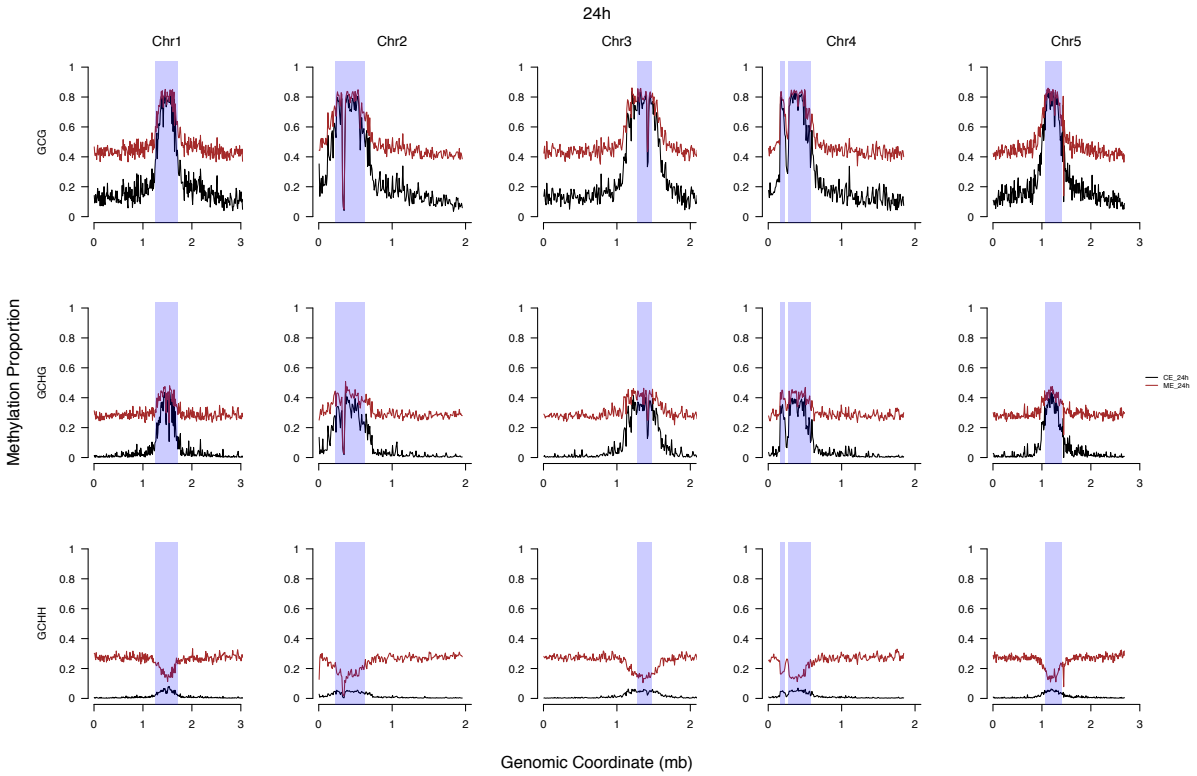

Figure 7.12: Methylation profile of all chromosome in root tissue in GCG, GCHG and GCHH contexts with 24-hour ß-estradiol induction. Cytosine methylation increased in all contexts at 24-hour induction time compared to control plants. The red line indicates methylation profile of CE samples (Control) and black line represents methylation profile of ß-estradiol induced line.

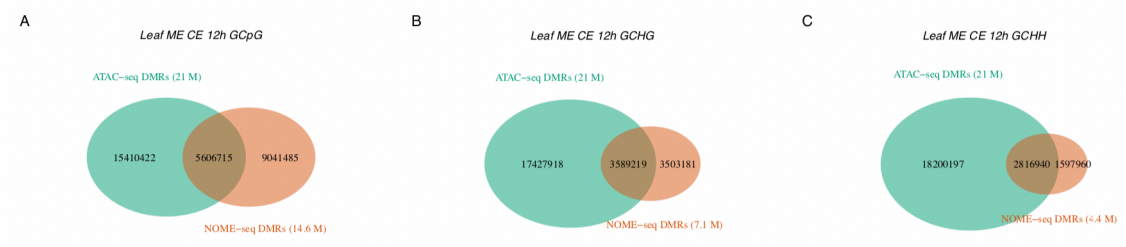

Figure 7.13: Accessible chromatin regions found by iNOMe compared to ATAC in leaf samples with 12-hour ß-estradiol induction. The orange circle indicates the length of accessible chromatin detected by iNOMe only (Mb) whereas the turquoise indicates accessible chromatin identified by ATAC only (Mb). The overlap of both circles shows the length of accessible chromatin found by both methods.

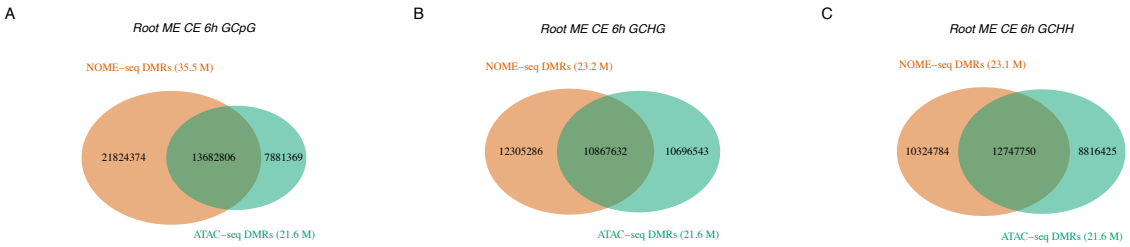

Figure 7.14: Accessible chromatin regions found by iNOMe compared to ATAC- in root samples with 6-hour ß-estradiol induction. The orange circle indicates the length of accessible chromatin detected by iNOMe only (Mb) whereas the turquoise indicates accessible chromatin identified by ATAC only (Mb). The overlap of both circles shows the length of accessible chromatin found by both methods.

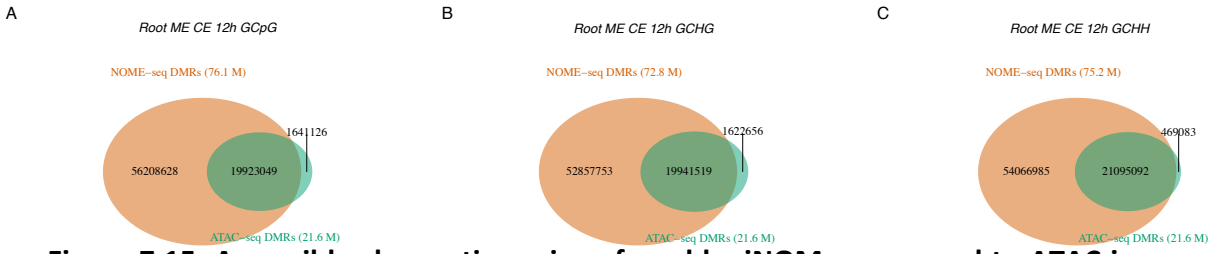

Figure 7.15: Accessible chromatin regions found by iNOMe compared to ATAC in root samples with 12-hour ß-estradiol induction. The orange circle indicates the length of accessible chromatin detected by iNOMe only (Mb) whereas the turquoise indicates accessible chromatin identified by ATAC only (Mb). The overlap of both circles shows the length of accessible chromatin found by both methods.

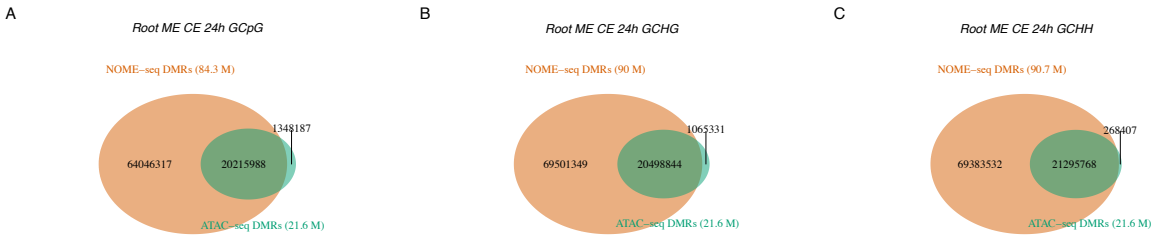

Figure 7.16: Accessible chromatin regions found by iNOMe compared to ATAC in root samples with 24-hour ß-estradiol induction. The orange circle indicates the length of accessible chromatin detected by iNOMe only (Mb) whereas the turquoise indicates accessible chromatin identified by ATAC only (Mb). The overlap of both circles shows the length of accessible chromatin found by both methods.