INTERPLAY BETWEEN TRANSCRIPTION AND CHROMATIN DYNAMICS

by

LI WANG

(Under the Direction of Xiaoyu Zhang)

ABSTRACT

Transcription of nuclear genes in eukaryotic organisms occurs in a chromatin environment, where the template DNA is wrapped around histone octamers to form nucleosomes. The interplay between transcription and the local chromatin structure is studied in this dissertation both at stress and normal conditions.

Exposure to heat triggers a conserved response of rapid alteration in gene expression. We studied the role of chromatin in the induction, maintenance and attenuation of temperature-dependent transcriptional changes. The results from genome-wide time course analyses of RNA abundance, RNA Polymerase II occupancy, histone variant distribution (H3.1 vs. H3.3) and a histone modification (H3K36me3) in Arabidopsis during heat stress indicated heat treatment led to the rapid and complete loss of both H3.1 and H3.3 as well as H3K36me3 signals, which was concomitant with the recruitment of Pol II to extraordinarily high levels. Attenuation of transcription (i.e. the loss of Pol II) was accompanied by the redeposition of new H3.3, but not H3.1. Histone modifications appeared to be re-established at a much later time. Our results also showed there was no evidence that heat-inducible genes were pre-deposited with any unusual chromatin state, or that H2A.Z may affect nucleosome theromostability in planta. Approximately 1/4 of the total Pol II signals relocalized from a specific set of genes to heat-inducible genes

during transcriptional activation, and from early inducible genes to late inducible genes during transcriptional attenuation.

Inevitably, all intra-nucleosomal DNA-histone interactions must be at least transiently disrupted during transcription. However, nucleosome organization on differentially expressed genes did not show any observable differences perhaps because the pattern of nucleosome positioning is determined by the combination of DNA sequence, nucleosome remodelers, DNA methylation, transcription factors including activators, components of the pre-initiation complex, elongating Pol II, histone modifications and histone variants.

This dissertation, for the first time, showed H3K4 tri-methyltransferase SDG2 interacts with a group of SMC like coiled-coil domain proteins that may direct it to chromatin for establishing the H3K4me3 pattern. It also showed the first genome-wide profiling of a new histone double mark H3K27me3S28ph, and suggested it may be involved in H3K27me3 target gene de-repression.

INDEX WORDS: Transcription, Chromatin Dynamics, Heat Stress, Nucleosome Positioning, Histone Variants, Histone Modifications, Histone Methyltransferase, H3K27me3S28ph

INTERPLAY BETWEEN TRANSCRIPTION AND CHROMATIN DYNAMICS

by

LI WANG

BS, Beijing Normal University, 2008

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

INTERPLAY BETWEEN TRANSCRIPTION AND CHROMATIN DYNAMICS

by

LI WANG

Major Professor:	Xiaoyu Zhang
Committee:	Katrien Devos
	Kelly Dawe
	Wolfgang Lukowitz
	Richard Meagher

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2015

DEDICATION

I would like to dedicate my dissertation to my beloved husband, Yaping Liu and my daughter, Sophia Liu. Without you, I would not be strong and courageous enough to complete my degree. I also dedicate this dissertation to my parents for their unconditional love.

ACKNOWLEDGEMENTS

I would like express my deepest gratitude to Dr. Xiaoyu Zhang, my advisor and mentor for his endless support, encouragement and most of all patience throughout the entire process. He has taught me both wet lab skills and computational data analysis skills. More importantly, he has taught me as a scientist, we should focus on what questions to ask rather than what we could do. Perhaps the most important thing I have learned from him is to "keep working on the right things without distractions", which is the sole truth to success. I have been greatly impressed and influenced by his passion, curiosity, criticism, insights and wisdom in science. Without his guidance and persistent help this dissertation would not have been possible.

I would also like to thank my committee members Drs. Kelly Dawe, Katrien Devos, Wolfgang Lukowitz and Richard Meagher for taking their time and efforts to guide me through my graduate years. In particular, Dr Kelly Dawe and Dr Wolfgang Lukowitz greatly helped me during the fourth year of my graduate study. They taught me to focus on "what's working" and never give up. Dr Katrien Devos has influenced me with her insightful comments and critiques, and she is a role model for me as a woman in science. Dr Richard Meagher has showed me various tips on biochemistry experiments. I feel very lucky to have all their knowledge and support in my research.

I am also incredibly lucky and honored to have the collaboration with Dr. Robert Schmitz, Dr. Ping Ma and Dr. Zach Lewis. They provided great support on conducting the experiments and data analysis. I really enjoyed and learned a lot at the joint group meetings with

their groups. Especially Xiaoxiao Sun, with whom I worked closely and carried out part of my data analysis for my dissertation.

I would like to thank my current and former lab mates who are really wonderful persons to work with. The time spent at the bench with all of them will forever be a cherishable memory for me.

The thesis would not have come to a successful completion, without the help I received from the staff of the Department of Plant Biology. I would like to thank Michael Boyd and Susan Watkins for all their support.

I would like to gratefully and sincerely thank Dr Brigitte Bruns. She has encouraged me to grow as an instructor and provided me with huge supports. The warm and cheerful words from her are truly one of the most important reasons that I did not give up.

I am very much indebted to my parents, Yonglin Wang and Jianqin Ge, my brother Chao Wang and my in-laws, Dongbao Liu and Xiaoling Wu, who supported me in every possible way to se the completion of this work.

Last but not the least, I would like to thank my husband and friend Dr. Yaping Liu, who loves me and supports me unconditionally, who teaches me patiently both in life and at work, who encourages me when time is hard, shares the joy when life is sweet, calms me down when I am overwhelmed, and tells me to let go when I am confused. He is my lucky one, and shows me the silver lining of the clouds. I am so grateful that I will have his companion through the rest of my life. He and our daughter Sophia are the inner peace of my heart.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

1. INTRODUCTION

1.1 Chromatin at a Glance

In eukaryotic cells, DNA is wrapped around a histone octamer consisting two copies of H2A, H2B, H3 and H4 to form nucleosomes, the basic units of chromatin. Although this organization likely evolved as a packaging mechanism, at least two lines of evidence suggest that it also provides a complex regulatory platform for multiple cellular processes such as gene transcription, DNA replication, recombination, and repair, by regulating the accessibility of DNA. First, a growing body of evidence from genomic studies showed that the structural properties of the chromatin are highly dynamic and non-even across the genome and different expression status is associated with drastically different characteristics. For example, actively transcribed genes contain regularly spaced nucleosomes with high levels of trimethylation on histone H3 lysine 4 (H3K4me3), H3K36me3 and H3/H4 hyperacetylation; In contrast, repressed genes are often hypoacetylated and associated with repressive marks like H3K27me3 (Lippman et al. 2004; Turck et al. 2007; Zhang et al. 2007a; Bernatavichute et al. 2008; Oh et al. 2008; Zhang 2008b; Charron et al. 2009; Zhang et al. 2009a; Jacob et al. 2010a); in addition, actively transcribed genes also display a strong association with some histone variants like H3.3 and H2A.Z (Deal et al. 2007; Jin et al. 2009; Ray-Gallet et al. 2011a; Adam et al. 2013; Chen et al. 2013b). Second, numerous chromatin modifying pathway mutants have been found to be defective in gene expression and development. For instance, mutants of histone lysine methyltransferases have

altered gene expression profiles and show various pleiotropic phenotypes throughout development (Grini et al. 2009; Guo et al. 2010; Jacob et al. 2010b). Such dynamic regulations usually do not involve changes to the DNA sequence itself, thus known as epigenetic regulations. As epigenetic regulation is the fundamental concept of this thesis, I would like to quote the definition of Epigenetics as described in the NIH Roadmap Epigenomics Project 2013(Bernstein et al. 2010):

Epigenetics is an emerging frontier of science that involves the study of changes in the regulation of gene activity and expression that are not dependent on gene sequence. For purpose of this program, epigenetics refers to both heritable changes in gene activity and expression (in the progeny of cells or of individuals) and also stable, long-term alterations in the transcriptional potential of a cell that are not necessarily heritable.

Current research in this field is mainly focused on DNA methylation, histone variants, histone modifications, nucleosome remodeling and RNA interference. Genomic approaches such as high-density microarrays and high-throughput sequencing have revolutionized epigenetic studies(Zhang 2008a; Schmitz and Zhang 2011). This chapter will cover the background knowledge of histone H3 variants, histone modifications and nucleosome organizations.

1.2 Nucleosome Organization

Nucleosome in eukaryotic cells is the basic repeating unit of chromatin. Each nucleosome comprises 145~147bp of DNA wrapped 1.65 times around histone octamer of four highly evolutionarily conserved histone proteins – H2A, H2B, H3 and H4 (Jiang and Pugh 2009). Nucleosome positioning refers to where the nucleosomes are located with respect to genomic DNA sequence, which can vary from perfect positioning that a nucleosome is located at a given position in all cell populations, to no positioning that nucleosome is located at all possible

genomic positions with equal frequency in a cell population. With the help of nucleosome mapping techniques on genome-wide scale, researchers have discovered the common patterns of nucleosome organization in several model organisms. Nucleosome positioning and phasing are more prominent at some genomic regions than others, in that the phasing and positioning is strongest flanking the NFR (nucleosome free region) upstream of transcription start sites, and becomes less and less pronounced downstream of the gene. Nucleosome positioning is determined by different factors including DNA sequences, ATP-dependent chromatin remodelers, DNA methylation, transcription factors and elongating RNA polymerase II, histone modifications and histone variants (Fernandez and Anderson 2007; Morozov 2011; Sadeh and Allis 2011; Hughes et al. 2012; Narlikar et al. 2013; Struhl and Segal 2013).

DNA sequence is the first cue for nucleosome organizations. One of the most important questions is does the DNA sequence affect nucleosome organization, and to what extend. As a first step towards understanding the DNA preferences of nucleosomes on chromosome scale, Yuan *et. al.* (Yuan et al. 2005) measured nucleosome occupancy along yeast chromosome III using a tiled microarray and found nucleosomes seem to favor poly-A or poly T sequences. Since then, the debate on the role of DNA sequence in nucleosome positioning *in vivo* and *in vitro* has been going on. There is evidence that intrinsic DNA sequence plays a central role in determining the organization of nucleosomes in different eukaryotic genomes(Segal et al. 2006; Kaplan et al. 2009; Kaplan et al. 2010; Tillo et al. 2010). Two major sequence determinants that affect nucleosome organization are: 1) The 10bp periodicity of AA/TT/TA dinucleotides that helps DNA bend around nucleosomes, and the exact position of the histone octamer with respect to the 10bp helical repeat is termed “rotational positioning” and 2) homopolymeric sequences poly (dA:dT) and poly (dG:dC) strongly inhibit nucleosome array formation, and are important for

nucleosome depletion(Struhl and Segal 2013). Elements named “container sequences” were discovered in human that GC dinucleotides are enriched in the center of highly positioned nucleosomes (dyad), while AT dinucleotides seem to define the boundaries of the nucleosome core and were suggested to position nucleosomes (Valouev et al. 2011). However, using a positioning stringency metric that quantifies the fraction of defined nucleosome positions within a given segment they only found less than 20% of the nucleosomes are highly positioned *in vivo*, suggesting that others factors might also function to establish accurate nucleosome positioning and phasing.

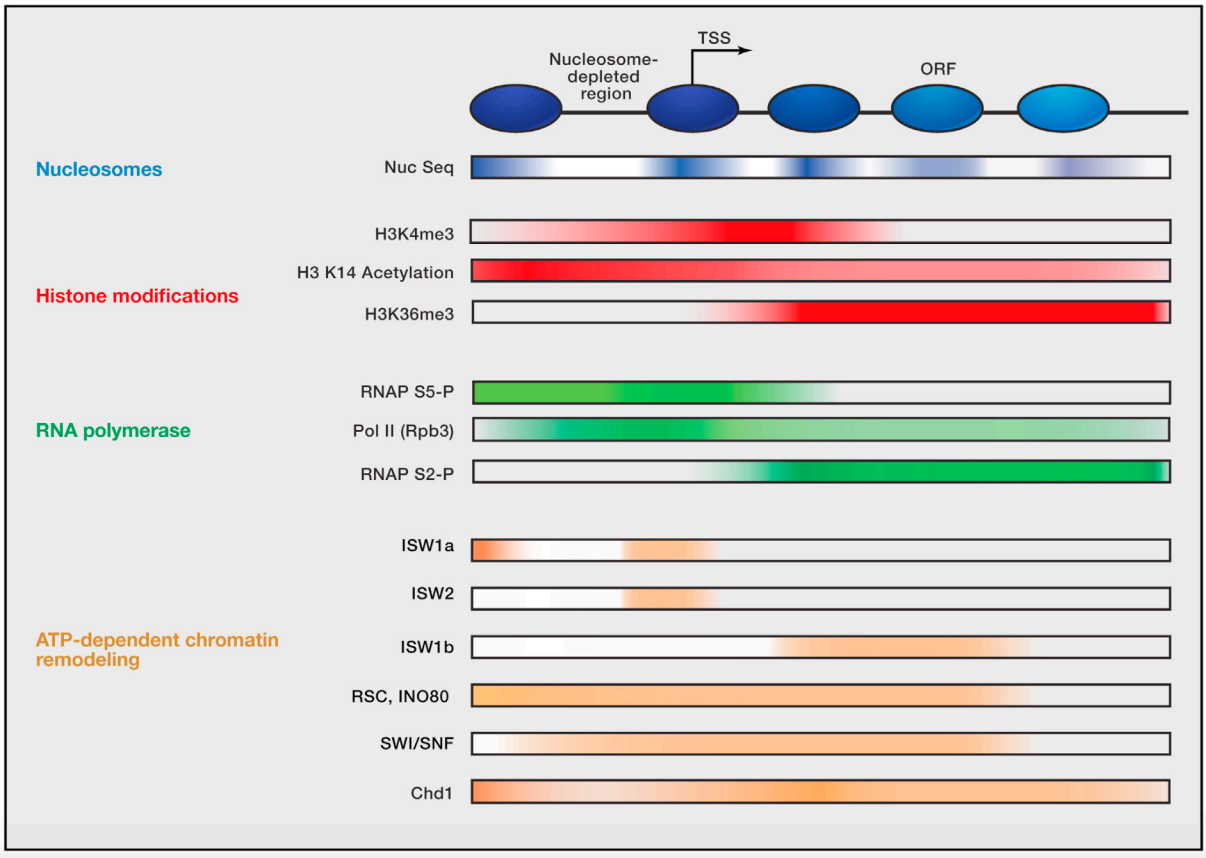

Fig. 1.1 A schematic representation of distribution of nucleosome, histone modifications, RNA polymerase II and ATP-dependent chromatin remodelers with respect to transcribed genes. When genes are aligned by transcription start site, the region just upstream of the promoter is depleted for nucleosomes and an array of ordered nucleosomes extends into the

coding region. Histone modifications, RNA polymerase II, ATP-dependent chromatin remodelers, histone variants H3.3 and H2A.Z (not shown) all exhibit distinct distributions with respect to transcribed genes, suggesting the interplay with nucleosome positioning. Adapted from (Narlikar et al. 2013). Note that metazoans differ from yeast in that they do not bury the TSS in the +1 nucleosome (Jiang and Pugh 2009).

ATPase-dependent chromatin remodelers are critical for nucleosome organization *in vivo*. Several aspects of the *in vivo* nucleosome positioning pattern can be reconstituted *in vitro* if a yeast crude extract (provides a source of chromatin-associated proteins) together with ATP are added to purified histones and DNA (Zhang et al. 2011). Specifically, they observed the organization of nucleosomes in which the region just upstream of promoter is depleted and an array of ordered nucleosomes extends into the coding regions. As this pattern is only observed at the presence of ATP, it suggests the important roles of ATP-dependent chromatin remodelers play in nucleosome organization. It is hypothesized that remodeling enzymes not only simply facilitate the movement of nucleosomes to preferred intrinsic positions, but also are important in the specificity of where nucleosomes are located (Struhl and Segal 2013). Even though the mechanism by which nucleosome remodelers override intrinsic DNA sequence preferences of histone octamers is unclear, it is possible that they either bind DNA specifically to position nucleosomes, or they uses nucleosome depleted regions as a measuring device to position nucleosomes. The catalytic subunits of ATP-dependent chromatin remodeling complexes are the helicase-related Snf2- or SWI/SNF- related proteins, and can be assigned into 24 distinct subfamilies (Flaus et al. 2006). They have been identified within the genomes of all eukaryotes

including *Arabidopsis*, suggesting conserved mechanisms for their functions (Narlikar et al. 2013).

It is still under debate where DNA methylation is correlated or anti-correlated with nucleosome occupancy from both computational and experimental sides. Taking advantage of NOMe-seq which includes both nucleosome positioning and DNA methylation data for individual DNA strands, researchers found the relationship between nucleosome occupancy and DNA methylation is context-specific depending on genomic location(Kelly et al. 2012; Liu et al. 2012; Lay et al. 2015), e.g. DNA methylation and nucleosome occupancy were strongly anti-correlated surrounding CTCF sites such that DNA methylation peaked in the linker regions between nucleosomes, and loss of DNA methylation at CGI promoters lead to acquisition of active and poised histone marks, but did not alter chromatin structure of non-CGI promoters.

RNA polymerase II transcription also plays critical roles in establishing nucleosome pattern *in vivo*. Two major aspects of *in vitro* reconstituted nucleosome patterns do not go faithfully with *in vivo* patterns: the precise location of the +1 nucleosomes matched poorly with their *in vivo* positions; and the degree of positioning of downstream nucleosomes is significantly lower(Zhang et al. 2011). These seem to be reconciled by general transcription factors that compose the pre-initiation complex and the elongating Pol II machinery. One striking feature of nucleosome organization in all eukaryotes is the presence of nucleosome-depleted region at promoter and a precisely positioned +1 nucleosome on transcribed genes. Transcriptional activators, via targeted recruitment of nucleosome remodelers, can generate nucleosome-depleted regions, and the pre-initiation complex plays a role in fine-tuning the position of the +1 nucleosome (Narlikar et al. 2013). The unidirectionality of nucleosome arrays suggests the Pol II elongation affects nucleosome positioning, too. For instance, in human CD4+ T cells, the + 1

nucleosome of expressed genes with elongating Pol II was located at + 40 bp downstream of the TSS, whereas genes with stalled Pol II had their + 1 nucleosome + 10 bp downstream of TSS, and well-positioned Pol II was not observed at inactive genes (Schones et al. 2008).

The roles of histone modifications and histone variants on nucleosome positioning have been explored by researches recently. As found in human and in *Arabidopsis*, histone modifications also co-localize with nucleosome positions (Schones et al. 2008; Chodavarapu et al. 2010), indicating nucleosome distribution likely to be the underlying basis of differential distribution of histone modifications. Direct protein-protein interactions identified recently may further suggest the importance of histone modifications in nucleosome positioning. For examples, H3K4m2/3 together with H4K16ac were able to recruit BPTF, a subunit of the NURF ATP-dependent chromatin-remodeling complex, to chromatin regions, suggesting that histone modifications might indeed affect nucleosome organization(Ruthenburg et al. 2011). New findings in yeast further support there are indeed direct interactions between nucleosome core particle and histone acetyltransferase complex NuA4 (Chittuluru et al. 2011), suggesting that acetylation of histone H4 and H2A tails may affect nucleosome positioning. Replacement of major histones by histone variants that modulate DNA/histone interactions and alter the inter-nucleosomal interactions mediated by the histone tails could also affect nucleosome positioning. For examples, the replacement of histone H3 with CENP-A (centromere specific H3 variant) directs global nucleosome array condensation and loosening of nucleosome superhelical termini (Panchenko et al. 2011). In all eukaryotes, the +1 nucleosome and a few downstream nucleosomes of active genes are enriched for the histone variant H2A.Z (Bonisch and Hake 2012). In ESCs, H2A.Z increases nucleosome accessibility at FoxA2-binding sites (Li et al. 2012). The incorporation of

H2A.Z affects the stability of +1 nucleosome thus influences the context-specific barriers formed by nucleosomes to transcription in *Drosophila* (Weber et al. 2014).

The genome-wide pattern of nucleosome positioning is determined by the combination of DNA sequence, nucleosome remodelers, DNA methylation, transcription factors including activators, components of the pre-initiation complex, and elongating Pol II, and histone modifications and histone variants. The organization of nucleosomes is a barrier to RNA polymerase II (Pol II) and affects transcription in various ways (Workman and Kingston 1998; Workman 2006; Hodges et al. 2009; Jin et al. 2010; Bintu et al. 2012; Teves et al. 2014). But the relationship *vice versa* has not been studied extensively so far. In my third chapter, I will mainly focus on the effects of transcription, if there is any, on nucleosome positioning in *Arabidopsis*.

1.3 Histone H3 Variants

Histones are a family of basic proteins that interact with DNA to form nucleosomes, which is the basic unit of chromatin. There are five major histone families: H1/H5, H2A, H2B, H3 and H4. Two copies of H2A, H2B, H3 and H4 each form the core of nucleosomes, while H1/H5 are known as linker histones. Despite their conserved role as general DNA packaging agents, it is now clear that they are crucial to confer variations in chromatin structure to ensure dynamic patterns of transcriptional regulation in eukaryotes. The incorporation of histone variants into nucleosomes marks specific chromatin states probably resulting from structural alterations in the core octamer, see reviews (Sarma and Reinberg 2005; Talbert and Henikoff 2010; Maze et al. 2014). Variants of the same histone differ from each other at their primary sequences that can vary from a few amino acid substitutions to large domains. They have distinct expression, deposition and regulatory mechanisms that could potentially affect the structure and function of chromatin (Sarma and Reinberg 2005; Maze et al. 2014). To date, only one H4 isoform has been

identified, and H3 variants are generally less diverse than those arising from the H1, H2A and H2B families. There are four types of histone H3: the canonical H3.1, the replacement H3.3, centromere-specific CenH3 (or CENP-A in human), and a male-gamete-specific variant AtMGH3 in *Arabidopsis* (Okada et al. 2005; Ingouff and Berger 2010) or H3t in human that is specifically found in testis(Witt et al. 1996; Talbert and Henikoff 2010; Szenker et al. 2011). My research focuses on H3.1 and H3.3 in *Arabidopsis*, particularly their interactive relationship with transcriptional dynamics.

H3.1 and H3.3 are highly similar at their primary sequences, and both are conserved in eukaryotes. Phylogenetic analysis strongly suggests H3.3 evolved from H3.1 independently in animals and plants (Malik and Henikoff 2003; Waterborg 2012). Sequence alignments among *Arabidopsis*, human and *Drosophila* showed that the differences are defined by four amino acid substitutions that have arisen numerous times along history (Talbert and Henikoff 2010; Shi et al. 2011; Szenker et al. 2011). The substitutions are at sites 31 (Ala vs. Ser), 87 (Ser vs. Ala), 89 (Val vs. Ile) and 90 (Met vs. Gly) in animals and it has been shown that amino acids 87, 89 and 90 are critical in the choice of incorporation pathways (Ahmad and Henikoff 2002; Goldberg et al. 2010; Elsasser et al. 2012). Particularly, the amino acid substitution from Ala to Ser at position 31 determines the specific recognition of H3K36me3 by its reader protein ZMYND11 that affects histone-variant-mediated transcription elongation as well as mRNA splicing(Guo et al. 2014; Wen et al. 2014). In *Arabidopsis*, the differences between H3.1 and H3.3 are also present at sites 31 (Ala vs. Thr), 89 (Ser vs.His) and 90 (Ala vs. Leu), but with different substitution residues. Instead of site 89, there is an *Arabidopsis* specific substitution at position 41 from Phe to Tyr, where it passes through the minor groove of the DNA double helix around the nucleosome (Shi et al. 2011). This substitution is conserved in both monocot and dicot plants

(Ingouff et al. 2010). The features of H3.1 and H3.3 primary sequences suggest that conserved and specific regulatory mechanisms may exist in plants and animals.

Despite the high sequence similarity, the synthesis, deposition and posttranslational modification mechanisms are very different for H3.1 and H3.3. In animals, genes encoding canonical H3.1 are intron-less and are organized in tandem clusters (Szenker et al. 2011). Bulk H3.1 is expressed specifically at S phase and is incorporated into chromatin in a replication-dependent manner (Osley 1991; Ahmad and Henikoff 2002; Schwartz and Ahmad 2005; Hamiche and Shuaib 2012). H3.1 could also be incorporated at DNA damage sites outside of S phase, but is still DNA synthesis coupled (Polo et al. 2006). In contrast, genes encoding H3.3 are scattered throughout the genome and have distinct untranslated regions (Szenker et al. 2011). H3.3 is expressed

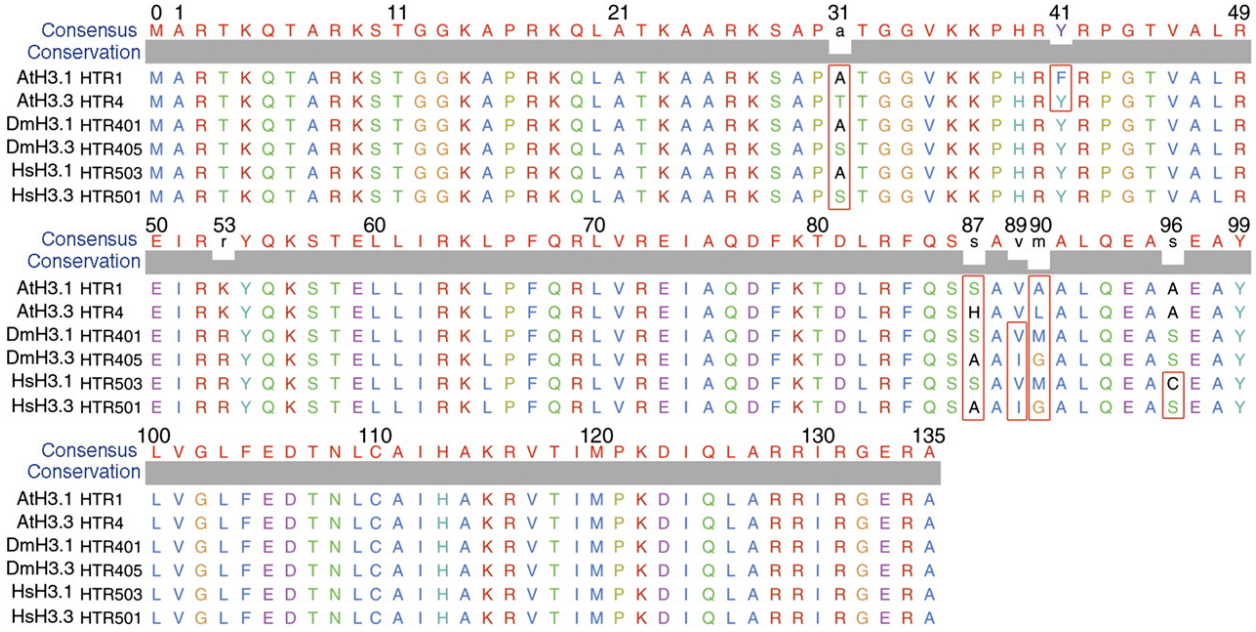

Fig. 1.2 Multiple alignments between H3.1 and H3.3 among *Arabidopsis thaliana* (At), *Drosophila melanogaster* (Dm), and *Homo sapiens* (Hs) showing the 4 amino acids substitutions in plant and in animals. Adapted from (Shi et al. 2011)

constitutively and is incorporated into chromatin in a replication independent pathway. The deposition of H3.1 is mediated by chromatin-assembly factor 1 (CAF1)(Gaillard et al. 1996),

whereas H3.3 is deposited by various factors including HIRA (histone regulator A)(Osley 1991; Ahmad and Henikoff 2002; Hollenbach et al. 2002; Ray-Gallet et al. 2002; Nakatani et al. 2004; Loyola and Almouzni 2007; Elsaesser and Allis 2010), ATRX (alpha thalassemia/mental retardation syndrome X-linked)(Goldberg et al. 2010), death-associated protein DAXX(Hollenbach et al. 2002; Drane et al. 2010; Elsaesser and Allis 2010; Goldberg et al. 2010; Elsasser et al. 2012) and DEK (DEK oncogene) (Sawatsubashi et al. 2010). Intriguingly, different sets of histone post translational modifications are found on the two H3s. Mass-spec studies show H3.1 is more often modified with repressive or silencing marks like H3K9me and H3K27me, while H3.3 is associated with active marks, like H3K36me3 and histone acetylation(Johnson et al. 2004; McKittrick et al. 2004; Loyola et al. 2006; Wu et al. 2009). Genome wide ChIP-seq results show H3.3 is co-localized with active transcription marks like serine 5 phosphorylated PolII and active histone modifications including H3K4me3, H3K36me3 and H3K4me1 in human ESC cells(Goldberg et al. 2010). Consistently, H3.3 is found to be enriched at transcriptionally active genes and regulatory regions in *Drosophila*(Ahmad and Henikoff 2002; Mito et al. 2005; Mito et al. 2007; Henikoff et al. 2009), human ES cells(Goldberg et al. 2010), and HeLa cell lines(Jin et al. 2009), and is positively associated with gene transcription levels, whereas H3.1 is primed at repressed genes and inversely co-related with gene transcription levels(Mito et al. 2005; Goldberg et al. 2010; Ooi et al. 2010; Ray-Gallet et al. 2011b).

Similar to their animal counterparts, plant H3.1 and H3.3 are also associated with distinct sets of modifications. Mass spec studies in *Arabidopsis* showed silencing related histone methylation marks like H3K27 methylation are associated with H3.1, while modifications found at transcribed regions of active genes like H3K36 methylation were enriched at H3.3(Johnson et al.

2004). In addition, the global distribution of H3.1 and H3.3 in plant genomes is also similar to that in animals, with H3.1 enriched at inactive chromatin regions and inversely associated with transcription, whereas H3.3 is primed at active chromatin regions, and positively associated with gene transcription(Stroud et al. 2012; Wollmann et al. 2012). Despite the similarities, the gene-level distribution of H3.3 appears to be different in plants and in animals. In *Arabidopsis*, the distribution of H3.3 gradually increases towards the 3' end of genes and reaches a maximum immediately upstream of transcription termination site (TTS)(Stroud et al. 2012; Wollmann et al. 2012). Even though preferential enrichment of H3.3 towards the gene end has also been reported in mammalian cells, the maximum abundance is observed beyond TTS (Ray-Gallet et al. 2011b). In *Drosophila*, H3.3 enrichment is biased in that it flanks the TSS and is gradually reduced throughout the transcribed regions(Mito et al. 2005). Moreover, H3.3 enrichment on cis-regulatory elements is not observed in plants. All these profiles are positively correlated with their corresponding PolII profiles(Mito et al. 2005; Wirbelauer et al. 2005; Ooi et al. 2010; Ray-Gallet et al. 2011b; Wollmann et al. 2012), suggesting deposition of H3.3 is a process correlated with transcription but may be regulated by distinct mechanisms.

1.4 Tri-methylation on Histone H3 Lysine 4: Pattern, Establishment and Function

Histones can be post translationally methylated at lysine and arginine residues, which can either promote or prevent the docking of key transcriptional effectors to regulate gene transcription. One of the most extensively studied modifications is methylation of lysine residues on histone H3 tails, namely lysine 4, lysine 9, lysine 27 and lysine 36. Histone lysine methylation is catalyzed by a group of proteins with a conserved SET domain, which is about 130-160aa long, and is named after Drosophila histone methyltransferases *Su(var)3-9*, *Enhancer of Zeste* [*E(z)*] and *trithorax* (*trx*) (Baumbusch et al. 2001), known as "writer" proteins. Whole genome level

profiling in *Arabidopsis* revealed strong associations of certain lysine methylation with active or repressed genes, in that transcribed genes are strongly associated with H3K4 or H3K36 methylation, while developmentally repressed genes often have H3K27 tri-methylation (Lippman et al. 2004; Turck et al. 2007; Zhang et al. 2007a; Bernatavichute et al. 2008; Oh et al. 2008; Zhang 2008b; Charron et al. 2009; Zhang et al. 2009a; Jacob et al. 2010a). In contrast, H3K9me2 are exclusively enriched on transposons and other repeats (Lippman et al. 2004; Turck et al. 2007). Histone modifications contribute to gene regulation through direct effects on chromatin structure and through the recruitment of effecter proteins known as "reader" proteins. In this chapter, I will focus on histone H3 lysine 4 trimethylation to discuss its distribution along genes, the "writer" and "reader" proteins and the possible functions.

H3K4me3 has been shown as a promoter-specific histone modification associated with active transcription. Chromatin immunoprecipitation followed by micro array shows that the enrichment of H3K4me3 along genes is positively correlated with gene transcription levels (Zhang et al. 2009a). In embryonic stem cells, H3K4me3 in found in bivalent domains with the repressive methylation mark H3K27me3 on developmentally relevant genes that are silent but 'poised' for activation (Bernstein et al. 2006; Mikkelsen et al. 2007b). But genome wide co-occurrence of H3K4me3 and H3K27me3 has not been identified in plants (Zhang et al. 2009b), or the seemingly co-localization is largely due to mixed tissue and cell types (Lafos et al. 2011). E.g. Fig. 1.3 shows lysince methylation on K4, K27 and K36 occupy different regions of genes, with H3K4me3 located at 5' end of genes and the most actively transcribed genes occupy the highest level of H3K4me3 in *Arabidopsis*. Histone H3 lysine 4 methylation is catalyzed by the highly evolutionarily conserved multi-protein complex known as Set1/ COMPASS or MLL/COMPASS-like complexes. The first HMT shown to catalyze mono, di and trimethylation

of H3K4 is SET domain-containing 1 (Set1) in *Saccharomyces cerevisiae*, forming a protein complex called COMPASS (complex proteins associated with Set1) (Miller et al. 2001). In mammals, six COMPASS-related complexes with essential and non-redundant functions have been identified that can methylate H3K4 through their conserved SET domain as reviewed in (Schuettengruber et al. 2011): SET1A- or SET1B-containing complexes and MLL1-4 complexes. *Arabidopsis* H3K4 tri-methyltransferases have been identified as ATX1/SDG27, ATX2/SDG30, ATXR7/SDG25 (Alvarez-Venegas and Avramova 2005; Saleh et al. 2008; Berr et al. 2009; Tamada et al. 2009) which only display locus-specific defects in H3K4me, whereas the SDG2 has been demonstrated to have global defects in H3K4me3 level, is able to methylate H3K4 in all three context mono- di- and tri, and thus is likely to be the major H3K4 tri-methyltransferase in *Arabidopsis* (Guo et al. 2010).

The functional outcome of H3K4me3 is largely determined by its "reader" proteins. So far, a large number of readers for methylated histones have been described in mammalian cells, which usually contain Plant Homeodomain (PHD), Tudor domain, Chromo domain, PWWP domain, WD40 repeats or the CW domain (Ruthenburg et al. 2007; Taverna et al. 2007; Adams-Cioaba and Min 2009; Vermeulen et al. 2010; Musselman et al. 2012). The functions of H3K4me3 have been shown based on the recognition of its reader proteins include but are not limited to 1) activation of transcription, which has been shown that the basal transcription factor TFIID directly binds to the H3K4me3 mark via the PHD finger of TAF3 (Vermeulen et al. 2007), facilitates global TFIID recruitment to the core promoters of active genes, and enhances PIC (pre-initiation complex) formation to regulate the expression of select p53 target genes (Lauberth et al. 2013). 2) Chromatin remodeling and histone acetylation, where H3K4me3 was recognized

by human BPTF via the PHD domain (Li et al. 2006) and is involved in ATP-dependent chromatin remodeling (Neely and Workman 2002); The ING proteins bind to H3K4me3 via

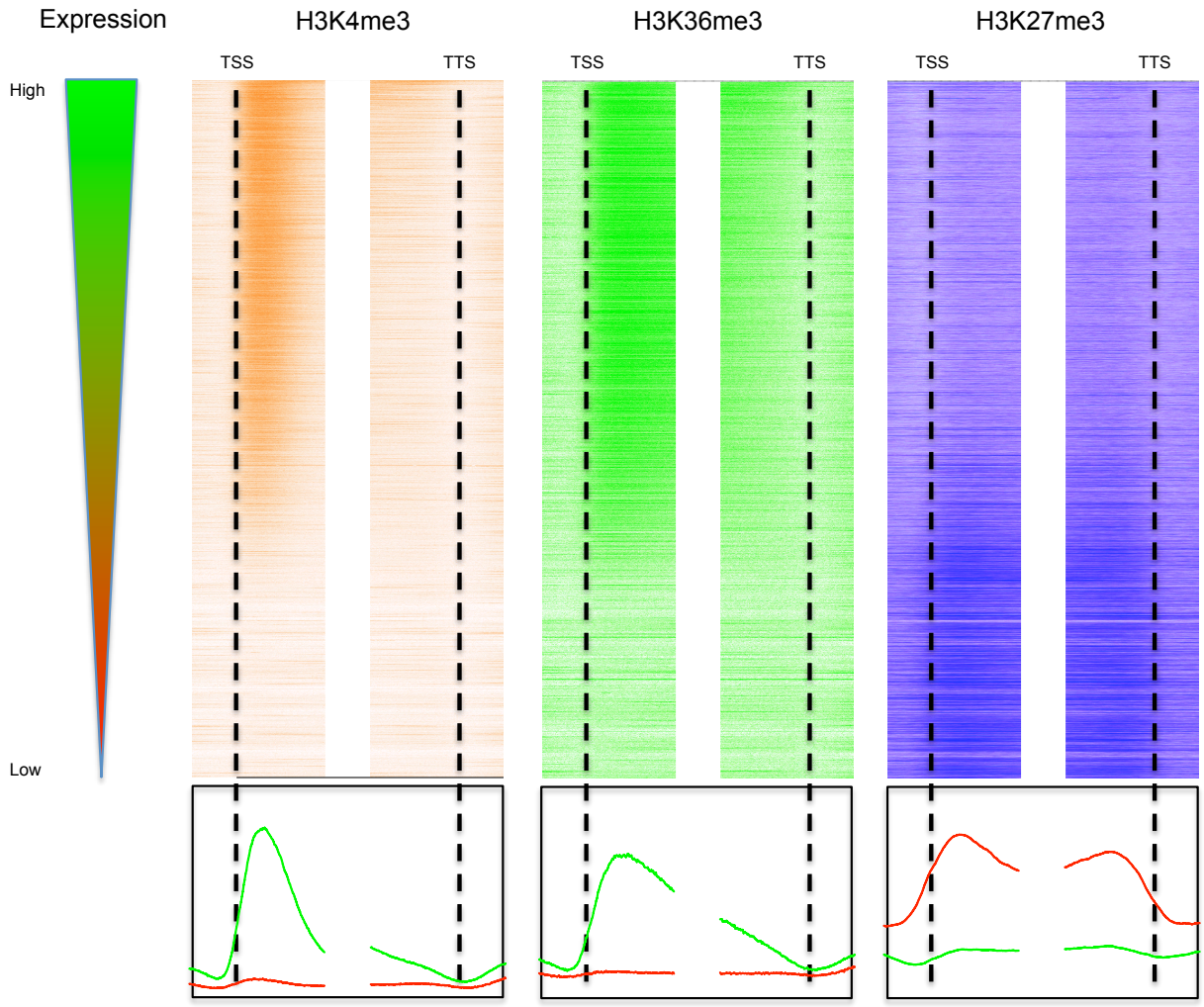

Fig. 1.3 Heatmaps depict H3K4me3, H3K36me3 and H3K27me3 ChIP-Seq at regions spanning all transcription start sites (TSS) and transcription termination sites (TTS), and their association with transcription levels.

PHD domain and are key components of specific HAT or HDAC multi-subunit complexes (Pena et al. 2006); The multisubunit SAGA coactivator complex binds to H3K4me3 via the double

Tudor domain in its subunit Sgf29, has both histone acetyltransferase and deubiquitination activities and remodels chromatin to allow transcription (Vermeulen et al. 2010). 3) Histone demethylation, where histone lysine demethylase JMJD2A recognizes H3K4me3 via its double Tudor domain (Huang et al. 2006; Lee et al. 2008) to demethylate lysine 9 or lysine 36 di- and tri-methylation (Couture et al. 2007); the zf-CW domain of ZCWPW1 also binds to H3K4me3 (He et al. 2010) though its function needs to be further explored. 4) Transcription repression, where the *Arabidopsis* CW domain containing proteins VAL1 and VAL2 have been shown to bind H3K4me3, and the *val1val2* double mutant fails to repress embryonic development during vegetative growth (Suzuki et al. 2007). 5) Facilitate rRNA expression, where the meiotic spindle-associated protein Spindlin1 recognizes H3K4me3 through its tandem Tudor-like domain and facilitate the expression of rRNA genes (Wang et al. 2011). 6) DNA recombination, where an essential component of the RAG1/2 V(D)J recombinase RAG2 recognizes H3K4me3 through its PHD domain and functions in mammalian DNA recombination (Matthews et al. 2007).

Given this functional diversity an important question is how specificity of H3K4me3 readers is achieved. The first mechanism could be sequence context specificity of reader proteins as exemplified by the different binding modes of H3K4me3 to the Tudor domains of JMJD2A, Sgf29 and Spindlin1 (Yang et al. 2012) as well as PHF8 and KIAA1718 both bind to H3K4me3 through PHD domain, but demethylate H3K9me2 or H3K27me2, respectively, resulted from different protein conformations the two enzyme adopt (Horton et al. 2010). Specificity can be achieved through cell type-specific expression. For instance, RAG2 is only expressed in lymphocytes that undergo active rearrangements of antigen receptor genes (Matthews et al. 2007). A third mechanism lies in the interaction of specific DNA sequence that are recognized by H3K4me3 readers or their complexes, which can be demonstrated in the TFIID/TAF3

example (Lauberth et al. 2013). Differential binding affinities of reader proteins to methyl-lysine also help establish specificity. For example, the binding of H3K4me2 and H3K4me1 to MLL5 were five- and 16-fold more weakly than H3K4me3 (Ali et al. 2013). Last but not least, specificity can be achieved through cross talk with other histone modifications. The interaction between the TAF3 PHD domain and H3K4me3 is eight- to twenty-fold lower when H3R2 is asymmetrically dimethylated (H3R2me2a), whereas the interaction between the ING2 and BPTF PHD finger and H3K4me3 seems less affected (Vermeulen et al. 2007; Iberg et al. 2008).

Most of the reader proteins of H3K4me3 have also been found in *Arabidopsis*. There are 83 PHD fingers found in 70 proteins in *Arabidopsis*, which play important roles in diverse types of cellular, developmental and environmental regulation and 9 of them may bind to H3K4me3 (Lee et al. 2009). 16 ZF-CW domain proteins including SDG8, VAL1/VAL2, SNF2, etc.. Even though VAL1/VAL2 have been shown to bind H3K4me3 (Suzuki et al. 2007), the CW domain of SDG8 likely bind to mono- or di- methylated H3K4 better (Hoppmann et al. 2011). However, unpublished data from our lab suggest this binding affinity may need closer examination.

It is of note that even though H3K4me3 is located preferentially at the 5' end of active genes, the breadth of its distribution along gene body varies. It has been shown that in human, the breadth of H3K4me3 is broader over genes that have more paused polymerase at their promoters and are characterized by increased marks of elongation, and these genes tend to be essential for cell identity (Benayoun et al. 2014). Similar distribution pattern has been identified in *Arabidopsis* (Fig 1.3), and the breadth of H3K4me3 is proposed to be linked to inducibility of a gene under stress and nucleosome distribution (unpublished data from our lab).

The underlying targeting mechanism of histone methyltransferase is also of great interest to researchers. As exemplified in Fig 1.3, the distributions of H3K4me3, H3K36me3 and

H3K27me3 are uneven along genes, and display distinct patterns from each other. Understanding how specificity is achieved for the recruitment of histone methyltransferases is important to interpret the different patterns of histone modifications. I am particularly interested in understanding the targeting mechanism of H3K4me3 methyltrasferase in *Arabidopsis*, and will discuss this further in chapter 4.

My dissertation studies the interplay between gene transcription and chromatin dynamics. In my second chapter, I studied the role of chromatin in the induction, maintenance and attenuation of temperature-dependent transcriptional changes. I will discuss the results from genome-wide time course analysis of RNA abundance, RNA polymerase II occupancy, histone variant distribution (H3.1 and H3.3) and a histone modification (H3K36me3) in *Arabidopsis* during heat stress. In my third chapter, I investigated the effects of gene transcription on nucleosome positioning. Nucleosome positioning patterns on differentially expressed genes between *Arabidopsis* seedlings and flowers will be shown. Then I will switch the gear to histone modifications, where in my fourth chapter, I will focus on one of the histone lysine 4 tri-methyltransferase SDG2 and characterize the recruitment mechanism for H3K4me3 establishment. I will go back to genome-wide profiling of a new histone bi-mark H3K27me3S28p in my fifth chapter.

CHAPTER 2

2. DYNAMIC HISTONE REPLACEMENT AND POL II RE-LOCALIZATION DURING HEAT INDUCED TRANSCRIPTIONAL ACTIVATION IN *ARABIDOPSIS THALIANA*

2.1 Introduction

Transcription of nuclear genes in eukaryotic organisms occurs in a chromatin environment, where the template DNA is wrapped around histone octamers to form nucleosomes. The interplay between transcription and the local chromatin structure is one of the most intensely studied topics in the past two decades. Chromatin modification pathways (such as chromatin remodeling, histone replacement and histone modification, etc) were once thought to be regulatory mechanisms that act upstream of transcription to either facilitate or inhibit the major steps of transcription, namely initiation, elongation, termination, and co-transcriptional RNA processing (Li et al. 2007; Rando and Winston 2012; Smolle et al. 2012; Thurman et al. 2012; Adam et al. 2013; Voss and Hager 2014). Consistent with this notion, numerous cases have been reported where mutations in chromatin modification pathways lead to severe defects in transcription (Grini et al. 2009; Wen et al. 2009; Guo et al. 2010; Jacob et al. 2010b). However, it has also become clear in recent years that transcription itself plays a central role in shaping the chromatin. Passage of the transcription apparatus transiently interrupts the histone-DNA interaction in nucleosomes, which in time can have profound effects on the organization as well as the composition of the nucleosome array (Naughton et al. 2013). In addition, RNA

polymerase II (Pol II) serves as the binding platform for numerous histone-modifying enzymes to add or remove covalent modifications during each round of transcription (Badeaux and Shi 2013). Therefore, in many cases, transcription and chromatin modifications can be viewed as different aspects of the same process.

Consistent with the close association between transcription and chromatin modifications, there is a nearly perfect correlation between the transcription status of a given gene and its chromatin states (described in detail in Chapter 1). In Arabidopsis, actively transcribed genes are associated with low levels of H3.1, uneven levels of of H3.3 and H2A.Z, and high levels of H3K4me1/2/3, H3K36me1/2/3 as well as H3/H4ac. In contrast, developmentally repressed genes are associated with high levels of H3.1, H2A.Z and H3K27me3 (Zhang et al. 2007b; Zhang et al. 2007c; Zhang 2008b; Zhang et al. 2009a; Goldberg et al. 2010; Ray-Gallet et al. 2011b; Coleman-Derr and Zilberman 2012; Huang et al. 2012; Stroud et al. 2012; Wollmann et al. 2012; Chapman-Rothe et al. 2013). During developmentally controlled transcriptional activation or repression, the transition from one chromatin state to another is a complicated and deliberate process that often involves the collective actions and multiple pathways. In some cases, the existing modifications are enzymatically removed (the active removal mode) and new modifications are established by a different set of pathways. In other cases, the maintenance pathways are inhibited at specific genes, such that the existing modifications are diluted through the incorporation of unmodified histones during multiple rounds of cell divisions (the passive loss mode), before the new modifications can be established.

In contrast to developmentally controlled genes that are activated or repressed at specific stages, plants also encode a large and diverse set of stress-inducible genes that can rapidly respond to a wide range of environmental stimuli. Very little is known regarding the interplay between

transcription and chromatin at these genes, and many outstanding questions remain. Studies showed that SDG8 functions to regulate the H3K36 methylation of histones associated with gene bodies in *Arabidopsis* (Xu et al. 2008), is required for the permissive expression of touch induced genes (Cazzonelli et al. 2014), as well as associated with high-level expression of a specific set of light and/or carbon responsive genes (Li et al. 2015). Are stress-inducible genes associated with a special chromatin state under normal conditions, which could poise them for activation? How is the rapid and high level of transcriptional activation achieved in a chromatin environment? How may the relatively short transcription activation and attenuation cycle affect the structure, composition and modification of the local chromatin?

To begin to address these questions, we performed a comprehensive analysis of transcriptional and chromatin changes during heat stress, including genome-wide time-course studies of the histone content and abundance, histone modifications, transcript level, and Pol II occupancy. We found that Arabidopsis responded to heat stress with a transcriptional burst that began with rapid and extraordinarily high levels of Pol II recruitment to heat-induced genes. Concomitant with Pol II recruitment, preexisting histones were evicted from heat-induced genes, and transcription proceeded in a nucleosome-free environment to produce nearly all the transcripts. Transcription was attenuated approximately an hour later, which was followed by the redeposition of newly synthesized and unmodified histones. Interestingly, Pol II recruitment during transcriptional activation and Pol II dissociation during transcriptional attenuation appeared to be facilitated by the redistribution of existing Pol II among several specific set of genes. Taken together, these results provided the first insight into the transcription-chromatin interplay during heat stress, and unveiled fundamental differences between developmentally and stress-induced transcriptional changes.

2.2 Results and Discussions

Early transcriptional and chromatin responses to severe heat stress

In order to induce transcriptional changes at specific genes, 12-day-old *Arabidopsis* seedlings that had been grown at 22ºC were subjected to heat stress at 45ºC, and subsets of plants were harvested at 16 different time points within the first 2 hours for RNA-seq experiments (see Methods). The results generated this way were highly consistent with previous microarray analysis (microarray on Col seedlings treated at 40ºC for 1hr, unpublished results from our lab). A total of 356 genes were found to be up-regulated by heat stress, which were further divided into 3 clusters based on the level of up-regulation (clusters 1, 4 and 8; Figure 2.1 A). Transcriptional activation was detected at the earliest time point following the onset of heat stress (5 min) and appeared to be completed within the first 20 minutes. In addition, 2171 genes were down-regulated (clusters 2, 3, 5 and 6; Figure 2.1 B). Compared to heat-induced genes, the decrease of transcript levels for at heat-repressed genes occurred slightly later and lasted longer (~40-50 min).

To detect chromatin changes during heat stress, H3.1 and H3.3 ChIP-seq experiments were performed at 4 time points: 0min, 30min, 60min and 120min. Under normal conditions, the distribution patterns of H3.1 and H3.3 detected here were consistent with previously published results (Stroud et al. 2012; Wollmann et al. 2012), as briefly summarized below. The level of H3.1 was found to be higher in the pericentromeric heterochromatin than the euchromatic arms of all 5 *Arabidopsis* chromosomes (Figure 2.2). In contrast, H3.3 was highly enriched in the euchromatin and depleted from the heterchromatin (Figure 2.2). At the local level, we found that H3.1 was relatively evenly distributed along transcribed regions of genes, and there was a strong negative correlation between H3.1 abundance and gene transcription level (Figure 2.3 A). In

contrast, the level of H3.3 was even and low along genes with low expression levels. At highly expressed genes, H3.3 was depleted from the 5' half but highly enriched near the 3' end (Figure 2.3 B).

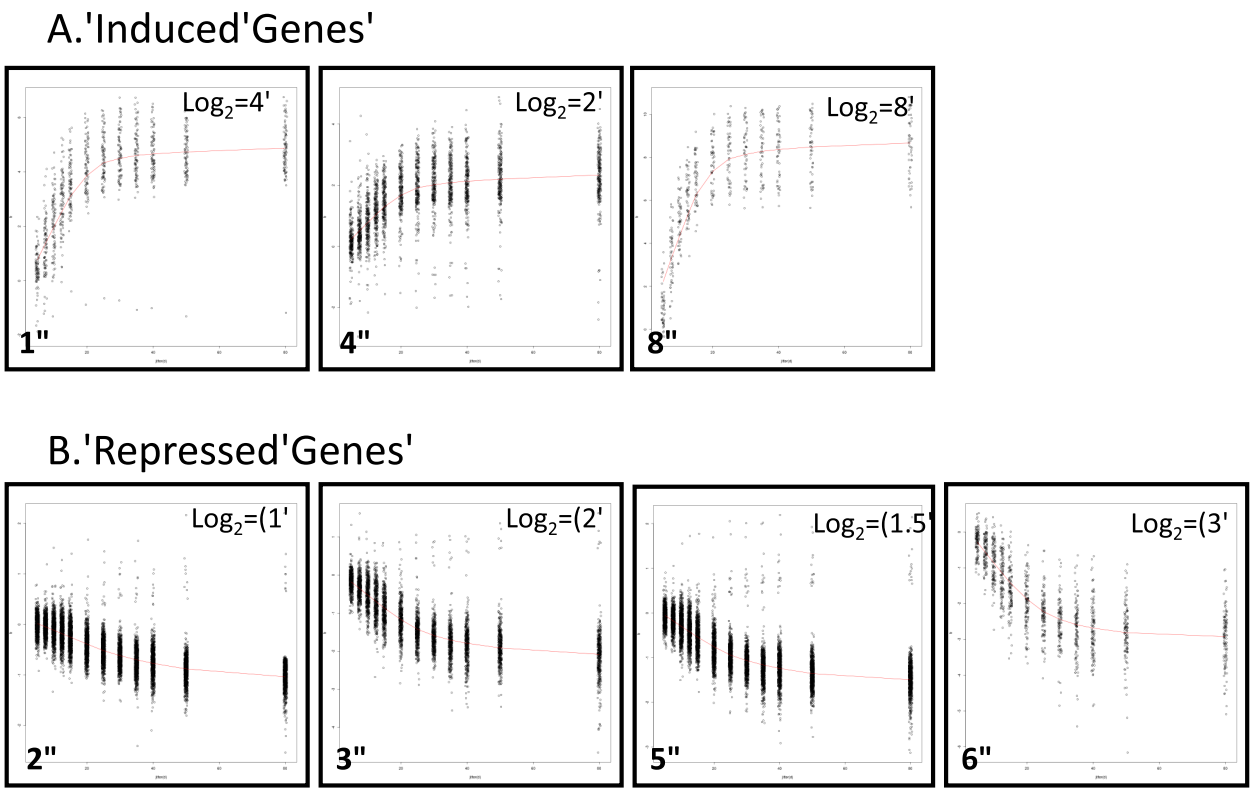

Fig. 2.1 Heat induced and heat repressed genes identified by RNA-seq. The heat responsive genes were then clustered into 8 distinct clusters based on their expression level fold change using k-means clustering, with genes from cluster 1, 4 and 8 being up regulated, while genes from 2, 3, 5 and 6 being down regulated.

During heat stress and at each of the three time points, the chromosomal distribution patterns of H3.1 and H3.3 were virtually identical to those under normal temperature. The local distribution of H3.1 and H3.3 at genes that were not affected by heat stress also remained the same (Fig. 2.4 C & F and additional descriptions below), indicating that the increase in temperature from 22°C

to 45°C did not cause disruption of the vast majority of nucleosomes in the genome. Interestingly, we also did not detect any significant change in either H3.1 or H3.3 at heat-repressed genes (Fig. 2.4 D & G), indicating that the attenuation of transcription alone may not be sufficient to convert the profiles of H3.1 and H3.3 from active to repressive. In stark contrast, we detected a significant decrease of H3.1 and H3.3 signals at genes in the most highly up-regulated cluster (Fig. 2.4 E & H), and to a less degree, at clusters 1 and 4. Notably, H3.1 and H3.3 loss was highly localized and limited to the transcribed regions as well as the immediate flanking regions of heat-induced genes. To further characterize the relationship between transcription and chromatin changes, we calculated abundance of H3.1 and H3.3 at each gene before and after heat stress, and found during heat stress, H3.1 and H3.3 loss predominantly occur at heat-induced genes.

The H3.1 and H3.3 ChIP-seq experiments were performed using transgenic *Arabidopsis* plants expressing N-terminal Myc-tagged H3.1 and H.3, respectively. Although the plants were phenotypically normal, the transcriptional responses were consistent with previous studies and the tagged histones were expressed at low levels, it was still formally possible that the H3.1 and H3.3 loss described above was an artifact caused by the epitope tag. Unfortunately, because of the high sequence similarity between H3.1 and H3.3, an antibody that could distinguish between these two variants was not available. We therefore performed an additional set of ChIP-seq experiments using a commercial antibody (ab0791; Abcam) that recognized both H3.1 and H3.3 to determine the distribution of all H3.1 and H3.3 (predominantly the endogenous H3.1 and H3.3). Similar to H3.1 and H3.3 ChIP-seq results, we found significantly lower of H3 signal at heat-induced genes, but not at heat-repressed or unaffected genes (Fig 2.4 I-K). Taken together,

these results unveiled the rapid, severe and local loss of nucleosomes specifically at heat-induced genes in the *Arabidopsis* genome that was otherwise remarkably stable at high temperature.

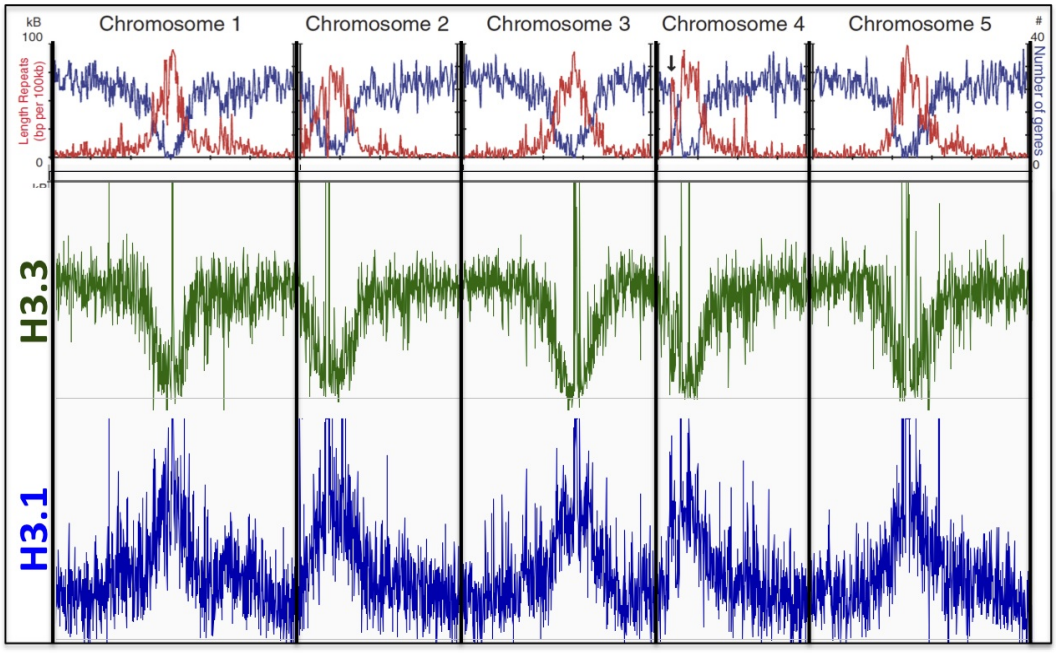

Fig. 2.2 Chromosomal distributions of H3.1 and H3.3 ChIP-seq reads under RT. Top row (adopted from (Zhang et al. 2009a)): total length of repetitive sequences (y-axis, left-side scale) and number of gens per 100kb (y-axis, right-side scale). Red line: bp of repeats per 100kb, blue line: number of genes per 100kb. Middle row: IGV view of H3.3 mapped reads. Bottom row: IGV view of H3.1 mapped reads.

Redeposition of H3.3 following the initial loss of H3.1 and H3.3

During the analyses of the heat stress data, we noted that essentially all transcription and chromatin changes appeared to have taken place in the first 30 minutes, and remained static for the remainder of the treatment (30min to 120min). It was possible that prolonged exposure of *Arabidopsis* to severe heat stress (45°C) may have compromised certain physiological process,

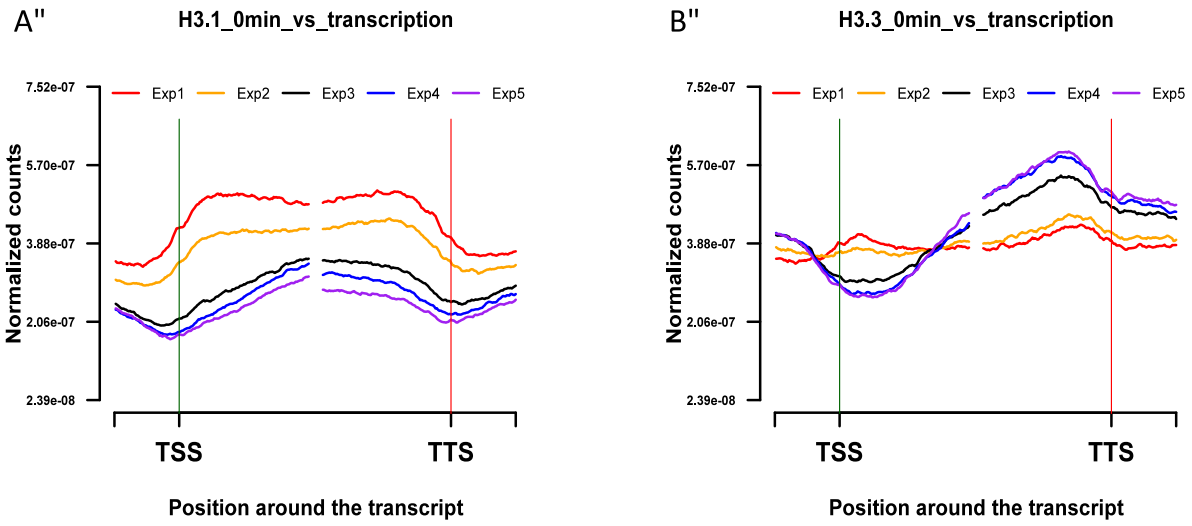

Fig. 2.3 Gene level distribution of H3.1 (A) and H3.3 (B) ChIP-seq reads under normal conditions. Correlation of H3.1 (A), H3.3 (B) and gene expression levels, where exp5 represents the most active transcribed genes and exp1 the least expressed ones. Average distribution of normalized reads in 10bp bins +500bp and -1kb of TSS, +1kb and -500bp of TTS were plotted.

such that the molecular changes that happened beyond the first 30 minutes were masked. To test this, *Arabidopsis* seedlings were treated at 45°C for various length and allowed to recover at 22°C afterwards. Seedlings that were treated for an hour or less were able to fully recover, whereas those treated for longer than an hour died. It therefore appeared that, by treating *Arabidopsis* at 45°C, we were only able to detect transcription and chromatin changes during the first 30 minutes to 1 hour.

In order to characterize later events during heat stress, we performed an additional set of experiments on *Arabidopsis* subjected to a milder heat stress at 38°C, since seedlings could be

treated under this condition for at least 8 hours and still recover. RNA-seq and ChIP-seq experiments were performed at 7 time points during the first 8 hours of heat stress (see methods).

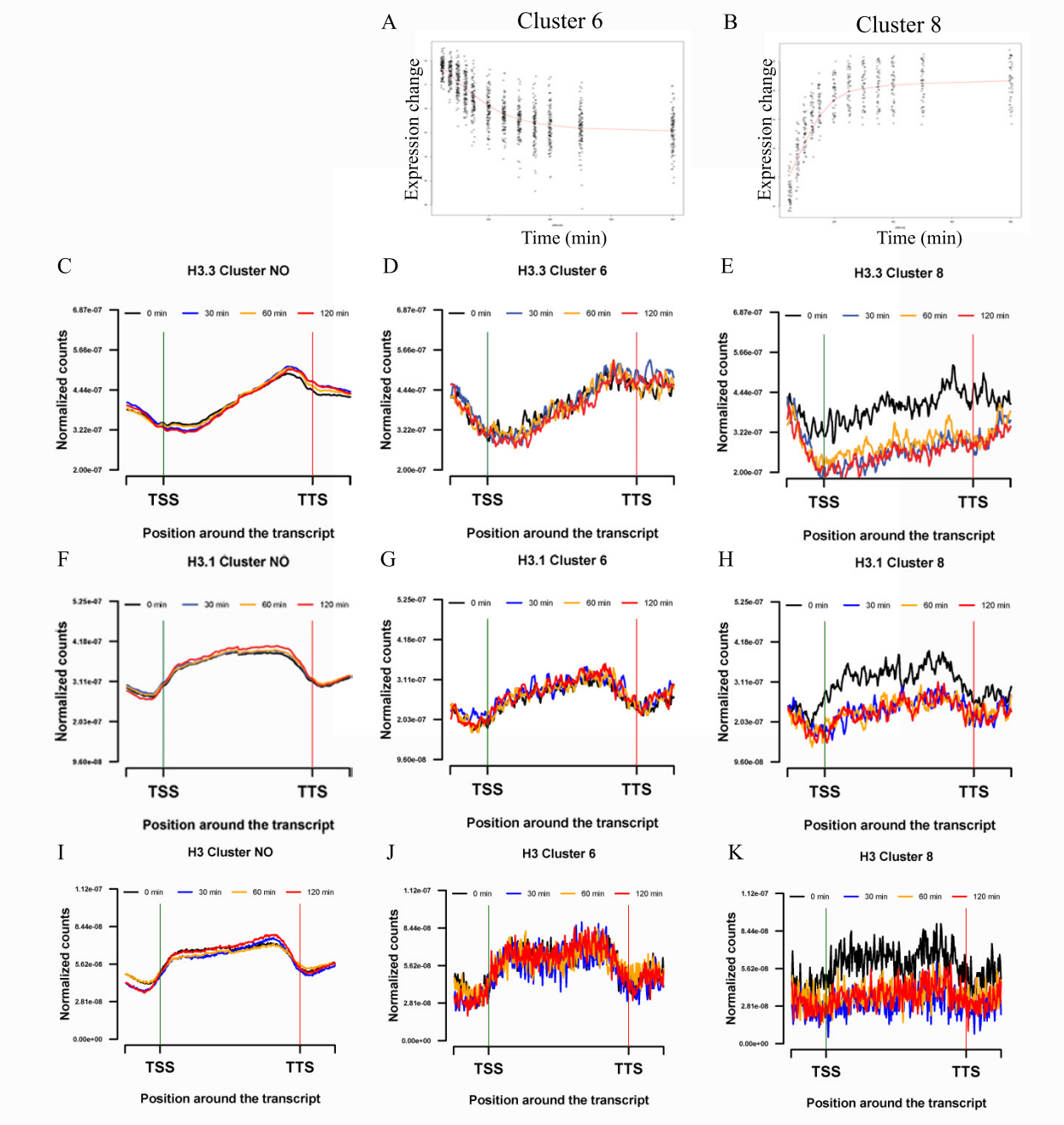

Fig. 2.4 Total H3, H3.1 and H3.3 changes in respond to transcriptional changes. A&B: transcriptional changes for a representative down regulation cluster 6 and an up regulation

cluster 8, respectively. C-E: H3.3 distribution changes over non-heat inducible genes, genes in cluster 6, and cluster 8; F-H: H3.1 distribution changes over non-heat inducible genes, genes in cluster 6, and cluster 8; I-K: total H3 distribution changes over non-heat inducible genes, genes in cluster 6, and cluster 8.

Comparison of the RNA-seq results at 45°C and 38°C showed that largely the same set of genes were up- or down-regulated, and that the highest level of transcripts were detected at around the same time (~30min; Figure 2.5). However, whereas transcript levels of heat-induced genes remained high for the remainder of the 45°C treatment, they decreased between the 30min and 2h at 38°C, to levels that were ~10-20 folds higher than before heat stress (compared to ~250x at 30 min). Such a decrease in transcript levels was consistent with previously described transcriptional attenuation following the initial heat induction as well as the rapid degradation of mRNAs (processes not detected in the 45°C experiments).

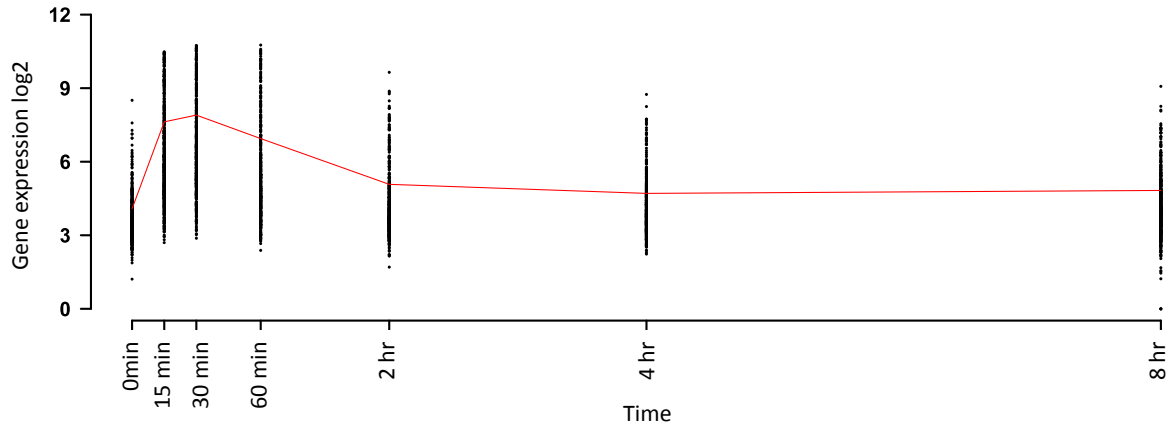

Fig 2.5 Expression pattern for heat induced genes at 38°C. Scatter plot of log2 (RPKM) for induced genes at 38°C shows transcriptional attenuation following the initial heat induction as well as the rapid degradation of mRNAs.

H3.1 ChIP-seq results at 38°C were remarkably similar to those at 45°C, in that a severe loss was detected at the first time point (15 min, Fig. 2.6 A), and that H3.1 levels remained at the basal level throughout the remainder of the 8-hour treatment (Fig. 2.6 B-F). No significant difference in H3.1 was detected at heat-repressed genes or those unaffected by heat stress. Considering the replication-coupled nature of H3.1 deposition, it was not surprising to find that H3.1 did not recover at heat-induced genes, since the vast majority of the cells did not undergo cell divisions during the heat stress.

Similar to results from the 45°C treatment, we also observed a rapid and severe loss of H3.3 during at the first time point of the 38°C treatment (15 min Fig. 2.7 A). However, in contrast to the 45°C results where H3.3 remained low, we found that H3.3 began to recover at the 2 hour time point to an intermediate level (Fig. 2.7 D), and fully recovered at 4 hours (Fig. 2.7 E). The eventual level of H3.3 was slightly higher than that before heat stress, perhaps to compensate the loss of H3.1 at heat-induced genes (Fig. 2.7 E-F).

Taken together, the results from the 38°C treatment not only validated the rapid histone loss during transcriptional activation observed at 45°C, but also showed that H3.3 (but not H3.1) was redeposited during transcriptional attenuation.

Newly synthesized histones were re-deposited following heat-induced nucleosome eviction

The disappearance of both H3.1 and H3.3 upon heat stress and the recovery of H3.3 beginning at ~2 hours post stress are consistent with the rapid loss of all "original" H3 (and presumably other histones) at heat-induced genes, followed by the deposition of newly synthesized H3.3. It is also possible, however, that the chromatin of heat-induced genes may adopt an intermediate

conformation (e.g. transient dissociation of histones from DNA) during transcriptional activation, which could interfere with the ChIP procedure (e.g., more difficult to crosslink). During transcriptional attenuation, the chromatin state returned to normal, resulting in the apparent redeposition of histones. It should be noted that the reappearance of only H3.3 (but not H3.1) is inconsistent with this latter scenario.

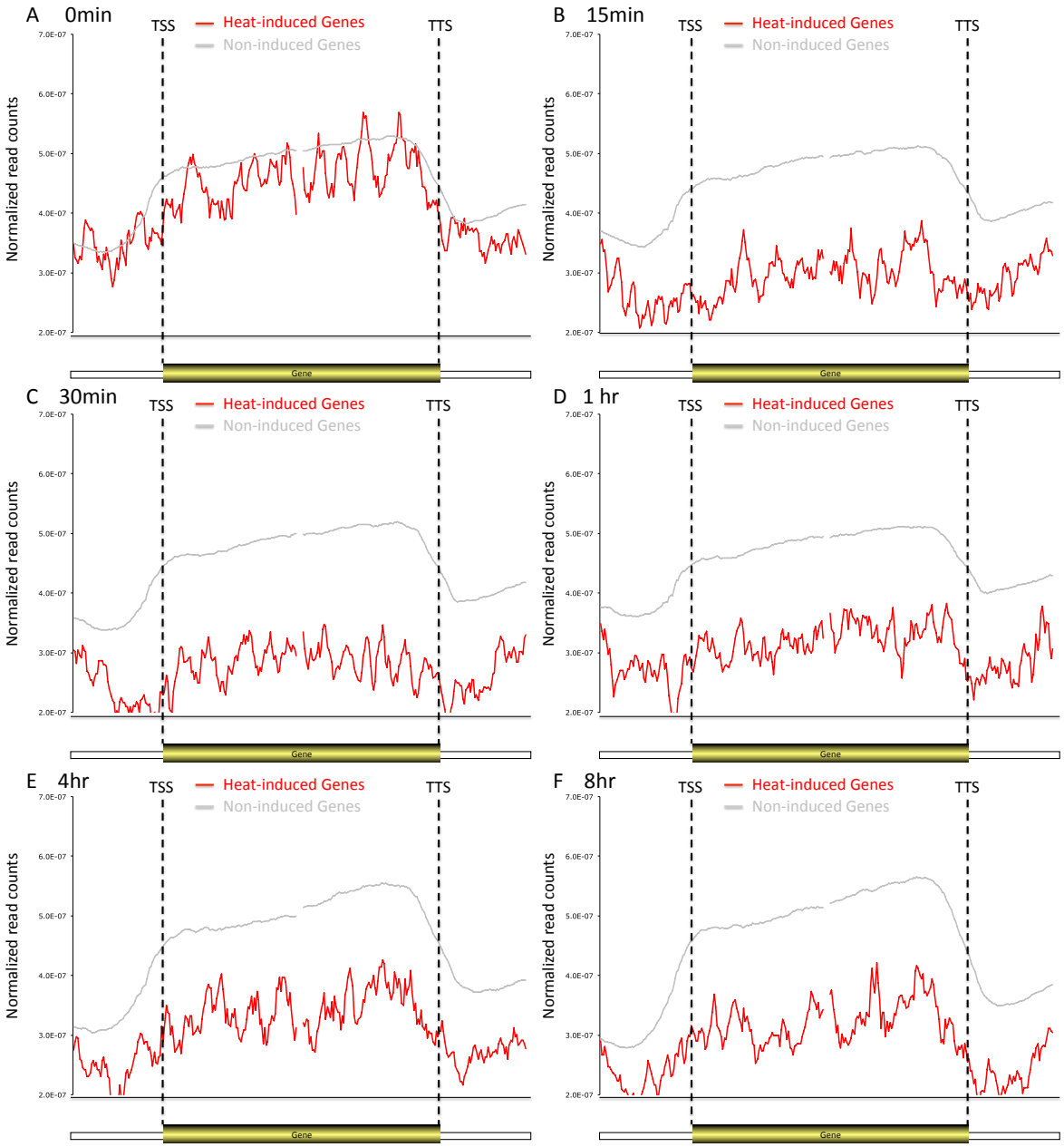

Fig. 2.6 Distribution of H3.1 over heat induced genes and non-induced genes at various time points treated at 38°C.

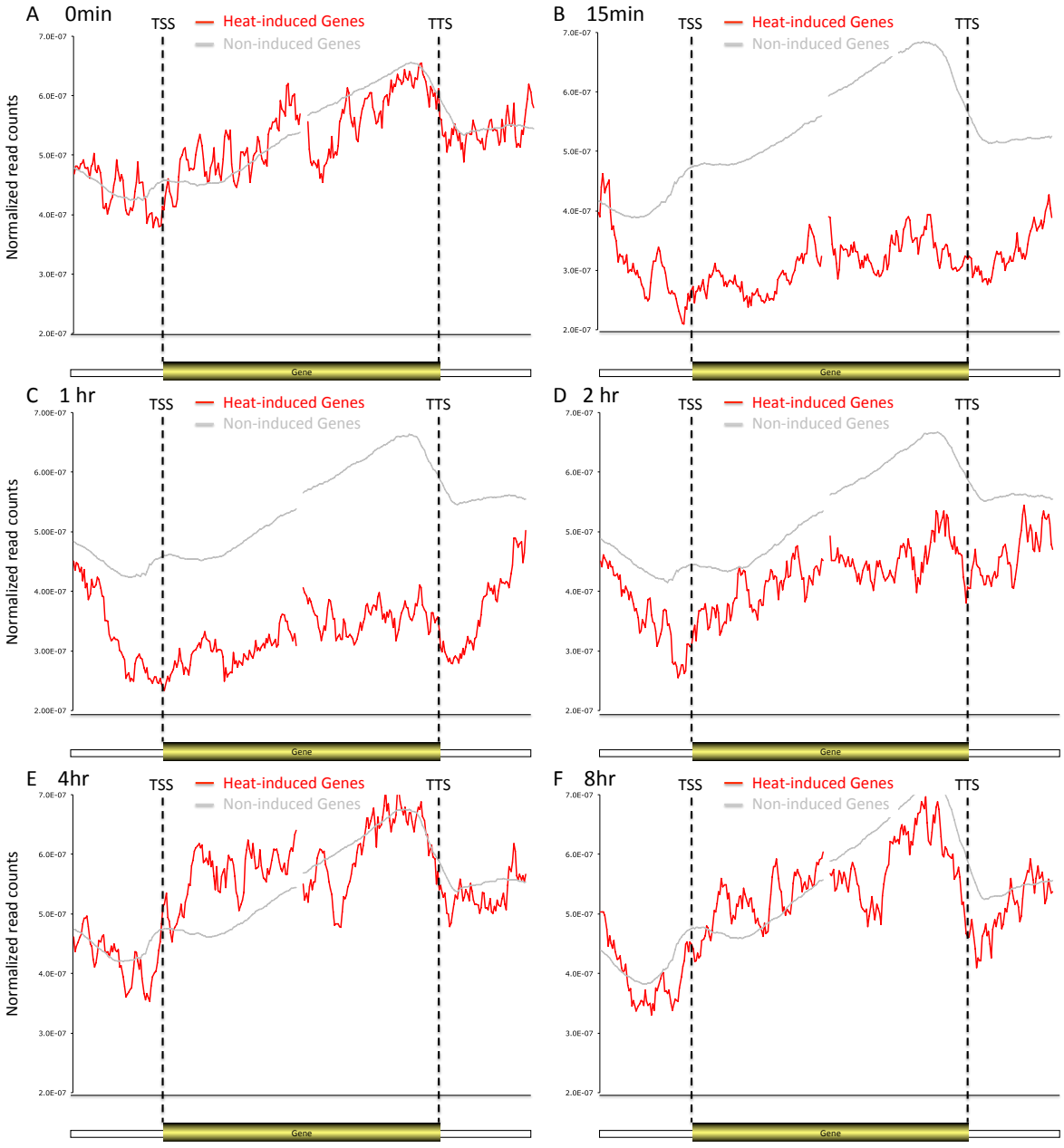

Fig. 2.7 Distribution of H3.3 over heat induced genes and non-induced genes at various time points treated at 38°C.

The following experiment was performed to further address whether the apparent loss of H3.1 and H3.3 and the recovery of H3.3 were caused by histone replacement or transient chromatin conformation changes. Previous studies in fungal and animal systems have shown that many covalent histone modifications are established after histones are deposited into the chromatin (Verreault et al. 1998; Masumoto et al. 2005; Loyola et al. 2006; Das et al. 2009; Campos et al. 2010). One such example is H3K36me3, which is catalyzed by the SET2 family of histone methyltransferases in a transcription-coupled manner (Du et al. 2008). That is, SET2 interacts with the elongating RNA polymerase II complex (Pol II) to methylate H3K36 on histones within the transcribed regions. Due to the basal level of transcription under normal temperature, heat-inducible genes already contained H3K36me3 prior to heat stress. We reasoned that, if the transient chromatin conformation change scenario was correct, one would expect H3K36me3 to recover at the same time as H3.3. In contrast, if the histone replacement scenario was correct, recovery of H3K36me3 should occur slower than that of H3.3. To directly test this, we performed H3K36me3 ChIP-seq experiments throughout the heat stress, using the same chromatin preparations as those for H3.1 and H3.3 ChIP. We found that, similar to H3.1 and H3.3, H3K36me3 was present at heat-inducible genes before heat stress (0min, Fig. 2.8), but disappeared at the first time point (15 min, Fig. 2.8). Importantly, while H3.3 signals began to recover at ~2 hour post stress (Fig. 2.7 D), H3K36me3 signals remained low during the entire treatment period of 8 hours (Fig. 2.8). Whether and how H3K36me3 would eventually be reestablished remain interesting and exciting questions. Regardless, the lack of H3K36me3 on H3.3 during the recovery period provided strong support for the notion that H3.1 and H3.3 were lost upon heat stress, and later replaced by newly synthesized H3.3 during transcriptional attenuation.

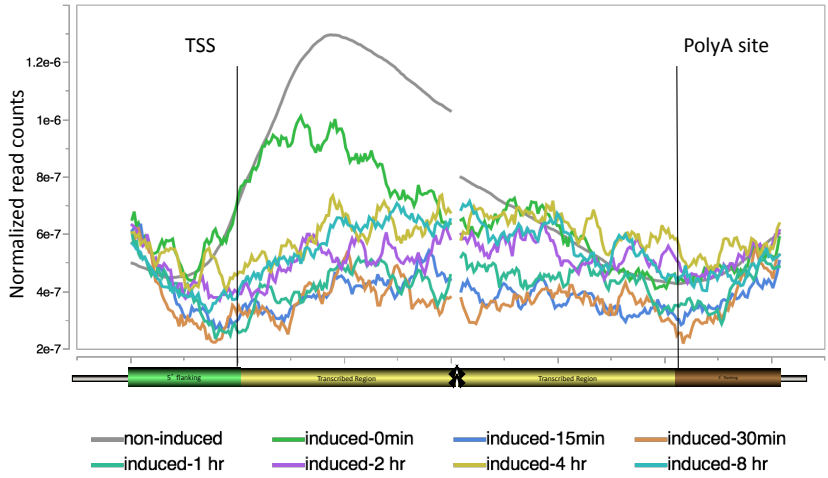

Fig. 2.8 Distribution of H3K36me3 over heat induced genes at various time points treated at 38°C and non-induced genes at normal conditions as comparison.

Thermostability of H2A- and H2A.Z-containing nucleosomes

In addition to the variants for histone H3, *Arabidopsis* also expresses several H2A variants, including the canonical H2A, H2A.W, H2A.X and H2A.Z (Talbert et al. 2012). H2A.W is located in the heterochromatin and involved in chromatin condensation (Yelagandula et al. 2014), whereas H2A.X marks the sites of double-stranded DNA breaks (Lang et al. 2012). H2A.Z is broadly distributed in genes, and mutants that are deficient in the production or incorporation of H2A.Z display severe developmental defects (Zilberman et al. 2008; Coleman-Derr and Zilberman 2012). The molecular function of H2A.Z is not yet clear, partly due to the fact that, although H2A and H2A.Z differ substantially in amino acid sequences, the overall structures of H2A- and H2A.Z-containing nucleosomes appear to be highly similar.

A simple and elegant model has been proposed to explain how H2A.Z may be involved in heat-induced chromatin and transcription changes (Kumar and Wigge 2010). In this model, H2A.Z-containing nucleosomes may be more stable than H2A-containing nucleosomes. As the

temperature rises, H2A-containing nucleosomes (but not H2A.Z-containing nucleosomes) dissociate from heat-inducible genes, which allows the transcription activation of these genes. This model is supported by the following findings: 1) in the *arp6* mutant defective in the deposition of H2A.Z into the chromatin, some heat-inducible genes are overexpressed under normal temperature; 2) overexpression of one such gene in *arp6* is correlated with lower nucleosome occupancy at the 5' end; 3) H2A.Z-containing nucleosomes appear to have higher thermostability than H2A-containing nucleosomes in *in vitro* experiments.

The overall chromatin stability and the dynamic changes at specific genes during heat stress described above appear to be inconsistent with the model by Kumar. To further characterize any potential differences between H2A- and H2A.Z-containing nucleosomes *in vivo*, we first determined the genome-wide distribution of H2A.Z. A monoclonal antibody against the endogenous H2A.Z (Deal et al. 2005; Deal et al. 2007) was used in ChIP experiments using the aerial part of *Arabidopsis* seedlings at the same developmental stage as those used in heat-shock experiments, and the resulting ChIP sample was analyzed by high-throughput sequencing. H3 ChIP-seq was performed using the same materials to control for nucleosome occupancy. In addition, both H2A.Z and H3 ChIP-seq experiments were also performed in the *arp6* mutants to control for antibody specificity.

At the chromosomal level, H2A.Z was predominantly found in the gene-rich euchromatin and depleted from the transposon-rich heterochromatin. This distribution reflected the fact that H2A.Z was located at virtually all endogenous genes, but absent from intergenic regions and transposons or other repeats. Comparison of genes with different expression levels unveiled two major types of H2A.Z distribution: at active genes, H2A.Z was primarily found near the 5' end

and absent from the 3' half; at repressed genes, H2A.Z level remained high throughout the entire transcribed regions (Fig. 2.9 A). These results are consistent with those from previous studies in

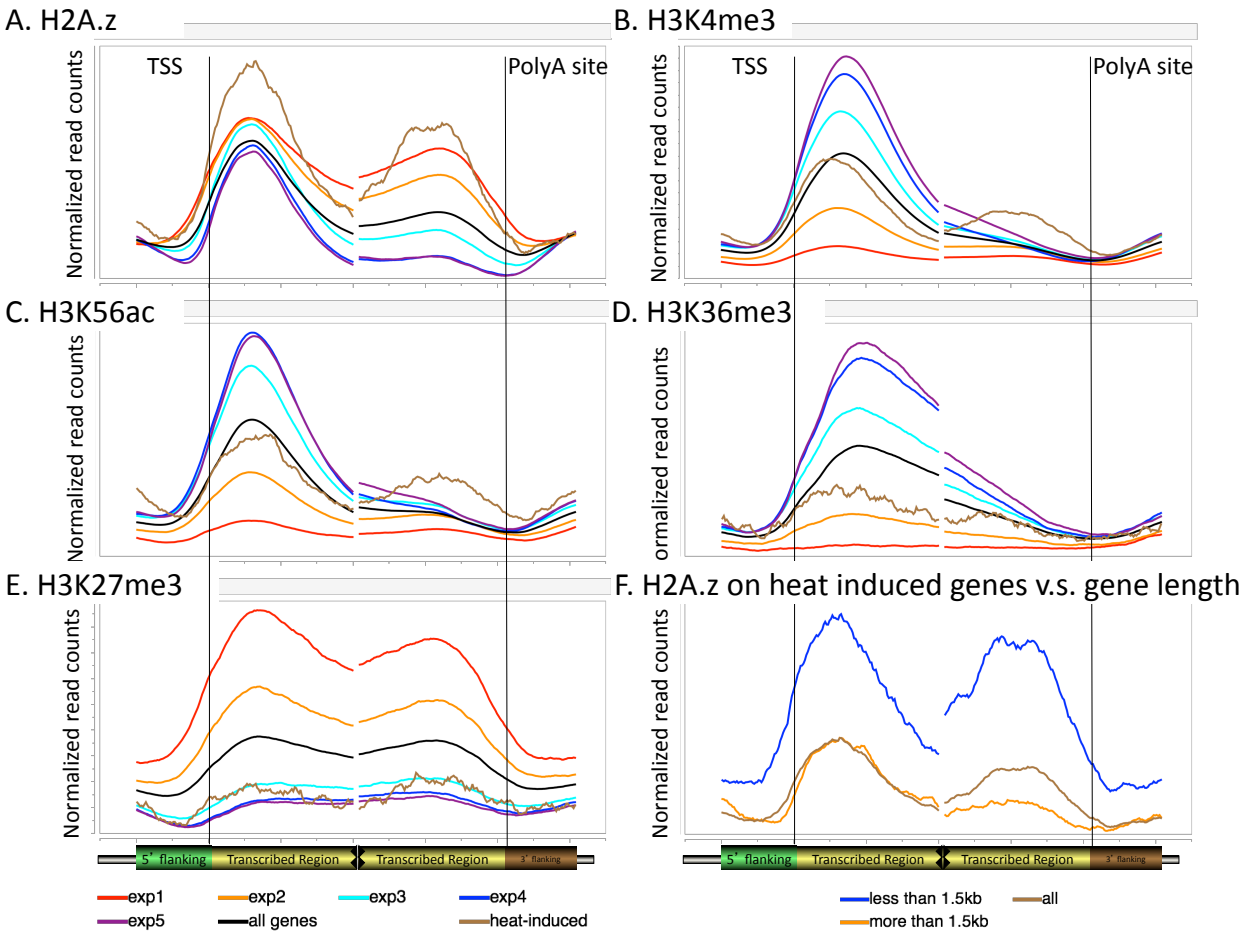

Fig. 2.9 Heat induced genes have normal perhaps more H2A.Z and other histone modifications under normal conditions. Distribution of H2A.Z and four other histone modifications A-E (H3K4me3, H3K56ac, H3K36me3 and H3K27me3) on heat induced genes (brown) are plotted, and their correlation with gene transcription levels are shown, where exp5 represents the most active transcribed genes and exp1 the least expressed ones. F. Distribution of H2A.Z on heat induced genes according to gene length. Average distribution of normalized reads in 10bp bins +500bp and -1kb of TSS, +1kb and -500bp of TTS were plotted.

Arabidopsis roots (Zilberman et al. 2008; Zemach et al. 2010), and the differential distribution of H2A.Z at active and repressed genes indicates that H2A.Z may play multiple roles in regulating gene expression.

If H2A.Z-containing nucleosomes are indeed more stable at high temperature than H2A-containing nucleosomes, one may expect heat-induced nucleosome loss to preferentially occur in regions depleted of H2A.Z. Furthermore, genes that undergo nucleosome loss upon heat stress (and only these gens) should contain minimal levels of H2A.Z throughout their entire length. However, neither of these predictions is consistent with our H3.1 and H3.3 ChIP-seq results during heat stress. First, we found no detectable heat-induced nucleosome loss (i.e. lower levels of H3.1 or H3.3 following heat stress) in intergenic regions that were H2A.Z-free. Similarly, despite the absence of H2A.Z in the 3' half of active genes, there was no detectable nucleosome loss from the 3' half of the vast majority of active genes upon heat stress (with the exception of the heat-inducible genes described earlier). Second, we found high levels of H2A.Z at genes that underwent heat-induced nucleosome eviction and redeposition (Fig. 2.9 A). Taken together, these results are inconsistent with the notion that H2A- and H2A.Z-containing nucleosomes have different thermostabilities *in vivo*. The presence of H2A.Z (instead of H2A) is neither necessary nor sufficient to render nucleosomes more stable at higher temperature.

In addition to H2A.Z, we also examined the distribution of several important histone modifications at heat-inducible genes prior to heat stress, including H3K4me3, H3K27me3, H3K36me3 and H3K56ac (Fig. 2.9 B-E). In each case, the modification profiles at heat-inducible genes were found to be similar to other genes expressed at comparable levels. Collectively, these results suggest that the rapid nucleosome eviction and redeposition at heat-

inducible genes occur without any chromatin-based predisposition in the form of chromatin composition or modification.

Relationship between histone replacement and transcription

The dynamic histone replacement at heat-induced genes is a novel and unexpected finding that raises a number of questions. Of particular interest is the relationship between the histone replacement process and transcription. Does histone eviction precede transcriptional activation to allow high levels of transcription, or is it a consequence of transcriptional activation? Similarly, does histone redeposition facilitate transcriptional attenuation, or is it only possible when transcriptional activity decreases to a certain level? A better understanding of the sequence of events (i.e. transcriptional activation and repression; histone eviction and redeposition) should provide important information to address these questions. Unfortunately, although detailed time course RNA-seq data were available, the increase of transcript level lags behind transcriptional activation and it is further complicated by gene-specific RNA degradation rates.

To better understand the transcription activities of heat-inducible genes during heat stress, we directly characterized the genome-wide occupancy of Pol II at 7 time points using the same materials as for H3.1/H3.3 ChIP-seq. An antibody against the non-phosphorylated form C-terminal domain (CTD) of Pol II (ab26721; Abcam) was used for CHIP, and the resulting samples were analyzed by high-throughput sequencing.

At the individual gene level, Pol II signals should be positively correlated with the number of Pol II complexes involved in active transcription, and inversely correlated with the movement speed of Pol II at the corresponding region. Under normal temperature, Pol II ChIP-seq signals were positively correlated with transcript abundance, although the dynamic range of Pol II signals was

far lower than that of transcript abundance (Fig. 2.10 A). Interestingly, we found that Pol II appeared to move much slower at the 3' half of genes, particularly near the polyadenylation site, perhaps to facilitate transcription termination and 3' RNA processing.

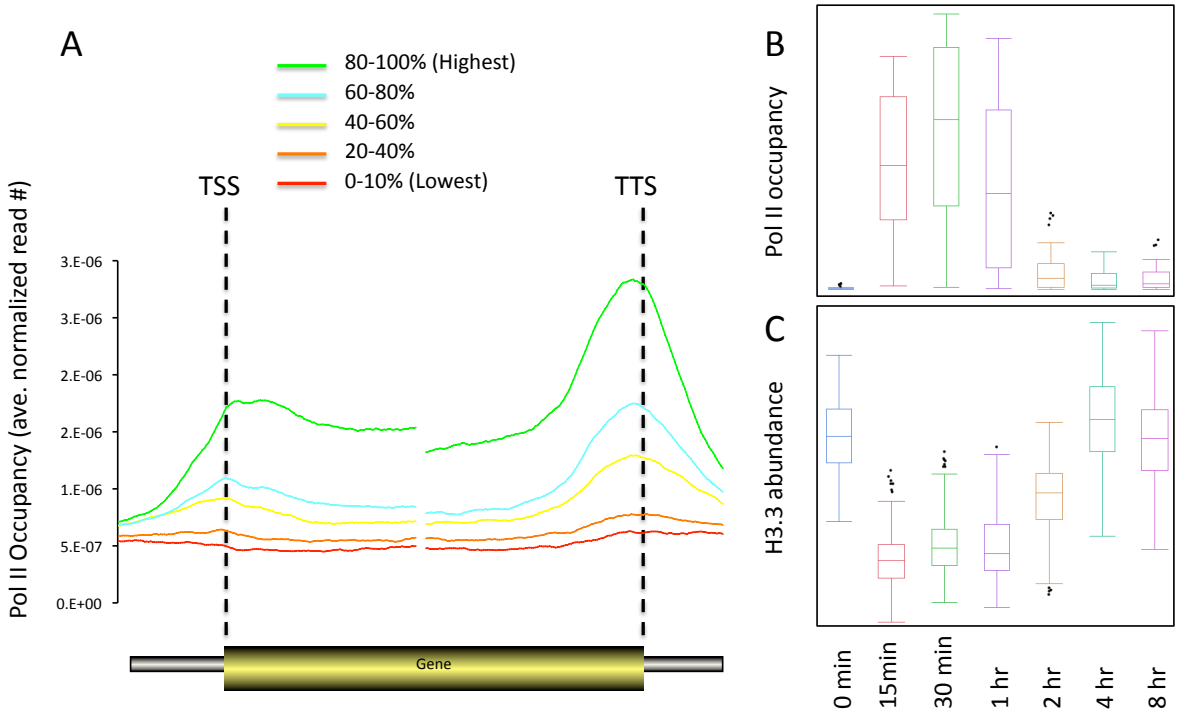

Fig. 2.10 Relationship between histone replacement and transcription. A. Correlation between Pol II occupancy and gene transcription levels is shown. B. Pol II occupancy over heat induced genes during heat treatment at 38°C. C. H3.3 abundance over heat induced genes during heat treatment at 38°C.

Heat stress led to rapid and extraordinarily high levels of Pol II recruitment to heat-induced genes. As shown in Figure 2.10 B, Pol II recruitment was nearly complete at the first time point (15 minute), and Pol II level remained high for the first hour. Importantly, Pol II level was drastically reduced at the 2 hour time point and remained low for the remainder of the heat

treatment (Fig. 2.10 B). Comparison of Pol II ChIP-seq and RNA-seq results showed that changes in Pol II occupancy indeed preceded changes in transcript levels. Taken together, these results indicated that the vast majority of transcripts from heat-induced genes were produced by a single "burst" of transcription that lasted for approximately an hour.

Comparison between changes in Pol II occupancy and histone abundance at heat-induced genes suggested that histone loss might occur slightly earlier than Pol II recruitment. As shown in Figure 2.10 B & C, histone eviction was complete in 15 minutes, whereas the highest level of Pol II occupancy was achieved at the 30 minute time point. It should be noted that the timing difference was subtle and, although it appeared more likely that histone eviction might precede Pol II recruitment, the opposite scenario that high level of Pol II might cause histone eviction could not be ruled out. The timing difference during transcriptional attenuation was more pronounced: Pol II level decrease was complete before the 2 hour time point, whereas H3.3 redeposition continued to occur until the 4 hour time point. It therefore appeared most likely that H3.3 redeposition could only occur after the transcription burst.

Dynamic Pol II redistribution during heat stress

The level of Pol II occupancy at heat-induced genes during the transcription burst was extraordinarily high. For example, the top 1% of heat-induced genes had 6.6% of all Pol II signals under normal temperature, but they accounted for 27.8% of all Pol II signals in the genome at the 30 minute time point. The Pol II complex is large with dozens of subunits, and the high level of Pol II at heat-induced genes only lasts for approximately an hour. It is therefore difficult to imagine a scenario in which an enormous amount of Pol II complexes (equivalent to >20% of total) are synthesized and assembled within minutes, and then degraded after the

transcription burst. Alternatively, existing Pol II complexes engaged in the transcription of other genes under normal temperature could be transiently redistributed to heat-induced genes upon heat stress to facilitate the transcription burst (the "redistribution model").

To test the redistribution model, we first determined the Pol II occupancy for each gene at all 7 time points. This dataset was then used for tight clustering analysis to identify groups of genes that shared the same pattern of Pol II occupancy changes. As expected, heat-induced genes formed a cluster that exhibited low-high-low Pol II occupancy during heat stress (Fig. 2.11 A).

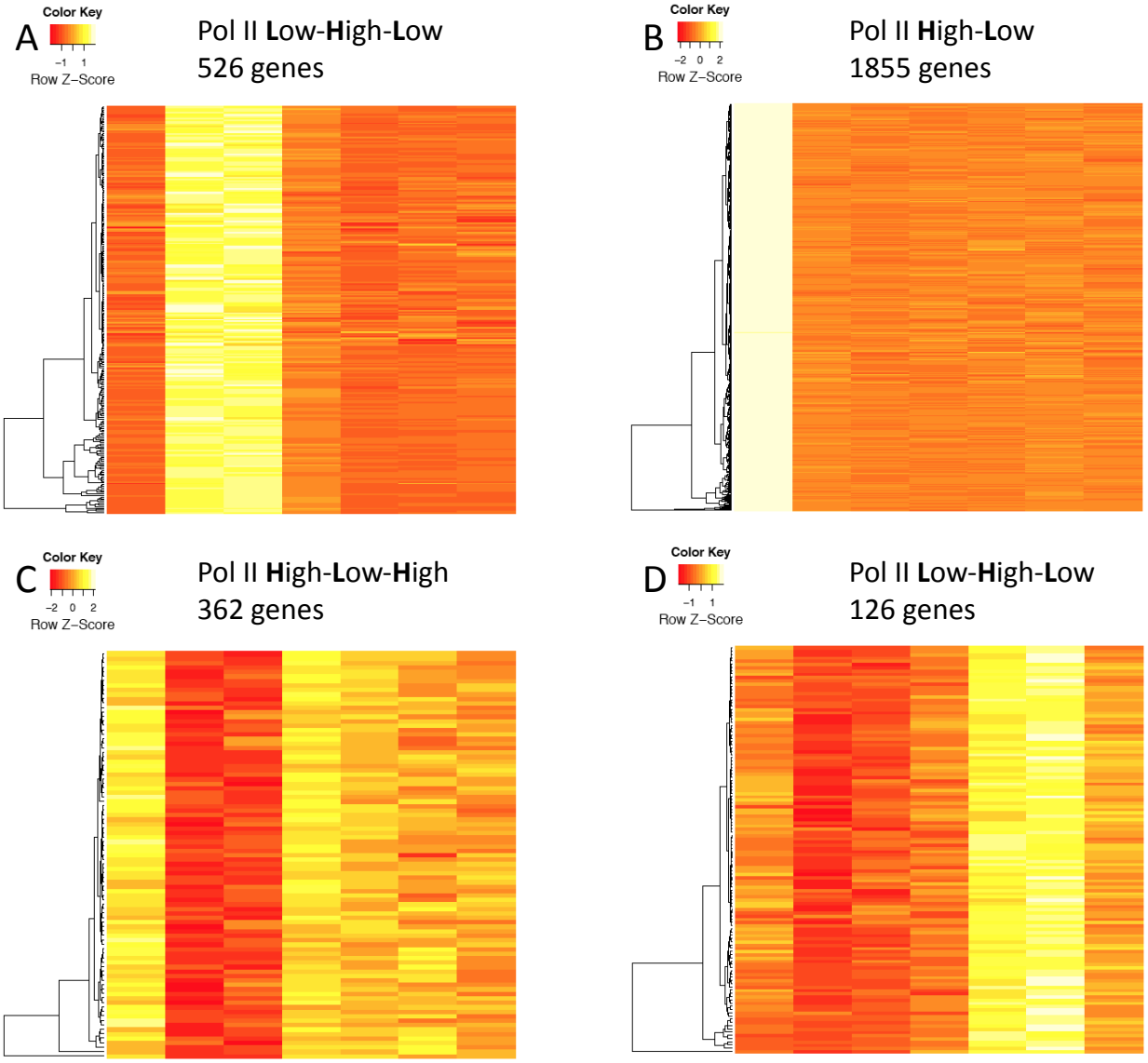

Fig. 2.11. Heat map of dynamic Pol II re-distribution during heat treatment at 38°C.

Two clusters of genes were identified as potential Pol II "donors". The first cluster ("HL cluster"; Fig. 2.11 B) had high levels of Pol II before heat stress, but rapidly lost Pol II upon heat stress and for the entire 8-hour treatment. The second cluster ("HLH cluster"; Fig. 2.11 C) also had high levels of Pol II before heat stress and lost Pol II upon heat stress. In contrast to the HL cluster, Pol II loss from the HLH cluster was transient and only lasted ~1 hour. Note that the Pol II occupancy changes in the HLH cluster were precisely anticorrelated with heat-induced genes. Interestingly, GO analysis showed that the HL cluster was highly enriched for genes involved in photosynthesis (p=2.3E-34), and HLH cluster was weakly enriched for genes involved translation (p=1.2E-3).

In addition to potential Pol II donors that lost Pol II during the early stage of heat stress, we also identified other groups of genes with interesting Pol II changes. For example, a cluster of 126 genes ("LHL cluster"; Fig. 2.11 D) had low Pol II signal under normal temperature and during the early stage of heat stress, but gained a very high level of Pol II from 2-4 hours, followed by a decrease of Pol II level at the 8 hour time point. Note that the timing of Pol II recruitment to LHL genes coincided with the loss of Pol II from heat-induced genes.

The results described above are consistent with a scenario in which both HL and HLH genes contribute Pol II to heat-induced genes to support the early transcription burst. Follow transcriptional attenuation, some Pol II from heat-induced genes return to HLH genes, and some pol II is recruited to LHL genes for transcriptional activation during the late stage of heat stress. Future studies should address the important question of how the specificity of Pol II redistribution is achieved.

Does histone replacement provide a transcriptional "memory" of heat stress?

As described above, both H3.1 and H3.3 are evicted upon the activation of heat-inducible genes, and newly synthesized H3.3 is deposited during transcriptional attenuation. This process leads to a change in chromatin composition that can persist in cells that do not undergo further divisions, since H3.1 is only synthesized and deposited during the S phase. In other words, a single round of heat treatment may cause permanent chromatin changes at a specific set of genes in the majority of cells in *Arabidopsis*. It is interesting to consider that such a chromatin composition change may affect the inducibility of the corresponding genes upon the next round of heat stress, thereby providing a transcriptional memory.

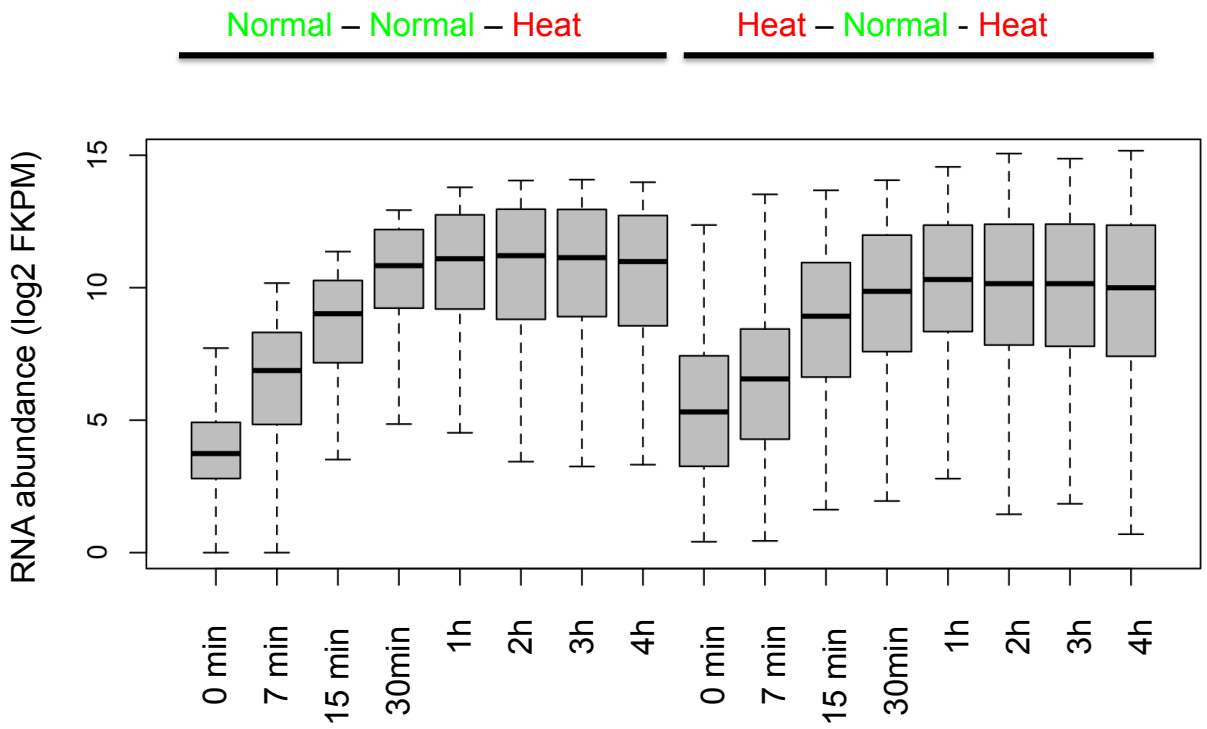

Fig. 2.12. RNA-seq analysis for heat induced genes under NNH or HNH treatment at various time points.

To test this possibility, we compared the dynamics of transcriptional activation of heat-inducible genes in plants with or without being exposed to prior heat stress. Specifically, *Arabidopsis* plants were grown on plates at 22°C for 12 days. Half of the plants ("heat-normal-heat; HNH group") were subjected to 38°C treatment for 4 hours to allow histone replacement, and recovered at 22°C for 24 hours, whereas the other half ("normal-heat; NH group") were kept at 22°C. Both groups were then treated at 38°C, and the transcriptional responses were assessed by RNA-seq at 7 different time points during the first four hours of heat stress.

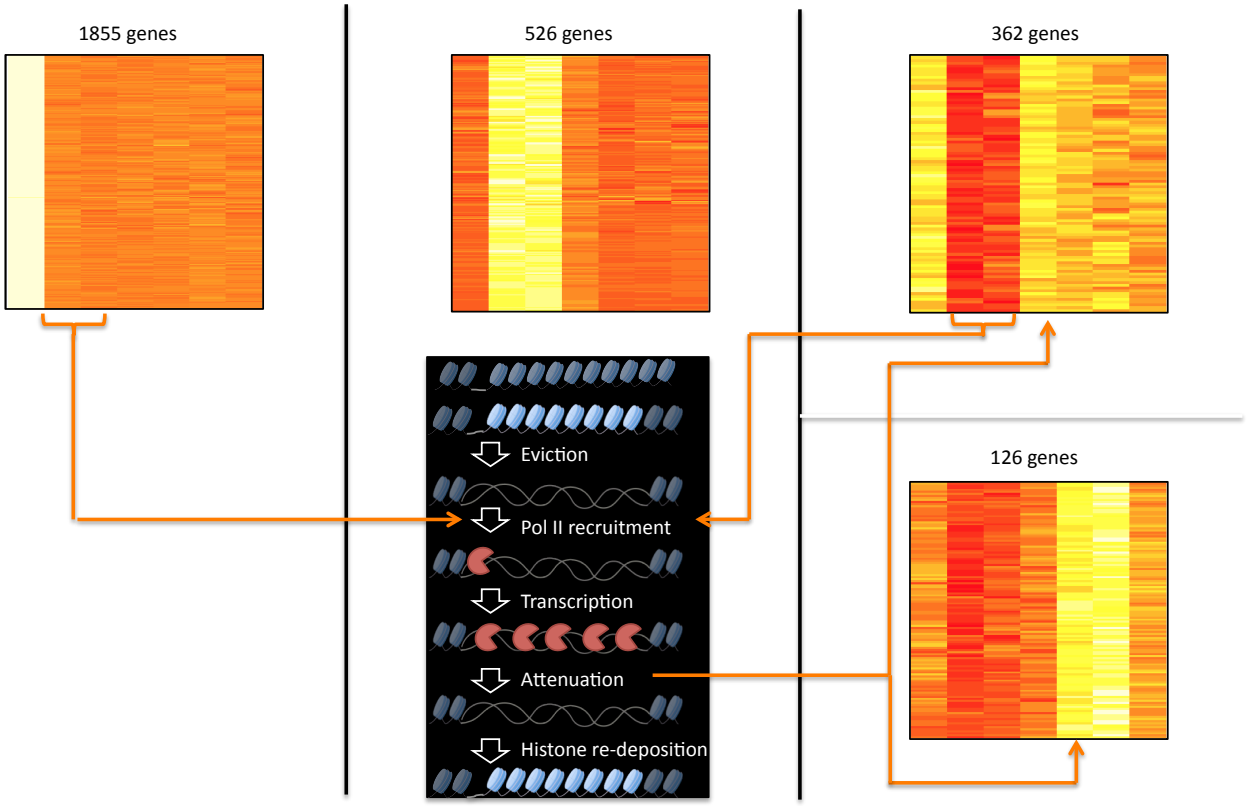

Fig. 2.13 A schematic model demonstrating dynamic histone replacement and Pol II re-localization during heat induced transcriptional activation in *Arabidopsis thaliana*.

If heat-induced histone replacement facilitates transcriptional response to the next round of heat stress, one may expect the heat-inducible genes to be activated more rapidly in HNH plants than

in NH plants, although the peak levels may be similar. Interestingly, we found that the transcript level of heat-inducible genes to be significantly higher in HNH plants than in NH plants before they were subjected to the second heat stress (Fig. 2.12). It is possible that transcription attenuation of heat-induced genes following the first round of heat stress was incomplete. It is also possible that a fraction of the transcripts produced during the first round of heat stress persisted for more than 24 hours. Importantly, the transcript levels of heat-inducible genes became virtually identical in HNH and NH plants at the first time point after heat stress (7 minutes), and remained highly similar throughout the four-hour treatment period (Fig. 2.12). It therefore appears that, although the transcripts of heat-inducible genes were present at an elevated level after the first heat stress, histone replacement at these genes had very little effect on the transcriptional activation in response to the second heat stress.

2.3 Conclusions

In summary, the data presented here provided a comprehensive and high-resolution view of the transcriptional and chromatin-based events during heat stress. There was no evidence that heat-inducible genes were predeposited with any unusual chromatin state, or that H2A.Z may affect nucleosome theromostability in planta. We found that *Arabidopsis* responded to heat stress with rapid and high-level recruitment of Pol II to heat-inducible genes, which was concomitant with the loss of preexisting histones. Transcription proceeded on nucleosome-free template DNA for approximately an hour, followed transcriptional attenuation and redeposition of newly synthesized histones. Importantly, large-scale redistribution of Pol II among specific sets of genes appeared to have contributed to the rapid transcriptional activation upon heat stress. The dynamic histone replacement and Pol II redistribution have not been described during the transcriptional activation of developmentally controlled genes. For example, a recent study of the

developmental transition between dark and light growth during germination showed no gross chromatin changes at activated genes (Sullivan et al. 2014). It is interesting to consider that both histone replacement and Pol II redistribution may be specific to processes involving rapid transcriptional activation to extraordinarily high levels. In this regard, we note that many other abiotic stresses are also capable of inducing transcriptional changes in manners similar to heat. Future studies should determine whether the histone replacement and Pol II redistribution observed here are shared by all stress-response processes.

2.4 Materials and Methods

Plant materials and heat treatment

Myc-tagged H3.1 and H3.3 seeds are generously given by the Gutierrez lab (Stroud et al. 2012). The presence of myc tag was confirmed by western blot with individual plants using anti-myc (clone 4A6, Millipore). Seeds for myc-H3.1 and myc-H3.3 were separately pooled, each about 40ml in volume. All experiments were done with the same generation seedlings from those seeds. Seeds were sterilized and imbibed in water at 4°C in darkness for 3 days, grown on MS plates at 22ºC under continuous light conditions for 12 days. Then the plates were moved to 45ºC in growth chamber for 0min, 5min, 7.5min, 10min, 12.5min, 15min, 20min, 25min, 30min, 35min, 40min, 50min, 60min, 80min 100min and 120min. To eliminate uncontrollable changes the plates were placed in liquid nitrogen immediately after each treatment. This allowed us to quickly harvest the aerial part of seedlings by using a razor blade. Or the plates were moved to 38ºC in growth chamber for 0min, 15min, 30min, 60min, 2hr, 4hr and 8hr, and aerial part of seedlings were harvested as described above. Two plates were collected for each time point.

X-linking frozen tissue

Mix 7.7mL GB buffer (0.4M sucrose, 10mM Tris 8.0 and 1mM EDTA 8.0) with 80uL 0.1M PMSF and 216uL 37% formaldehyde in a 50mL beaker for each sample. About 0.5g frozen tissue for each time point was quickly dumped in the beaker and incubated for 5min under vacuum (Savant gel pump) at room temperature. Release vacuum and gently mix, then incubate for 5min under vacuum again. Release vacuum and add 800uL 1M glycine (BP381-1), gently mix and incubate for 5min under vacuum to stop X-linking. Quickly rinse tissue with ddH_2O for 3 times. Remove excess moisture by letting the tissue surface dry on regular facial tissue. Freeze tissue in liquid nitrogen, proceed to ChIP assay or store at -80°C.

ChIP-seq library generation

Grind the X-linked tissue in liquid nitrogen to a fine powder; add the powder to 10 volumns of Extraction Buffer 1 (0.4M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl, 0.1mM PMSF, 2 Roche mini tablets/ 100 ml, 1mM EDTA) in a 50 ml falcon tube. Leave on ice for 5 min. Filter the solution through Miracloth into a 50 ml falcon tube. Repeat for a total of two filters, and then spin the filtered solution for 20 min at 4000 rpm in a centrifuge at 4 degrees. Gently remove supernatant and resuspend the pellet in 1 ml of Extraction Buffer 2 (0.25M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl2, 1% Triton X-100, 5mM BME, 0.1mM PMSF, 1 Roche complete mini/10ml buffer, 1mM EDTA). Transfer the solution to 1.5 ml eppendorf tube and centrifuge at 12,000 g for 10 min at 4 degrees. Remove supernatant and resuspend pellet in 300 ul of Extraction Buffer 3 (1.7M sucrose, 10mM Tris-HCl pH 8, 0.15% Triton X-100, 2mM MgCl2, 1mM EDTA, 5mM BME, 0.1mM PMSF, 1 Roche complete mini/10ml buffer). In a clean

eppendorf tube, add 300ul of Extraction Buffer 3. Take the 300ul resuspended pellet from previous step and carefully layer it on top of the clean 300ul of Extraction Buffer 3. Spin for 1 hour at 16,000 (full speed) at 4 degrees. Remove the supernatant, and resuspend the pellet in 150uL nuclei lysis buffer (50mM Tris 8.0, 10mM EDTA, 1% SDS, 1 Roche complete mini tablet per 5ml total volume, 0.1mM PMSF). This is isolated chromatin, save 10uL to check sonication efficiency. Proceed to sonication or freeze at -20°C.

Sonicate the chromatin in a Bioruptor, save 10uL sonicated chromatin as input. Then 100uL sonicated chromatin was diluted to 1mL using ChIP dilution buffer (1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris 8.0, 167mM NaCl, 0.1mM PMSF, 1 Roche complete mini tablet per 10ml total volume), and immunoprecipitated by rotating over night at 4°C with 10ug of anti-Myc antibody (clone 4A6; Millipore) or 1ug anti-H3 antibody (ab1791; Abcam) that are pre-immobilized with 50uL or 25uL dynabeads proten G (10004D; Novex), respectively. For PolII ChIP, 10ug anti-PolII CTD (ab26721; Abcam) were immobilized with 50uL dynabeads proten G (10004D; Novex). For H3K36me3 ChIP, 5ug anti-H3K36me3 antibody (ab9050; Abcam) were immobilized with 50uL dynabeads proten G (10004D; Novex). Beads were washed subsequently with low salt wash buffer (150mM NaCl, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris 8.1), high salt wash buffer (500mM NaCl, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris 8.1), LiCl wash buffer (0.25M LiCl, 1% NP40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris 8.1) twice and TE buffer (10mM Tris 8.0, 1mM EDTA) once. Beads were eluted for 15min with gentle agitation using 250uL elution buffer (1% SDS, 0.1M $NaHCO_3$) twice at 65°C for a total of 500uL. IP samples and input were reverse X-linked with 0.2M NaCl for 6hr, and DNA was recovered by phenol/chloroform. The pellet was diluted in 50uL water. Illumina

libraries for sequencing were generated, and the libraries were sequenced using the HiSEq 2000 Sequencing System (Illumina), following the manufacturer's instructions.

RNA-seq library preparation

Total RNA was extracted from 0.1 g tissue using a standard TRIzol method. Samples were sent to BerryGenomics for sequencing in China. Alternatively, total RNA was subjected to poly(A) purification using the Dynabeads mRNA Purification Kit (Invitrogen). Illumina libraries were generated, and the libraries were sequenced using the HiSEq 2000 Sequencing System (Illumina), following the manufacturer's instructions.

Recovery test

32 plates of seedlings were grown as described above for 12 days, 16 of them were left at 22ºC and 16 moved to 38ºC in growth chamber for an initial treatment of 4hr, then moved back to 22ºC and let recover for 24hr. Then the plates were treated together at 38ºC in growth chamber for 0min, 7.5min, 15min, 30min, 60min, 120min, 180min and 240min. Two plates for each time point with or without initial treatment were placed immediately in liquid nitrogen and aerial part of seedlings were harvested using a razor blade. RNA-seq library preperation was done as described above.

Data analysis

To map the ChIP-seq and RNA-seq reads, *A. thaliana* reference genome sequence and gene models were retrieved from The A. thaliana Information Resource (TAIR10; http://www.arabidopsis.org). ChIP-seq reads were mapped to the genome with bowtie2

(Langmead and Salzberg 2012) (--phred33 --end-to-end --very-sensitive). These generated .bam files were then converted to .bw files (Li et al. 2009; Quinlan and Hall 2010) for direct visualization using the Integrative Genomic Viewer (IGV) (Thorvaldsdottir et al., 2013). The gene level profiles were generated by counting reads in 10bp windows at the 5' and 3' transcriptional boundaries, and then aligned by TSS (transcription start site) and TTS (transcription termination site) annotated in tair10. Transcript levels of annotated genes were determined by mapping the RNA-seq reads to the genome with TopHat (version 2.0.14) (Trapnell et al. 2009), and were shown as FPKM (Fragments Per Kilobase per Million fragments mapped) calculated with Cufflinks (version 2.2.1) (Trapnell et al. 2012; Trapnell et al. 2014).

CHAPTER 3

3. EFFECTS OF TRANSCRIPTION ON NUCLEOSOME POSITIONING

3.1 Introduction

Nucleosome in eukaryotic cells is the basic repeating unit of chromatin. Each nucleosome comprises 145~147bp of DNA wrapped 1.65 times around histone octamer of four highly evolutionarily conserved histone proteins – H2A, H2B, H3 and H4 (Jiang and Pugh 2009). Nucleosomes regulate access of cellular machinery to DNA and are modified by incorporation of functionally distinct histone variants (Bintu et al. 2012; Weber and Henikoff 2014; Weber et al. 2014). Nucleosome positioning refers to where the nucleosomes are located with respect to genomic DNA sequence, and the genome-wide pattern of nucleosome positioning is determined by the combination of DNA sequence, nucleosome remodelers, DNA methylation, transcription factors including activators, components of the pre-initiation complex, and elongating Pol II, and histone modifications and histone variants (reviewed in chapter 1.2).

It is intriguing to understand how nucleosomes survive during transcription. Eukaryotic transcription by RNA polymerase II (Pol II) is highly regulated and occurs in the context of nucleosomes. Inevitably, all intra-nucleosomal DNA-histone interactions must be at least transiently disrupted during transcription. On many eukaryotic genes, Pol II pauses after transcribing the initial 50–100 bp of the template. This pause frequently occurs after encountering one of the nucleosomes positioned within the early transcribed region (Churchman

and Weissman 2011). It's suggested that a key role of paused Pol II is to compete with nucleosomes for occupancy of highly regulated promoters, thereby preventing the formation of repressive chromatin architecture (Gilchrist et al. 2010). The pausing of Pol II coordinates recoiling of upstream DNA on histone octamer to form the loop that allows nucleosome survival, and uncoiling of downstream DNA allowing further transcription (Kulaeva et al. 2013). Direct evidence *in vitro* showed that transcription through a nucleosome involves transfer of the core histones behind the transcribing polymerase via a transient DNA loop (Hodges et al. 2009). It is suggested that only H2A/H2B are displaced during every round of transcription (Kulaeva et al. 2007) at moderately active genes because of the observation that the exchange of H3/H4 histones occurs at least 20-fold slower than that of H2A/H2B. It has been shown that FACT (facilitates chromatin transcription) interacts with DNA-binding surfaces of H2A/H2B dimers, facilitating uncoiling of DNA from the histone octamer. Such interactions play a key role in overcoming the nucleosomal barrier by Pol II and promoting nucleosome survival during transcription (Hsieh et al. 2013). The observed hexasome survival at the original position during transcription through a nucleosome *in vitro* and structures of the intermediates formed during this process (Kulaeva et al. 2009; Kulaeva and Studitsky 2010) suggest that H3/H4 are not exchanged *in vivo* because they are never fully displaced from the DNA during loop mediated nucleosome survival.

The question remains whether or not transcription would affect nucleosome organization *in vivo*. Several aspects of the *in vivo* nucleosome positioning pattern can be reconstituted *in vitro* if a yeast crude extract (provides a source of chromatin-associated proteins) together with ATP are added to purified histones and DNA (Zhang et al. 2011). However, two major aspects of *in vitro* reconstituted nucleosome patterns do not go faithfully with *in vivo* patterns: the precise location of the +1 nucleosomes matched poorly with their *in vivo* positions; and the degree of positioning

of downstream nucleosomes is significantly lower (Zhang et al. 2011). These seem to be reconciled by general transcription factors that compose the pre-initiation complex and the elongating Pol II machinery. Transcriptional activators, via targeted recruitment of nucleosome remodelers, can generate nucleosome-depleted regions, and the pre-initiation complex plays a role in fine-tuning the position of the +1 nucleosome (Narlikar et al. 2013).

Even though studies in human suggested T cell activation induced nucleosome reorganization, the major differences for nucleosome levels were at promoter regions between induced and repressed genes (Schones et al. 2008), which could be caused by favorable DNA sequences for induced v.s. repressed genes rather than transcription activation. Even though nucleosome level was indeed higher downstream of TSS at induced genes after T cell activation, a similar phenomenon was present for repressed genes, too. Thus this difference is likely to be a basal difference in cells before and after activation. Though correlation between gene expression and the genomic nucleosomal remodeling in response to salicylic acid (SA) treatment was published in *A. thaliana* (Singh et al. 2015), the authors still was not able to rule out the effect of DNA motifs found at promoters of SA responsive genes.

Here we propose to study the effects of transcription on nucleosome organization by examining nucleosome positioning over genes that are differentially expressed in *Arabidopsis* seedlings v.s. flowers. Essentially we are looking at the same genes, rather than comparing different genes based on their expression. Moreover, studying in a developmental context also helps to rule out the possible complication caused by DNA motifs found at promoters of stress-inducible genes. This is not studying tissue specific nucleosome organizations but using differentially expressed genes in such context to investigate the effects of transcription on nucleosome organization. Our results indicate that mononucleosomes from floral tissue were somehow more readily released

from chromatin than those from seedling tissue. The overall nucleosome occupancy in flowers was significantly lower than those in seedlings, making it inconclusive to say if differential gene transcription actually affected nucleosome organization. Interestingly, we did not observe any obvious difference in nucleosome phasing on differentially expressed genes, indicating transcription either does not influence nucleosome positioning to an extend that is readily observable, or other factors may have buried the differences caused by transcription.

3.2 Materials and Methods

Plant materials

Arabidopsis thaliana Col-0 seeds were sterilized and imbibed in water at 4°C in darkness for 3 days for stratification, grown on MS plates or in soil at 22ºC under continuous light conditions. The 12-day-old whole seedlings were harvested from MS plates and immediately immersed in liquid nitrogen. To assess the effects of transcription on nucleosome positioning, inflorescences from plants grown in soil were collected and also immediately immersed in liquid nitrogen.

Chromatin Extraction

Grind the seedlings or inflorescences tissue in liquid nitrogen to a fine powder, add the powder to 10 volumns of Extraction Buffer 1 (0.4M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl, 0.1mM PMSF, 2 Roche mini tablets/ 100 ml, 1mM EDTA) in a 50 ml falcon tube. Leave on ice for 5 min. Filter the solution through Miracloth into a 50 ml falcon tube. Repeat for a total of two filters, and then spin the filtered solution for 20 min at 4000 rpm in a centrifuge at 4 degrees. Gently remove supernatant and resuspend the pellet in 1 ml of Extraction Buffer 2 (0.25M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl2, 1% Triton X-100, 5mM BME, 0.1mM PMSF

1 Roche complete mini/10ml buffer, 1mM EDTA). Transfer the solution to 1.5 ml eppendorf tube and centrifuge at 12,000 g for 10 min at 4 degrees. Remove supernatant and resuspend pellet in 300 ul of Extraction Buffer 3 (1.7M sucrose, 10mM Tris-HCl pH 8, 0.15% Triton X-100, 2mM MgCl2, 5mM BME, 0.1mM PMSF, 1 Roche complete mini/10ml buffer, no EDTA!). In a clean eppendorf tube, add 300ul of Extraction Buffer 3. Take the 300ul resuspended pellet from previous step and carefully layer it on top of the clean 300ul of Extraction Buffer 3. Spin for 1 hour at 16,000 (full speed) at 4 degrees. Remove the supernatant, add 1ml digestion buffer (0.32M sucrose, 50mM Tris pH=8.0, 4mM MgCl2, 1mM CaCl2, 0.1mM PMSF) to gently wash the pellet, careful not to disturb the pellet and pipette the buffer out. Resuspend the pellet in 200uL digestion buffer. This is isolated chromatin, proceed to next step or freeze at -20°C.

Isolation and quantification of gDNA

20uL of isolated chromatin was diluted with digestion buffer to 200ul, mixed with equal volumn of proteinase K buffer (10ul 20mg/ml proteinase K per 100ul lysis buffer (0.1M Tris=8.5, 0.1M NaCl, 50mM EDTA, 1%SDS)). Add equal volume phenol: chloroform (=1:1), votex 30s. Spin 5min, transfer supernatant to a new tube, about 400uL. Add equal volume chloroform and votex 30s. Spin 5min, transfer supernatant to a new tube, about 360uL. Add 4ul 5M NaCl per 100ul supernatant, 2.5 volume 100% EtOH, precipitate at -20°C for 1 hour. Spin for 10min. Wash with 1 ml 75% EtOH. Dissolve pellet in 40-50ul RNase A water (20ug/ml) for 30min at 37°C. Quantify with Qubit using broad range dsDNA assay kit.

MNase digestion

Based on Qubit quantification, aliquot 600ng chromatin for each reaction. Bring volume to 200uL with digestion buffer. Pre-warm chromatin at 37°C for 10min. Digest warm chromatin with 3U MNase (Takara 2910A) per reaction at 37°C for 10min, 30min and 60min. After digestion, add equal volume proteinase K buffer, incubate at 65°C for 45min, mix occasionally. Recover DNA as described above. Run 10uL digested DNA as well as gDNA on a 2% agarose gel to check result.

Isolation of total RNA

Total RNA was extracted using the TRizol reagent (Invitrogen). Three replicates of seedlings and inflorescences tissue were extracted side by side.

Library prep and sequencing

MNase samples were subjected to Illumina paired end sequencing libraries according to manufacturers instructions. Total RNA samples were subjected to Illunima single end sequencing libraries according to manufacturers instructions.

Data analysis

To map the MNase-seq and RNA-seq reads, *A. thaliana* reference genome sequence and gene models were retrieved from The A. thaliana Information Resource (TAIR10; http://www.arabidopsis.org). PE MNase reads were mapped to the genome with bowtie2 (Langmead and Salzberg 2012) (--phred33 --end-to-end --very-sensitive). These generated .bam files were then converted to .bw files (Li et al. 2009; Quinlan and Hall 2010) for direct

visualization using the Integrative Genomic Viewer (IGV) (Thorvaldsdottir et al., 2013). The gene level profiles were generated by counting reads in 10bp windows at the 5' and 3' transcriptional boundaries, and then aligned by TSS (transcription start site) and TTS (transcription termination site) annotated in tair10. Transcript levels of annotated genes were determined by mapping the RNA-seq reads to the genome with TopHat (version 2.0.14) (Trapnell et al. 2009), and were shown as FPKM (Fragments Per Kilobase per Million fragments mapped) calculated with Cufflinks (version 2.2.1) (Trapnell et al. 2012; Trapnell et al. 2014). Custom scripts were developed and used to count only the end of reads (5' of sense reads, and 3' of antisense reads) in 10bp bins around TSS and TTS so that the edges of nucleosomes are plotted in subsequent analysis, and nucleosome phasing was determined by the DANPOS method (Chen et al. 2013a).

3.3 Results and Discussions

The inflorescences tissue seems easier to digest than seedlings

The MNase digestion strategy is widely used to map nucleosome position and occupancy, based on its ability to produce sequence-dependent cleavages in DNA preferentially within linker regions. To determine the changes in the nucleosome positioning caused by differential expression, 2-week old Col seedlings and inflorescences from mature plants were used for mononucleosomal DNA preparation. As shown in Fig. 3.1, the inflorescences tissue displayed a relatively enriched mono-nucleosome band, suggesting that mononucleosomes from floral tissue were somehow more readily released from chromatin than those from seedling tissue (Gent et al. 2014) since there was equal amount of chromatin for each reaction to begin with, As the digestion patterns are comparable between flowers-10min and seedlings-30min, the mononucleosome DNA at around 150bp were used in subsequent ChIP-seq library preparation.

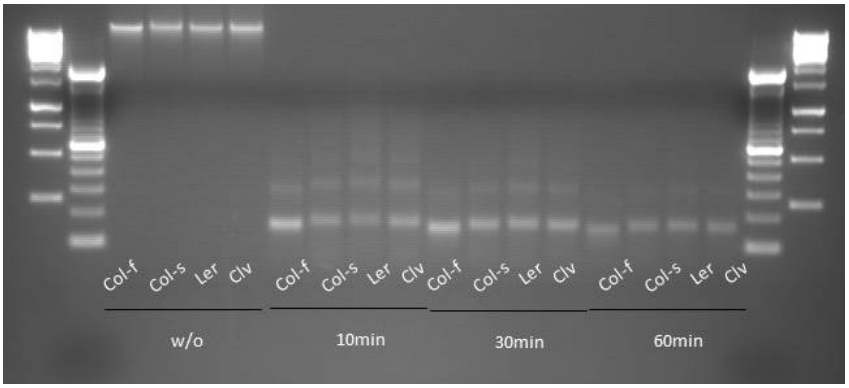

Fig. 3.1 gDNA and MNase digested DNA samples were run on 2% agarose gel. The w/o samples indicate gDNA. Two other ecotypes of arabidopsis Ler and Clv were also included for comparison. The 10min digestion for flowers and 30min digestion for seedlings were subjected to library prep.

Gene level nucleosome distribution in flowers v.s. seedling

MNase preferentially cuts within linker DNA (Noll 1974), so the size-selected fragments indicate the positions of nucleosomes when they are sequenced and mapped back to the genome (Kaplan et al. 2009). Custom scripts were developed and used to count only the end of reads (5' of sense reads, and 3' of antisense reads) in 10bp bins around TSS and TTS so that the edges of nucleosomes are plotted in subsequent analysis. As shown in Fig. 3.2, nucleosome organization displays a nucleosome depleted region at promoter regions for all genes in flowers and in seedlings. Nucleosomes are relatively well positioned downstream of TSS, but the precision of nucleosome positioning decays with increasing distance from the +1 nucleosome. Even though the nucleosome bands were comparable for these two samples on the gel, the occupancy of nucleosome in seedlings appears to be higher than that of in flowers. This is probably because the chromatin from flowers was much cleaner during extraction, and is favorable for the MNase activity. We need to be cautious of this in subsequent analysis. This affected nucleosome

occupancy, but the positions of each nucleosome especially at the 5’ end of genes are the same (Fig. 3.2).

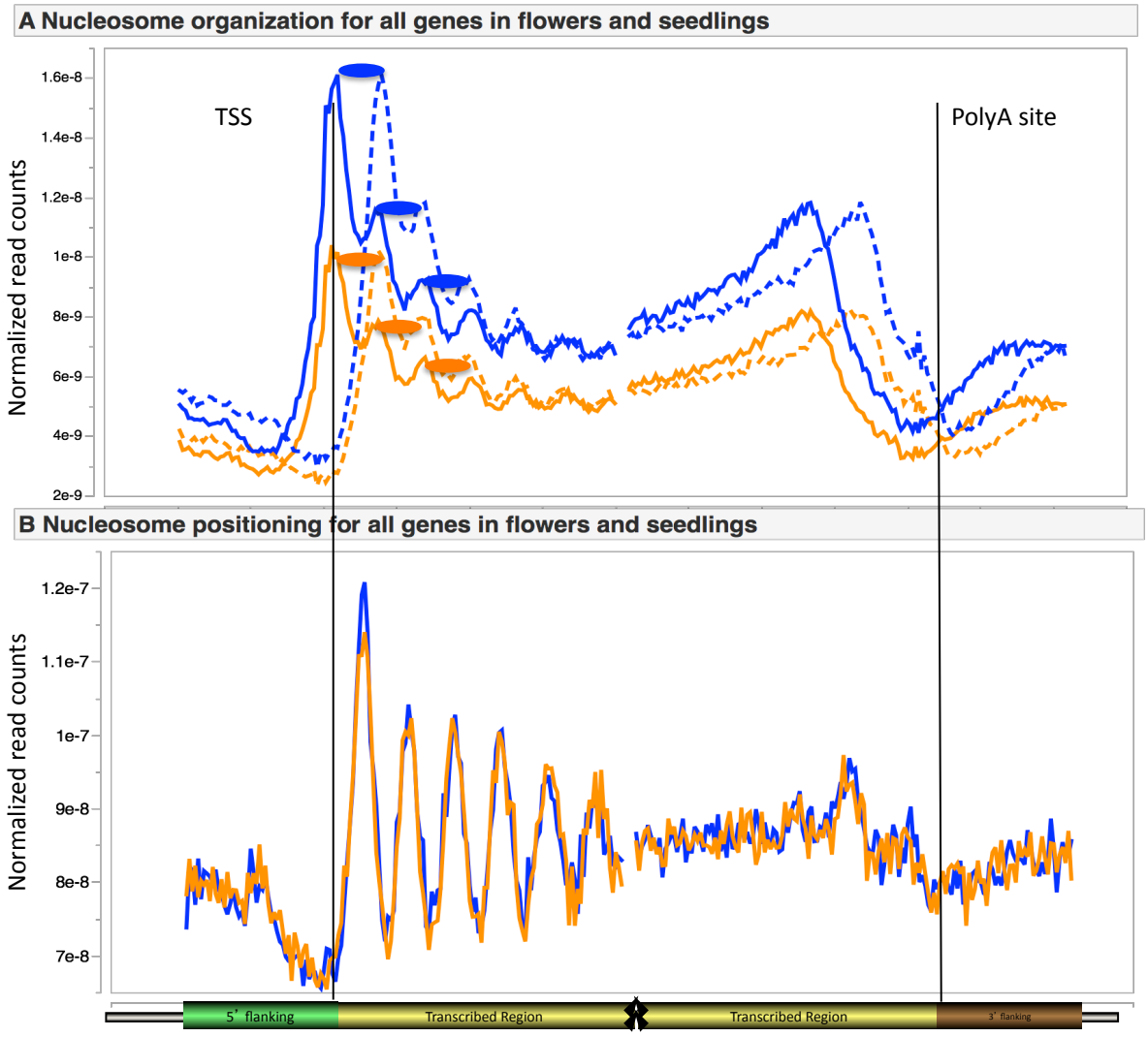

Fig. 3.2 Gene level nucleosome distribution for all genes in flowers (orange) and seedlings (blue). A. Normalized read counts by customized scripts showing the 5’ (solid line) and 3’ (dashes line) of nucleosomes. Filled ovals represent phased nucleosomes. B. DANPOS analysis of nucleosome positioning/phasing.

Identification of differentially expressed genes

In order to identify differentially expressed genes in seedlings and flowers, RNA-seq was performed using matching tissues with that used in MNase digestion. It needs to be pointed out

that the up-regulated genes from cufflinks output are actually a mixture of three scenarios: 1) genes that were not expressed (or expressed at low levels) in seedlings, but became actively

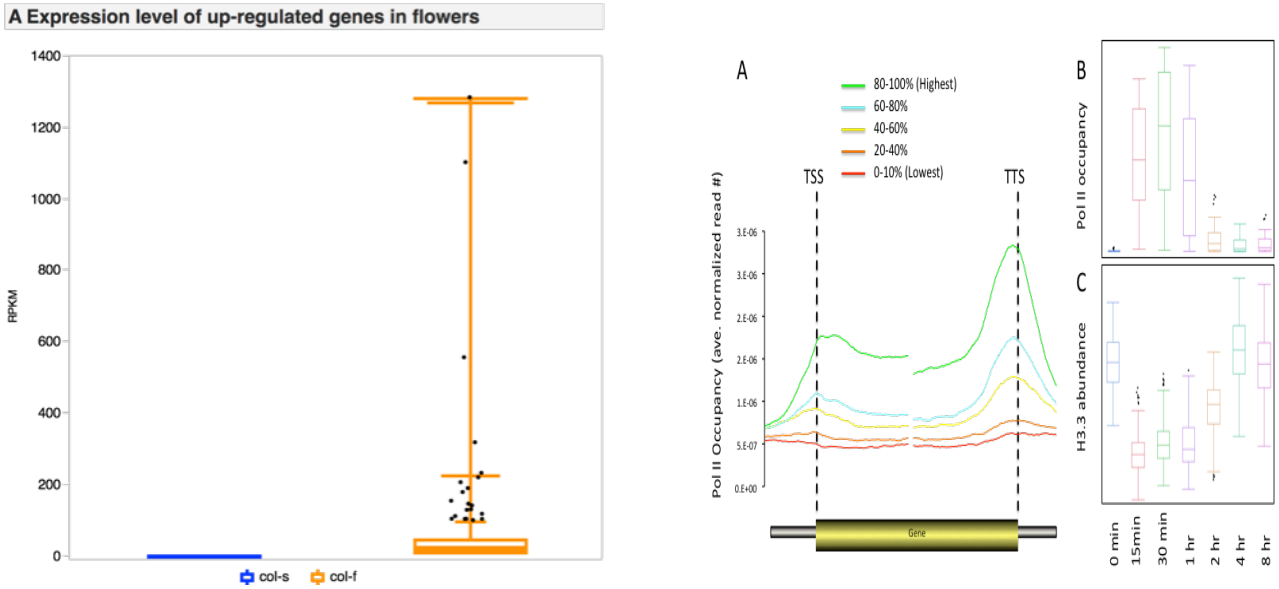

Fig. 3.3 Identification of differentially expressed genes in flowers (orange) and seedlings (blue). Box plots of RPKM of up (A) or down (B) regulated genes in flowers compared to seedlings. About 210 and 225 genes passed our criteria and were considered up or down regulated genes, respectively.

expressed in flowers; 2) genes were expressed at medium level turned to high levels in flowers and 3) genes were already expressed at high levels and became even higher in flowers. Only the first senario would give us the opportunity to examine the effects of transcription on nucleosome organization. Similarly, only those that were expressed at a high level in seedlings and turned to very low or even no expression in flowers were considered down regulated genes. To select real differentially expressed genes, we took all significantly changed genes and picked those whose expression changed by at least 2 fold. Upregulated genes were picked whose initial expression level were smaller than the median RPKM in seedlings but higher than the median RPKM in flowers, which ended up with 210 genes. Similarly, downregulated genes were picked whose initial expression level were higher than the median RPKM in seedlings but lower than the

mdedian RPKM in flowers, which ended up with 228 genes. Their RPKM were plotted in Fig 3.3 to show the expression of those genes were indeed significantly changed.

No observable changes in nucleosome phasing/positioning on differentially expressed genes

After identifying the true up- or down regulated genes, nucleosome distribution was examined as did previously. Nucleosome occupancy for both up- and down regulated genes was lower in flowers compared to that in seedlings (Fig. 3.4 A&B). Even though the difference might be slightly smaller for down regulated genes, it would be too ambitious to make the conclusion that transcription affects nucleosome organization to an observable extend. Moreover, nucleosome phasing (Fig. 3.4 C&D) is comparable for up- and down regulated genes.

One of the reasons at technical level would be nucleosomes might be re-organized to favorable DNA sequences during MNase digestion. The fact that several aspects of the *in vivo* nucleosome positioning pattern can be reconstituted *in vitro* if a yeast crude extract (provides a source of chromatin-associated proteins) together with ATP are added to purified histones and DNA (Zhang et al. 2011) may support this. Applying a X-link step before chromatin extraction and MNase digestion would likely help to rule out this possibility. However, if this is indeed the case, we may need to revisit our understandings interpreted from MNase digestion results in all organisms.

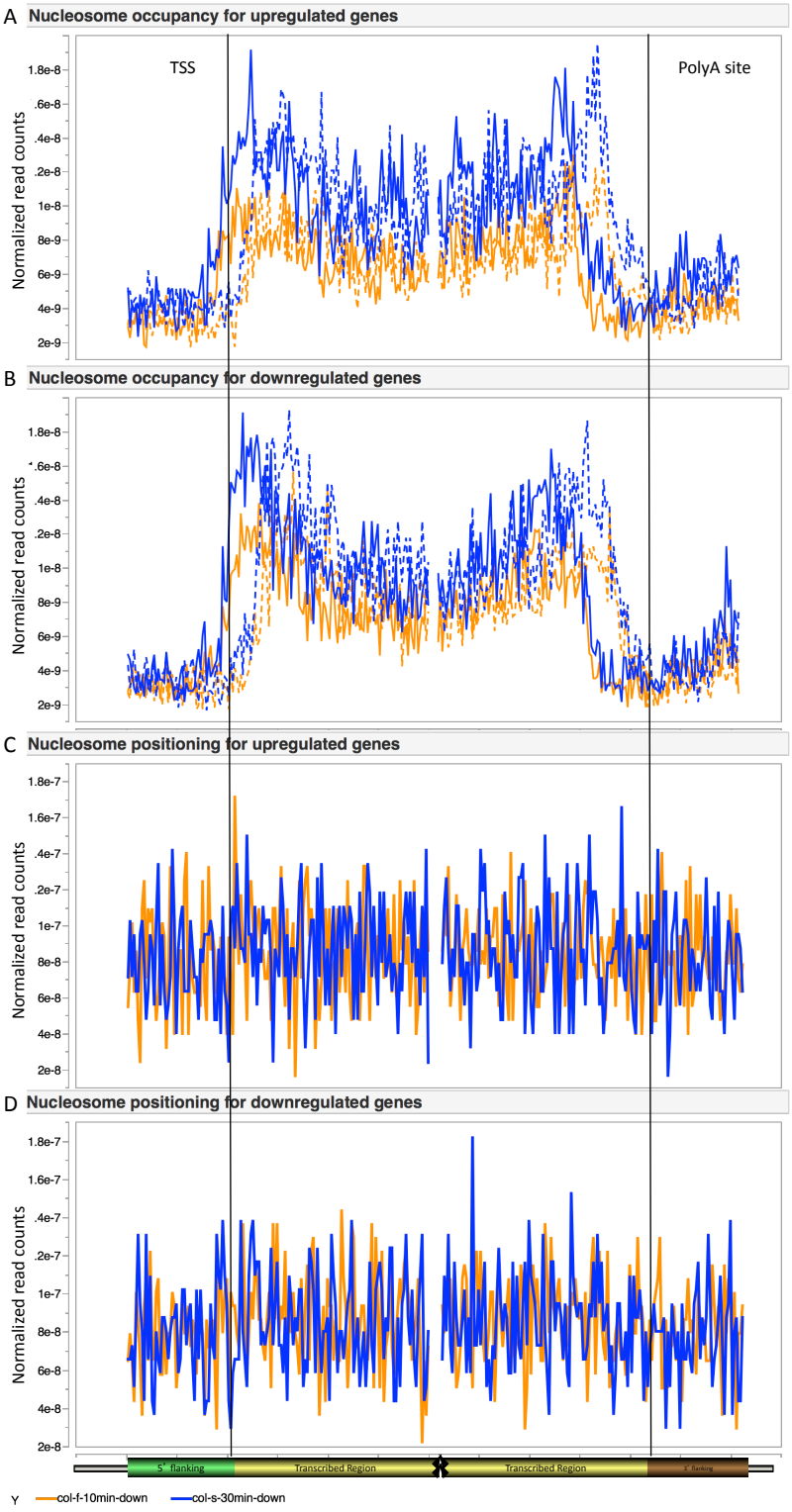

Fig. 3.4 Nucleosome distribution for differentially expressed genes

CHAPTER 4

4. TARGETING MECHANISMS OF THE H3K4 TRI-METHYLTRANSFERASE SET DOMAIN GROUP PROTEIN 2 IN ARABIDOPSIS

4.1 Introduction

Nucleosomes in eukaryotes are composed of DNA wrapped around histone octamers of two copies of H2A, H2B, H3 and H4. Histones are subjects to posttranslational modifications (PTMs) such as methylation, phosphorylation, acetylation and ubiquitination. One of the most extensively studied modifications is methylation of lysine residues on histone tails, which plays an important role in regulating chromatin structure and gene expression. Whole genome level profiling in *Arabidopsis* revealed strong associations of certain lysine methylation with active or repressed genes, in that transcribed genes are strongly associated with H3K4 or H3K36 methylation, while developmentally repressed genes often have H3K27 methylation (Lippman et al. 2004; Turck et al. 2007; Zhang et al. 2007a; Bernatavichute et al. 2008; Oh et al. 2008; Zhang 2008b; Charron et al. 2009; Zhang et al. 2009a; Jacob et al. 2010a). Intriguingly, these modifications occupy different chromatin regions with distinct boundaries. For instance, tri-methylation of histione H3 lysine 4 (H3K4me3) is preferentially located at the 5' regions of most actively transcribed genes in *Arabidopsis* (Zhang et al. 2009a). The underlying targeting mechanism of histone methyltransferase is of great interest to researchers.

Histone lysine methylation is catalyzed by a group of proteins with a conserved SET domain, which is about 130-160aa long, and is named after Drosophila histone methyltransferases

Su(var)3-9, *Enhancer of Zeste* [*E(z)*] and *trithorax* (*trx*) (Baumbusch et al. 2001). There are 39 genes in *Arabidopsis* encoding a SET or interrupted SET domain, and they are divided into five classes based on sequence similarity to their animal and/or fungal counterparts(Baumbusch et al. 2001; Springer et al. 2003). Four out of the five classes are shared among organisms: class I SDGs are involved in H3K27me2/3, class II in H3K36me2/3, class III in H3K4me2/3, and class V in H3K9me2. Class IV seems to be specific to plants and is responsible for H3K27me1. Although HMTase activities have been found in plants, the targeting mechanism has yet to be identified.

Specifically, histone H3 lysine 4 methylation is catalyzed by the highly evolutionarily conserved multiprotein complex known as Set1/ COMPASS or MLL/COMPASS-like complexes. The first HMT shown to catalyse mono, di and trimethylation of H3K4 is SET domain-containing 1 (Set1) in *Saccharomyces cerevisiae*, forming a protein complex called COMPASS (complex proteins associated with Set1) (Miller et al. 2001). In mammals, six COMPASS-related complexes with essential and non-redundant functions have been identified that can methylate H3K4 through their conserved SET domain as reviewed in (Schuettengruber et al. 2011): SET1A- or SET1B-containing complexes and MLL1-4 complexes. It is proposed that COMPASS complexes are specifically targeted to 5' of transcribed genes via the following mechanisms(Schuettengruber et al. 2011): 1) MLL can directly interact with CpG-rich sequences via its CXXC domain, a type of zinc-finger that can recognize unmethylated CpG sequence; 2) MLL can be targeted to DNA by interacting with sequence-specific transcription factors or the polymerase-associated factor 1 (PAF1) elongation complex; 3) non-coding RNA such as HOXA transcript at the distal tip (HOTTIP) can recruit MLL via their interaction with the adaptor protein WD repeat-containing 5 (WDR5); 4) other histone marks such as di-methylated H3K4 or H2B mono-ubiquitination is

likely to assist MLL recruitment for proper H3K4me3 establishment at 5' genic regions. Alternatively, the enrichment of H3K4me3 at 5' transcribed genes could be the result of counteract effects of HMTase and lysine demethylase, where the targeting of HMTases is not specific and tri-methylate lysine 4 all along the transcribed gene, but lysine demethylase is specifically targeted to remove methyl group from tri-methylated K4 in the body of transcribed genes. This is supported to some extend in human embryonic stem cells (ESC) that the H3K4me3 demethylase KDM5b is located exclusively from H3K4me3 peaks(Xie et al. 2011). The identification of a unique class III protein SET DOMAIN GROUP PROTEIN 2 (SDG2) by our group is crucial for elucidating the targeting mechanism of H3 lysine 4 methyltransferase in plants. It is so far the major H3K4 tri-methyltransferase in *Arabidopsis*, since the single mutant shows a global loss of H3K4me3 that leads to severe and pleiotropic developmental abnormalities (Guo et al. 2010) while mutants in other three characterized class III HMTase genes (ATX1/SDG27, ATX2/SDG30, ATXR7/SDG25) (Alvarez-Venegas and Avramova 2005; Saleh et al. 2008; Berr et al. 2009; Tamada et al. 2009) only display locus-specific defects in H3K4me. This is probably because out of the seven *Arabidopsis* class III SDG proteins computationally identified as putative H3K4 HMTases, only SDG2 contains a Phe residue at the Phe/Tyr switch; the re- maining six SDGs contain a Tyr residue at this position (Fig. 4.1), a characteristic of HMTases that preferentially catalyze mono- and dimethylation (Collins et al. 2005). Given the facts that 1) more than two thirds of *Arabidopsis* genes contain at least one type of H3K4me (Zhang et al. 2009a), 2) *In vitro* HMTase assay demonstrated that SDG2 is capable of catalyzing lysine 4 methylation in all three contexts: mono, di and tri (Guo et al. 2010), and 3) distribution patterns of H3K4me1, H3K4me2 and H3K4me3 show distinct boundaries in the

genome (Zhang et al. 2009a), understanding how SDG2 is recruited to chromatin is critical to study the establishment of this modification.

```
ATX4            LEYRGEQ--------VRGSIADLREARYRR--VGKDCYLFKIS------EEVVVDATDKG 955
ATX5            LEYRGEQ--------VRGIIADLREARYRR--EGKDCYLFKIS------EEVVVDATEKG 971
ATX3            IEYRGVK--------VRRSVADLREANYRS--QGKDCYLFKIS------EEIVIDATDSG 945
ATX1            IEYTGEL--------VRPSIADKREQLIYNSMVGAGTYMFRID------DERVIDATRTG 970
ATX2            IEYTGEL--------VRPPIADKREHLIYNSMVGAGTYMFRID------NERVIDATRTG 991
ATXR7           IEYVGEL--------IRSSISEIRERQYEKMGIGS-SYLFRLD------DGYVLDATKRG 1337
SDG2            VEFLGEVYPVWKWFEKQDGIRSLQENKTDPAPEFYNIYLERPKGDADGYDLVVVDAMHMA 1858
                :*: *           :  : . :*            *: : .      :  *:**   .

ATX4            NIARLINHSCTPNCYARIMSVGD-EESRIVLIAKANVAVGEELTYDYLFDPDE-AEELKV 1013
ATX5            NIARLINHSCMPNCYARIMSVGD-DESRIVLIAKTTVASCEELTYDYLFDPDE-PDEFKV 1029
ATX3            NIARLINHSCMPNCYARIVSMGDGEDNRIVLIAKTNVAAGEELTYDYLFEVDE-SEEIKV 1004
ATX1            SIAHLINHSCVPNCYSRVITVNG--DEHIIIFAKRHIPKWEELTYDYRFFS----IGERL 1024
ATX2            SIAHLINHSCEPNCYSRVISVNG--DEHIIIFAKRDVAKWEELTYDYRFFS----IDERL 1045
ATXR7           GIARFINHSCEPNCYTKIISVEG--KKKIFIYAKRHIDAGEEISYNYKFPLE----DDKI 1391
SDG2            NYASRICHSCRPNCEAKVTAVDG--HYQIGIYSVRAIEYGEEITFDYNSVTESKEEYEAS 1916
                . *  * *** *** ::: :: .  . :* : :   :   **:::::*

ATX4            PCLCKAPNCRKFMN---------------------------------------------- 1027
ATX5            PCLCKSPNCRKFMN---------------------------------------------- 1043
ATX3            PCLCKAPNCRKFMN---------------------------------------------- 1018
ATX1            SCSCGFPGCRGVVNDTEAEEQHAKICVPRCDLIDWTAE---------------------- 1062
ATX2            ACYCGFPRCRGVVNDTEAEERQANIHASRCELKEWTES---------------------- 1083
ATXR7           PCNCGAPNVYCFCEQVPWIAKLKRRTWFSRRN---------------------------- 1423
SDG2            VCLCGSQVCRGSYLNLTGEGAFQKVLKDWHGLLERHRLMLEACVLNSVSEEDYLELGRAG 1976
                 * *
```

Fig. 4.1 Multiple sequence alignment of 7 Arabidopsis class III H3K4 HMTases. Protein sequences were retrieved from www.arabidopsis.org and aligned with ClustalW2 (McWilliam et al. 2013). The Phe/Tyr switch was indicated in red box.

Recent studies have revealed that similar targeting mechanisms for H3K4me3 establishment may exist in *Arabidopsis*. It has been shown that the *Arabidopsis* Paf1c is required for the accumulation of H3K4me3 in the 5' regions of actively transcribed genes in *Arabidopsis* genome (Oh et al. 2008). The *Arabidopsis* homologs of the human COMPASS-like complex core components Ash2 and RbBP5, together with WDR5a, form a nuclear subcomplex for H3K4me3 at specific loci, and loss of Ash2R led to global loss of H3K4me3 in the genome (Jiang et al.

2011). But the catalytic component of this complex was not shown in this complex except WDR5a interacts with multiple putative *Arabidopsis* H3K4 Methyltransferases via yeast two-hybrid. It is of note that WDR5a has not been shown to interact with SDG2, indicating the real targeting mechanism remains to be elucidated.

SDG2 is a large (~2,414aa) and highly conserved protein with the catalytic domain (the "SET domain") located near the C-terminus. Interestingly, several blocks of amino acid residues outside the SET domain are highly conserved in plants but share no homology with any known proteins. We hypothesized that these conserved domains may be involved in mediating protein-protein interactions between SDG2 and other proteins, and that such interactions may be responsible for targeting SDG2 to the chromatin. I used three approaches to try to get the SDG2 interacting protein complex: 1) I first tried to pull down the complex *in vivo* using polyclonal antibodies generated against endogenous SDG2 in mouse; 2) I fused a 9*myc + BLRP tag and a YFP + HA to the C terminus of genomic copy of SDG2, respectively. Then made transgenic lines in hope of pulling down the SDG2 containing complex (es); 3) I applied yeast two-hybrid against an *Arabidopsis* cDNA library to try to get individual SDG2 interacting proteins, and re-constitute the protein complex *de novo*. The polyclonal anti-SDG2 antibodies failed to pull down any complexes, and the C-terminal tagged SDG2 failed to rescue the mutant phenotype. I was successful with the last approach and identified 5 putative genes that interact with truncated carboxyl fragment of SDG2 containing the SET domain and pre-SET conserved domains. All 5 genes share an SMC (structural maintenance of chromosomes) like coiled coil domain, the family of which has been suggested to function as intermolecular and intramolecular linkers of DNA (Jeppsson et al. 2014). Our results indicate SDG2 might be directed to chromatin via interacting with DNA recognition proteins.

4.2 Materials and methods

Plant materials

All Arabidopsis plants used in this study are of the Columbia (Col-0) accession. T-DNA insertion lines were obtained from the Arabidopsis Biological Research Center. Plants were grown at 22 °C under continuous light.

C terminal tag construct

Genomic copy of SDG2 was amplified with primers zp1995 and zp1996 and subcloned to pENTR-D/TOPO (sigma) vector, sequenced and recombined to pEarleyGate 101 vector with Gateway LR Clonase (11791-020: Invitrogen) to obtain the YFP+HA C-tag. A 9*myc plus BLRP tag was inserted to the KpnI and SpeI sites (KpnI partial digestion) of pEarleyGate 101. Then a BirA cassette was inserted to the SpeI site. The modified pEarleyGate 101 vector will now have both the 9*myc+BLRP and BirA gene and is ready for recombination. Confirmed vectors were transformed to sdg2 heterozygous plants via floral dip method.

Immunoprecipitation

Prepare beads/antibody by adding 10uL anti-SDG2 to 100uL argarose beads per sample and rotating for 1hr rotating at 4°C then wash four times with 1mL IP buffer. 0.5g inflorescence tissue from N-tag SDG2 transgenic plants, wild type Col and *sdg2-/-* were grinded to a fine powder with 2.5mL IP buffer (50mM Tris pH 7.6, 150mM NaCl, 5mM MgCl2, 10% glycerol, 0.1% NP40 and for 50mL IP buffer: add 1 Roche complete protease tab, 50uL of 1M DTT, 500uL of 0.1M PMSF and 50uL of 10mg/ml pepstatin). Have the IP buffer dropped in mortar containing liquid nitrogen and grind together with tissue. Continue grinding while the

powder/buffer thaws in the mortar. Transfer solution to homogenizer and dounce (first with looser plunger, then tighter one). Spin at max speed for 10min at 4°C. Transfer supernatant to a new tube, spin at max speed for 10min. transfer supernatant into single 1.5ml tubes with prepared beads. Rotating at 4°C for 4hrs. Wash beads with fresh IP buffer with inhibitors for 5min and a total of 5 wathes. Spin down again to pellet beads. Add 50ul sample buffer to beads and boil for 10min. Load samples immediately to 3%-8% pre-cast Tris-Acetate gels (EA03752BOX; NuPAGE) for electrophoresis followed by anti-myc (clone 4A6) western blot.

BASTA screening for C-tag transgenic plants

T1 seeds were bulk harvested, and half of the seeds (5ml in volume) were planted in soil and placed at 4°C for 3 days to break dormancy then grown under continuous light at 22°C. About 4 days after germination, spray seedlings with 120mg/L BASTA solution. After 4 days repeat BASTA treatment, for a total of 3 treatments. Non-transformed plants should show yellowing of cotyledons by now. The plants that passed BASTA screening were transplanted to new soil and grown under continuous light at 22°C.

Constructing bait strains

Primers were designed to amplify the conserved blocks of SDG2 based on multiple sequence alignment. Specifically, primers zp903 and zp1709 were used to generate the N terminus fragment; zp905 and zp1271 were used to generate the middle fragment; zp906 and zp1709 were used to generate the C terminus fragment (see table 4.1). The fragments were first cloned to pCR®2.1-TOPO from Invitrogen, sequenced and finally sub-cloned into the SmaI SalI restriction sites of the pGBKT7 bait vector. Confirmed plasmids were transformed into yeast

AH109 according to the Clontech Yeast Handbook. Expression of individual baits in the yeast cells was confirmed by western blot (anti-myc 9E10).

Table 4.1 Primers used for constructs

Primer Name	Sequence
zp903 F	cccgggaAGCGATGGGGGTGTC (ATG)
zp1709 R	gtcgacAGTGGAAGCTGTTGATTGGG (aa992)
zp905 F	cccgggaGAGAGAAATCAGACGAGGGGAGT (aa921)
zp1271 R	gtcgacCCTGGAATGTAATTCACCTGATCG (aa1361)
zp906 F	cccgggaCGATCAGGTGAATTACATTCCAGG (aa1362)
zp286 R	gtcgacCTAACTATCCCATGTCGCTTG (end)
zp1471	TAATACGACTCACTATAGGGC (pGBKT7 sequencing)
zp1993 F	caccCCAAGGTCGCAGAGAGAAAC (-54 of ATG on gSDG2)
zp1995 F	caccATGAGCGATGGGGGTGTC (ATG)
zp1996 R	ACTATCCCATGTCGCTTGGA (genomic end no TAG)

Screening the activation library

The cDNA library fused with AD vector was obtained directly from ABRC stocks center. Before transform of the library, the bait strains were put on SD/Trp-/His- plates to test if they have intrinsic transcription ability. Adding certain amount of 3-AT can inhibit this intrinsic transcription. Library transformation was carried out as previously did. Healthy growing colonies were picked from SD/Trp-/Leu-/His-/Ade- +5mM 3-AT plates after 4 days of incubation at 30°C. Then they were transferred onto SD/Trp-/Leu-/His-/Ade- +X-α-Gal plates to test for α-galactosidase activity.

positive colonies and sequencing

Plasmids will be recovered from yeast colonies that turn blue on SD/Trp-/Leu-/His-/Ade- +X-α-Gal plates, and transformed back into *E.coli*. The transformants will be cultured on LB plates

Table 4.2 Primers used for sequencing

Primer Name	Sequence
M13 F	GTAAAACGACGGCCAG
M13 R	CAGGAAACAGCTATGAC
zp152	CTTCCAAATACTCTTGGGGTGC
zp242	GAACGCTCCTCCTCCGCT
zp172	GGAAATTCTTAAGCACAATCTAGAAG
zp2111	TTGGCATGGTCTGCTGGAA
zp171	CTTCTAGATTGTGCTTAAGAATTTCC
zp507	GAGATTCAAGTAGCTGCCTCGAC
zp1009	GAGACTTACCAAGTATGAATGAGCAG
zp359	GGAACcGATCCCATCTTGC
zp237	CAGGCATTGAATCAAGTAAGAGG
zp999	CCGCTCAAAGCGTATATCTTGAG
zp1000	CCACAAATTGTTAATTCCTCACACC
zp1001	CGATTTATCTTCCTTTGACCTCG
zp1002	CAGATGATTGTTCCCGCTCTC
zp1003	GAGGGATTGTCAGACTGGTACG
zp1004	GCAGTGACACAGCTCGTAAATCC
zp1005	CCATGAAAAGGGCGATCGA
zp1006	GGAGACAACCACCCAAGTTCAG
zp1007	GAGCATCCCTGGATAGTATGGC
zp1010	AGGAGAAGGTTTTCGGAGCC
zp1011	CTGACGTGAAAGGAACTGATGG
zp1013	CCTAGGAGTGTTTACAGATGTGTCGG
zp1019	CCAGCCATCAAGCGAGCTC
zp1020	GAGCCACAGCAGTGTCTTCAG

containing Ampicillin so that only cells with the AD vector will be selected. Then I will use mini-prep to recovery the library plasmids and have them sequenced using primer zp1471. BLAST will be done against *Arabidopsis* tair10 transcripts to identify genes.

4.3 Results and Discussion

Anti-SDG2 failed to pull down any visible complexes *in vivo*

We first tested the efficiency of anti-SDG2 antibodies using recombinant SDG2C, which by itself was able to catalyze all three context of H3K4 methylation: mono- di- and tri (Guo et al. 2010). Three out of six mouse polyclonal antibodies successfully recognizes the recombinant SDG2C (results not shown). The three antibodies were mixed together in IP with inflorescence tissues from col, N-tagged (9*myc+BLRP) sdg2 that rescued the mutant phenotype and *sdg2* mutant. Anti-myc (clone 4A6) was used in western to test IP efficiency. The SDG2 band (~250kD) was clearly shown in input samples from N-tagged sdg2 as shown in Fig. 4.2. However, IP itself did not seem to have worked. Alternatively, IP was done with anti-myc (clone 4A6) and checked with anti-SDG2. No reliable band was visible on the film (result not shown). As anti-SDG2 used were generated with a fragment of SDG2 aa921-aa1301. It is possible that the antigen was not exposed during the folding of mature SDG2 protein *in vivo*. It is shown that an anti-SDG2 generated with N-terminal region of SDG2 from amino acid 1 to 160 was successful in chromatin immuno-precipitation (Kim et al. 2013), using this very anti-SDG2 may help in identifying SDG2 containing complexes. Alternatively, as SDG2 is highly likely to localize in the nucleus, thus a nuclei extraction before IP will probably help in detecting putative SDG2 containing protein complexes.

Neither myc+BLRP nor YFP+HA C-tag SDG2 transgenic plants rescued the mutant phenotype

As homozygous *sdg2-/-* is sterile, the C-tag SDG2 transgene was transformed to *sdg2+/-* background. After BASTA treatment, a total of 161 myc+BLRP C-tag SDG2 plants and 108 YFP+HA C-tag SDG2 plants survived, respectively. As these are from a segregating population, a quarter of them should be *sdg2-/-* background harboring the transgene. If the transgene is able to rescue the mutant phenotype, they should be phenotypically indistinguishable from wild type. However, 8 plants from the myc+BLRP transformation and 2 plants from the YFP+HA

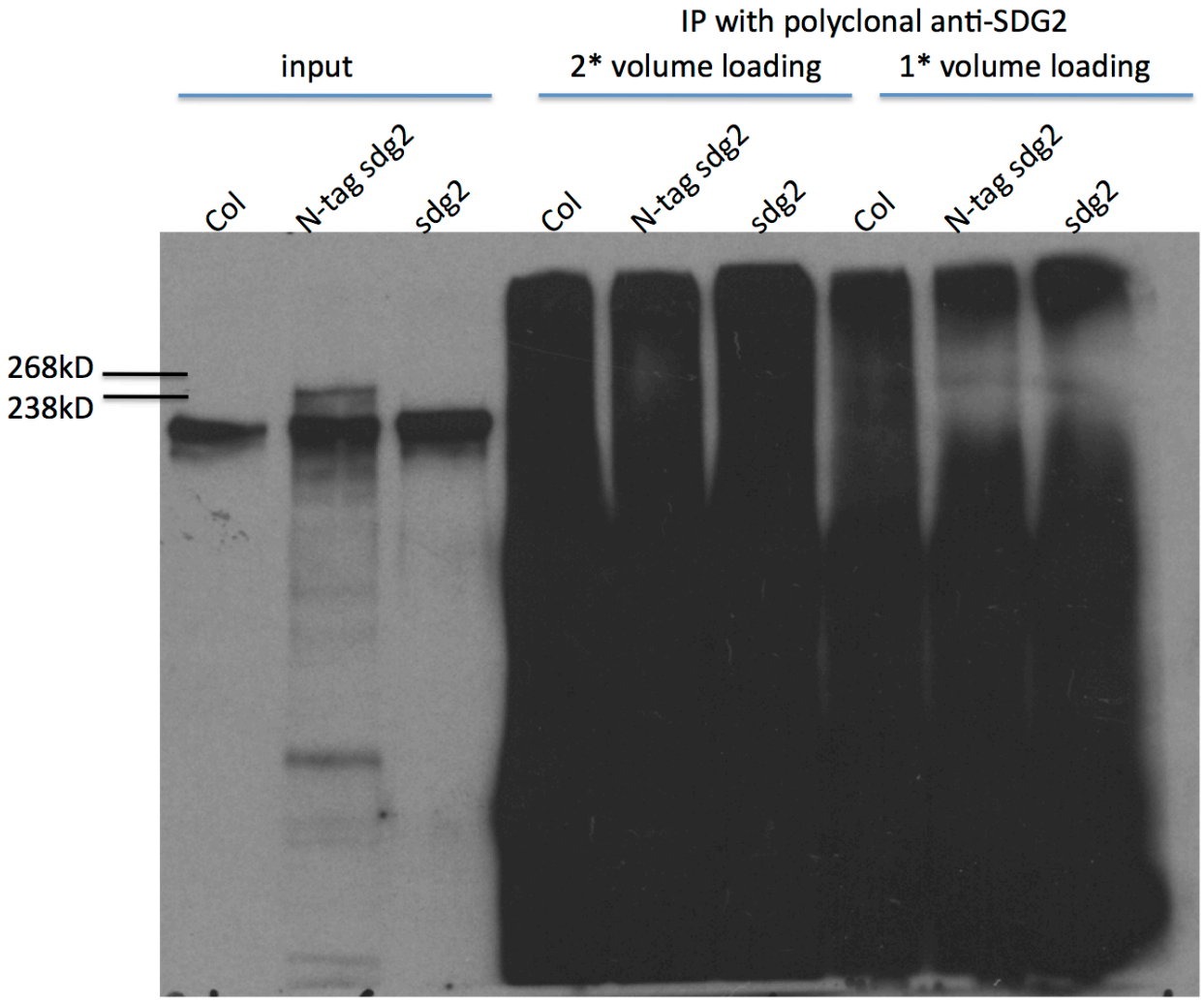

Fig. 4.2 Anti-SDG2 pull-down followed by anti-myc western. If IP has had worked, an enriched band (or smear) should appear in the IP N-tag SDG2 lanes.

transformation exhibited strong, pleiotropic *sdg2-/-* phenotype, even though these plants did have the transgene. After genotyping with the rest 259 BASTA resistant plants that look normal as wild type, none of them was *sdg2-/-* background. Therefore, it is unlikely the C-tag SDG2 transgene was functional *in vivo*, and I was unable to pursue subsequent IP experiment.

Yeast two-hybrid identified five SDG2 interacting genes

To identify *Arabidopsis* proteins that interact with SDG2, we performed yeast two-hybrid screens with N terminus, Middle and C terminus truncated SDG2 fragments as baits and were individually fused to the Gal4 DNA-binding domain (BD) on the bait plasmid pGBKT7 and transformed into yeast strain AH109. Next, an *Arabidopsis* cDNA library is fused to the Gal4 activation domain (AD) and transformed into the bait strains. Yeast colonies containing the cDNAs that encode SDG2-interacting proteins will be identified as those that are able to grow on media in the absence of Trp, Leu, His and Ade. We focused on putative interaction partners that were found at least twice in the screen or were known to be involved in chromatin regulation. The positive colonies from screening were re-plated again with 3-AT to confirm the interactions. Using the C terminus construct, At2g34730 was identified seven times out of 39 positive colonies, At4g14760 four times and At3g22790 two times. Using the middle construct, At1g75950 was identified 24 times out of 135 positive colonies. The N terminus construct exhibited intrinsic auto-activation activity, and none was identified. At2g34730 was the first SDG2 interacting protein identified, thus was named SIP1. At4g14760 and At3g22790 are annotated as kinase interacting proteins, and were named KIP1a and KIP1b, respectively. At1g75960 is annotated as Arabidopsis-SKP1-like (ASK) gene. At1g28520 is annotated as VOZ1 (vascular plant one- zinc finger 1).

Confirmation of SDG2 interacting proteins were done with three steps: first the positive colony was re-plated with serial dilution, and if the growth was comparable with positive control, the cells was plated on minimum media containing 3-AT or X-α-Gal. All five genes have passed all three tests and were analyzed further. An example of the representative result for confirmation of interaction between SIP1 and SDG2 was shown in Fig. 4.3.

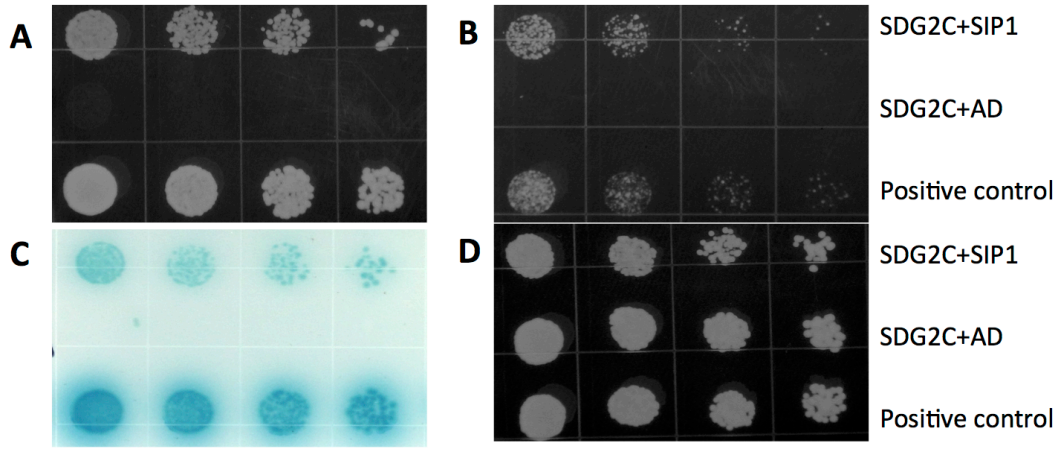

Fig. 4.3 Interaction between SIP1 and SDG2. SIP1 was screened from a yeast two-hybrid screen using a truncated C terminus of SDG2 containing the SET domain and pre-SET domain. Growth of serial dilutions on A) selective SD/His-/Leu-/Trp-media was detected when SDG2 was co-transformed with SIP1 but not with AD empty vector. B) selective SD/His-/Leu-/Trp- media with 30mM 3-AT C) selective SD/His-/Leu-/Trp- media plus X-α-Gal. Growth was not affected on D) SD/Leu-/Trp- media.

Interestingly, all five genes contain a SMC (Structural Maintenance of Chromosomes) like coiled coil domain. Generally SMC proteins are critical for a wide range of processes including chromosome structure and dynamics, gene regulation and DNA repair. Eukaryotes have three SMC complexes, consisting of hetero-dimeric pairs of six different SMC proteins along with several specific regulatory subunits. In addition to their other functions, all three SMC

complexes play distinct roles in DNA repair. Cohesin (SMC1–SMC3) is involved in DNA double-strand break repair, condensin (SMC2–SMC4) participates in single-strand break (SSB) repair, and the SMC5–SMC6 complex functions in various DNA repair pathways (Griese et al. 2010). However, unlike the SMC proteins, the five identified SDG2-interacting proteins only have two or more coiled-coiled domains but not the ATPase domain. A similar scenario in another SET domain protein CLF (curly leaf) that catalyzes H3K27me3 has been shown to interact with BLISTER, which also contains the SMC like coiled coil domain (Schatlowski et al. 2010). Another protein named DMS3 also contains this coiled coil domain, and has been shown to function in a complex with the potential for driving a dynamic component of the RNA-directed DNA methylation machinery (Lorkovic et al. 2012). Therefore, it is very promising that SDG2 is been targeted to chromatin via protein-protein interactions with these SMC like domain containing proteins.

Genetic analysis of SDG2 interacting protein mutants

We got T-DNA insertion lines for all genes except At1g75960, for which no T-DNA insertion line is available. Unfortunately, none of the T-DNA line caused a developmental phenotype. We also got the T-DNA insertion line for BLI (Schatlowski et al. 2010), but failed to observe the described phenotype, either. The double mutant between KIP1a and KIP1b that did exhibit a sever phenotype showed a segregating ratio of 1:3 in T2, indicating a T-DNA may have landed in a different loci independent of the two genes. Even though only VOZ1 was detected from yeast two-hybrid, the double mutant between VOZ1 and its homologue VOZ2 exhibited a slight late flowering phenotype rather than early flowering mimicking *sdg2-/-*, and either genomic copy of VOZ1 or VOZ2 rescued the double mutant phenotype, indicating VOZ1 and VOZ2 regulate

flowering time redundantly (Yasui et al. 2012). Therefore, a closer phenotypic analysis may be needed to identify possible mutant phenotypes with SDG2 interacting genes. Also, crosses among SDG2 and interacting protein genes, may need to be done to further elucidate the underlying mechanism of SDG2 targeting.

One of the genes identified by yeast two-hybrid At1g75960 may play a role in directing SDG2 degradation. This gene is annotated as ASK1 and it belongs to a well characterized E3 ligase SCF (SKP1, cullin/CDC53, F-box protein) complex, which is typically composed of a CULLIN1 (CUL1) scaffold protein; a catalytic RING domain protein, RING-BOX1 (RBX1); an adaptor protein called S-PHASE KINASE-ASSOCIATED PROTEIN1 (SKP1) in animals and ARABIDOPSIS SKP1 HOMOLOG1/2 (ASK1/2) in *Arabidopsis* (Zhao et al. 2003). Recent study showed that posttranslational regulation of H3K27 methyltransferase CURLY LEAF (CLF) by the F-box protein, UP-WARD CURLY LEAF1 (UCL1) does exist in *Arabidopsis* (Jeong et al. 2011). Similarly, ASK1 identified here may as well lead to posttranslational regulation of SDG2. However, the *Arabidopsis* genome encodes 21 Skp1-related genes thus the functional redundancy remains an issue.

4.4 Conclusions

We propose that the SMC like coiled-coil domain specifically interacts with SDG2 and possibly functions to either recruiting SDG2 to its target chromatin location or directing post-translational regulation of SDG2. The genes identified in this yeast two-hybrid screening share structure similarity but are distinct from BLI that is identified in CLF yeast two-hybrid, indicating the functional specificity of SMC like coiled-coil domain. Further analysis will be required to confirm the interactions *in vivo.* Various crosses need to be made and the mutants need to be both phenotypically and at the molecular level.

CHAPTER 5

5. CHARACTERIZATION AND PROFILING OF A NEW HISTONE H3 DOUBLE MARK H3K27ME3S28P IN *ARABIDOPSIS* GENOME

5.1 Introduction

Nucleosomes are composed of DNA wrapped around histone octamers of two copies of H2A, H2B, H3 and H4. Histones are subjects to posttranslational modifications (PTMs) such as methylation, phosphorylation, acetylation and ubiquitination (Portela and Esteller 2010). Whole genome level profiling revealed strong associations of certain PTMs with active or repressed genes. In fact, the histone modifications or combinations of histone modifications a gene occupies could readily predict its expression status (Mikkelsen et al. 2007a; Wang et al. 2009; Roudier et al. 2011; Dong et al. 2012). To ensure proper developmental progression, gene expression must be spatially and temporally regulated, which in turn is associated with dynamic regulation of histone modifications. H3 Lysine 27 trimethylation (H3K27me3) is one of the modifications that play important roles in regulating normal development of animals and plants. As has been demonstrated in plants, insects and animals, H3K27me3 generally functions to suppress the expression of its target genes through Polycomb group (PcG) complexes. This repressive mark is catalyzed by the PRC2 complex through its enzymatic subunits EZH1 and EZH2, which then could be bind by Pc component in PRC1 complex that recruits other factors such as HP1 to further maintain suppression. Details of PRC1 and PRC2 mediated gene repression have been reviewed in multiple literatures (Schuettengruber et al. 2007; Simon and

Kingston 2009; Margueron and Reinberg 2011; Holec and Berger 2012; Schwartz and Pirrotta 2013; Kim and Sung 2014) and will not be discussed here.

Particularly, tri-methylation on lysine 27 of H3 were found to be largely restricted to transcriptionally repressed genes in *Arabidopsis*, and is required for proper expression of developmentally important genes in plants. H3K27me3 does not occupy large genomic regions and the maintenance is largely independent of DNA methylation or RNA interference, suggesting different mechanisms for establishment and spreading of this mark in plants and in animals. (Zhang et al. 2007b; Zhang et al. 2007c). However, the core components of PRC2 are conserved and there is strong genetic and biochemical evidence to suggest that they associate to form at least three similar complexes in Arabidopsis: the FIS (FERTILIZATION INDEPENDENT SEED) complex which controls seed development, the VRN (VERNALIZATION) complex which mediates the vernalization response and the EMF (EMBRYONIC FLOWERING) complex which represses precocious flowering and flower development (Schatlowski et al. 2008). Homologues to the catalytic component E(z) are identified as CLF (curly leaf, SDG1)(Chanvivattana et al. 2004), SWN (Swinger, SDG10) as well as the imprinting gene MEA (Medea, SDG5)(Wang et al. 2006).

The mechanism that how repressed genes marked by H3K27me3 could be activated or *vise versa* during development is largely unknown. But PcG mediated silencing is generally antagonized by Trithorax-group genes (Trx-G) that catalyze active chromatin marks like H3K4me3 and H3K36me3 (Schuettengruber et al. 2007; Schatlowski et al. 2008; Schmitges et al. 2011; Yuan et al. 2011). Studies in ES (embryonic stem) cells have identified a bivalent state of H3K4me3 and H3K27me3 at target genes that are mainly involved in cell fate determination (Bernstein et al. 2006; Mikkelsen et al. 2007b) when they are mostly resolved in one way or the other. The

presence of bivalent chromatin domains has been further supported by the fact that H3K27me3 often co-exist with H3K4me3 or H3K36me3 on tails of two H3s from the same mono-nucleosome (van Rossum et al. 2012). But genome wide co-occurrence of H3K4me3 and H3K27me3 has not been identified in plants (Zhang et al. 2009b), or the seemingly co-localization is largely due to mixed tissue and cell types(Lafos et al. 2011).

A closer examination of the histone H3 tail shows that methylated lysine often co-localizes with serine or threonine phosphorylation (Portela and Esteller 2010). As a matter of fact, the concept "physopho-methyl switch" has been first proposed as the dynamic addition of a covalent modification of a site adjacent to a previously set covalent mark that has engaged an effecter module leads to consecutive loss of binding to that factor (Fischle et al. 2003). Two independent studies showed the kinase activity of Aurora B on H3S10 phosphorylation causes HP1 dissociation from chromatin in mitosis correlating with H3K9me3(Fischle et al. 2005; Hirota et al. 2005). Similarly, the removal of H3K27me3 thus activation of developmental or tissue-specific genes may be induced by upstream kinase activities that catalyzes serine 28. A recent study in human suggested that cross talk between H3K27me3 and phosphorylation of H3 serine28 likely functions in gene de-repression and activation of H3K27me3 target genes(Gehani et al. 2010). The authors showed that the kinase MSK mediated H3S28p at the presence of H3K27me3 leads to displacement of PcG proteins of the PRC1 and PRC2 complexes.

Taking advantage of the antibody Gehani et al. generated that recognizes co-localization of H3K27me3S28p, we explored the de-repression mechanism of PcG target gens in *Arabidopsis*. We first tested if the double mark of H3K27me3S28p does exist in *Arabidopsis*, and then applied ChIP-seq to examine its localization in the genome. Our results confirmed that the antibody is highly specific and only recognizes the histone H3 peptide that is modified both with K27me3

and S28ph. Western result showed the occurrence of this double mark in *Arabidopsis* (result not shown) and immunofluorescence staining indicated its distribution was largely dispersed rather than restricted to heterochromatic regions in the nucleus. ChIP-seq results showed the double mark distribution is distinct from that of H3K27me3. More intriguingly, there seems to be a peak right around TSS (transcription start site), and the level is positively correlated with transcription levels, indicating it might function in transcription activation. This study is the first to show the presence and distribution of this double H3K27me3S28ph mark in *Arabidopsis*, and it provides valuable insights in elucidating the de-repression mechanism of H3K27me3 target genes.

5.2 Results and Discussion

H3K27me3S28p antibody specifically recognizes only the co-modified peptide

We first tested the specificity of H3K27me3S28p antibody by dot blot. As shown in Figure 5.1, this antibody recognizes the peptide only when it contains the modifications of interest at both positions K27me3 and S28ph.

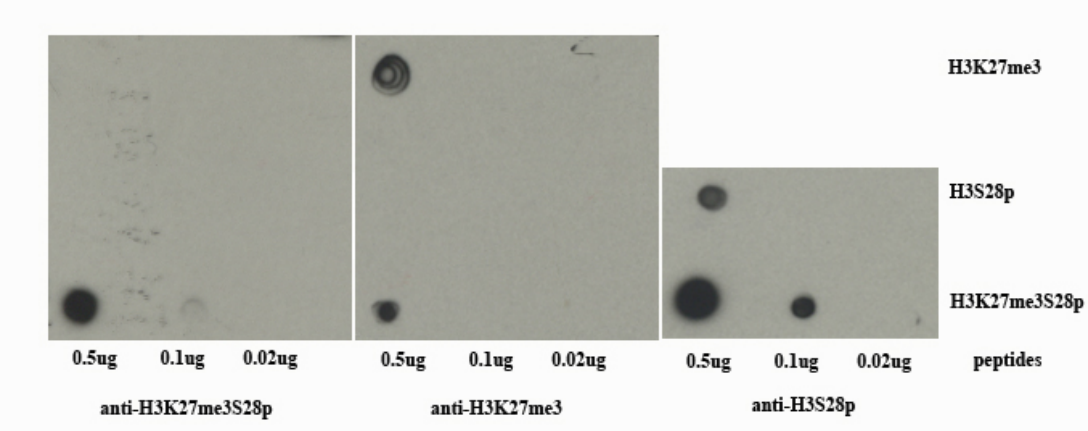

Fig. 5.1. A Dot Blot analysis of antibodies against H3K27me3S28p, H3K27me3 and H3S28p with peptides of modified histone H3. Peptide positions are indicated on the

right side, antibodies are indicated bellow with the amount showing for each spot. Catalog numbers for each antibody and peptide are available upon request.

Similar to the results in Gehani paper (Gehani et al. 2010), H3S28p antibody does show cross reactivity with H3S28p alone and H3K27me3S28p double mark. However, our result indicated H3K27me3 antibody also recognizes both H3K27me3 and H3K27me3S28p, even though the authors showed negative results for H3K27me3 antibody with the double modified peptide. This problem will be studied when it becomes necessary in the future.

H3K27me3S28p double mark does exist in *Arabidopsis* cells

After we confirmed the specificity of the antibody, we went on to test if H3K27me3 does co-exit with H3S28p in *Arabidopsis*. I first carried out a western analysis with basic proteins extracted from Col inflorescence. The double modification band showed up at the expected size as compared to calf thymus histones (results not shown).

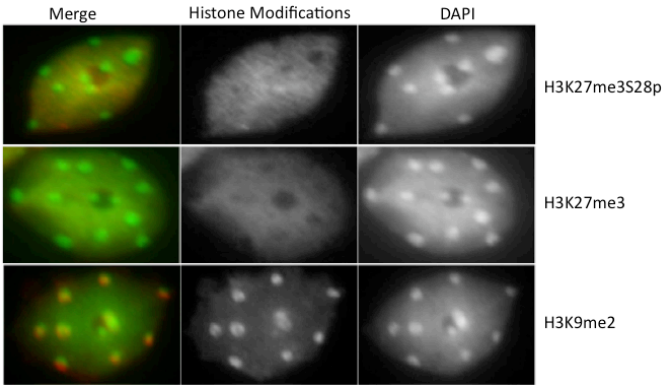

Fig. 5.2. Immunofluorescence staining for antibodies against H3K27me3S28p, H3K27me3 and H3K9me2 in *Arabidopsis* leaf cell nuclei. Both H3K27me3S28p and

H3K27me3 are localized to euchromatin regions. H3K9me2 is used as a control for heterochromatin mark. DAPI was artificially colored with green in the merged images.

As this double modification was indicated to function in activating gene transcription in human(Gehani et al. 2010), it should roughly localize to euchromatin regions in the *Arabidopsis* genome. Promisingly, immunofluorescence staining (Fig. 5.2) clearly showed that H3K27me3S28p localized to euchromatin regions, similar to H3K27me3, which was shown to be restricted to repressed genes(Zhang et al. 2007c). It did not co-localize with DAPI staining, unlike H3K9me2, which is believed to be a heterochromatin mark in animals (Wen et al. 2009) and in plants (Bernatavichute et al. 2008; Zhang 2008b).

Gene level distribution of H3K27me3S28p

We got high quality sequencing results with mapping ratio all above 70% (table 5.1) uniquely mapped to *Arabidopsis* genome (Bailey et al. 2013). We acquired genome-wide maps of H3K27me3S28ph and H3K27me3 in *Arabidopsis* seedlings using ChIP-seq. We observed H3K27me3S28ph enrichment along gene bodies with a high basal background level (Fig5.3 A), which is distinct from H3K27me3 enrichment (Fig 5.3 B).

H3K27me3 enrichment is clearly anti-correlated with gene transcription levels, and mainly found at repressed genes. Both were normalized to total H3 ChIP and intriguingly, a peak at around TSS (transcription start site) of actively transcribed genes showed up for H3K27me3S28ph (Fig 5.3 C). This result echoes the finding that H3K27me3S28ph enriches at promoter regions of the PcG target and IE gene *ATF3* (Gehani et al. 2010), and seems to suggest the active role of the double mark in gene transcription. It is of note that the antibody against the double mark does not

cross-react with H3K27me3 (Fig. 5.1), thus the peak is unlikely signals from tri-methylated lysine 27. As a matter of fact, the antibody against H3K27me3 also recognized the double mark (Fig 5.1), and the minor peak in Fig 5.3 D for expression groups 3, 4 and 5 is likely contamination from the double mark.

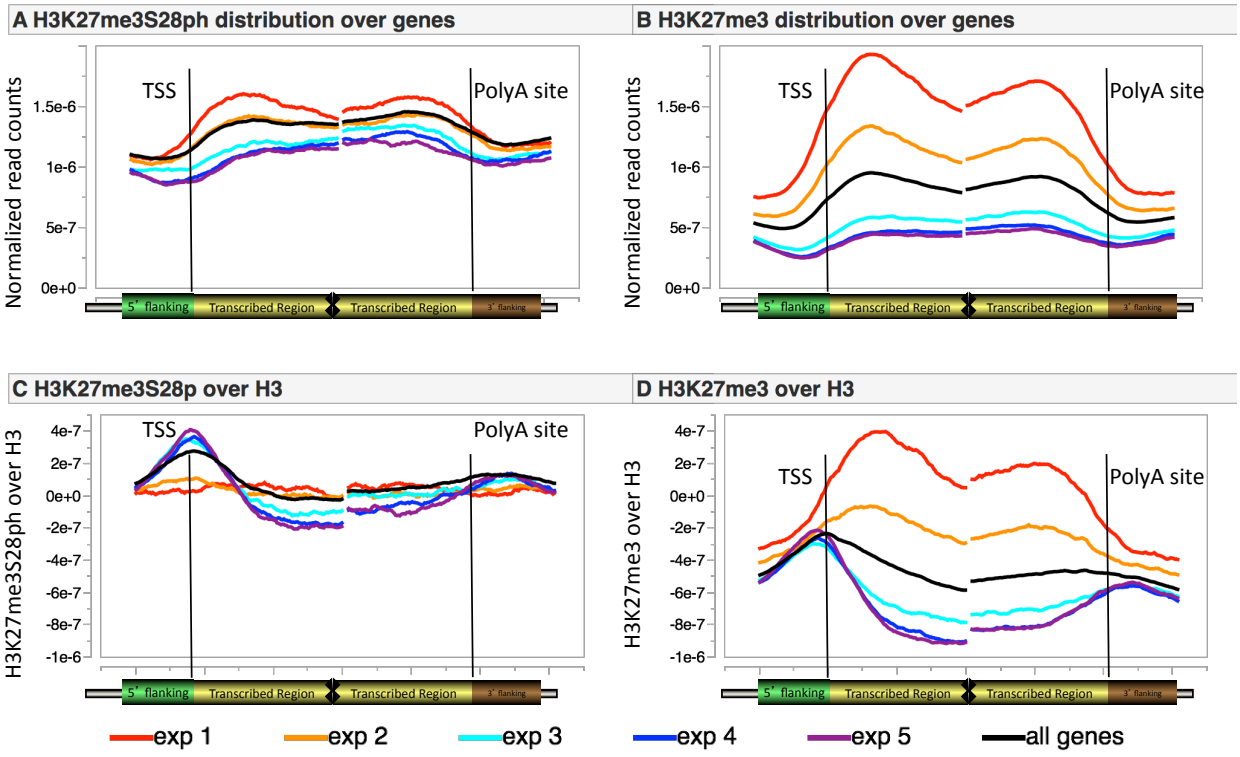

Fig. 5.3 Gene level distribution of H3K27me3S28ph shows a 5' peak at actively transcribed genes. Correlation of H3K27me3S28ph (A) and H3K27me3 (B) with gene expression levels, with exp5 representing the most actively transcribed genes and exp1 the least expressed ones. To further identify patterns, both H3K27me3S28ph and H3K27me3 were normalized again with total H3 ChIP. Average distribution of normalized reads in 10bp bins +500bp and -1kb of TSS, +1kb and -500bp of TTS were plotted.

Table 5.1 Mapping statistics for H3K27me3S28ph ChIP-seq

Sample Name	Total Raw Reads	Mapping Ratio	Total Mapped Reads
H3K27me3	9613900	90.53%	8703796
H3K27me3S28p	9511603	74.28%	7065076
H3	15270831	96.94%	14803894

To further confirm the enrichment of H3K27me3S28ph at TSS regions, two approaches could be taken. First, chromatin can be pre-treated with phosphotase followed by ChIP-seq. This treatment is expected to remove all phospo groups, resulting in decreased or even absense of the double mark signal. Alternatively, one could take advantage of the H3K27 tri-methyltransferase mutant clf/swn, which shows a global decrease in H3K27me3 level, and see if the absence of H3K27me3 will lead to the absence of the double mark.

Identification of kinases that are likely to phosphorylate H3S28 in *Arabidopsis*

The most affinitive way to elucidate the de-repression mechnism would be to identify kinas(es) that phosphorylate H3S28 at the presence of H3K27me3. After a literature search, two kinases caught our attention: Arora3 and MSK-like proteins. Aurora proteins play critical roles in mitosis. Arabidopsis genome encodes three aurora proteins (Fig. 5.4 left): AUR1 and AUR2 share a high sequence similarity and are referred as the alpha group, while AUR3 belongs to the beta group(Demidov et al. 2005). The two groups of aurora proteins are functionally divergent and occupy different subcellular localizations during cell divisions(Demidov et al. 2005; Kurihara et al. 2006; Kurihara et al. 2008; Van Damme et al. 2011). AUR1 is able to phosphorylate H3S10 *in vitro* (Demidov et al. 2009), while AUR3 has been shown to be able to phosphorylate both

H3S10 and S28 in tobacco BY-2 cells(Kurihara et al. 2006). All these evidences suggest that AUR3 should be verily likely a cell-cycle dependent H3S28 phosphorylation kinase.

In human, MSK mediated H3K27me3S28 phosphorylation leads to polycomb group protein displacement and is believed to be able to activate a subset of PcG target genes (Gehani et al. 2010). The authors showed that unlike H3S28p alone, which accumulates strongly during G2/M phase in the cell cycle, this double mark is present all through the cell cycle. This indicates that MSK is probably a cell-cycle independent H3 phosphorylation kinase. A BLAST search identified several MSK-like proteins in the *Arabidopsis* genome, and Fig.3 right shows the relationships among MSKs in human and *Arabidopsis*. *Arabidopsis* genome encodes two proteins that are closely related to human MSK1 and MSK2. Interesting enough, these two genes are tandem duplications to each other, which may indicate their potential significances.

One aspect of the phospho-methyl switch is that the kinase activity on serine residues often leads to dissociation of methyl lysine binding proteins (Fischle et al. 2005; Hirota et al. 2005; Gehani et al. 2010). It is of particular interest to see if the *Arabidopsis* H3K27me3 binding protein LHP1 (Turck et al. 2007; Zhang et al. 2007b) would fail to bind at the presence of phosphorylated S28. If this is the case, we need to reconsider the demethylation of H3K27me3 as the key mechanism for gene derepression during development. If demethylation activity is blocked by S28 phosphorylation, the H3K27me3S28 phosphorylation mechanism for Polycomb group protein displacement speaks for a more transient nature of derepression as compared to regions where H3K27me3 is removed by JmjC proteins. We also need to show if the presence of this double mark indeed leads to gene activation. The fact that unlike in animals, anti-H3K27me3 antibody did recognize a portion of the double mark makes is urgent to further examine the target genes of

both modifications. It may be complicated to understand the genome wide cross talk between H3K27me3 and H3S28ph in *Arabidopsis* because the anti-H3S28ph antibody did not work well in plants. All in all, this study is the first step in understanding the de-repression mechanism of PcG target gens in *Arabidopsis*.

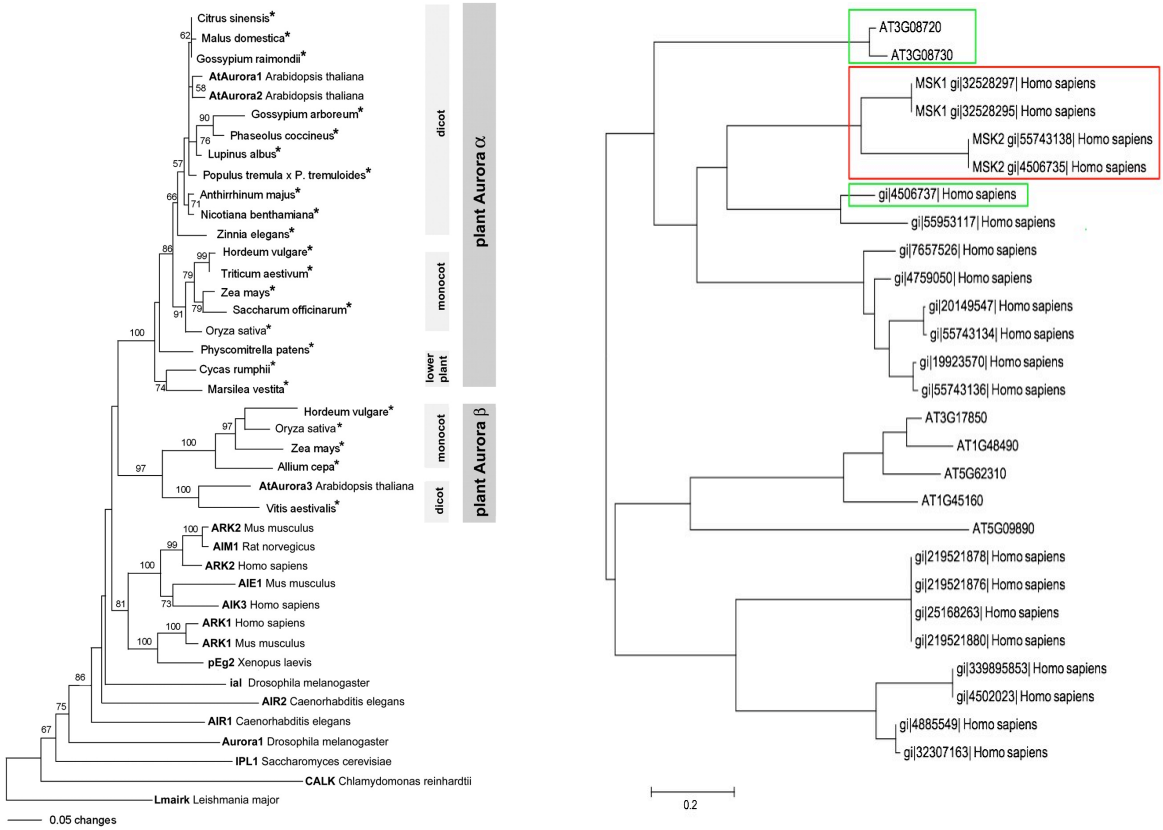

Fig. 5.4. Putative H3S28 kinases. Left: Neighbor-Joining analysis shows the relationships among Kinase domains of Aurora Kinases (Demidov et al. 2005). Right: Maximum Likelyhood tree shows there are two putative MSK proteins in *Arabidopsis*. Genes in the red box are the four MSK proteins from Gehani paper (Gehani et al. 2010). Genes in the green box are reciprocal best BLAST hits to each other.

5.3 Materials and Methods

Plant Material

Arabidopsis Col seeds were planted in soil and placed at 4°C for 3 days to break dormancy then grown under continuous light at 22°C. The green part of three-week old seedlings were harvested and X-linked for ChIP. Fresh leaves from seedlings were used in immuno-staining. Inflorescences are collected for basic protein extraction used in western blot.

Dot blot

Draw 3*3 1cm grid by pencil on nitrocellulose membrane to indicate the region to blot. Using narrow-mouth pipette tip, spot a series of commercial histone H3 peptides onto the nitrocellulose membrane at the center of the grid. Minimize the area that the solution penetrates by applying it slowly. Let the membrane dry. Block non-specific sites by soaking in 5% BSA in TBS-T (0.5-1 hr, RT) in a 10cm Petri Dish for reaction chamber. Incubate with primary antibody dissolved in BSA/TBS-T for 30 min at RT: 1:200 anti-H3K27me3S28ph antibody (CS-091-100; diagenode), 1:500 anti-H3K27me3 antibody (07-449; Millipore), and 1:500 anti H3S28ph antibody (07-145: Millipore). Wash three times with TBS-T (3 x 5 min). Incubate with secondary antibody conjugated with HRP: 1:3000 anti-mouse () and 1:2500 anti-rabbit () for 30 min at RT. Wash three times with TBS-T (15 min x 1, 5 min x 2), then once with TBS (5 min). Incubate with ECL reagent for 1 min, cover with Saran-wrap (remove excessive solution from the surface), and expose X-ray film in the dark room.

Basic protein extraction and western blot

Histones were extracted from Col inflorescence tissue using a standard acid extraction protocol. Grind ~3g of Col inflorescence tissue in liquid nitrogen to fine powder; transfer to 10ml (can be

scaled down based on how much tissue) chilled NIB buffer (0.25M sucrose, 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 1mM $CaCl_2$, 15mM PiPes, 0.8% Triton X-100, 1mM EDTA, add 1 Roch complete mini (4693159001; Roche) per 10mL buffer, 100mM PMSF and 0.7mg/mL pepstatin right before use) on ice; mix well. Filter twice through a single layer of Mira cloth on ice, first into a 30 ml corex tube and then into a 15 ml corex tube; Centrifuge @ 10,000g for 20 minutes at 4°C; discard supernatant. Add 0.5 ml 0.4M H_2SO_4 (chilled on ice, do not scale down) to the pellet and resuspend; leave on ice for 1 hour (this extracts basic proteins). Transfer to a 1.5 Eppendorf tube and centrifuge in a microfuge @ top speed for 5 minutes at 4°C; transfer supernatant to a chilled 15 ml corex tube. Repeat H_2SO_4 extraction but without the 1 hour incubation on ice; pool the supernatant into the 15 ml corex tube (total= 1ml); add 12 ml of acetone. Precipitate overnight at -20°C. Centrifuge @ 7,000 rpm for 15 minutes at 4°C; remove supernatant and air dry pellet in a fume hood. Resuspend pellet in 4M urea (100-200 ul) and quantify total protein (Bradford, should be in the range of 5-10 ug/ul). Add 2× Laemmli buffer to solution; 65-70 °C for 3 minutes; store at -20°C or run samples on SDS-PAGE gel for electrophoresis and proceed to western blot.

Immunofluorescence staining

For the detection of chromatin proteins, in most cases isolated nuclei fixed in formaldehyde are required to preserve the targeted antigenic epitopes.

Cut 4-6 leaves from 3 week old plants (or other tissues) and place in 10 ml of ice-cold 4% paraformaldehyde in Tris buffer for 20 min (rotating 4°C in a 15 ml tube). Blot on filter paper. Wash the sample in 10 ml of ice-cold Tris buffer (10 mM Tris-HCl pH 7.5, 10 mM Na EDTA, 100 mM NaCl) for 2 x 10 min (rotating 4°C). Blot on filter paper between washes. Chop tissue

with new razor blade in glass petri dish in 200 ul cold LB01 buffer (15 mM Tris-HCl pH 7.5, 2 mM NaEDTA, 0.5 mM spermine-4HCl, 80 mM KCl, 20 mM NaCl, 0.1% Triton X-100) until very fine. Strain through 5 ml tube with cell-strainer cap (Fisher). Wash chopped tissue with additional 100-200 ul LB01 and add to strainer. Pipet 5 ul of nuclei suspension directly in 12 ul of sorting buffer (100 mM Tris-HCl, 50 mM KCl, 2 mM MgCl2, 0.05% Tween 20, 5% sucrose) on slide. Air dry approximately 30 minutes. Proceed to antibody detection (below), or store in tightly sealed box at –20°C.

Remove slides from freezer and let warm to room temperature without opening slide box. Etch around spots before fixing. Post-fix the preparation on slides in 4% paraformaldehyde in 1x PBS at RT for 15-20 min. Slide holder holds entire 150 ml. Remove slides from paraformaldehyde and put in slide holder with 150 ml PBS for 5 min. Repeat wash. Remove slides, wipe bottom, blot side on filter paper. Add 25 ul of Block to sample spot on slide. Quickly cover with small square of parafilm (do not let dry). Put slides in moist box and cover. Incubate 37°C 30 min. Remove parafilm, blot and quickly add 25 ul of diluted primary antibody in antibody dilution buffer (1% BSA, 10% horse serum, 0.1% tween20 in 1x PBS). Cover with parafilm, place in moist box and store overnight at 4°C. Next day, rinse in 1x PBS at RT for 3x5 min. Add 25 ul of diluted secondary antibody to sample spot on slide and cover with parafilm. Incubate at RT in moist chamber for 1.5 hours in dark. Rinse in 1x PBS at RT for 3x5 min (incubate in slide holder in dark). Add 10 ul of Vectashield mounting medium containing DAPI (diluted). Cover with glass coverslip. Press dry between filter papers and seal with clear nail polish. Let dry and store in light-tight microscope slide storage box at 4oC until analysis. At the microscope, put one drop of oil on cover slip where sample is. Find nuclei using DAPI filter, magnification 10X. Then switch to 100X and refocus. Change filter to look for fluorescence.

ChIP-seq libarary prep

Grind the X-linked tissue in liquid nitrogen to a fine powder; add the powder to 10 volumns of Extraction Buffer 1 (0.4M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl, 0.1mM PMSF, 2 Roche mini tablets/ 100 ml, 1mM EDTA) in a 50 ml falcon tube. Leave on ice for 5 min. Filter the solution through Miracloth into a 50 ml falcon tube. Repeat for a total of two filters, and then spin the filtered solution for 20 min at 4000 rpm in a centrifuge at 4 degrees. Gently remove supernatant and resuspend the pellet in 1 ml of Extraction Buffer 2 (0.25M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl2, 1% Triton X-100, 5mM BME, 0.1mM PMSF, 1 Roche complete mini/10ml buffer, 1mM EDTA). Transfer the solution to 1.5 ml eppendorf tube and centrifuge at 12,000 g for 10 min at 4 degrees. Remove supernatant and resuspend pellet in 300 ul of Extraction Buffer 3 (1.7M sucrose, 10mM Tris-HCl pH 8, 0.15% Triton X-100, 2mM MgCl2, 1mM EDTA, 5mM BME, 0.1mM PMSF, 1 Roche complete mini/10ml buffer). In a clean eppendorf tube, add 300ul of Extraction Buffer 3. Take the 300ul resuspended pellet from previous step and carefully layer it on top of the clean 300ul of Extraction Buffer 3. Spin for 1 hour at 16,000 (full speed) at 4 degrees. Remove the supernatant, and resuspend the pellet in 300uL nuclei lysis buffer (50mM Tris 8.0, 10mM EDTA, 1% SDS, 1 Roche complete mini tablet per 5ml total volume, 0.1mM PMSF). This is isolated chromatin, save 10uL to check sonication efficiency. Proceed to sonication or freeze at -20°C.

Sonicate the chromatin in a Bioruptor, save 10uL sonicated chromatin as input. Then the rest sonicated chromatin was diluted to 3mL using ChIP dilution buffer (1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris 8.0, 167mM NaCl, 0.1mM PMSF, 1 Roche complete mini tablet per 10ml total volume), then 1ml diluted chromatin was immunoprecipitated by rotating over night at 4°C

with 10ug of anti-H3K27me3S28ph antibody (CS-091-100; diagenode), 5ug anti-H3K27me3 antibody (07-449; Millipore) and 1ug anti-H3 antibody (ab1791; Abcam), respectively, that are pre-immobilized with 50uL dynabeads proten G (10004D; Novex). Beads were washed subsequently with low salt wash buffer (150mM NaCl, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris 8.1), high salt wash buffer (500mM NaCl, 0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris 8.1), LiCl wash buffer (0.25M LiCl, 1% NP40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris 8.1) twice and TE buffer (10mM Tris 8.0, 1mM EDTA) once. Beads were eluted for 15min with gentle agitation using 250uL elution buffer (1% SDS, 0.1M $NaHCO_3$) twice at 65°C for a total of 500uL. IP samples and input were reverse X-linked with 0.2M NaCl for 6hr, and DNA was recovered by phenol/chloroform. The pellet was diluted in 50uL water. Illumina libraries for sequencing were generated, and the libraries were sequenced using the HiSEq 2000 Sequencing System (Illumina), following the manufacturer's instructions.

ChIP-seq data analysis

To map the ChIP-seq reads, *A. thaliana* reference genome sequence and gene models were retrieved from The A. thaliana Information Resource (TAIR10; http://www.arabidopsis.org). ChIP-seq reads were mapped to the genome with bowtie2 (Langmead and Salzberg 2012). The gene level profiles were generated by counting reads in 10bp windows at the 5' and 3' transcriptional boundaries, and then aligned by TSS (transcription start site) and TTS (transcription termination site) annotated in tair10.

CHAPTER 6

6. CONCLUSIONS

In summary, the data presented in chapter 2 provided a comprehensive and high-resolution view of the transcriptional and chromatin-based events during heat stress, which is summarized in table 6.1. We found that Arabidopsis responded to heat stress with rapid and high-level

Table 6.1 Interplay between transcription and chromatin state

	Inactive	Active	"Super Active"
Nucleosomes	Disorganized	Well-positioned	None
H3.1	High	Low	None
H3.3	Low	High (3' end)	None
H2A.Z	High (broad)	High (5' end)	None
H3K4me3	Low	High (5' end)	None
H3K27me3	High (broad)	Low	None
H3K36me3	Low	High (broad)	None
H3/H4ac	Low	High (5' end)	None

recruitment of Pol II to heat-inducible genes, which was concomitant with the loss of preexisting histones. Transcription proceeded on nucleosome-free template DNA for approximately an hour, followed transcriptional attenuation and redeposition of newly synthesized histones. Importantly, large-scale redistribution of Pol II among specific sets of genes appeared to have contributed to the rapid transcriptional activation upon heat stress.

The dynamic histone replacement and Pol II redistribution have not been described during the transcriptional activation of developmentally controlled genes. For example, a recent study of the developmental transition between dark and light growth during germination showed no gross chromatin changes at activated genes. It is interesting to consider that both histone replacement and Pol II redistribution may be specific to processes involving rapid transcriptional activation to extraordinarily high levels. In this regard, we note that many other abiotic stresses are also capable of inducing transcriptional changes in manners similar to heat. Future studies should determine whether the histone replacement and Pol II redistribution observed here are shared by all stress-response processes.

The combined genetic, biochemical, and informatics analyses performed by many laboratories demonstrate that the genome-wide pattern of nucleosome positioning is determined by the combination of DNA sequence, nucleosome remodelers, DNA methylation, transcription factors including activators, components of the pre-initiation complex, and elongating Pol II, and histone modifications and histone variants. Here we studied the effects of transcription on nucleosome organization by examining nucleosome positioning over genes that are differentially expressed in *Arabidopsis* seedlings v.s. flowers. Essentially we are looking at the same genes, rather than comparing different genes based on their expression. Moreover, studying in a developmental

context also helps to rule out the possible complication caused by DNA motifs found at promoters of stress-inducible genes. This is not studying tissue specific nucleosome organizations but using differentially expressed genes in such context to investigate the effects of transcription on nucleosome organization. Our results indicate that mononucleosomes from floral tissue were somehow more readily released from chromatin than those from seedling tissue. The overall nucleosome occupancy in flowers was significantly lower than those in seedlings, making it inconclusive to say if differential gene transcription actually affected nucleosome organization. Interestingly, we did not observe any obvious difference in nucleosome phasing on differentially expressed genes, indicating transcription either does not influence nucleosome positioning to an extend that is readily observable, or other factors may have buried the differences caused by transcription.

We studied the targeting mechanisms for H3K4me3 establishment in *Arabidopsis* by characterizing the major tri-methyltransferase SDG2. We found that the SMC like coiled-coil domain specifically interacts with SDG2 and possibly functions to either recruiting SDG2 to its target chromatin location or directing post-translational regulation of SDG2. The genes identified in this yeast two-hybrid screening share structure similarity but are distinct from BLI that is identified in CLF yeast two-hybrid, indicating the functional specificity of SMC like coiled-coil domain. Further analysis will be required to confirm the interactions *in vivo.* Various crosses need to be made and the mutants need to be both phenotypically and at the molecular level.

Lastly, we tried to reveal the mechanism of de-repression of H3K27me3 target genes by profiling a new double histone mark H3K27me3S28ph. ChIP-seq results showed the double

mark distribution is distinct from that of H3K27me3. More intriguingly, there seems to be a peak right around TSS (transcription start site), and the level is positively correlated with transcription levels, indicating it might function in transcription activation.

REFERENCES

Adam S, Polo SE, Almouzni G. 2013. Transcription Recovery after DNA Damage Requires Chromatin Priming by the H3.3 Histone Chaperone HIRA. *Cell* **155**(1): 94-106.

Adams-Cioaba MA, Min JR. 2009. Structure and function of histone methylation binding proteins. *Biochem Cell Biol* **87**(1): 93-105.

Ahmad K, Henikoff S. 2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* **9**(6): 1191-1200.

Ali M, Rincon-Arano H, Zhao W, Rothbart SB, Tong Q, Parkhurst SM, Strahl BD, Deng LW, Groudine M, Kutateladze TG. 2013. Molecular basis for chromatin binding and regulation of MLL5. *P Natl Acad Sci USA* **110**(28): 11296-11301.

Alvarez-Venegas R, Avramova Z. 2005. Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive Arabidopsis genes and in atx1 mutants. *Nucleic Acids Research* **33**(16): 5199-5207.

Badeaux AI, Shi Y. 2013. Emerging roles for chromatin as a signal integration and storage platform. *Nature Reviews Molecular Cell Biology* **14**(4): 211-224.

Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li QH, Liu T, Madrigal P, Taslim C, Zhang J. 2013. Practical Guidelines for the Comprehensive Analysis of ChIP-seq Data. *Plos Comput Biol* **9**(11).

Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB. 2001. The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Research* **29**(21): 4319-4333.

Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, Devarajan K, Daugherty AC, Kundaje AB, Mancini E et al. 2014. H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. *Cell* **158**(3): 673-688.

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. 2008. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PloS one* **3**(9): e3156.

Bernstein BE, Mikkelsen TS, Xie XH, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**(2): 315-326.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR et al. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**(10): 1045-1048.

Berr A, Xu L, Gao J, Cognat V, Steinmetz A, Dong AW, Shen WH. 2009. SET DOMAIN GROUP25 Encodes a Histone Methyltransferase and Is Involved in FLOWERING LOCUS C Activation and Repression of Flowering. *Plant Physiology* **151**(3): 1476-1485.

Bintu L, Ishibashi T, Dangkulwanich M, Wu YY, Lubkowska L, Kashlev M, Bustamante C. 2012. Nucleosomal elements that control the topography of the barrier to transcription. *Cell* **151**(4): 738-749.

Bonisch C, Hake SB. 2012. Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic Acids Res* **40**(21): 10719-10741.

Campos EI, Fillingham J, Li G, Zheng H, Voigt P, Kuo WH, Seepany H, Gao Z, Day LA, Greenblatt JF et al. 2010. The program for processing newly synthesized histones H3.1 and H4. *Nat Struct Mol Biol* **17**(11): 1343-1351.

Cazzonelli CI, Nisar N, Roberts AC, Murray KD, Borevitz JO, Pogson BJ. 2014. A chromatin modifying enzyme, SDG8, is involved in morphological, gene expression, and epigenetic responses to mechanical stimulation. *Front Plant Sci* **5**.

Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon YH, Sung ZR, Goodrich J. 2004. Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development* **131**(21): 5263-5276.

Chapman-Rothe N, Curry E, Zeller C, Liber D, Stronach E, Gabra H, Ghaem-Maghami S, Brown R. 2013. Chromatin H3K27me3/H3K4me3 histone marks define gene sets in high-grade serous ovarian cancer that distinguish malignant, tumour-sustaining and chemo-resistant ovarian tumour cells. *Oncogene* **32**(38): 4586-4592.

Charron JB, He H, Elling AA, Deng XW. 2009. Dynamic landscapes of four histone modifications during deetiolation in Arabidopsis. *Plant Cell* **21**(12): 3732-3748.

Chen KF, Xi YX, Pan XW, Li ZY, Kaestner K, Tyler J, Dent S, He XW, Li W. 2013a. DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res* **23**(2): 341-351.

Chen P, Zhao J, Wang Y, Wang M, Long H, Liang D, Huang L, Wen Z, Li W, Li X et al. 2013b. H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. *Genes & development* **27**(19): 2109-2124.

Chittuluru JR, Chaban Y, Monnet-Saksouk J, Carrozza MJ, Sapountzi V, Selleck W, Huang J, Utley RT, Cramet M, Allard S et al. 2011. Structure and nucleosome interaction of the yeast NuA4 and Piccolo‚ÄìNuA4 histone acetyltransferase complexes. *Nat Struct Mol Biol* **18**(11): 1196-1203.

Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**(7304): 388-392.

Churchman LS, Weissman JS. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**(7330): 368-+.

Coleman-Derr D, Zilberman D. 2012. Deposition of Histone Variant H2A.Z within Gene Bodies Regulates Responsive Genes. *PLoS genetics* **8**(10).

Collins RE, Tachibana M, Tamaru H, Smith KM, Jia D, Zhang X, Selker EU, Shinkai Y, Cheng X. 2005. In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem* **280**(7): 5563-5570.

Couture JF, Collazo E, Ortiz-Tello PA, Brunzelle JS, Trievel RC. 2007. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat Struct Mol Biol* **14**(8): 689-695.

Das C, Lucia MS, Hansen KC, Tyler JK. 2009. CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature* **459**(7243): 113-U123.

Deal RB, Kandasamy MK, McKinney EC, Meagher RB. 2005. The nuclear actin-related protein ARP6 is a pleiotropic developmental regulator required for the maintenance of FLOWERING LOCUS C expression and repression of flowering in Arabidopsis. *Plant Cell* **17**(10): 2633-2646.

Deal RB, Topp CN, McKinney EC, Meagher RB. 2007. Repression of Flowering in Arabidopsis Requires Activation of FLOWERING LOCUS C Expression by the Histone Variant H2A.Z. *Plant Cell* **19**(1): 74-83.

Demidov D, Hesse S, Tewes A, Rutten T, Fuchs J, Ashtiyani RK, Lein S, Fischer A, Reuter G, Houben A. 2009. Aurora1 phosphorylation activity on histone H3 and its cross-talk with other post-translational histone modifications in Arabidopsis. *Plant J* **59**(2): 221-230.

Demidov D, Van Damme D, Geelen D, Blattner FR, Houben A. 2005. Identification and dynamics of two classes of Aurora-like kinases in arabidopsis and other plants. *Plant Cell* **17**(3): 836-848.

Dong XJ, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, Gingeras TR, Gerstein M, Guigo R, Birney E et al. 2012. Modeling gene expression using chromatin features in various cellular contexts. *Genome Biology* **13**(9).

Drane P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. 2010. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes & Development* **24**(12): 1253-1265.

Du HN, Fingerman IM, Briggs SD. 2008. Histone H3 K36 methylation is mediated by a trans-histone methylation pathway involving an interaction between Set2 and histone H4. *Genes & Development* **22**(20): 2786-2798.

Elsaesser SJ, Allis CD. 2010. HIRA and Daxx constitute two independent histone H3.3-containing predeposition complexes. *Cold Spring Harb Symp Quant Biol* **75**: 27-34.

Elsasser SJ, Huang H, Lewis PW, Chin JW, Allis CD, Patel DJ. 2012. DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. *Nature* **491**(7425): 560-565.

Fernandez AG, Anderson JN. 2007. Nucleosome positioning determinants. *Journal of molecular biology* **371**(3): 649-668.

Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, Hunt DF, Funabiki H, Allis CD. 2005. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**(7071): 1116-1122.

Fischle W, Wang Y, Allis CD. 2003. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**(6957): 475-479.

Flaus A, Martin DMA, Barton GJ, Owen-Hughes T. 2006. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Research* **34**(10): 2887-2905.

Gaillard PHL, Martini EMD, Kaufman PD, Stillman B, Moustacchi E, Almouzni G. 1996. Chromatin assembly coupled to DNA repair: A new role for chromatin assembly factor I. *Cell* **86**(6): 887-896.

Gehani SS, Agrawal-Singh S, Dietrich N, Christophersen NS, Helin K, Hansen K. 2010. Polycomb Group Protein Displacement and Gene Activation through MSK-Dependent H3K27me3S28 Phosphorylation. *Mol Cell* **39**(6): 886-900.

Gent JI, Madzima TF, Bader R, Kent MR, Zhang XY, Stam M, McGinnis KM, Dawe RK. 2014. Accessible DNA and Relative Depletion of H3K9me2 at Maize Loci Undergoing RNA-Directed DNA Methylation. *Plant Cell* **26**(12): 4903-4917.

Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao YA, Li LP, Adelman K. 2010. Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. *Cell* **143**(4): 540-551.

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo XY, Li X et al. 2010. Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* **140**(5): 678-691.

Griese JJ, Witte G, Hopfner KP. 2010. Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins. *Nucleic Acids Research* **38**(10): 3454-3465.

Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, J √ Πrstad TS, Wilson ZA, Aalen RB. 2009. The ASH1 HOMOLOG 2 (ASHH2) Histone H3 Methyltransferase Is Required for Ovule and Anther Development in <italic>Arabidopsis</italic>. *PloS one* **4**(11): e7817.

Guo L, Yu Y, Law JA, Zhang X. 2010. SET DOMAIN GROUP2 is the major histone H3 lysine 4 trimethyltransferase in Arabidopsis. *Proceedings of the National Academy of Sciences* **107**(43): 18557-18562.

Guo R, Zheng LJ, Park JW, Lv RT, Chen H, Jiao FF, Xu WQ, Mu SR, Wen H, Qiu JS et al. 2014. BS69/ZMYND11 Reads and Connects Histone H3.3 Lysine 36 Trimethylation-Decorated Chromatin to Regulated Pre-mRNA Processing. *Molecular Cell* **56**(2): 298-310.

Hamiche A, Shuaib M. 2012. Chaperoning the histone H3 family. *Bba-Gene Regul Mech* **1819**(3-4): 230-237.

He FH, Umehara T, Saito K, Harada T, Watanabe S, Yabuki T, Kigawa T, Takahashi M, Kuwasako K, Tsuda K et al. 2010. Structural Insight into the Zinc Finger CW Domain as a Histone Modification Reader. *Structure* **18**(9): 1127-1139.

Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. 2009. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* **19**(3): 460-469.

Hirota T, Lipp JJ, Toh BH, Peters JM. 2005. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature* **438**(7071): 1176-1180.

Hodges C, Bintu L, Lubkowska L, Kashlev M, Bustamante C. 2009. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* **325**(5940): 626-628.

Holec S, Berger F. 2012. Polycomb group complexes mediate developmental transitions in plants. *Plant physiology* **158**(1): 35-43.

Hollenbach AD, McPherson CJ, Mientjes EJ, Iyengar R, Grosveld G. 2002. Daxx and histone deacetylase II associate with chromatin through an interaction with core histones and the chromatin-associated protein Dek. *Journal of Cell Science* **115**(Pt 16): 3319-3330.

Hoppmann V, Thorstensen T, Kristiansen PE, Veiseth SV, Rahman MA, Finne K, Aalen RB, Aasland R. 2011. The CW domain, a new histone recognition module in chromatin proteins. *Embo J* **30**(10): 1939-1952.

Horton JR, Upadhyay AK, Qi HH, Zhang X, Shi Y, Cheng XD. 2010. Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nature Structural & Molecular Biology* **17**(1): 38-U52.

Hsieh FK, Kulaeva OI, Patel SS, Dyer PN, Luger K, Reinberg D, Studitsky VM. 2013. Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *P Natl Acad Sci USA* **110**(19): 7654-7659.

Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. 2006. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**(5774): 748-751.

Huang Y, Min S, Lui Y, Sun J, Su X, Liu Y, Zhang Y, Han D, Che Y, Zhao C et al. 2012. Global mapping of H3K4me3 and H3K27me3 reveals chromatin state-based regulation of human monocyte-derived dendritic cells in different environments. *Genes and immunity* **13**(4): 311-320.

Hughes AL, Jin Y, Rando OJ, Struhl K. 2012. A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol Cell* **48**(1): 5-15.

Iberg AN, Espejo A, Cheng D, Kim D, Michaud-Levesque J, Richard S, Bedford MT. 2008. Arginine methylation of the histone H3 tail impedes effector binding. *J Biol Chem* **283**(6): 3006-3010.

Ingouff M, Berger F. 2010. Histone3 variants in plants. *Chromosoma* **119**(1): 27-33.

Ingouff M, Rademacher S, Holec S, Soljlc L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F. 2010. Zygotic Resetting of the HISTONE 3 Variant Repertoire Participates in Epigenetic Reprogramming in Arabidopsis. *Current Biology* **20**(23): 2137-2143.

Jacob Y, Stroud H, Leblanc C, Feng S, Zhuo L, Caro E, Hassel C, Gutierrez C, Michaels SD, Jacobsen SE. 2010. Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* **466**(7309): 987-991.

Jeong CW, Roh H, Dang TV, Choi YD, Fischer RL, Lee JS, Choi Y. 2011. An E3 ligase complex regulates SET-domain polycomb group protein activity in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*.

Jeppsson K, Kanno T, Shirahige K, Sjogren C. 2014. The maintenance of chromosome structure: positioning and functioning of SMC complexes. *Nature Reviews Molecular Cell Biology* **15**(9): 601-614.

Jiang CZ, Pugh BF. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics* **10**(3): 161-172.

Jiang DH, Kong NC, Gu XF, Li ZC, He YH. 2011. Arabidopsis COMPASS-Like Complexes Mediate Histone H3 Lysine-4 Trimethylation to Control Floral Transition and Plant Development. *Plos Genetics* **7**(3).

Jin CY, Zang CZ, Wei G, Cui KR, Peng WQ, Zhao KJ, Felsenfeld G. 2009. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature genetics* **41**(8): 941-U112.

Jin J, Bai L, Johnson DS, Fulbright RM, Kireeva ML, Kashlev M, Wang MD. 2010. Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. *Nature structural & molecular biology* **17**(6): 745-752.

Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE. 2004. Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Research* **32**(22): 6511-6518.

Kaplan N, Moore I, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, Hughes TR, Lieb JD, Widom J, Segal E. 2010. Nucleosome sequence preferences influence in vivo nucleosome organization. *Nature Structural & Molecular Biology* **17**(8): 918-920.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**(7236): 362-U129.

Kelly TK, Liu YP, Lay FD, Liang GN, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res* **22**(12): 2497-2506.

Kim DH, Sung S. 2014. Polycomb-Mediated Gene Silencing in Arabidopsis thaliana. *Mol Cells* **37**(12): 841-850.

Kim S, Lee J, Yang JY, Jung C, Chua NH. 2013. Arabidopsis histone methyltransferase SET DOMAIN GROUP2 is required for regulation of various hormone responsive genes. *Journal of Plant Biology* **56**(1): 39-48.

Kulaeva OI, Gaykalova DA, Pestov NA, Golovastov VV, Vassylyev DG, Artsimovitch I, Studitsky VM. 2009. Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. *Nat Struct Mol Biol* **16**(12): 1272-1278.

Kulaeva OI, Gaykalova DA, Studitsky VM. 2007. Transcription through chromatin by RNA polymerase II: Histone displacement and exchange. *Mutat Res-Fund Mol M* **618**(1-2): 116-129.

Kulaeva OI, Hsieh FK, Chang HW, Luse DS, Studitsky VM. 2013. Mechanism of transcription through a nucleosome by RNA polymerase II. *Bba-Gene Regul Mech* **1829**(1): 76-83.

Kulaeva OI, Studitsky VM. 2010. Mechanism of histone survival during transcription by RNA polymerase II. *Transcription* **1**(2): 85-88.

Kumar SV, Wigge PA. 2010. H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. *Cell* **140**(1): 136-147.

Kurihara D, Matsunaga S, Kawabe A, Fujimoto S, Noda M, Uchiyama S, Fukui K. 2006. Aurora kinase is required for chromosome segregation in tobacco BY-2 cells. *Plant J* **48**(4): 572-580.

Kurihara D, Matsunaga S, Uchiyama S, Fukui K. 2008. Live cell imaging reveals plant Aurora kinase has dual roles during mitosis. *Plant and Cell Physiology* **49**(8): 1256-1261.

Lafos M, Kroll P, Hohenstatt ML, Thorpe FL, Clarenz O, Schubert D. 2011. Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. *PLoS genetics* **7**(4): e1002040.

Lang J, Smetana O, Sanchez-Calderon L, Lincker F, Genestier J, Schmit AC, Houlne G, Chaboute ME. 2012. Plant gamma H2AX foci are required for proper DNA DSB repair responses and colocalize with E2F factors. *New Phytologist* **194**(2): 353-363.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**(4): 357-U354.

Lauberth SM, Nakayama T, Wu XL, Ferris AL, Tang ZY, Hughes SH, Roeder RG. 2013. H3K4me3 Interactions with TAF3 Regulate Preinitiation Complex Assembly and Selective Gene Activation. *Cell* **152**(5): 1021-1036.

Lay FD, Liu YP, Kelly TK, Witt H, Farnham PJ, Jones PA, Berman BP. 2015. The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome Res* **25**(4): 467-477.

Lee J, Thompson JR, Botuyan MV, Mer G. 2008. Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nature Structural & Molecular Biology* **15**(1): 109-111.

Lee WY, Lee D, Chung WI, Kwon CS. 2009. Arabidopsis ING and Alfin1-like protein families localize to the nucleus and bind to H3K4me3/2 via plant homeodomain fingers. *Plant J* **58**(3): 511-524.

Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* **128**(4): 707-719.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.

Li HT, Ilin S, Wang WK, Duncan EM, Wysocka J, Allis CD, Patel DJ. 2006. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**(7098): 91-95.

Li Y, Mukherjee I, Thum KE, Tanurdzic M, Katari MS, Obertello M, Edwards MB, McCombie WR, Martienssen RA, Coruzzi GM. 2015. The histone methyltransferase SDG8 mediates the epigenetic modification of light and carbon responsive genes in plants. *Genome Biol* **16**: 79.

Li ZY, Gadue P, Chen KF, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH. 2012. Foxa2 and H2A.Z Mediate Nucleosome Depletion during Embryonic Stem Cell Differentiation. *Cell* **151**(7): 1608-1616.

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**(6998): 471-476.

Liu YP, Siegmund KD, Laird PW, Berman BP. 2012. Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biology* **13**(7).

Lorkovic ZJ, Naumann U, Matzke AJM, Matzke M. 2012. Involvement of a GHKL ATPase in RNA-Directed DNA Methylation in Arabidopsis thaliana. *Current Biology* **22**(10): 933-938.

Loyola A, Almouzni G. 2007. Marking histone H3 variants: How, when and why? *Trends Biochem Sci* **32**(9): 425-433.

Loyola A, Bonaldi T, Roche D, Imhof A, Almouzni G. 2006. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Mol Cell* **24**(2): 309-316.

Malik HS, Henikoff S. 2003. Phylogenomics of the nucleosome. *Nat Struct Biol* **10**(11): 882-891.

Margueron R, Reinberg D. 2011. The Polycomb complex PRC2 and its mark in life. *Nature* **469**(7330): 343-349.

Masumoto H, Hawke D, Kobayashi R, Verreault A. 2005. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436**(7048): 294-298.

Matthews AGW, Kuo AJ, Ramon-Maiques S, Han SM, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN et al. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**(7172): 1106-U1118.

Maze I, Noh KM, Soshnev AA, Allis CD. 2014. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nature Reviews Genetics* **15**(4): 259-271.

McKittrick E, Gaften PR, Ahmad K, Henikoff S. 2004. Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proceedings of the National Academy of Sciences of the United States of America* **101**(6): 1525-1530.

McWilliam H, Li WZ, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research* **41**(W1): W597-W600.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153): 553-560.

Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A. 2001. COMPASS: A complex of proteins associated with a trithorax-related SET domain protein (vol 98, pg 12902, 2001). *P Natl Acad Sci USA* **98**(26): 15393-15393.

Mito Y, Henikoff JG, Henikoff S. 2005. Genome-scale profiling of histone H3.3 replacement patterns. *Nature genetics* **37**(10): 1090-1097.

Morozov AV. 2011. Sequence determinants of histone-DNA binding preferences: comment on "Cracking the chromatin code: precise rule of nucleosome positioning" by Edward N. Trifonov. *Physics of life reviews* **8**(1): 62-63; discussion 69-72.

Musselman CA, Lalonde ME, Cote J, Kutateladze TG. 2012. Perceiving the epigenetic landscape through histone readers. *Nature Structural & Molecular Biology* **19**(12): 1218-1227.

Nakatani Y, Ray-Gallet D, Quivy JP, Tagami H, Almouzni G. 2004. Two distinct nucleosome assembly pathways: Dependent or independent of DNA synthesis promoted by histone H3.1 and H3.3 complexes. *Cold Spring Harb Sym* **69**: 273-280.

Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. 2013. Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell* **154**(3): 490-503.

Naughton C, Avlonitis N, Corless S, Prendergast JG, Mati IK, Eijk PP, Cockroft SL, Bradley M, Ylstra B, Gilbert N. 2013. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* **20**(3): 387-395.

Neely KE, Workman JL. 2002. The complexity of chromatin remodeling and its links to cancer. *Bba-Rev Cancer* **1603**(1): 19-29.

Noll M. 1974. Subunit Structure of Chromatin. *Nature* **251**(5472): 249-251.

Oh S, Park S, van Nocker S. 2008. Genic and Global Functions for Paf1C in Chromatin Modification and Gene Expression in Arabidopsis. *PLoS genetics* **4**(8): e1000077.

Okada T, Endo M, Singh MB, Bhalla PL. 2005. Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. *Plant J* **44**(4): 557-568.

Ooi SL, Henikoff JG, Henikoff S. 2010. A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Research* **38**(4).

Osley MA. 1991. The Regulation of Histone Synthesis in the Cell-Cycle. *Annu Rev Biochem* **60**: 827-861.

Panchenko T, Sorensen TC, Woodcock CL, Kan ZY, Englander W, Hansen JC, Black BE. 2011. Replacement of Histone H3 with CENP-A Directs Global Nucleosome Array Condensation and Loosening of Nucleosome Superhelical Termini. *Mol Biol Cell* **22**.

Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG. 2006. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**(7098): 100-103.

Polo SE, Roche D, Almouzni G. 2006. New histone incorporation marks sites of UV repair in human cells. *Cell* **127**(3): 481-493.

Portela A, Esteller M. 2010. Epigenetic modifications and human disease. *Nature biotechnology* **28**(10): 1057-1068.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.

Rando OJ, Winston F. 2012. Chromatin and Transcription in Yeast. *Genetics* **190**(2): 351-387.

Ray-Gallet D, Quivy JP, Scamps C, Martini EMD, Lipinski M, Almouzni G. 2002. HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Mol Cell* **9**(5): 1091-1100.

Ray-Gallet D, Woolfe A, Vassias I, Pellentz C, Lacoste N, Puri A, Schultz DC, Pchelintsev NA, Adams PD, Jansen LE et al. 2011. Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol Cell* **44**(6): 928-941.

Roudier F, Ahmed I, Berard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L et al. 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *Embo J* **30**(10): 1928-1938.

Ruthenburg AJ, Allis CD, Wysocka J. 2007. Methylation of lysine 4 on histone H3: Intricacy of writing and reading a single epigenetic mark. *Molecular Cell* **25**(1): 15-30.

Ruthenburg AJ, Li H, Milne TA, Dewell S, McGinty RK, Yuen M, Ueberheide B, Dou Y, Muir TW, Patel DJ et al. 2011. Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell* **145**(5): 692-706.

Sadeh R, Allis CD. 2011. Genome-wide "Re"-Modeling of Nucleosome Positions. *Cell* **147**(2): 263-266.

Saleh A, Alvarez-Venegas R, Yilmaz M, Le O, Hou GC, Sadder M, Al-Abdallat A, Xia YN, Lu GQ, Ladunga I et al. 2008. The highly similar Arabidopsis homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions. *Plant Cell* **20**(3): 568-579.

Sarma K, Reinberg D. 2005. Histone variants meet their match. *Nature Reviews Molecular Cell Biology* **6**(2): 139-149.

Sawatsubashi S, Murata T, Lim J, Fujiki R, Ito S, Suzuki E, Tanabe M, Zhao Y, Kimura S, Fujiyama S et al. 2010. A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. *Genes & Development* **24**(2): 159-170.

Schatlowski N, Creasey K, Goodrich J, Schubert D. 2008. Keeping plants in shape: polycomb-group genes and histone methylation. *Seminars in cell & developmental biology* **19**(6): 547-553.

Schatlowski N, Stahl Y, Hohenstatt ML, Goodrich J, Schubert D. 2010. The CURLY LEAF Interacting Protein BLISTER Controls Expression of Polycomb-Group Target Genes and Cellular Differentiation of Arabidopsis thaliana. *The Plant Cell Online* **22**(7): 2291-2305.

Schmitges FW, Prusty AB, Faty M, Stutzer A, Lingaraju GM, Aiwazian J, Sack R, Hess D, Li L, Zhou SL et al. 2011. Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol Cell* **42**(3): 330-341.

Schmitz RJ, Zhang X. 2011. High-throughput approaches for plant epigenomic studies. *Curr Opin Plant Biol* **14**(2): 130-136.

Schones DE, Cui KR, Cuddapah S, Roh TY, Barski A, Wang ZB, Wei G, Zhao KJ. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**(5): 887-898.

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. 2007. Genome regulation by polycomb and trithorax proteins. *Cell* **128**(4): 735-745.

Schuettengruber B, Martinez AM, Iovino N, Cavalli G. 2011. Trithorax group proteins: switching genes on and keeping them active. *Nature Reviews Molecular Cell Biology* **12**(12): 799-814.

Schwartz BE, Ahmad K. 2005. Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes & Development* **19**(7): 804-814.

Schwartz YB, Pirrotta V. 2013. A new world of Polycombs: unexpected partnerships and emerging functions. *Nature reviews Genetics* **14**(12): 853-864.

Segal E, Fondufe-Mittendorf Y, Chen LY, Thastrom A, Field Y, Moore IK, Wang JPZ, Widom J. 2006. A genomic code for nucleosome positioning. *Nature* **442**(7104): 772-778.

Shi LL, Wang J, Hong F, Spector DL, Fang YD. 2011. Four amino acids guide the assembly or disassembly of Arabidopsis histone H3.3-containing nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America* **108**(26): 10574-10578.

Simon JA, Kingston RE. 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature reviews Molecular cell biology* **10**(10): 697-708.

Singh M, Bag SK, Bhardwaj A, Ranjan A, Mantri S, Nigam D, Sharma YK, Sawant SV. 2015. Global nucleosome positioning regulates salicylic acid mediated transcription in Arabidopsis thaliana. *Bmc Plant Biology* **15**.

Smolle M, Venkatesh S, Gogol MM, Li H, Zhang Y, Florens L, Washburn MP, Workman JL. 2012. Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol* **19**(9): 884-892.

Springer NM, Napoli CA, Selinger DA, Pandey R, Cone KC, Chandler VL, Kaeppler HF, Kaeppler SM. 2003. Comparative Analysis of SET Domain Proteins in Maize and Arabidopsis Reveals Multiple Duplications Preceding the Divergence of Monocots and Dicots. *Plant physiology* **132**(2): 907-925.

Stroud H, Otero S, Desvoyes B, Ramirez-Parra E, Jacobsen SE, Gutierrez C. 2012. Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **109**(14): 5370-5375.

Struhl K, Segal E. 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**(3): 267-273.

Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP et al. 2014. Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in A-thaliana. *Cell Rep* **8**(6): 2015-2030.

Suzuki M, Wang HHY, McCarty DR. 2007. Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiology* **143**(2): 902-911.

Szenker E, Ray-Gallet D, Almouzni G. 2011. The double face of the histone variant H3.3. *Cell Res* **21**(3): 421-434.

Talbert PB, Ahmad K, Almouzni G, Ausio J, Berger F, Bhalla PL, Bonner WM, Cande WZ, Chadwick BP, Chan SWL et al. 2012. A unified phylogeny-based nomenclature for histone variants. *Epigenet Chromatin* **5**.

Talbert PB, Henikoff S. 2010. Histone variants - ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology* **11**(4): 264-275.

Tamada Y, Yun JY, Woo SC, Amasino RM. 2009. ARABIDOPSIS TRITHORAX-RELATED7 Is Required for Methylation of Lysine 4 of Histone H3 and for Transcriptional Activation of FLOWERING LOCUS C. *Plant Cell* **21**(10): 3257-3269.

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**(11): 1025-1040.

Teves SS, Weber CM, Henikoff S. 2014. Transcribing through the nucleosome. *Trends Biochem Sci* **39**(12): 577-586.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**(7414): 75-82.

Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, Lieb JD, Widom J, Segal E, Hughes TR. 2010. High Nucleosome Occupancy Is Encoded at Human Regulatory Sequences. *Plos One* **5**(2).

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**(9): 1105-1111.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**(3): 562-578.

Turck F, Roudier Fß, Farrona S, Martin-Magniette M-L, Guillaume E, Buisine N, Gagnot Sv, Martienssen RA, Coupland G, Colot V. 2007. Arabidopsis TFL2/LHP1 Specifically Associates with Genes Marked by Trimethylation of Histone H3 Lysine 27. *PLoS genetics* **3**(6): e86.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. *Nature* **474**(7352): 516-U148.

Van Damme D, De Rybel B, Gudesblat G, Demidov D, Grunewald W, De Smet I, Houben A, Beeckman T, Russinova E. 2011. Arabidopsis alpha Aurora Kinases Function in Formative Cell Division Plane Orientation. *Plant Cell* **23**(11): 4013-4024.

van Rossum B, Fischle W, Selenko P. 2012. Asymmetrically modified nucleosomes expand the histone code. *Nat Struct Mol Biol* **19**(11): 1064-1066.

Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG et al. 2010. Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* **142**(6): 967-980.

Vermeulen M, Mulder KW, Denissov S, Pijnappel WWMP, van Schaik FMA, Varier RA, Baltissen MPA, Stunnenberg HG, Mann M, Timmers HTM. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**(1): 58-69.

Verreault A, Kaufman PD, Kobayashi R, Stillman B. 1998. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Current Biology* **8**(2): 96-108.

Voss TC, Hager GL. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews Genetics* **15**(2): 69-81.

Wang D, Tyson MD, Jackson SS, Yadegari R. 2006. Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **103**(35): 13244-13249.

Wang WX, Chen Z, Mao Z, Zhang HH, Ding XJ, Chen S, Zhang XD, Xu RM, Zhu B. 2011. Nucleolar protein Spindlin1 recognizes H3K4 methylation and stimulates the expression of rRNA genes. *Embo Rep* **12**(11): 1160-1166.

Wang Z, Schones DE, Zhao K. 2009. Characterization of human epigenomes. *Current opinion in genetics & development* **19**(2): 127-134.

Waterborg JH. 2012. Evolution of histone H3: emergence of variants and conservation of post-translational modification sites. *Biochem Cell Biol* **90**(1): 79-95.

Weber CM, Henikoff S. 2014. Histone variants: dynamic punctuation in transcription. *Genes & development* **28**(7): 672-682.

Weber CM, Ramachandran S, Henikoff S. 2014. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell* **53**(5): 819-830.

Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP. 2009. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature genetics* **41**(2): 246-250.

Wen H, Li YY, Xi YX, Jiang SM, Stratton S, Peng DN, Tanaka K, Ren YF, Xia Z, Wu J et al. 2014. ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* **508**(7495): 263-+.

Wirbelauer C, Bell O, Schubeler D. 2005. Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes & Development* **19**(15): 1761-1766.

Witt O, Albig W, Doenecke D. 1996. Testis-specific expression of a novel human H3 histone gene. *Exp Cell Res* **229**(2): 301-306.

Wollmann H, Holec S, Alden K, Clarke ND, Jacques PE, Berger F. 2012. Dynamic Deposition of Histone Variant H3.3 Accompanies Developmental Remodeling of the Arabidopsis Transcriptome. *PLoS genetics* **8**(5).

Workman JL. 2006. Nucleosome displacement in transcription. *Genes & development* **20**(15): 2009-2017.

Workman JL, Kingston RE. 1998. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annual review of biochemistry* **67**: 545-579.

Wu T, Yuan TZ, Tsai SN, Wang CM, Sun SM, Lam HM, Ngai SM. 2009. Mass spectrometry analysis of the variants of histone H3 and H4 of soybean and their post-translational modifications. *Bmc Plant Biology* **9**.

Xie LQ, Pelz C, Wang WS, Bashar A, Varlamova O, Shadle S, Impey S. 2011. KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *Embo Journal* **30**(8): 1473-1484.

Xu L, Zhao Z, Dong AW, Soubigou-Taconnat L, Renou JP, Steinmetz A, Shen WH. 2008. Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana. *Mol Cell Biol* **28**(4): 1348-1360.

Yang N, Wang WX, Wang Y, Wang MZ, Zhao Q, Rao ZH, Zhu B, Xu RM. 2012. Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1. *P Natl Acad Sci USA* **109**(44): 17954-17959.

Yasui Y, Mukougawa K, Uemoto M, Yokofuji A, Suzuri R, Nishitani A, Kohchi T. 2012. The phytochrome-interacting vascular plant one-zinc finger1 and VOZ2 redundantly regulate flowering in Arabidopsis. *Plant Cell* **24**(8): 3248-3263.

Yelagandula R, Stroud H, Holec S, Zhou K, Feng SH, Zhong XH, Muthurajan UM, Nie X, Kawashima T, Groth M et al. 2014. The Histone Variant H2A. W Defines Heterochromatin and Promotes Chromatin Condensation in Arabidopsis. *Cell* **158**(1): 98-109.

Yuan GC, Liu YJ, 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.

Yuan W, Xu M, Huang C, Liu N, Chen S, Zhu B. 2011. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *The Journal of biological chemistry* **286**(10): 7983-7989.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science* **328**(5980): 916-919.

Zhang X. 2008. The epigenetic landscape of plants. *Science (New York, NY)* **320**(5875): 489-492.

Zhang X, Bernatavichute Y, Cokus S, Pellegrini M, Jacobsen S. 2009. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome biology* **10**(6): R62.

Zhang X, Germann S, Blus BJ, Khorasanizadeh S, Gaudin V, Jacobsen SE. 2007a. The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat Struct Mol Biol* **14**(9): 869-871.

Zhang XY, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J, Jacobsen SE. 2007b. Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *Plos Biol* **5**(5): 1026-1035.

Zhang Z, Wippo CJ, Wal M, Ward E, Korber P, Pugh BF. 2011. A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome. *Science* **332**(6032): 977-980.

Zhao DZ, Ni WM, Feng BM, Han TF, Petrasek MG, Ma H. 2003. Members of the Arabidopsis-SKP1-like gene family xhibit a variety of expression patterns and may play diverse roles in Arabidopsis. *Plant Physiology* **133**(1): 203-217.

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S. 2008. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**(7218): 125-129.

APPENDIX

A. Mapping Statistics for ChIP samples in Chapter 2

38°C heat

sample name		details	total mapped reads	mapping rate	reads after filter
HTR13-1		0min	13398895	95.58%	9997099
HTR13-2		15min	11343253	95.61%	8700956
HTR13-3		30min	11595476	95.75%	8901687
HTR13-4		1hr	10275934	95.65%	7987341
HTR13-5		2hr	9012872	90.87%	4510825
HTR13-6		4hr	10213071	94.94%	8106732
HTR13-7	H3.1 myc-	8hr	11151178	95.19%	8242508
HTR13-8	ChIP	24hr	9757387	95.76%	7510096
HTR5-1	H3.3 myc-	0min	12869281	97.26%	10957086
HTR5-2	ChIP	15min	13984316	96.39%	11447794
HTR5-3		30min	11859571	97.28%	10206490
HTR5-4		1hr	13500675	97.69%	11491864
HTR5-5		2hr	12326472	96.99%	10456955
HTR5-6		4hr	12322653	97.56%	10596668

HTR5-7		8hr	14564877	97.51%	12048540
HTR5-8		24hr	13214888	97.22%	10667470
total H3-1		0min	15032621	94.11%	5651270
total H3-2		15min	16655222	92.63%	6099272
total H3-3		30min	13875839	93.95%	6573141
total H3-4		1hr	9146851	94.09%	6569178
total H3-5		2hr	14341235	95.86%	9343764
total H3-6		4hr	16740358	92.98%	7028024
total H3-7		8hr	15053827	92.43%	4826785
total H3-8	total H3 ChIP	24hr	15022601	97.06%	11137331
PolII-1		0min	5760652	68.28%	4903729
PolII-2		15min	4678374	61.26%	3835247
PolII-3		30min	5037065	74.82%	4162853
PolII-4		60min	3899610	42.61%	3158531
PolII-5		2hr	3829599	47.74%	3238585
PolII-6	PolII CTD	4hr	6035616	51.01%	5306796
PolII-7	ChIP	8hr	799262	45.48%	760725
K36me3-1	H3K36me3	0min	8435649	87.60%	8009834
K36me3-2	ChIP	15min	7737279	90.00%	7345207
K36me3-3		30min	4438603	88.57%	4241833
K36me3-4		60min	6716500	85.05%	6374648
K36me3-5		2hr	6437915	84.29%	6118601
K36me3-6		4hr	7267647	85.73%	6845018

K36me3-7		8hr	9476864	86.55%	8928499
45C heat					
CH-LG-1		0min	17095236	90.40%	13746657
CH-LG-2		30min	18891954	92.27%	14894656
CH-LG-3	H3.3 myc-	1hr	15169848	92.57%	12502467
CH-LG-4	ChIP	2hr	19223314	91.35%	15121268
CH-LG-5		0min	20830161	95.17%	15472985
CH-LG-6		30min	19752964	92.35%	14268273
CH-LG-7	H3.1 myc-	1hr	20071314	93.01%	14702565
CH-LG-8	ChIP	2hr	18169526	95.15%	13847071

B Mapping statistics for RNA-seq samples in chapter 2

Paried end RNA-seq matching ChIP-seq samples

sample name	details	total mapped reads	aligned pairs	concordant pair alignment rate
1	0min	19584751	8069513	75.20%
2	15min	20276909	9173933	78.40%
3	30min	21454941	8738099	74.10%
4	1hr	20627608	9381083	78.70%
5	2hr	20435837	9046741	77.70%
6	4hr	19329007	8918859	83.60%
7	8hr	19530169	8896661	83.40%
8	24hr	21383386	9153724	78.00%

Recover test RNA-seq

sample name	details	total mapped reads	mapping rate
set1-0min	without treatment	10212869	95.30%
set1-7.5min	treat 7.5min	14328872	98.20%
set1-15min	treat 15min	8710074	97.40%
set1-30min	treat 30min	7842454	96.70%
set1-1hr	treat 1hr	13555425	96.50%
set1-2hr	treat 2hr	8455532	96.80%
set1-3hr	treat 3hr	8423714	97.60%

set1-4hr	treat 4hr	12452277	95.70%
	treat 4hr recover 24hr then no		
set3-0min	treat	9384808	95.20%
	treat 4hr recover 24hr then treat		
set3-7.5min	7.5min	14371569	95.80%
	treat 4hr recover 24hr then treat		
set3-15min	15min	15567489	97.80%
	treat 4hr recover 24hr then treat		
set3-30min	30min	12293436	96.30%
	treat 4hr recover 24hr then treat		
set3-1hr	1hr	8992703	89.80%
	treat 4hr recover 24hr then treat		
set3-2hr	2hr	10184568	96.60%
	treat 4hr recover 24hr then treat		
set3-3hr	3hr	13721423	96.50%
	treat 4hr recover 24hr then treat		
set3-4hr	4hr	12484126	96.30%