GENOME-WIDE PROFILING OF HISTONE METHYLATION IN POPULUS TREMULA X

ALBA 717-1B4

by

LAURA LEVIN

(Under the Direction of Xiaoyu Zhang)

ABSTRACT

This study focuses on describing the distribution of histone modifications in the model system for woody perennial trees, *Populus.tremula x alba* 717-1B4. We examined three histone modifications, H3K36me3, H3K4me3 and H3K27me3. Our results suggest that the distribution of these modifications is very similar to that of *Arabidopsis thaliana*. All three modifications are enriched on euchromatic chromosome arms within genes rather than the heterochromatic centromere region. While H3K36me3 and H3K4me3 are enriched on the most highly expressed genes, H3K27me3 is enriched on genes that have the lowest expression levels. Additionally, we identified differentially methylated H3K27me3 genes in *Populus.tremula x alba* 717-1B4 and *Arabidopsis thaliana*. The results of this study have helped us understand that histone modifications distribution and function are conserved in both annual and perennial plant species.

INDEX WORDS: histone modification, H3K36me3, H3K4me3, H3K27me3, *Arabidopsis thaliana*, *Populus tremula x alba* 717-1B4

GENOME-WIDE PROFILING OF HISTONE METHYLATION IN POPULUS TREMULA X

ALBA 717-1B4

by

LAURA LEVIN

B.S, University of Florida, 2012

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

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2016

GENOME-WIDE PROFILING OF HISTONE METHYLATION IN POPULUS TREMULA X

ALBA 717-1B4

by

LAURA LEVIN

Major Professor: Xiaoyu Zhang
Committee: Robert Schmitz
                     Zachary Lewis

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

LITERATURE REVIEW

Populus tremula x alba 717-1B4 as a model system

The _Populus_ L. genus is a member of the Salicacea family that consists of deciduous trees; including aspens, poplars and cottonwoods[1]. There are about 30 species in the genus, which are distributed across the Northern Hemisphere and in parts of Africa[2]. Poplar (term used for all species in _Populus_ genus) are long-lived, dioecious, wind-pollinated trees with the capacity for reproduction both sexually and asexually[2]. Rapid growth rates and their capacity for vegetative propagation have led poplar to become one of the most economically important groups of trees[1,2]. They are cultivated worldwide commercially for many purposes including for pulp, lumber, and bioenergy[2]. These features combined with a small genome size have led poplar to become a model system for all woody perennials[3].

The _Populus trichocarpa_ genome, published in 2006, has been the primary resource to study poplar genetics, cellular, and molecular biology[4]. However, research has been limited by how difficult it is to transform the species. Researchers instead utilize a hybrid poplar species named, _P. tremula x alba_ 717-1B4 (sPta717) because of the ease of transformation[3]. The following study focused on poplar epigenetics has been performed in the more widely used hybrid in order for more researchers to be able to utilize the results.

Chromatin modifications, such as histone modifications and DNA methylation, may play an important role in how perennial tree species interact with their environment[5]. It has been

suggested that epigenetic variation contributes to how perennial species adapt to both long- and short-term changes in environmental conditions[5–7]. Despite the potential for epigenetics to play a central role in perennial tree adaptability, the study of study of epigenetic phenomena in poplar is extremely limited. One study, published in 2013, showed that histone acetylation immunofluorescence signal in the phloem was high in the summer and low in the winter, while the DNA methylation signal displayed an opposite pattern[8]. In order to further study poplar epigenetic phenomena, my thesis is focused on determining the pattern of specific histone modifications throughout the poplar genome. This information will be used as a resource for future researchers interested in poplar epigenetics

Introduction to chromatin structure

Chromatin is a dynamic polymer, composed of repeating nucleosome subunits[9]. The nucleosome is made up of ~150bp of DNA wrapped around an octamer of histone proteins, two copies each of H2A, H2B, H3, and H4[9,10]. There are two main forms of chromatin, euchromatin and heterochromatin. Euchromatin is lightly condensed, gene rich, and is the site of active transcription; while heterochromatin is densely packed, repeat rich, and is mostly inactive in terms of transcription.

The packaging of DNA into chromatin is mostly thought of as a mechanism to store DNA in the nucleus of eukaryotic cells. However, chromatin is the platform for a large number of biological processes; including DNA replication, DNA damage repair, transcription, and RNA processing[11,12]. Chromatin can be covalently modified, through DNA methylation and histone modification. These modifications can significantly affect the processes that occur in the chromatin environment[11,12].

In this review, epigenetics is defined as the study of changes to gene expression and other biological processes in the chromatin environment without affecting the underlying DNA sequence. DNA methylation and histone modifications have been called "epigenetic" due to their potential role in regulating transcriptional activity and other biological processes. Epigenetic modifications have the potential to be inherited by daughter cells through either meiotic or mitotic divisions[13-15]. The passage of epigenetic modifications is thought to allow daughter cells to maintain similar transcriptional activity to their parent cells resulting in distinct cell types. While there is greater understanding of how DNA methylation is inherited, there is emerging research suggesting that histone modifications may be inherited as well[13-15].

DNA methylation

DNA methylation occurs when a methyl group is covalently attached to the C5 position of the cytosine ring. This form of chromatin modification has been associated with gene and transposon silencing in many eukaryotic systems[16,17]. DNA methylation mutants are often lethal in mammalian systems and display severe pleiotropic phenotypes in plants highlighting the importance of this modification[17].

In mammals, the majority of cytosine methylation occurs in the CG context and almost all of those cytosines are methylated with the exception of those located within CpG islands [18,19]. These regions, found upstream of genes, remain unmethylated, despite higher than average GC content[19]. This pattern differs from plant genomes in which DNA methylation occurs in three main contexts; CG, CHG, and CHH, where H=A, C, or T)[20-23]. DNA methylation in all sequence contexts is highest in the pericentromeric heterochromatin region, on transposons and repetitive

DNA elements(Zhang et al. 2006). While methylation of transposons leads to silencing, gene body methylation in the CG context often have high transcription levels(Zhang et al. 2006).

Four DNA methyltransferases, DRM2, MET1, CMT2 and CMT3, are responsible for establishing and maintaining these methylation contexts in *Arabidopsis thaliana*[24,25]. DRM2 is a *de novo* methyltransferase in all three contexts and maintains methylation in the CHH context. CMT2 is responsible for *de novo* methylation or maintenance in CHH and CHG contexts, while CMT3 maintains the methylation status at CHG sites. MET1 maintains methylation at CG sites. It is probable that other plant species utilize similar mechanisms to establish and maintain DNA methylation.

Introduction to histones and histone modifications

Histones are small basic proteins responsible for organization of DNA inside nucleus of eukaryotic cells. DNA wraps around a histone octamer, two copies of each of the 4 core histones: H2A, H2B, H3, and H4[9]. Core histones are some of the most highly conserved proteins across all eukaryotic species, especially in their flexible N- and C- terminal tail regions. The N- and C-terminal tails exhibit flexible conformations that protrude away from the nucleosome and are subject to many covalent post-translational modifications including methylation, acetylation, sumoylation, ubiquitination, and phosphorylation, as well as other forms of modification[11,26]. While the majority of histone modifications occur on their N- and C- terminal tails, a number of modifications occur in the globular domain as well[12,27,28].

With so many types of chromatin modifications and histone residues with modification potential, it is plausible that almost every nucleosome could have a unique combination of modifications[29]. However, studies of the genome-wide locations of histone modifications have

shown that this is not the case. Each histone modification studied so far has been associated with distinct genomic features, including transposons, actively transcribed genes, or repressed genes[29]. The functional role that histone modifications play at these locations has been a central focus of the epigenetic field[11,30].

Histone modifications carry out their function in three main ways. First, they can influence chromatin structure directly. For example: Histone lysine acetylation or phosphorylation reduces the positive charge of histones, which can disrupt the interaction with negatively charged DNA[26]. This increases nucleosome mobility and leads to a more open chromatin environment. Second, histone modifications can act as docking stations for proteins and complexes with histone modification binding domains[26]. This provides a platform for proteins to localize to a specific genomic location in order to carry out their function in the chromatin environment. A nice example of this is HP1, the linker histone that is important for chromatin structure throughout the eukaryotes. HP1 has a chromodomain that recognizes and specifically binds Histone 3 lysine 9 tri-methylation (H3K9me3) in order to localize to heterochromatin[31].

Another mechanism by which histone modifications can function is through cross talk with other chromatin modifications. There are many forms of cross talk; including, competitive antagonism between modifications and dependence of chromatin modifications on each other. Competitive antagonism occurs when modification pathways compete with each other to target the same residue. Some residues can be acetylated or methylated, but not both[26]. A second mode of cross talk takes place when histone modifications are dependent on one another. In mammalian cells, histone 3 serine 10 phosphorylation helps facilitate histone 3 lysine 14 acetylation by enzyme, Gcn5[32,33]. It is likely that similar cross talk mechanisms exist in plants as well between H3K36me3 and H3K4me1. SET DOMAIN GROUP 8 (SDG8), a H3K36

methyltransferase, has a CW domain that interacts specifically with H3K4 methylation. At the same time, a H3K4 methyltransferase, ATX3, has a PWWP domain, a domain known for interacting with H3K36me3. Since H3K4 and H3K36 methylation are both present on actively transcribed genes, it is likely that they either facilitate establishment or maintenance of the other mark[29,34].

Histone lysine methylation in plants

Histone methyltransferase proteins, all of which contain a SET domain, catalyze histone lysine methylation. The SET domain is named after three Drophila histone methyltransferases, *Su(var) 3-9*, *Enhancer-of-zeste*, and *Trithorax*[35]. Due to an increased duplication rate, there are more SET domain proteins in plants compared to other eukaryotes [35]. Arabidopsis contains 32 SET domain genes; while mouse contains 17 and yeast contains only 4[35]. Many plant SET domain genes have other domains that facilitate protein-protein interaction; indicating that they may be prone to working in complexes[35].

SET domain genes can mono-, di-, or tri-methylate lysine residues. Each level of methylation is considered to be a distinct histone modification with a unique genome-wide distribution and function[29]. Thus far, four histone lysines are known to be methylated in plants and all are present on histone 3 (H3). These lysines are H3K4, H3K9, H3K27, and H3K36[36,37]. Three histone modifications known to be important for Arabidopsis growth and development are histone 3 lysine 36 trimethylation (H3K36me3), histone 3 lysine 4 trimethylation (H3K4me3), and histone 3 lysine 27 trimethylation (H3K27me3). This study will focus on the distribution of these three forms of histone methylation in poplar, sPta717. There is little to no study of these

modifications in poplar, therefore this literature review will focus on what is known in mammals and Arabidopsis.

H3K4me3

H3K4me3 is found on actively transcribed genes in yeast, mammals, and plants. The standard H3K4me3 distribution is a prominent peak around the transcription start site that extends into the promoter[38–40]. In Arabidopsis, H3K4me3 is catalyzed by SDG2 and SDG27, which are homologs to the mammalian H3K4 methyltransferase called TRITHORAX (TRX)[35]. A potential mechanism for H3K4me3 deposition has been discovered in yeast, in which the H3K4 methyltransferase interacts with the initiating form of PolII[41]. It is possible that the same mechanism exists in plants considering that the H3K4me3 distribution is similar in yeast and Arabidopsis.

H3K4me3 takes part in many biological processes by interacting with histone reader proteins. These readers have H3K4 methyl-binding domains, which allow H3K4me3 to carry out its function and cross talk with other modifications. A number of reader domains have been found to interact with H3K4me3, including the chromodomain, zf-CW, and the PHD finger. H3K4me3 binding proteins have been implicated in the assembly of the transcription preinitiation complex, recruitment of mRNA processing factors, DNA recombination, DNA damage response, telomere silencing, and cross-talk with H3 acetylation[42–46]. No evidence suggests that H3K4me3 affects higher-order chromatin structure directly.

Although most of what is known about H3K4me3 function comes from mammalian systems, Arabidopsis mutants lacking H3K4me3 help us understand the importance of the modification. SDG27 and SDG2 both have *in vitro* H3K4 methyltransferase activity as well as

decreased levels H3K4 di- and tri-methylation *in vivo*[29,47,48]. The *sdg2* mutant has pleiotropic

developmental defects[48]. It is dwarfed, sterile, early flowering, and has a terminal flowering

phenotype. The other H3K4 methyltransferase mutant, *atx1*, has a less dramatic phenotype than

sdg2, but nevertheless is small, early flowering, and has floral morphological defects[47]. The

presence of H3K4me3 on practically all actively transcribed genes in plants has led to the

implication that the function of H3K4me3 is to activate transcription[40]. However, decreased

levels of genic H3K4me3 does not always lead to down regulation of gene expression[48]. This

indicates that although H3K4me3 plays an important role in plants, it may be a consequence of

transcription, not a causative factor. In the future, studies of H3K4me3 reader proteins will

further elucidate its functional role in plants.

H3K36me3

Like H3K4me3, H3K36me3 is located on actively transcribed genes in eukaryotic

species. H3K36me3 is distributed across the majority of the gene body and reaches its highest

levels slightly downstream of the transcription start site[29,49,50]. There are two H3K36me3

methyltransferases in Arabidopsis, SDG7 and SDG8, which are homologous to mammalian

H3K36me3 methyltransferase, ASH1[35,51–53]. The mechanism used to deposit H3K36me3 in

plants is unknown. However, it is possible that SDG7 or SDG8 interact with the elongating form

of PolII to deposit H3K36me3 as is found in yeast[54]. This deposition mechanism would help

explain the H3K36me3 distribution across the gene body.

H3K36me3 has been linked to essential biological processes in yeast and mammals

through interactions with different reader domain proteins. The mark has been associated with

transcription elongation, pre-mRNA processing, nucleosome turnover, cryptic transcription, and

nucleosome turnover[55–59]. While the functional role of H3K36me3 is not as well understood in plants, the modification is essential for proper Arabidopsis development and is likely important in other plants as well.

The phenotypes of the H3K36me3 methyltransferase mutants point to the importance of the modification. The decreased levels of H3K36 di- and tri-methylation in *sdg8* lead to an early flowering, irregular branching, and stunted growth phenotype[51–53]. The other methyltransferase mutant, *sdg7*, has no obvious phenotype or decrease in H3K36 methylation. However, the *sdg7 sdg8* double mutant, mutant has pleiotropic developmental defects and lacks all H3K36 di- and tri-methylation (Zhang lab, unpublished). It has extremely student growth, reduced fertility, and abnormal leaf shape. It will be necessary to study H3K36me3 reader proteins in the future to understand exactly how H3K36me3 effects plant development.

H3K27me3

H3K27me3 is uniformly distributed across the body of repressed genes, beginning slightly upstream of the transcription start site in plants and mammals[60–62]. The Polycomb repressive complex 2,PRC2, is responsible for the majority of H3K27 di-, and tri-methylation in animals[63]. The PRC2 catalytic subunit is a SET domain protein, called is EZH2 (enhancer of zeste homologue 2), [64–67]. There are three EZH2 homologs in Arabidopsis, all of which are class 1 SDG genes[35]. These are: SDG1 (clf), SDG5 (mea), SDG10 (swn).

H3K27me3 acts to regulate gene expression in response to developmental transitions. It is present on developmentally important genes in both animals and plants. In mammals, H3K27me3 establishes and maintains the repression of genes. Genes that remain H3K27me3 methylated after differentiation have silencing reinforced by DNA methylation[63,68,69].

H3K27me3 does not seem to be responsible for establishing repressed states in plants, but is important for maintaining them[70]. H3K27me3 regions in plants are shorter than that of animals, usually covering only one gene[62]. Additionally, plant genes with H3K27me3 have high tissue specificity, meaning that genes are purposefully repressed in some tissues(Zhang et al. 2007; Wang et al. 2009; François Roudier et al. 2011a).

H3K27me3 methyltransferase mutants have pleiotropic developmental defects including problems with the female gametophyte, endosperm, and embryo[73]. Partial loss of H3K27me3 in the *clf* mutant results in pleiotropic defects(Zhang 2012). A greater loss of H3K27me3 in the *clf swn* double mutant leads to an immortal callus that is unable to differentiate(Zhang 2012). Additionally, loss of H3K27me3 leads to overexpression of genes that had H3K27me3 in the first place (Zhang lab, unpublished). It is clear that H3K27me3 function is important for repression of genes required for proper development.

Objective of thesis

The objective of my thesis is to understand how histone methylation is distributed across the poplar genome. My research will help us understand how well the pattern of histone modifications is conserved between an annual plant species, like Arabidopsis, to perennial species, such as poplar. Understanding genome wide location of poplar histone modifications provides the platform to assess the functional role of plant histone modifications in the future.

The future of epigenetic studies in a perennial species, like poplar, is bright. Studying epigenetics in a perennial species will allow us to learn about epigenetic variations that have occurred over a long period and whether that epigenetic variation has functional consequences.

Additionally, the nature of poplar reproduction makes it possible to study long-lived questions about the transgenerational inheritance of histone modifications[11].

CHAPTER 2

HISTONE METHYLATION IN POPULUS TREMULA X ALBA 717-1B4

Materials and methods

Chromatin Immunoprecipitation (ChIP

ChIP was performed on young poplar leaves and 12 day-old *Arabidopsis thaliana* (Arabidopsis)

seedlings using the following antibodies: H3K4me3 (Millipore, 07-473), H3K36me3 (abcam,

ab9050), H3K27me3 (Millipore, 07-449), and H3 (abcam, ab1791) and the protocol described in

Chapter 3. A library was also made from input in both Arabidopsis and poplar as described in

Bewick et.al, 2016[74]. ChIP-seq libraries were sequenced on the Illumina platform and 50 base

pair, single end reads were produced. The reads were trimmed using Trimmomatic v0.32, then

mapped to the Populus tremula x alba 717-1B4 (sPta717) genome with Bowtie2 v2.2.3 with the

default parameters. The SAM file output from mapping was converted to a BAM file with

Samtools v1.2. Clonal reads were then removed with Samtools v0.1.9. The BAM file was

converted to BED format with Bedtools v2.21.1.

RNA-seq

Single-end, 50 base pair, RNA-seq libraries were created from the same tissue used to create the

ChIP-seq libraries and were sequence on the Illumina platform. The reads were trimmed with

Trimmomatic v0.32, then were mapped to the sPta717 or TAIR10 genome with tophat v2.2.1

with special parameters of max intron length 20000, library type fr-firststrand, and bowtie2

sensitive. Expression values were obtained using Cuffnorm with the sPta717 or TAIR10 reference annotation and library type fr-firststrand.

Chromosome wide plots

Genes were computed as the number of genes that overlap with each 100 kb bin. Repeats were the sum of annotated repeat base pairs within each 100 kb bin. For each ChIP-seq library, the read coverage was calculated for each base pair in the genome. Iterating across all of the base pairs within a 100 kb bin, the coverage values were added together then normalized by the number of million mapped base pair in the ChIP library. Values were computed for each library separately then enrichment was computed as chromatin modification value divided by input value for each bin.

Absolute distance heatmaps

For each gene that was at least 500 base pairs long, the upstream and downstream regions were broken up into 25 base pair bins. Using a priori count of 1 per bin, the read coverage of each base pair within a bin was added together and then normalized by number of million mapped base pair in the library. Across all genes and bins, enrichment was calculated as the chromatin modification value divided by the input value. To improve visualization, the 95% quantile value t was calculated for all enrichment values $\mathbf{Y}$. For each value y in $\mathbf{Y}$, if $y > t$, $y \leftarrow t$.

Genic Metaplots

To determine ChIP enrichment values, read coverage was calculated for each base pair 1 kb upstream and downstream from the transcription start site and transcription termination site. The

coverage value at each base pair was normalized by the number of million mapped base pairs in the ChIP library and .01 was added to each value. The ChIP enrichment value was then divided by the input value at each base pair.

GO term analysis with AgriGO

GO term enrichment analysis was performed using AgriGO, Singular Enrichment Analysis, with the Complete GO gene ontology type[75]. Significant GO terms were identified using the Hypergeometric statistical test and Yekutieli multi-test adjustment method with a significance level of 0.05.

Peak finding with SICER

Peak finding software, SICER, was utilized to find H3K27me3, H3K4me3, and H3K36me3 peaks in both *Arabidopsis thaliana* and poplar[76]. Input was used as the control library in both cases. The chosen parameters were window size of 100 base pairs, gap size of 200 base pairs, effective genome fraction of 75%, and a false discovery rate of 0.05.

Calculating base pairs covered by peaks

Across the genome, each base pair was assigned as genic, promoter, repeat, or intergenic. Genic and transposon base pairs were assigned based on the sPta717 annotation file. The promoter region was defined as 1kb upstream of the nearest gene, unless that region overlapped with the adjacent gene. If the region overlapped with the adjacent gene, it was defined as genic. All of the other base pairs were considered to be intergenic. For each peak region, the number of base pairs were counted and assigned to a category. The category values were added together for all

of the peaks. The category percentages were fractions of counts for each category divided by the sum of the counts for all the categories multiplied by 100.

Differential H3K27me3 methylation in Arabidopsis and Poplar

We sought to identify genes enriched for H3K27me3 in Poplar or Arabidopsis as well as genes that were enriched in both species. To do so, 12,954 orthologous gene pairs were identified by reciprocal best BLAST using the nucleotide coding sequence, and an e-value cutoff of ≤1E-0.6 in Poplar and Arabidopsis. The three separate gene groups were named "pop only", "ara only", and "both".

The "Both" group as identified as those genes that had H3K27me3 peaks found in both the Arabidopsis and Poplar orthologs.

"Poplar Only" genes were those genes that had H3K27me3 peaks (peaks called by SICER) in the two Poplar H3K27me3 ChIP replicates. H3K27me3 peaks were not identified in the Arabidopsis orthologs of these genes. This group was further narrowed down using conservative criteria. First, only genes that had low gene expression (in the bottom 40% compared to all genes) were kept. The reason being that H3K27me3 is most enriched on genes with low expression. Second, after finding the ChIP RPKM, we selected those genes that had Poplar ChIP RPKM ratios for H3K27me3/ Input that were greater than 1.

The "Arabidopsis Only" group were those genes that had H3K27me3 peaks found in Arabidopsis, but not in Poplar. The group was narrowed down by selecting the genes that were in the lowest 40% of expression in Arabidopsis. The group was further narrowed down by choosing the genes with Poplar ChIP RPKM ratios for H3K27me3/Input that were less than 1.

15

Genome-wide distribution of poplar histone methylation

We sought to identify how histone methylation is distributed across a poplar genome. For this reason, chromatin immunoprecipitation and next-generation sequencing (ChIP-seq) was performed on young poplar leaves with antibodies against three modifications; H3K4me3, H3K36me3, and H3K27me3. ChIP-seq was performed with two biological replicates for each modification. The ChIP mapping statistics can be found in Table 1; where each poplar library had greater than 62% mapping percentage to the sPta717 genome[3].

We inspected the distribution of the modifications across the poplar chromosomes to learn if there was a consistent genome-wide distribution profile. The enrichment for each histone modification was calculated then divided by the input enrichment for each bin. We hypothesized that the distribution patterns would be similar to those found in Arabidopsis; where H3K36me3, H3K4me3, and H3K27me3 are most prevalent on euchromatic chromosome arms and are depleted at the pericentromeric repeat region[29]. Indeed, all three modifications appeared to prefer the euchromatic arms. On the majority of chromosomes, the number of genes in the region dictated the level of histone modification enrichment. Regions with higher numbers of genes had higher levels of modification enrichment. Typically, the repeat enriched pericentromeric heterochromatic region had low levels of the modifications.

Ten out of nineteen chromosomes display the expected pattern with H3K6me3, H3K4me3 and H3K27me3 enriched on euchromatic arms and depleted in the pericentromeric heterochromatin (chromosomes 2, 4, 5, 8-10,14-16, 18). Examples of this pattern are shown in Figure 1A and 1B. Chromosome 13, however, does not follow the typical pattern. (Figure 1C). On chromosome 13, all three histone modifications have the lowest enrichment around 15 MB,

where gene number is high. On other chromosomes, one or two of the histone modifications display the typical pattern of euchromatin enrichment (chromosome 3, 6, 7, 19), while the other modification/s show no enrichment pattern (Figure 1D). The rest of the chromosomes either have histone modification distributions with no discernable patterns or the modification has the euchromatin enriched distribution in only one of the replicates (chromosomes 1, 11, 12, 17) (Figure 1E).

The euchromatic chromosome arms contain genes, promoters, intergenic regions, and to a lesser extent transposons. We wanted to determine if the modifications were enriched specifically in any of these regions, since our evidence suggests that the modifications are present in the euchromatin (Figure 1). To carry out this analysis, H3K27, H3K4me3, and H3K36me3 ChIP enriched peaks were identified in both replicates with SICER using input as a control library[76]. About 80 million peaks were identified for each ChIP-seq library and for each modification the biological replicates had similar results.

To determine where the peaks were distributed throughout the genome the percentage of base pairs covered by peaks was calculated for each histone modification. The percentage of the genome covered by each genomic feature (promoter, gene, intergenic region, transposon) is represented in Figure 2A. Histone modifications were called as enriched for a genomic feature if the base pairs covered by the peaks were ≥ 5% more than the feature's genomic percentage. If the base pairs covered by the peaks were < 5% different from the feature's genomic percentage then the histone modification was not considered to be enriched in the region. Finally, if the feature's genomic percentage was ≥ 5% more than the base pairs covered by the peaks than the histone modification was underrepresented in that region.

H3K36me3 was enriched in genic regions, was not enriched in promoters, and was underrepresented in both intergenic regions and transposons (Figure 2B). H3K4me3 not only followed the same enrichment pattern as H3K36me3 with genes being the only enriched feature and had almost identical percentages to H3K36me3 (Figure 2C). These results indicate that H3K36me3 and H3K4me3 likely follow the pattern identified in other eukaryotic species, in which these modifications are primarily located on actively transcribed genes(Zhang 2012; Barski et al. 2007b; Santos-Rosa et al. 2002; Zhang et al. 2009; Roudier et al. 2011b). H3K27me3 was also enriched in genic regions, but to a lesser degree than the other modifications (Figure 2D). Additionally, H3K27me3 was underrepresented in transposons and was not enriched in promoter or intergenic regions. In other species, H3K27me3 is most often associated with developmentally repressed genes, although the modification has been associated with intergenic regions and transposons as well(Pan et al. 2007; Zhao et al. 2007; Rosenfeld et al. 2009; Leeb et al. 2010). Poplar H3K27me3 on the other hand is mostly enriched in genic regions.

Genic histone modification patterns

Although all three histone modifications were enriched in genic regions, we did not know how the histone modifications were distributed across individual genes. To visualize the distribution of histone modifications across individual genes, heatmaps were created to display modification enrichment normalized to input across poplar genes with the genes ranked by expression. From the heatmap, H3K36me3 seems to be located on most poplar genes (~2/3) and located on genes with higher expression levels (Figure 3A). Additionally, consistent with studies performed in Arabidopsis, H3K36me3 is enriched towards the 5' end of genes. This

18

differentiates plant H3K36me3 distribution from the distribution in yeast and humans, in which K36me3 starts downstream of the transcription start site (TSS) and is highest in the 3' end of genes[38,57]. H3K4me3 is enriched at the 5' end of actively transcribed genes, but does not seem to be enriched on lowly transcribed genes (Figure 3B). The distribution of H3K4me3 near the TSS is reminiscent of the H3K4me3 distribution in all eukaryotes studied thus far[38–40]. Although the H3K27me3 enrichment does not appear to be particularly strong, it is highest on genes with lower expression levels (Figure 3C). This finding is consistent with other eukaryotic systems, in which H3K27me3 is found on developmentally important, repressed genes[29,63,68].

We also sought to determine which genes were preferably enriched for different histone modifications. Poplar genes were split into five expression groups (Top 20%, 2nd 20, 3rd 20, 4th 20, Bot 20) and metaplots were created to visualize the average histone modification distribution across each group of genes (Figure 4). H3K36me3 and H3K4me3 were enriched on the most highly expressed genes (Top 20%, 2nd 20) and had the least enrichment on genes with lowest expression (Figure 4A, B). Additionally, the metaplots display that H3K36me3 and H3K4me3 are biased toward the 5' end of genes and peak around 500 bp downstream of the TSS. H3K27me3 had an opposite gene preference; it is most highly enriched on genes with the lowest expression and is least enriched on genes that are highly expressed (Figure 4C). The H3K27me3 distribution pattern across genes does not display the same 5' bias seen for the H3K4me3 and H3K36me3 modifications. Instead, it seems to be distributed across the entire gene body.

19

Characteristics of histone modification targeted genes

Based on analysis of genic histone modification patterns (Figure 4), it was determined that H3K27me3 is preferentially enriched on repressed genes, while H3K4me3 and H3K36me3 are present on actively transcribed genes. However, it was still unknown whether the histone modifications were associated with specific biological processes or molecular functions. Genes enriched for H3K6me3, H3K4me3, and H3K27me3 were identified, then Gene Ontology (GO) terms were found through agriGO, an agricultural GO database[75].

There were 4,187 genes identified as being enriched for H3K36me3 and these genes had slightly higher H3K36me3 enrichment as genes in the top 20% expression group (Figure 5A). GO terms are classified into one of three categories: cellular component, molecular function, and biological process. Genes with H3K36me3 were significantly enriched for a total of 37 GO terms: 22 in the cellular component category, 1 in molecular function, and 14 in biological processes category (Figure 6A). The GO terms in the cellular component category indicate that H3K36me3 gene products are active in many parts of the cell; including the nucleus, endoplasmic reticulum, cytoplasm, and ribosome. The single GO term that fell into the molecular function category was structural molecule activity (GO:0005198); which means that the 136 genes in this category may contribute to the structural integrity of complexes. In terms of cellular processes, genes with H3K36me3 are involved with macromolecule biosynthetic processes, protein metabolic processes, and gene expression. The gene expression category indicates that some H3K36me3 genes may function to regulate transcription.

We identified 5,072 H3K4 tri-methylated genes in poplar. These genes were slightly more enriched for H3K4me3 than genes in the top 20% expression group (Figure 5B). 32 GO terms were identified for these genes; with 18 cellular component terms, 1 molecular function,

and 13 biological process terms. Interesting, 30 of the 32 GO terms were the same as those identified with the H3K36me3 genes (Figure 6B). This indicates that the same genes are H3K36me3 and H3K4me3 methylated, which is not surprising when considering that H3K4me3 and H3K36me3 are both present on actively transcribed genes. Indeed, there are 3,197 overlapping genes that contain both H3K4me3 and H3K36me3. The two GO terms that are specific for H3K4me3 genes are cellular component organization and primary metabolic process.

2,875 poplar genes were enriched for H3K27me3. The H3K27me3 genes had higher H3K27me3 levels than the bottom 20% expression group indicating that we identified the correct group of genes (Figure 5C). 21 GO term categories were enriched with 0 in the cellular component category, 4 molecular function, and 18 in the biological process category (Figure 6C). These GO terms had almost no overlap with those identified with the H3K36me3 and H3K4me3 genes, consistent with the preference of H3K27me3 for genes with low transcriptional levels. The biological process and molecular function GO terms indicate that H3K27me3 genes are important for regulating many processes within the cell, including transcriptional regulation via transcription factors. H3K27me3 genes also are enriched for a biological process GO terms involved with response to stimuli indicating that these genes may be involved with responding to environmental conditions.

Differential H3K27me3 methylation in Poplar and Arabidopsis affects gene expression

Although Arabidopsi*s* and poplar are both species within the plant kingdom, they have wildly different growth habits and lifespans. Knowing that H3K27me3 is associated with developmentally important genes in both species, led us to wonder whether any genes are differentially H3K27 tri-methylated between the two species. To answer this question, we first

obtained a list of gene orthologs by reciprocal best BLAST using nucleotide sequence from poplar and Arabidopsis. We then identified two groups of orthologous genes pairs. The first group was enriched for H3K27me3 in both species and the second group was enriched for H3K27me3 in one species or the other. According to our analysis, 520 orthologous gene pairs had H3K27me3 in both species, 391 genes had H3K27me3 only in poplar (pop only), and 103 genes had H3K27me3 in Arabidopsis only.

Genes with H3K27me3 only in poplar (pop only) had high levels of H3K27me3 in poplar (Figure 7A) and low levels of H3K27me3 in Arabidopsis (Figure 7B). The pop only genes had higher H3K27me3 levels than genes in the poplar bottom 20% expression group. Together, these results indicate that we identified genes with H3K27me3 only in poplar. We also confirmed that genes with H3K27me3 in Arabidopsis only (ara only) had high levels of H3K27me3 in Arabidopsis (Figure 7B) and low levels in poplar (Figure 7A). The orthologs with H3K27me3 in both species had high levels of H3K27me3 in both Arabidopsis and poplar, greatly exceeding the H3K27me3 levels of genes enriched in one species or the other (Figure 7A, B).

We wanted to know whether differences in H3K27me3 levels between the three gene groups affected gene expression. Pop only genes lower expression in poplar than in Arabidopsis(Figure 7C,D). Ara only gene had low expression levels in Arabidopsis and higher expression in Poplar. Genes with H3K27me3 in both species had an intermediate expression level between "pop only" and "ara only" genes in poplar and Arabidopsis. These results indicate that H3K27me3 has a functional consequence in poplar and Arabidopsis, acting to repress transcription. Interestingly, even though the genes with H3K27me3 in both species have higher H3K27me3 levels, they do not have expression levels as low as the pop only or ara only genes in their respective backgrounds.

Since Arabidopsis is a weedy annual plant and poplar is a perennial tree species, we hypothesized that the poplar only and Arabidopsis only genes would be enriched for GO terms related to their living habits. The both genes were expected to be associated with standard H3K27me3 GO terms, as seen in Figure 6C. Contrary to our hypothesis, the poplar only and Arabidopsis only genes were not enriched for any GO terms. The both genes were enriched for 22 GO terms, most of which were the same as those found in Figure 6C. There were a few GO terms that had not been enriched previously, including 2 cellular component GO terms, and 1 molecular function term. The cellular component GO terms are both related to the cell wall (GO:0005618, GO:0030234), while the molecular function GO term is enzyme regulator activity (GO:0030234). The both GO terms were consistent with role of H3K27me3 to mark genes involved with development.

Discussion

We have described the distribution of three histone modifications, H3K36me3, H3K4m3, and H3K27me3 in the poplar species, *P.tremula x alba 717*. This is one of the first studies focused on histone methylation in poplar and will provide a basis for future epigenetic research in the species. Our genome-wide profiling approach allowed us to study how histone methylation is distributed across chromosomes. All three histone modifications were enriched on euchromatic chromosome arms and depleted in the pericentromeric heterochromatin region. This finding led us to ask whether the modifications are enriched specific parts of the euchromatin, such as promoter, genic, or intergenic regions. Each modification was enriched in the genes, but did not seem to be enriched in the intergenic space or promoter region. These results are

consistent with Arabidopsis; in which the three modifications are enriched on euchromatin arms and located within genes[29].

The finding that H3K4me3, H3K36me3, and H3K27me3 were located within genes led us to ask whether the modifications had a similar distribution across genes to that of Arabidopsis and other eukaryotic species. We also wanted to know whether the modifications had similar gene preferences, with H3K4me3 and H3K36me enriched on actively transcribed genes and H3K27me3 preferring genes with low transcriptional activity. The three modifications had very similar distributions and gene preferences to that of Arabidopsis[29,40,62]. We found that H3K4me3 was present on actively transcribed genes and had a prominent peak slightly downstream of the transcription start site. H3K36me3 was also located on actively transcribed genes and was enriched across the gene body with a preference for the 5' end of the gene. H3K27me3 was enriched across the entire gene body with a preference for genes with low transcription levels. These results suggested that poplar and Arabidopsis histone modifications likely serve similar functions because they have the same distribution and gene preference.

It was unknown whether H3K4me3, H3K36me3, or H327me3 were enriched on specific types of genes. We identified genes with high levels of H3K36me3, H3K27me3, and H3K4me3 then used AgriGO to determine whether the three groups of genes were enriched for any GO terms[75]. H3K36me3 and H3K4me3 are located on many of the same genes, which led to similar GO term results. The majority of the GO terms enriched for H3K4me3 and H3K36me3 genes were part of the cellular component GO category indicating that gene products with these modifications are active in many different cellular organelles and in the cytoplasm. Twenty-nine H3K36me3 genes were significantly enriched for "endoplasmic reticulum", while this GO term was not enriched in the H3K4me3 data. Genes with H3K4me3 and H3K36me3 were not

significantly enriched in GO terms involving transcriptional regulation. These results support the notion that H3K4me3 and H3K36me3 may be a consequence of transcription rather than activating marks themselves.

H3K27me3 genes were associated with transcription factors and the regulation of gene expression. This result was consistent with the findings from Arabidopsis, that H3K27me3 is responsible for the repression of developmentally important genes. It was also intriguing that H3K27me3 genes were enriched for "response to stimulus". This result could indicate that H3K27me3 is also responsible for repressing genes in response to environmental stimuli in poplar. The difference in Arabidopsis and poplar life strategies led us to wonder whether there were changes to which genes were H3K27me3 regulated in the two species. We found that 520 genes had H3K27me3 in both species. 391 genes had H3K27me3 in poplar, while the orthologs of the same genes did not have H3K27me3 in Arabidopsis and 103 genes had the modification in Arabidopsis only. Genes with H3K27me3 in one species and not the other showed transcriptional repression in that species. These results indicated that H3K27me3 is repressing those genes. Although some genes had H3K27me3 in one species and not the other, it is possible that if we sampled more cell types the genes may have had H3K27me3 in both species. H3K27me3 is often present on genes that are expressed in a specific tissue type(Zhang et al. 2007).

This research highlighted the similarities in histone modification distribution between Arabidopsis and poplar. Our research indicates that the modifications likely perform very similar, if not the same functions in both species. Since the modification distribution is so similar, future epigenetic studies in poplar could focus on research that would be difficult to perform in Arabidopsis. Poplar is a long-lived perennial species with the ability to reproduce

sexually and clonally. Researchers could utilize the life strategy of poplar to learn whether environmental stresses induce changes in genes marked by histone modifications. For example: Differential H3K27me3 gene methylation on poplar trees grown in distinct environmental conditions may indicate that the development of those trees has been fundamentally altered. Researchers could also discover whether inheritance of epigenetic variations help plants adapt to the environment. Progeny of the trees with differential H3K27me3 methylation may be better adapted to an environment similar to that which they came from.

Table 1. ChIP mapping statistics for ChIP-seq libraries used in the study.

sample	replicate	overall mapping %	% remaining	input reads	remaining reads
h3	sPta717 1	63.48	55.94	1.40E+07	7.86E+06
h3k27me3	sPta717 1	66.42	58.53	2.04E+07	1.20E+07
h3k36me3	sPta717 1	72.02	66.68	9.79E+06	6.52E+06
h3k4me3	sPta717 1	71.49	65.41	1.35E+07	8.83E+06
input	sPta717 1	63.76	58.84	4.31E+06	2.54E+06
h3	sPta717 2	64.33	55.89	1.73E+07	9.69E+06
h3k27me3	sPta717 2	66.71	58.39	2.55E+07	1.49E+07
h3k36me3	sPta717 2	75.86	69.65	1.77E+07	1.24E+07
h3k4me3	sPta717 2	74.92	69.55	1.20E+07	8.36E+06
input	sPta717 2	64.54	58.59	6.06E+06	3.55E+06
h3k27me3	Arabidopsis 1	92.26	73.88	1.18E+07	8.70E+06
input	Arabidopsis 1	98.04	68.47	2.68E+07	1.83E+07

A

Chromosome 02

Chromosome 08

30

C Chromosome 13
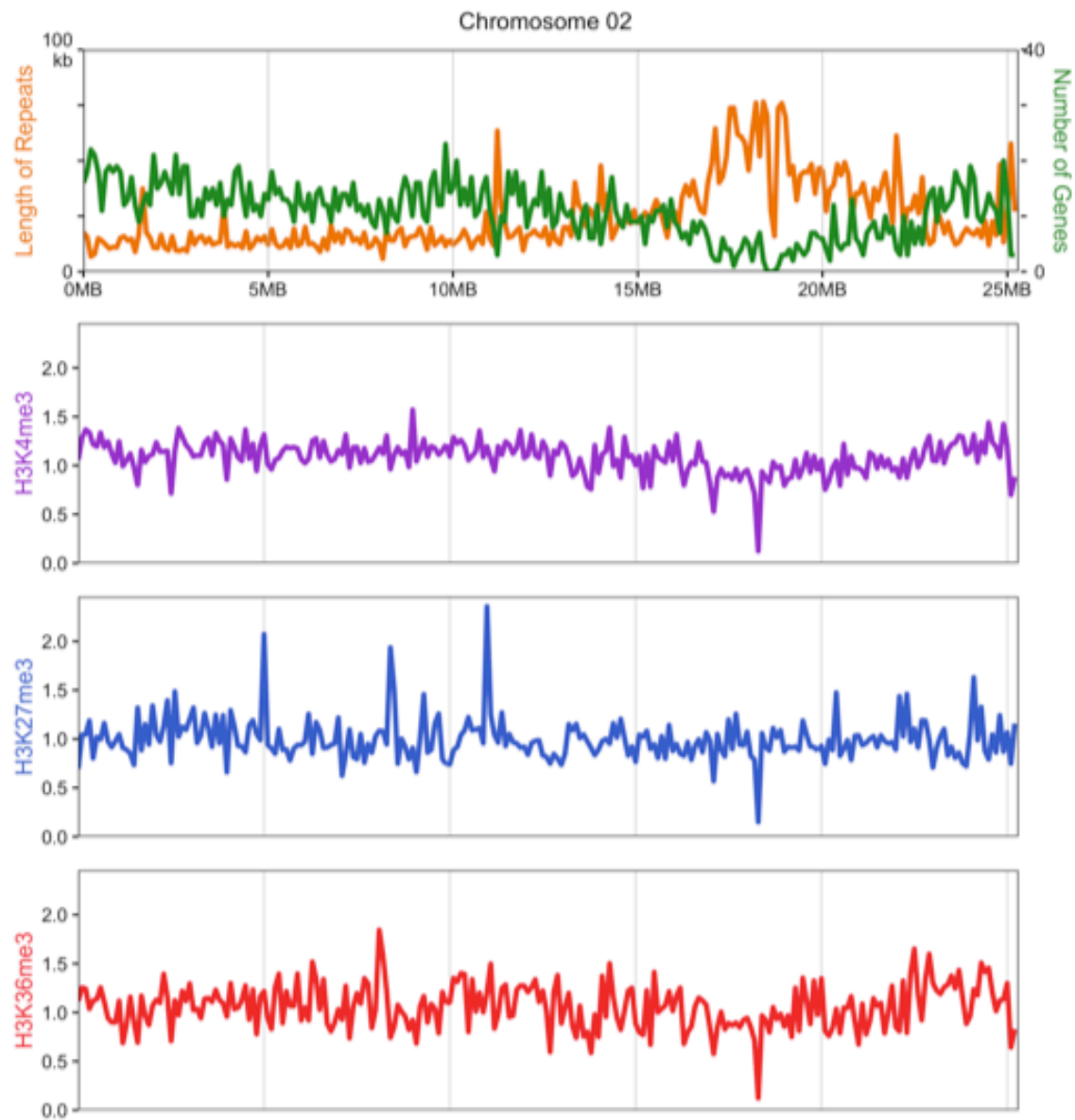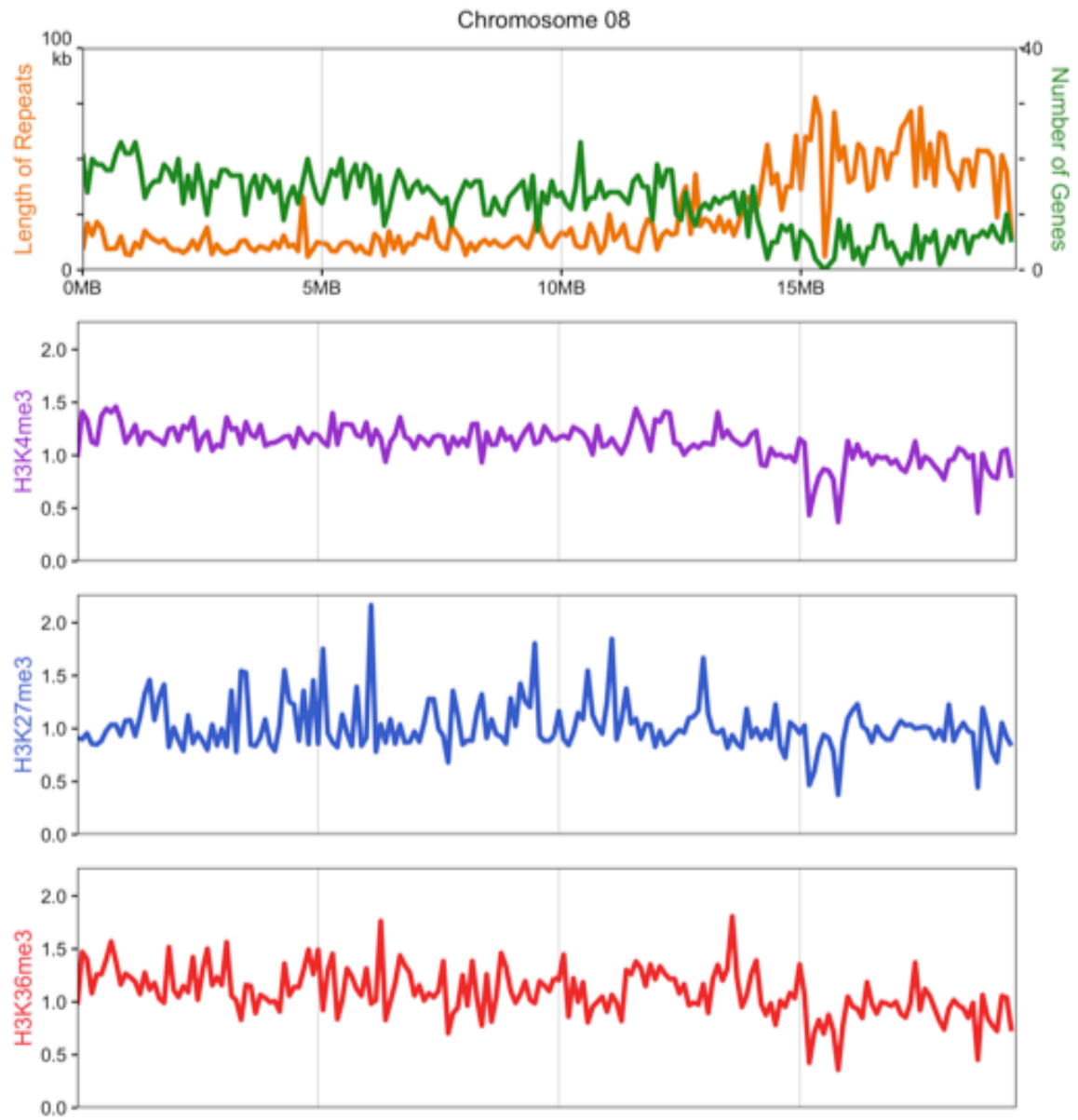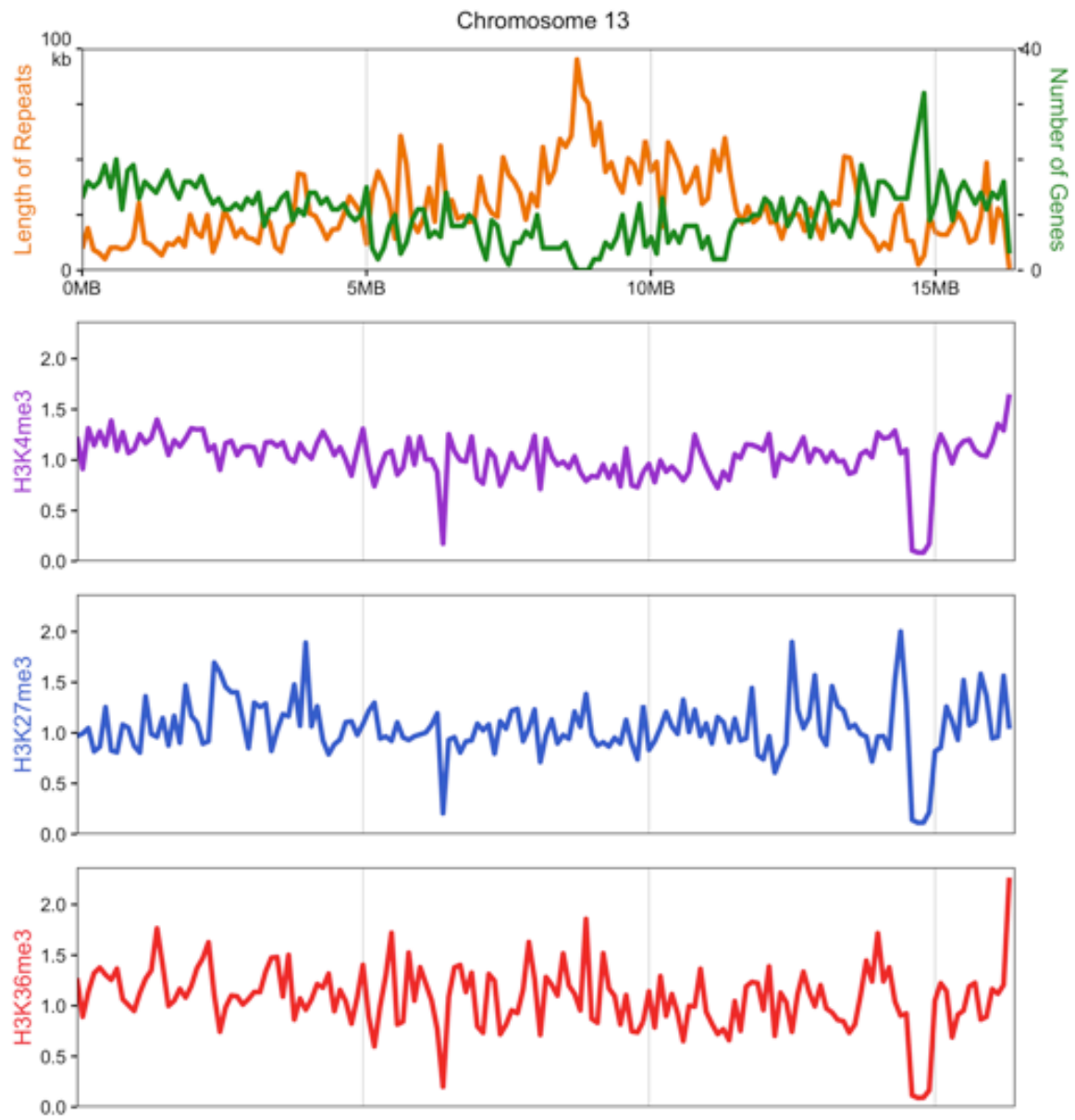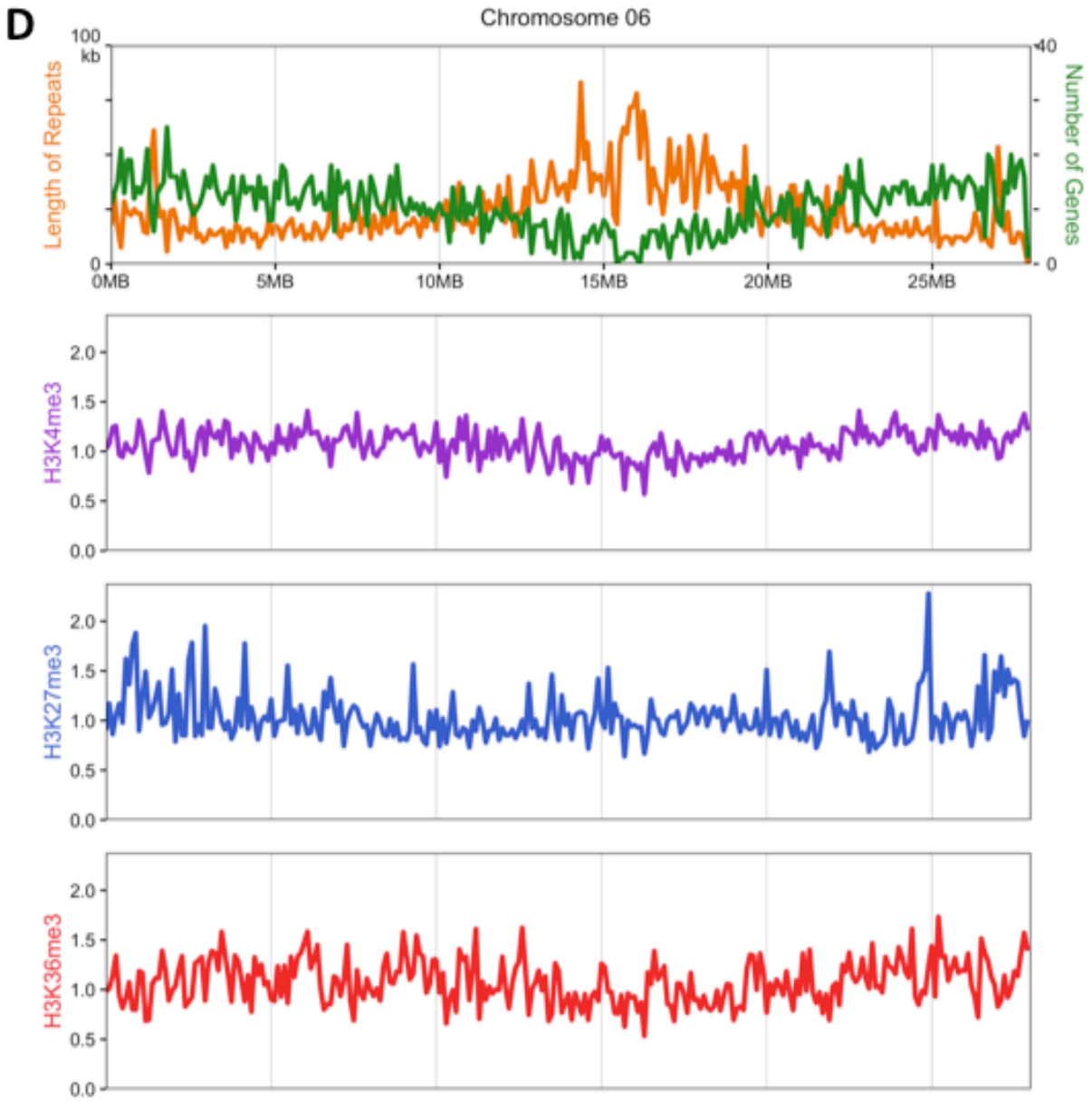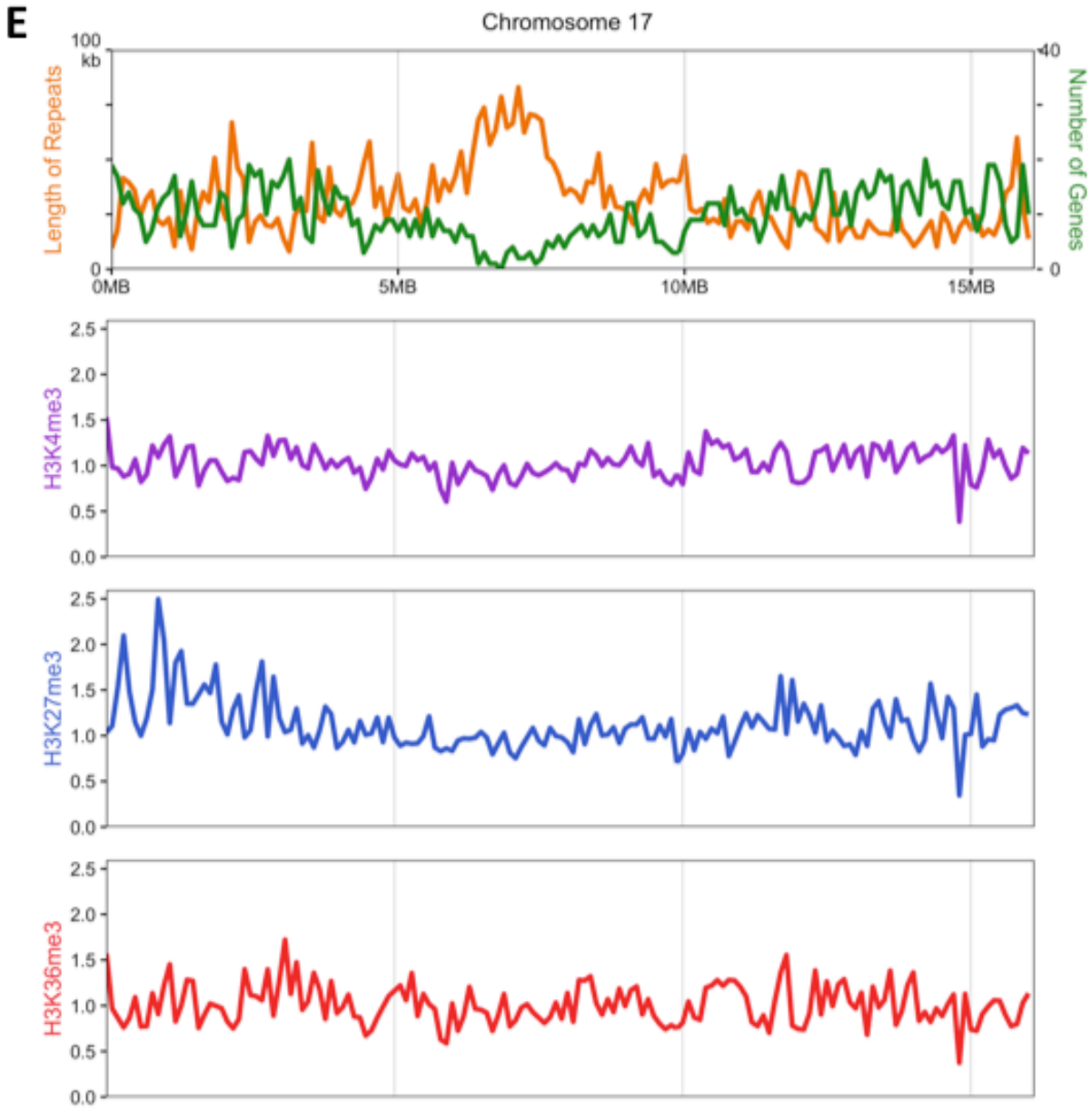
D Chromosome 06

Figure 1. Chromosomal distribution of histone methylation

Top Panel: the length of repetitive sequences (y-axis, left side) and number of genes per 100kb bin (y-axis, right side). Purple line represents H3K4me3 ChIP enrichment values divided by input enrichment values per 100kb bin. Blue line represents H3K27me3 ChIP enrichment values divided by input enrichment values per 100kb bin. Red line represents H3K36me3 ChIP enrichment values divided by input enrichment values per 100kb bin.
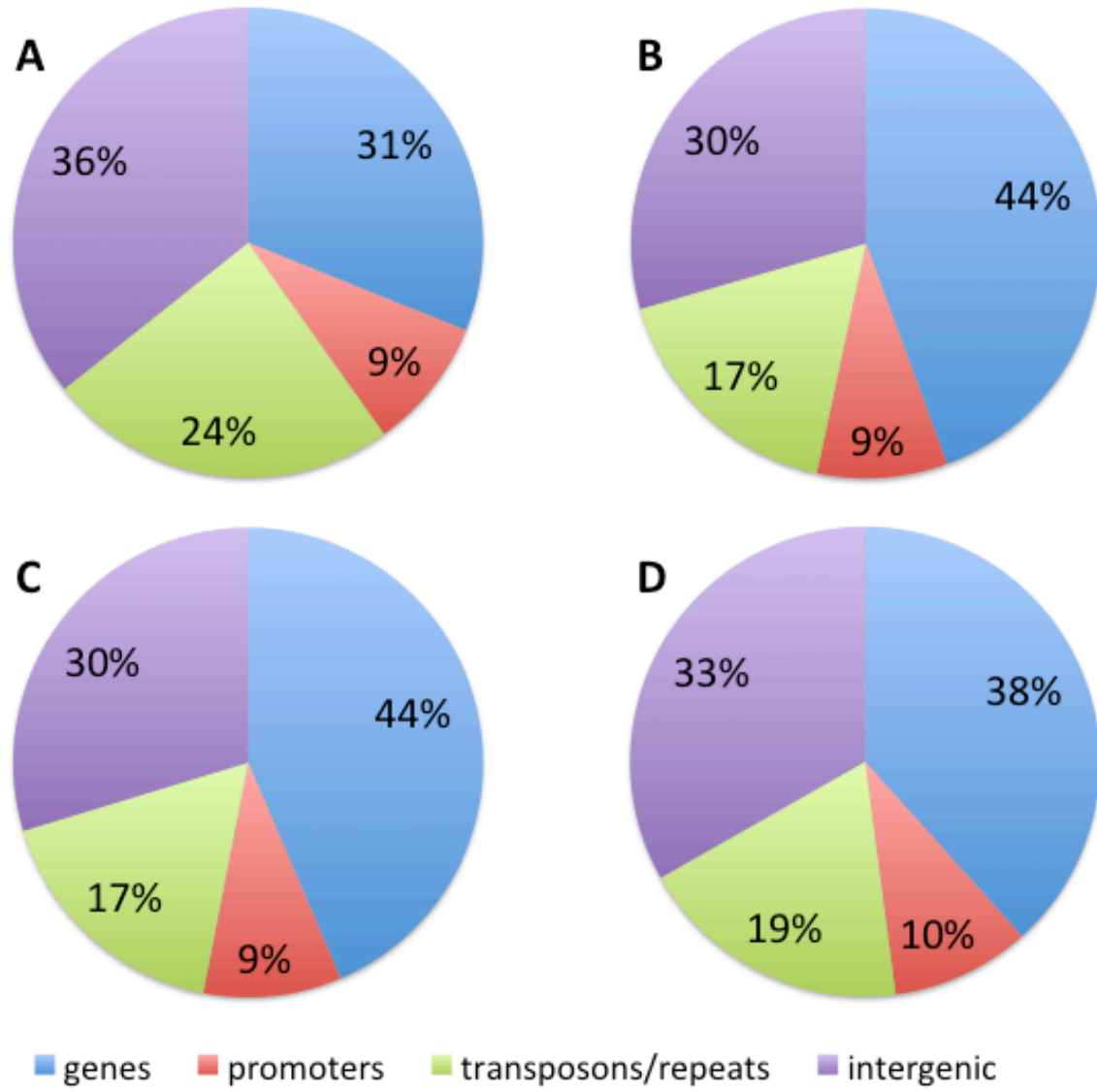
Figure 2. H3K36me3, H3K37me3, and H3K4me3 are associated with genic regions in poplar

A) Percentages of sPta717 genome covered by genomic features: genes, promoters, transposons, and intergenic region

B) Percentage of base pairs covered by H3K36me3 peaks per genomic feature

C) Percentage of base pairs covered by H3K4me3 peaks per genomic feature

D) Percentage of base pairs covered by H3K27me3 peaks per genomic feature

A H3K36me3

B H3K4me3
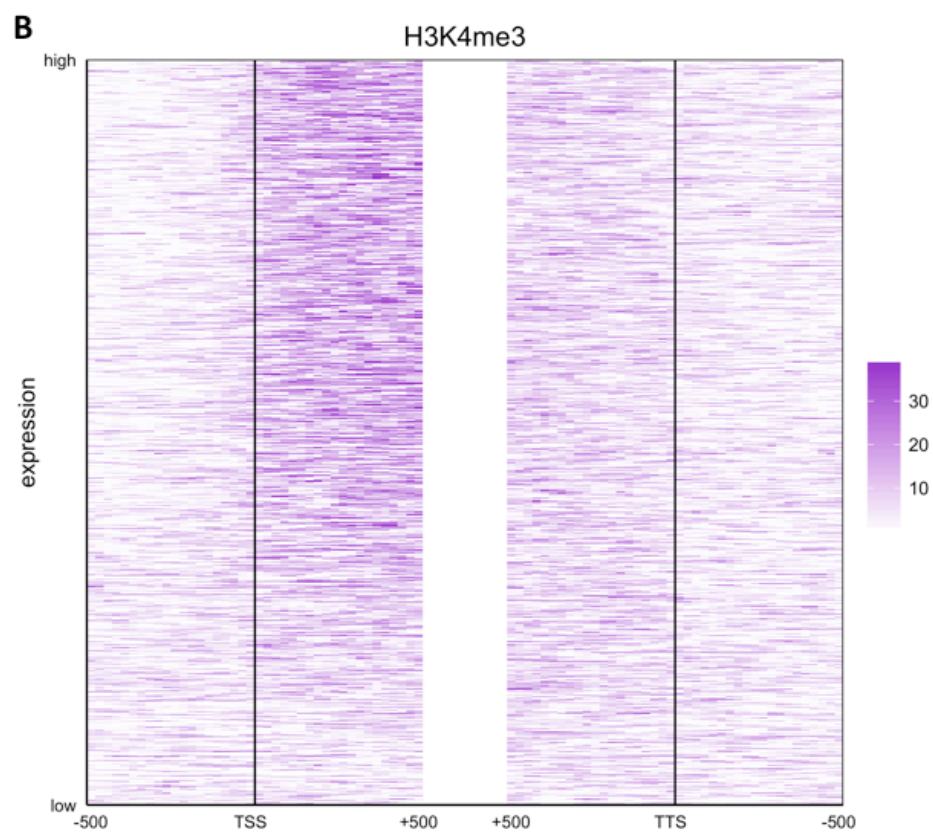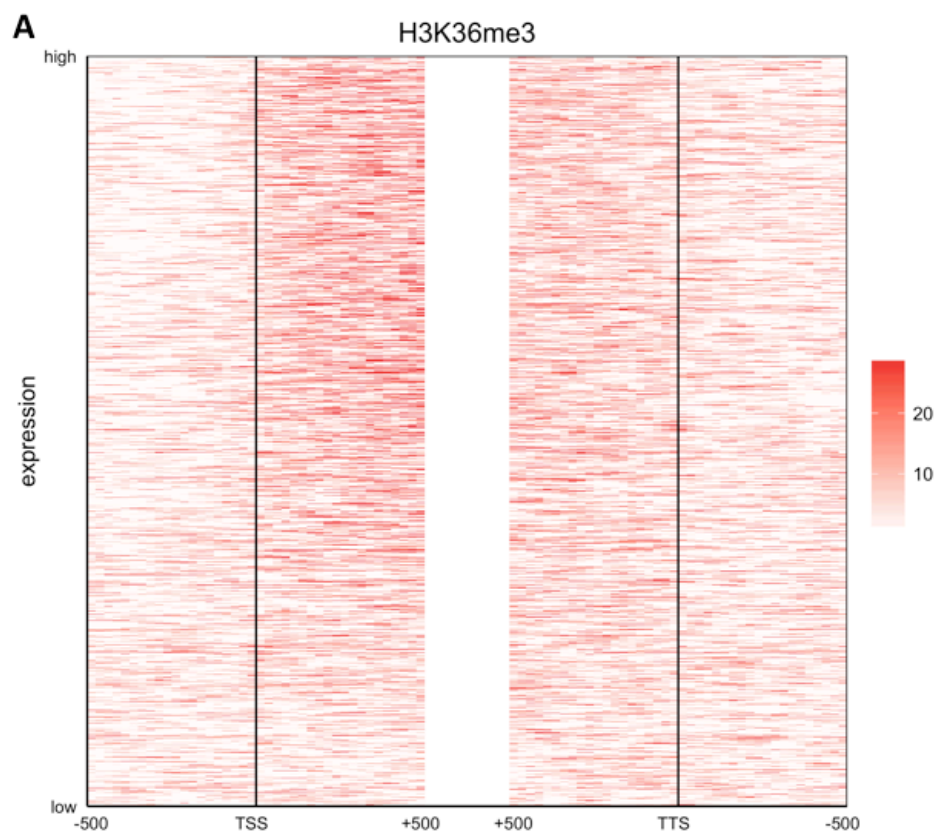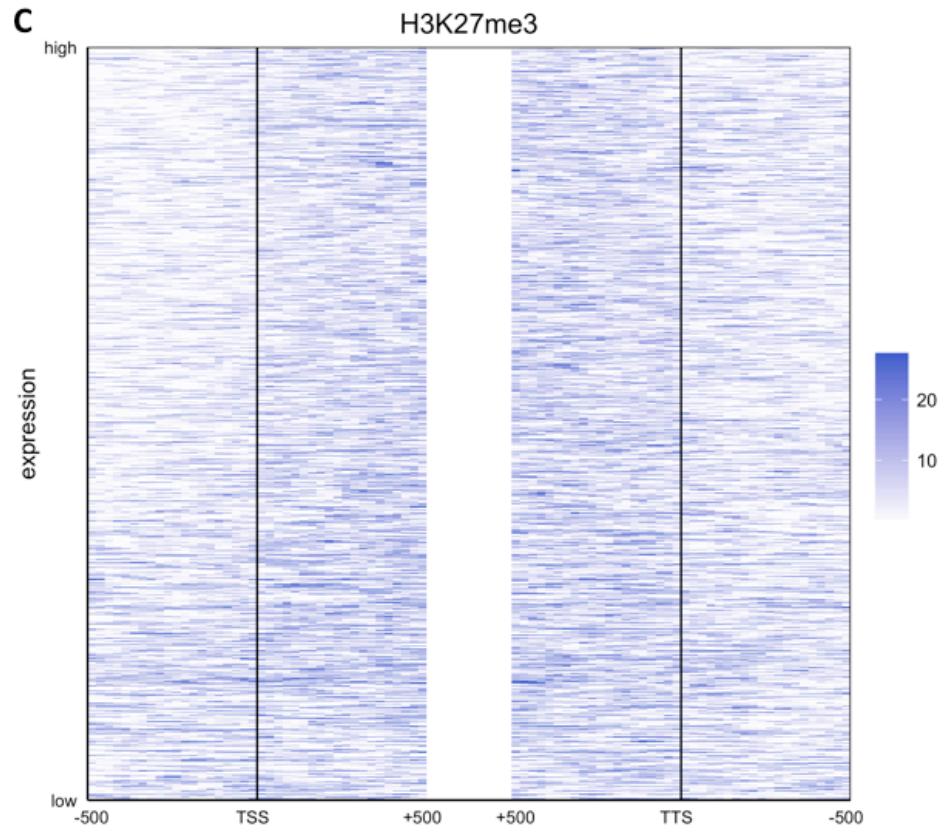
Figure 3. Distribution of histone methylation across poplar genes

Histone modification enrichment normalized by input, 500bp up and downstream of TSS and TTS. Each row represents 1 gene. Genes were ordered from highest to lowest expression (y-axis). A) H3K36me3 heatmap B) H3K4me3 heatmap C) H3K27me3 heatmap
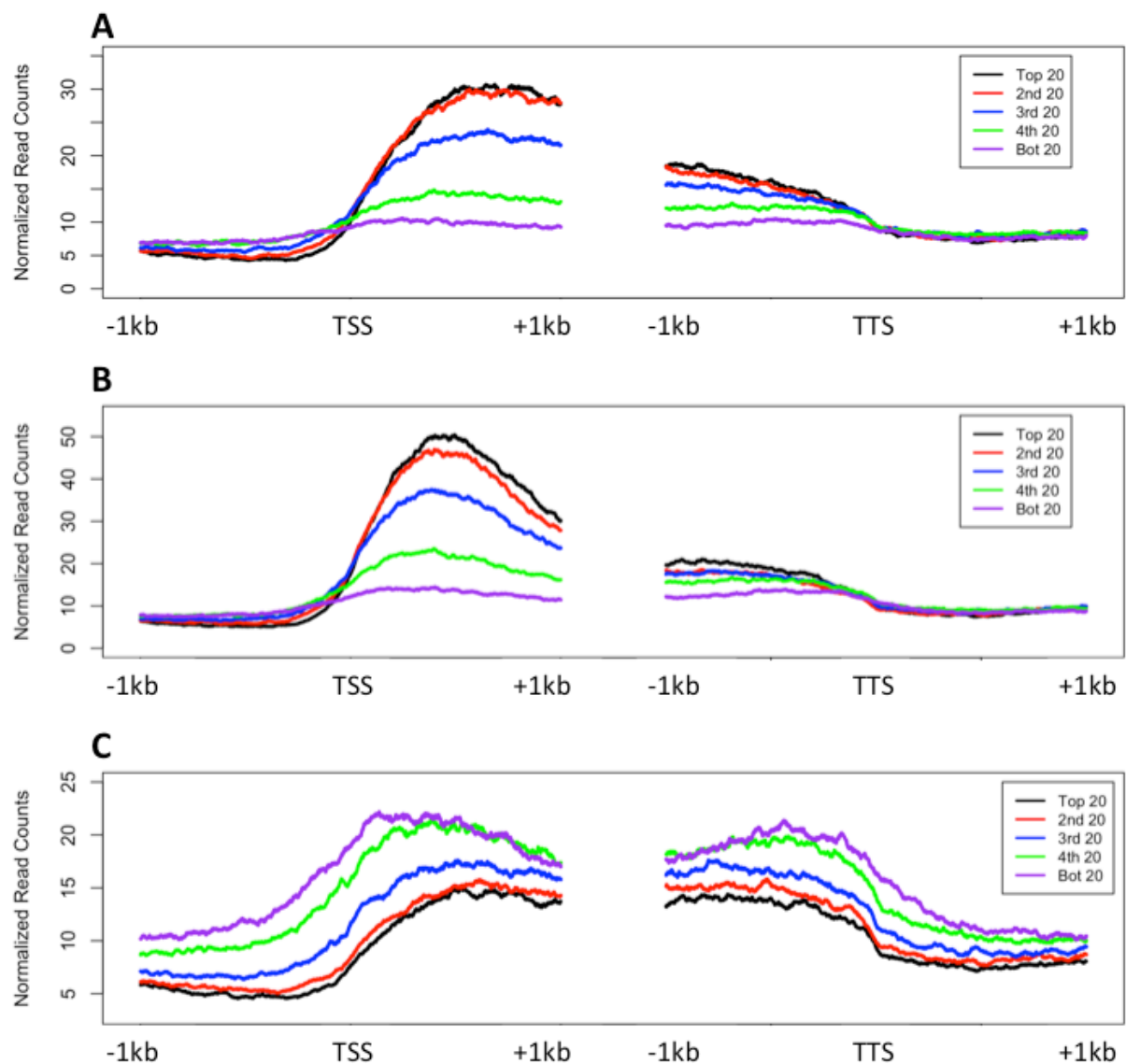
Figure 4. Association of histone methylation with gene expression

Metaplots depicting histone modification enrichment normalized by input, 1kb up and downstream of TSS and TTS. Genes were grouped by expression level with "Top 20" being the 20% highest expressed genes, "Bot 20" the 20% lowest expressed genes, and the other groups in between. A) H3K36me3 metaplot B) H3K4me3 metaplot C) H3K27me3 metaplot
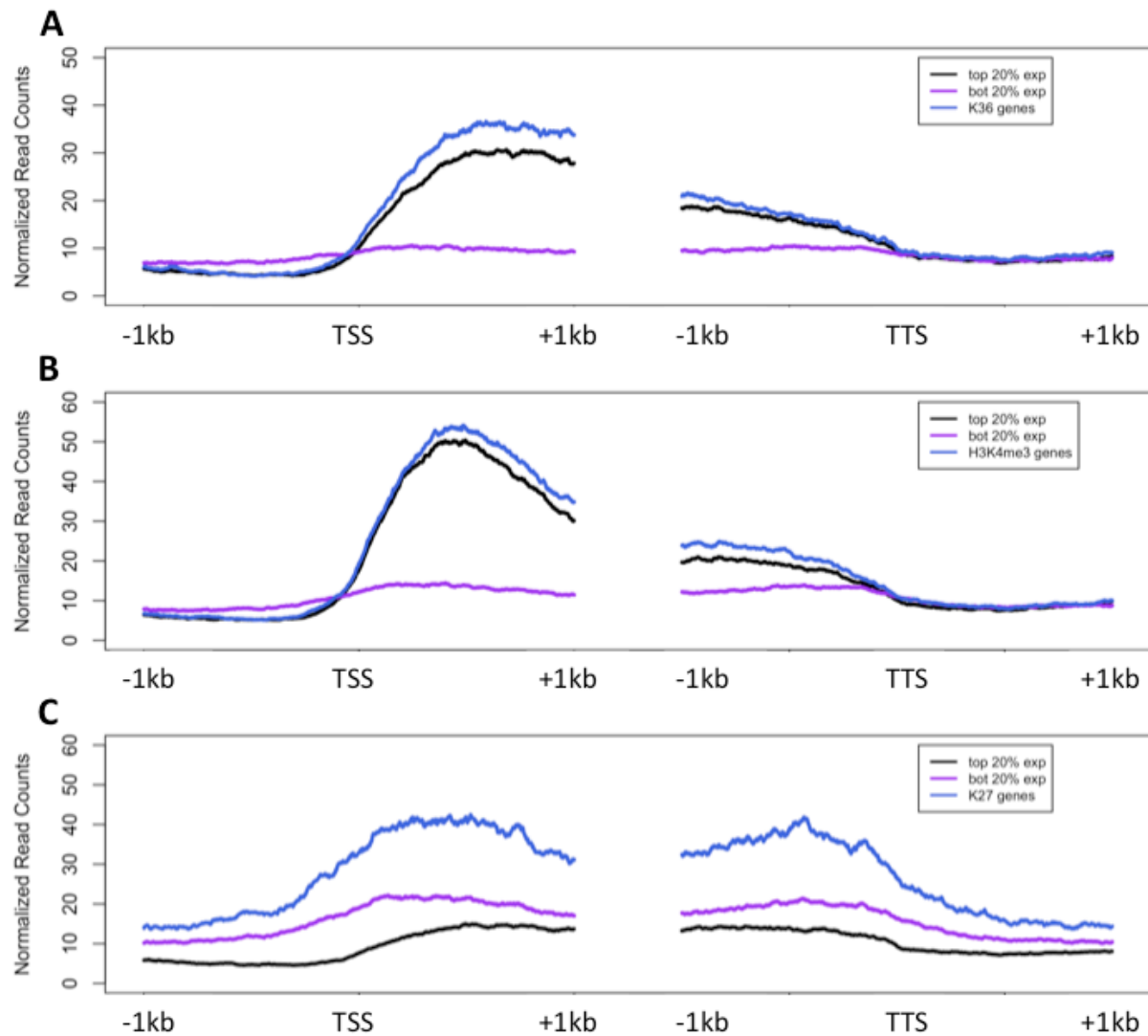
Figure 5. Identification of poplar genes with H3K36me3, H3K4me3, and H3K27me3

Metaplots depicting histone modification enrichment normalized by input, 1kb up and

downstream of TSS and TTS.

A) H3K36me3 genes have higher levels of H3K36me3 than "Top 20" expression group

 B) H3K4me3 genes have higher levels of H3K4me3 than "Top 20" expression group

C) H3K4me3 genes have higher levels of H3K27me3 than "Bot 20" expression group

A

B
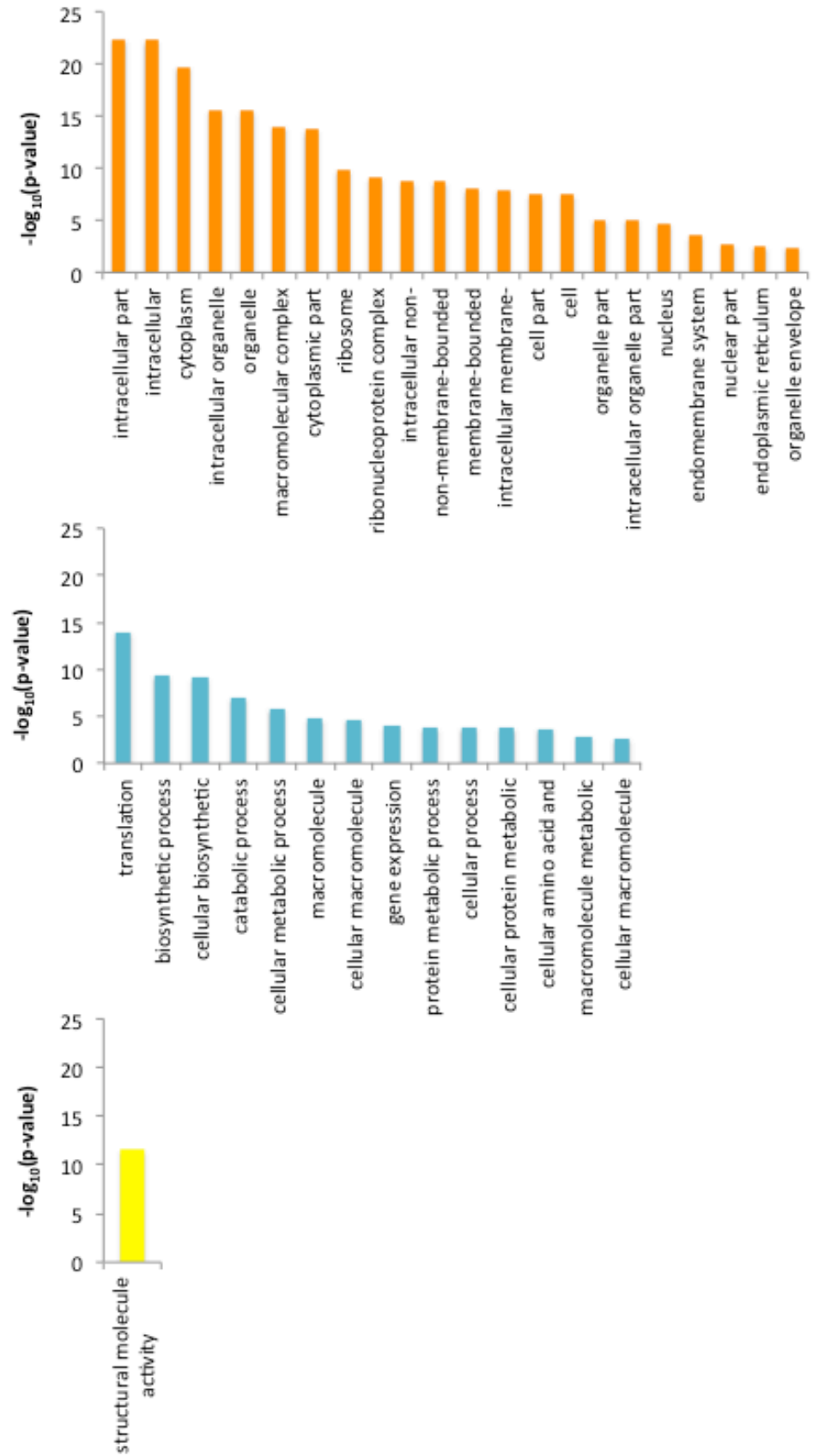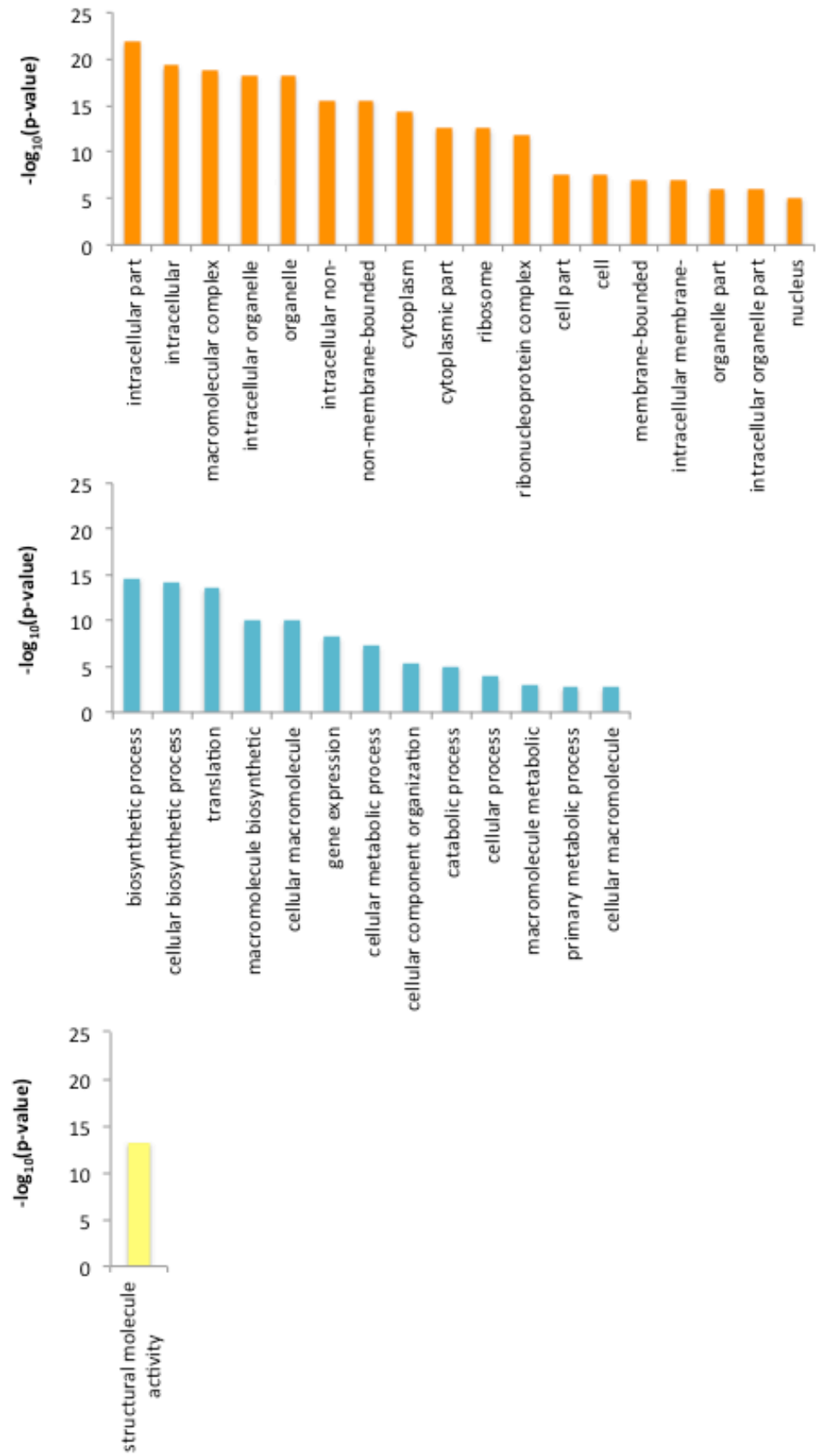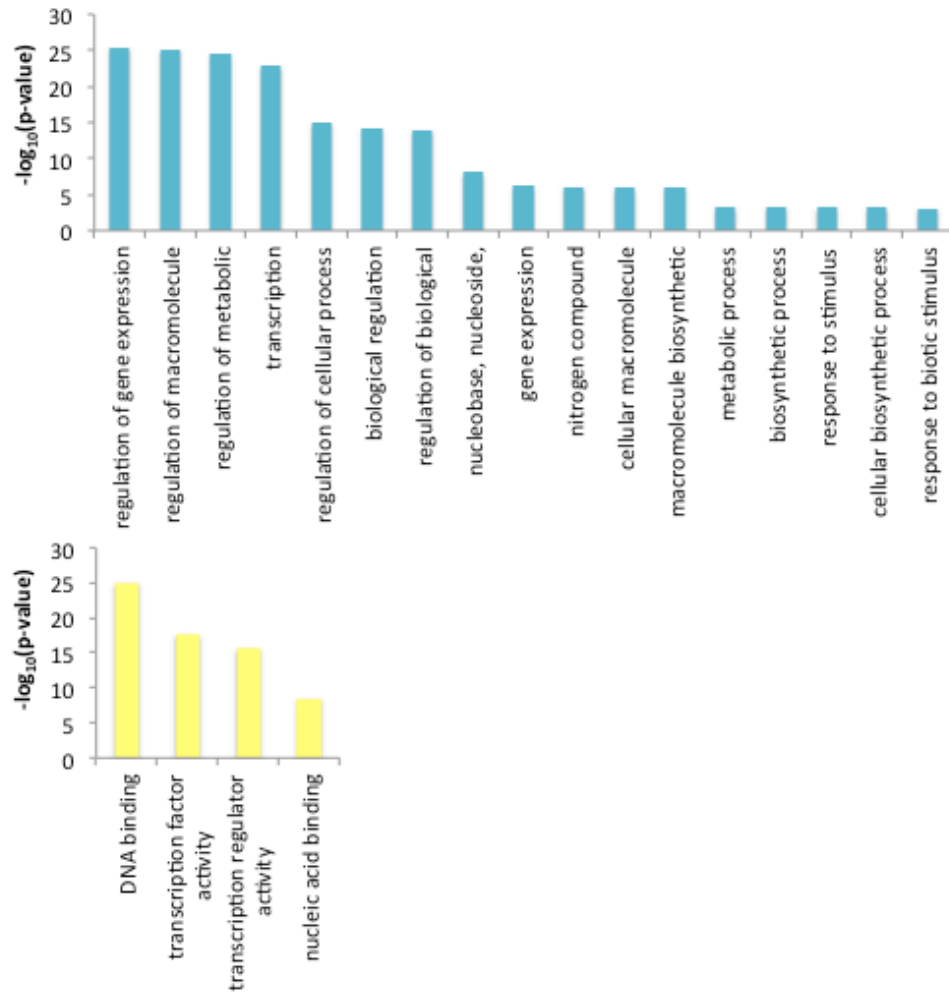
Figure 6. GO terms enrichment for A) H3K36me3, B) H3K4me3, and C) H3K27me3 genes. The panels with orange, blue, and yellow bars are Cellular Component, Biological Process, and Molecular Function GO terms, respectively. H3K27me3 genes were not enriched for Cellular Component GO terms.
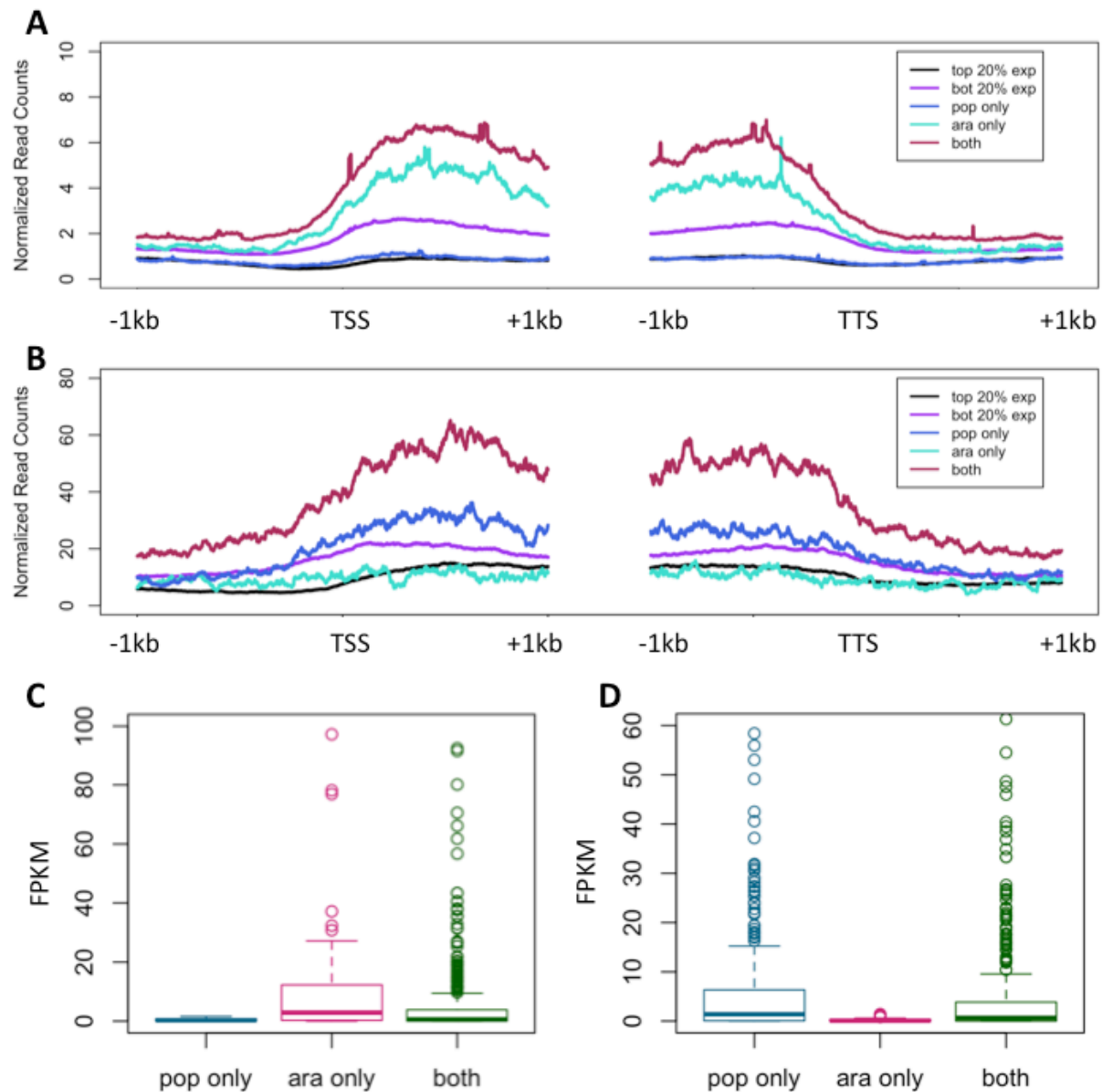
Figure 7. Differential H3K27me3 methylation in Arabidopsis and poplar

A) H3K27me3 enrichment in "Pop only", "Ara only", and "Both" gene groups in sPta717

B) H3K27me3 enrichment in "Pop only", "Ara only", and "Both" groups in Arabidopsis

C) Expression levels of "Pop only", "Ara only", and "Both" groups using sPta717 RNA-seq

D) Expression levels of "Pop only", "Ara only", and "Both" groups using Arabidopsis RNA-seq

CHAPTER 3

PROFILING PLANT HISTONE MODIFICATIONS

Introduction

Eukaryotic chromatin is made up of DNA-histone complexes, called nucleosomes.

Histones are subject to a wide array of chemical modification to conserved residues on their N-

terminal tails. Histone modifications can affect the nature of histone-DNA interaction and

chromatin packaging[26]. They also can serve as recruitment platforms for chromatin interacting

proteins involved in processes that occur on the chromatin template; including DNA replication,

RNA processing, and chromatin remodeling[11,80–82].

ChIP is an essential methodology to understand the distribution of histone modifications

and other chromatin-associated proteins throughout the genome. Performing ChIP in plant cells

requires special care to isolate chromatin due to the rigid plant cell wall. This protocol has been

utilized in our lab to perform ChIP of histone modifications in multiple plant species; including

Arabidopsis thaliana, *Populus tremulua x alba* 717-1B4, and *Eutrema penlandii*(L. Guo et al.

2010; Zhang 2008). In comparison to Native ChIP methodologies, this protocol utilizes

formaldehyde to cross-link the plant tissue[84].

In this protocol, plant nuclei are purified and chromatin is isolated from the nuclei. The

chromatin is then sheared by sonication to enrich for DNA fragments between 300 base pairs

(bp) and 1kilobase pairs (kb). After sonication, the sheared chromatin is incubated with

antibodies that recognize specific histone modifications. The chromatin fragments that associate

with the antibody are immunoprecipitated using protein A/G beads. The chromatin attached to

the beads will be enriched for the histone modification of interest. DNA is then isolated from the

chromatin that associated with the beads. After DNA extraction, it is recommended to perform

qPCR before creating libraries for ChIP sequencing, which is reviewed in Haring et. al[85]. ChIP-

sequencing libraries can be made using Illumina Tru-seq ChIP Sample Prep Kit.

Materials

1-2 grams of plant tissue

37% formaldehyde

.1 M PMSF (see recipe)

Complete Protease Inhibitor tablets, EDTA-free (Roche)

Complete Mini Protease Inhibitor tablets, EDTA-free (Roche)

Cross-linking buffer (see recipe)

1 M glycine (see recipe)

Protein A or G magnetic beads (?brand)

Extraction Buffer 1(see recipe)

Extraction Buffer 2 (see recipe)

Extraction Buffer 3 (see recipe)

Nuclear Lysis Buffer (see recipe)

ChIP Dilution Buffer (see recipe)

Elution Buffer (see recipe)

Low Salt Wash Buffer (see recipe)

High Salt Wash Buffer (see recipe)

LiCl Wash Buffer (see recipe)

TE Buffer (see recipe)

5 M NaCl

1 M Tris-Cl pH 6.5

.5 M EDTA

2.5 M Kac

20 mg/mL proteinase K

Phenol, buffered pH 8.0

Chloroform

Ethanol

glycogen

Special Equipment:

Bioruptor® UCD-200 (or other sonicator)

Vacuum centrifuge

Magnetic rack for 1.5 ml eppendorf tubes

Protocol

Chromatin Cross-linking

1. Mix 15.4 ml of Cross-linking Buffer with 160 µl of 0.1 M PMSF and 432 µl of 37%

   formaldehyde.

   Add formaldehyde inside hood. Add PMSF just before use as it only lasts 30 minutes in

   aqueous solution.

2. Harvest 1-2 g of above ground plant tissue.

3. In a 25 ml beaker, combine solution with plant tissue and incubate for 10 minutes under vacuum. Release the vacuum and mix after 5 minutes, then reapply vacuum).

   Use a vacuum centrifuge with the rotor removed for this step.

4. Release the vacuum and add 1.6 ml 1M glycine. Mix and reapply the vacuum for 5 minutes.

   Glycine stops the cross-linking by neutralizing the formaldehyde.

5. Rinse leaves with H_2O five times, then remove excess H_2O by blotting with paper towel. Flash freeze cross-linked tissue in liquid nitrogen and store at -80°C .

   The more water left on the tissue, the more difficult the tissue will be to grind in future steps. Cross-linked tissue can be stored at -80°C until ready to begin the ChIP experiment.

Prepare Antibody-Coated Beads

6. Prepare 3-4 tubes of antibody-coated beads per 1-2 g of cross-linked starting material. To prepare beads: place 25 µl of protein A or G magnetic beads in a 1.5 ml eppendorf tube. Wash the beads 3 times with ChIP dilution buffer. Add 100 µl of ChIP dilution buffer and 1-2 µg of antibody to the washed beads. Incubate by rotating at 4°C for 3 hr (could rotate overnight).

   All steps should be performed at 4°C or on ice. 1 to 2 g of cross-linked tissue will produce enough chromatin to perform ChIP with 3-4 antibodies. To wash, place tubes with beads in magnetic rack and take off supernatant. Add 1 ml of ChIP dilution buffer. Then, remove tubes from magnetic rack and resuspend beads via inversion. When all the beads are resupsended, place tube on magnetic rack. Repeat wash two more times

Chromatin Isolation *all steps should be performed at 4°C or on ice.*

7. Grind 1-2 g of cross-linked tissue in liquid nitrogen to a very fine powder.

8. Add powder to 30 ml of Extraction Buffer 1 in a 50 ml falcon tube. Resuspend the powder with the vortex and by inverting the tube.

9. Filter the solution through miracloth into a 50 ml falcon tube.

10. Centrifuge the filtrate for 20 minutes at 4000 rpm , 4°C

    This step pellets the nuclei and other organelles, without bursting the nuclei.

11. Remove supernatant and resuspend pellet in 1 ml of Extraction Buffer 2.

    This step bursts the chloroplasts, but does not contain enough detergent to burst the nuclei.

12. Centrifuge 10 minutes at 12000 x g, 4°C.

13. Remove supernatant and resuspend pellet in 300 µl Extraction Buffer 3 with a pipet.

    Pellet will be very difficult to resuspend. It is okay if a small amount of the pellet cannot be resupsended.

14. In a new 1.5 ml eppendorf tube, add 300 µl of Extraction Buffer 3. Carefully layer the 300 µl solution (resupsended pellet) from step 12 on top of the 300 µl of Extraction Buffer 3.

    T*his step creates a sucrose gradient, in which the chloroplasts remain in the supernatant and the nuclei in the pellet.*

15. Spin for 1 hour at 16,000 x g, 4°C.

16. Remove the supernatant and resuspend the pellet in 300 µl Nuclei Lysis Buffer. Resuspend the pellet via vortex and pipet. Keep 2-5 µl of this sample as unsonicated control and freeze in -20°C.

This step bursts the nuclei. Keep solution cold between vortexing and try to avoid getting

too many bubbles in the solution.

17. Sonicate the chromatin solution in a time dependent on the species and type of sonicator.

    The goal size of DNA after sonication is 300-1000 bp. Different plant species require

    different sonication times to reach that goal. For Arabidopsis thaliana, our lab sonicates

    for 15 minutes twice. The sample is put on ice for 5 minutes between rounds of

    sonication.

18. Centrifuge sonicated chromatin max speed, 4°C. Put supernatant into new 1.5 ml

    eppendorf tube. Keep 11 µl as input control and freeze in -20°C.

    Following this step, chromatin can be frozen in -20°C. If working with a plant species

    other than Arabidopsis, it is highly recommended to extract DNA (see DNA extraction

    section) from the unsonicated control (step #16) and the input control (step #18) before

    continuing with Chromatin Immunoprecipitation.

Chromatin Immunoprecipitation

19. Place rest of supernatant into a 15 ml falcon tube and bring up to 3 ml with ChIP dilution

    buffer.

20. Wash prepared antibody coated beads 3x with ChIP dilution buffer.

    To wash, place tubes with antibody coated beads in magnetic rack and take off

    supernatant. Add 1 ml of ChIP dilution buffer. Then, remove tubes from magnetic rack

    and resuspend beads via inversion. When all the beads are resupsended, place tube on

    magnetic rack. Repeat wash two more times.

21. Add 750 µl (for 4 samples) or 1 ml (for 3 samples) of the chromatin solution to the

    antibody coated beads and incubate rotating overnight at 4°C.

22. Attach beads to magnet and remove supernatant.

23. Wash beads two times with each buffer listed below. For the 1ˢᵗ wash, add 1 ml of buffer to beads. Remove beads from magnet and resuspend the beads via inversion. Place beads back on magnet and take off supernatant. For the 2ⁿᵈ wash, add 1 ml of buffer then rotate for 5 minutes, 4°C. After 5 minutes, take off supernatant.

    a. Low Salt Wash Buffer.

    b. High Salt Wash Buffer

    c. LiCl Wash Buffer

24. Wash antibody coated beads with TE Buffer one time with no rotation. Leave beads attached to magnetic rack after take off supernatant.

25. Add 250 µl of Elution Buffer to beads. Vortex beads to mix and incubate at 65°C for 15 minutes.

26. Place the tube in the magnetic rack and carefully transfer the supernatant to another 1.5 ml tube and repeat elution. Combine the two eluates.

27. Add 20 µl of 5M NaCl to eluate and reverse cross-link at 65°C overnight. Defrost unsonicated control (step 16) and input control (step 18). Add 500 µl of Elution Buffer and 20 µl of 5M NaCl to the unsonicated control and input control and place at 65°C with the eluate samples.

28. Add 10 µl of 0.5M EDTA, 20 µl Tris-Cl pH 6.5, and 1 µl of 20 mg/ml proteinase K to elutate and control samples. Incubate for 1 hr at 45°C.

DNA Extraction

29. Add 250 µl of phenol and 250 µl of chloroform to the samples. Vortex, then centrifuge 5 min at maximum speed, RT.

30. Transfer supernatant to 1.5 ml tube. Add 26 µl KaC 2.5 M, ~1 ml of 100% Ethanol, and 2 µl of glycogen. Place tubes in -20°C for 2 hr or overnight.

31. Centrifuge 30 minutes at maximum speed, 4°C.

    There is usually a white pellet after this step. If there is not a pellet, it is okay. Continue with the protocol.

32. Remove supernatant and add 1 ml of 70% ethanol.

33. Centrifuge the sample for 1 minute at maximum speed, RT.

34. Take off the supernatant and allow the pellet to air dry.

35. Resuspend the pellet in 50 µl of H_2O.

    This is the final ChIP DNA that can be used to make DNA libraries for sequencing in the future. It is highly recommended to continue to Support Protocol 1.

36. Run 5 µl of unsonicated and input DNA samples on 1% agarose gel to check that sonication worked properly.

Reagents and Solutions

<u>ChIP Dilution Buffer</u>

	For 10 ml:
1.2 mM EDTA	24 µl 0.5 M
167 mM NaCl	334 µl 5 M
16.7 mM Tris-Cl, pH 8.0	167 µl 1 M
1.1% (v/v) Triton X-100	1.1 ml 10% (v/v)
0.1 mM PMSF	20 µl 0.1 M
Complete mini tablet	1 tablet

H₂O to volume

Mix well and store at 4°C (check for contamination before use). Add BME, PMSF, and Complete Protease Inhibitor tablets just before use. Resuspend Complete Protease Inhibitor tablets in small volume of cold nuclease-free water before adding to ChIP Dilution Buffer. Protease inhibitors do not need to be added when using the ChIP dilution buffer to wash the beads.

<u>Cross-linking buffer</u>

	For 100 ml:
1 mM EDTA, pH 8.0	0.2 ml .5 M
0.4 M sucrose	13.7 g
10mM Tris-Cl, pH 8.0	1 ml 1M
	H$_2$O to volume

Mix well and store up to 3 months at 4°C (check for contamination before use).

<u>Elution Buffer</u>

	For 20 ml:
1% (v/v) SDS	2 ml 10% (w/v)
0.1 M NaHCO$_3$	0.168 g
	H$_2$O to volume

Store at room temperature (20 - 26°C).

<u>Extraction Buffer 1 (EB1)</u>

	For 100 ml:
1 mM EDTA	200 μl 0.5 M
10 mM MgCl$_2$	1 ml 1.0 M

0.4 M sucrose	20 ml 2 M
10 mM Tris-Cl pH 8.0	1 ml 1 M
5mM BME	35 μl 14.3 M
0.1 mM PMSF	200 μl 0.1 M
Complete tablets	2 tablets
	H_2O to volume

Mix well and store at 4°C (check for contamination before use). Add BME, PMSF, and Complete Protease Inhibitor tablets just before use. Resuspend Complete Protease Inhibitor tablets in small volume of cold nuclease-free water before adding to Extraction Buffer 1.

Extraction Buffer 2

	For 10 ml:
1 mM EDTA	200 μl 0.5 M
10 mM $MgCl_2$	100 μl 1 M
0.25 M sucrose	1.25 ml 2M
10 mM Tris-Cl, pH 8.0	100 μl 1M
1% (v/v) Triton X-100	1 ml 10% (v/v)
0.1 mM PMSF	20 μl 0.1 M
5 mM BME	3.5μl 14.3 M
Complete mini tablets	1 tablet
	H_2O to volume

Mix well and store at 4°C (check for contamination before use). Add BME, PMSF, and Complete Protease Inhibitor tablets just before use. Resuspend Complete Protease Inhibitor tablets in small volume of cold nuclease-free water before adding to Extraction Buffer 2.

Extraction Buffer 3

	For 10ml:
1 mM EDTA	20 µl 0.5M
2 mM $MgCl_2$	100 µl 1 M
1.7 M sucrose	8.5 ml 2M
10 mM Tris-Cl, pH 8	100 µl 1M
0.15% (v/v) Triton X-100	150 µl 10% (v/v)
0.1 mM PMSF	20 µl 0.1 M
Complete mini tablet	1 tablet
	H_2O to volume

Mix well and store at 4°C (check for contamination before use). Add BME, PMSF, and

Complete Protease Inhibitor tablets just before use. Resuspend Complete Protease Inhibitor

tablets in small volume of cold nuclease-free water before adding to Extraction Buffer 3.

High Salt Wash Buffer

	For 50 ml:
2mM EDTA	200 µl 0.5 M
500 mM NaCl	5 ml 5M
0.1% (v/v) SDS	0.5 ml 10% (w/v)
20 mM Tris-Cl, pH 8	1 ml 1M
1% (v/v) Triton X-100	5 ml 10% (v/v)
	H_2O to volume

Mix well and store at 4°C (check for contamination before use).

LiCl Wash Buffer

	For 50 ml:
1 mM EDTA	100 µl 0.5 M
0.25 M LiCl	3.125 ml 4 M
1% (w/v) sodium deoxycholate	0.5 g
1% (v/v) NP-40	2.5 ml 20% (v/v)
10 mM Tris-Cl, pH 8	0.5 ml 1M
	H$_2$O to volume

Mix well and store at 4°C (check for contamination before use).

Low Salt Wash Buffer

	For 50 ml:
2 mM EDTA	200 µl 0.5 M
150 mM NaCl	1.5 ml 5 M
0.1% (v/v) SDS	0.5 ml 10% (w/v)
20 mM Tris-Cl, pH 8	1 ml 1 M
1% (v/v) Triton X-100	5 ml 10% (v/v)
	H$_2$O to volume

Mix well and store at 4°C (check for contamination before use).

Nuclear Lysis Buffer

	For 5 ml:
10 mM EDTA	100 µl 0.5 M
1% (v/v) SDS	0.5 ml 10% (w/v)
50 mM Tris-Cl, pH 8	0.25 ml 1 M

0.1 mM PMSF	20 µl 0.1M
Complete Mini tablet	½ tablet
	H_2O to volume

Mix well and store at 4°C (check for contamination before use). Add PMSF, and Complete

Protease Inhibitor tablet just before use. Resuspend Complete Protease Inhibitor tablet in small

volume of cold nuclease-free water before adding to Nuclear Lysis Buffer.

<u>0.1M PMSF</u>

	For 50 ml:
PMSF	0.87 g
	Isopropanol to volume

Dissolve PMSF in isopropanol, store 200-µl aliquots up to 1 year at -20°C.

<u>TE Buffer</u>

	For 50 ml:
10 mM Tris-Cl, pH 8	0.5 ml 1M
1 mM EDTA	100 µl 0.5M
	H_2O to volume

Mix well and store at 4°C (check for contamination before use).

Background Information

The development of the modern ChIP protocol was preceded by the discovery UV and later

formaldehyde could be used to cross-link DNA to proteins. This discovery allowed researchers

to study the distribution of RNA Polymerase II in bacteria and later in *Drosophila*

melanogaster[86,87]. In 1988, formaldehyde was first used to crosslink DNA to proteins and was

utilized to study the distribution of H4 on heat shock genes in in *Drosophila melanogaster*[88].
Since then formaldehyde and histone specific antibodies have been used to study the distribution
of histone modifications on genes and across the genome. With the advent of modern
sequencing techniques, studying the distribution of histone modifications throughout the genome
of any species has become widely available.

Critical Parameters and Troubleshooting

1. Antibodies for ChIP

    Choosing antibodies for ChIP is one of the most important considerations when designing
    an experiment. Commercial histone modification antibodies are not typically raised against
    plant proteins. For that reason, it is recommended that you chose antibodies that are raised
    against an invariant domain of the histone protein. Since histone proteins so highly
    conserved, many of the same antibodies that work for mammalian species work in plant
    species.

    Often times, you will have multiple antibodies to chose from when deciding what to use
    for your ChIP experiment. However, you many have varying levels of success due to due to
    differences in epitope recognition by the antibody. Polyclonal antibodies recognize several
    epitopes of the target, so they may perform better in a ChIP experiment. It is our experience
    that antibodies that work in *Arabidopsis thaliana* typically work in other plant species. The
    antibodies our lab most commonly uses in Arabidopsis are: H3K4me3 (Millipore, 07-473),
    H3K36me3 (abcam, ab9050), H3K27me3 (Millipore, 07-449), and H3 (abcam, ab1791).
    These antibodies have worked in poplar as well.

2. Sonication

Sonication should produce fragments between 300 bp and 1000 bp. Chromatin should be

kept at 4°C throughout sonication. In our hands, different plant species require distinct

sonication times to reach the desirable degree of sonication.

3. Quality of ChIP DNA

qPCR should be utilitzed to check whether the quality of ChIP DNA is sufficient before

performing ChIP sequencing or drawing any conclusions[85].

Anticipated Results

After making ChIP-seq libraries, ChIP results can be viewed on an IGV browser to

determine if there is ChIP signal enrichment over the input control. If the ChIP experiment did

not work, the signal will look like a background, with no or very few peaks. Distribution of

ChIP signal is dependent on which histone target you use. For example: H3K36me3 distribution

in *Arabidopsis thaliana* should be within the transcribed genes across the gene body(Zhang

2008). Peak calling programs, like SICER, can be utilized to detect regions of histone

modification enrichment[76].

Time Considerations

The protocol should take 4-5 days. There are points specified in the protocol that you can

stop for the day if you do not take these recommendations.

If working with a species and tissue type that have worked before:

On day 1, perform crosslinking (steps 1-5). On day 2, perform the chromatin isolation

and incubate the antibody-coated beads with the sonicated chromatin overnight (steps 6-21). On

day 3, wash the beads and reverse cross–link overnight (steps 22-27). On day 4, finish the ChIP

protocol and extract the DNA. Run the agarose gel to make sure that your sonication worked

(steps 28-36). This timeline assumes that you already know the right sonication conditions for

the species or tissue type that you are working with.

If working with a plant species that you have not worked with before:

We recommend in this case that you make sure that your sonication worked properly,

before continuing with the ChIP experiment. Do this by isolating ChIP DNA from the

unsonicated control and the input samples and running the agarose gel. On day 1, perform

crosslinking (steps 1-5). On day 2, skip preparing the antibody-coated beads (step 6) and proceed

to isolate the chromatin (step 6-18). Reverse crosslink the unsonicated control and input DNA

overnight (step 27). On day 3, extract DNA from the unsonicated control and input (steps 28-36).

Run the agarose gel to make sure that the sonicated worked to a desirable degree. If not, sonicate

the chromatin samples longer to ensure proper sonication. After finishing sonication, prepare the

antibody-coated beads (step 6). Then, incubate the chromatin with antibody-coated beads

overnight (steps 19-21). On day 4, wash the beads and reverse cross-link overnight (steps 22-27).

On day 5, finish the ChIP protocol and extract the DNA. Run the agarose gel to make sure that

your sonication worked (steps 28-36).

REFERENCES

1.   Hamzeh M, Dayanandan S. Phylogeny of Populus (Salicaceae) Based on Nucleotide

      Sequences of Chloroplast trnt-trnf Region and Nuclear rDNA. *Am J Bot*.

      2004;91(9):1398-1408.

2.   Bradshaw HD, Ceulemans R, Davis J, Stettler R. Emerging model systems in plant

      biology: Poplar (Populus) as a model forest tree. *J Plant Growth Regul*. 2000;19(3):306-

      313. doi:10.1007/s003440000030.

3.   Xue LJ, Alabady MS, Mohebbi M, Tsai CJ. Exploiting genome variation to improve next-

      generation sequencing data analysis and genome editing efficiency in Populus tremula ??

      alba 717-1B4. *Tree Genet Genomes*. 2015;11(4). doi:10.1007/s11295-015-0907-5.

4.   Tuskan G a., Difazio S, Jansson S, et al. The genome of black cottonwood, Populus

      trichocarpa. *Science (80-)*. 2006;313(5793):1596. doi:10.1126/science.1128691.

5.   Bräutigam K, Vining KJ, Lafon-Placette C, et al. Epigenetic regulation of adaptive

      responses of forest tree species to the environment. *Ecol Evol*. 2013;3(2):399-415.

      doi:10.1002/ece3.461.

6.   Bossdorf O, Richards CL, Pigliucci M. Epigenetics for ecologists. *Ecol Lett*.

      2008;11(2):106-115. doi:10.1111/j.1461-0248.2007.01130.x.

7.   Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PCG.

      Epigenetic variation in mangrove plants occurring in contrasting natural environment.

      PLoS One. 2010;5(4):1-8. doi:10.1371/journal.pone.0010326.

8.   Allona I, Conde D, Gonza P. Poplar stems show opposite epigenetic patterns during

winter dormancy and vegetative growth. 2013:311-320. doi:10.1007/s00468-012-0800-x.

9.   Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*.

        1974;184(139):868-871. doi:10.1126/science.184.4139.868.

10.   Luger K, Maeder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the

        nucleosome core particle at 2.8Aa resolution. *Nature*. 1997;389(6648):251-260.

        http://dx.doi.org/10.1038/38444.

11.   Kouzarides T. Chromatin Modifications and Their Function. 2007:693-705.

        doi:10.1016/j.cell.2007.02.005.

12.   Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications.

        Biochim Biophys Acta. 2014. doi:10.1016/j.bbagrm.2014.03.001.

13.   Finnegan EJ, Genger RK, Peacock WJ, Dennis ES. Dna Methylation in Plants. *Annu Rev

        Plant Physiol Plant Mol Biol*. 1998;49:223-247.

14.   Calarco JP, Borges F, Donoghue MTA, et al. Reprogramming of DNA methylation in

        pollen guides epigenetic inheritance via small RNA. *Cell*. 2012;151(1):194-205.

        doi:10.1016/j.cell.2012.09.001.

15.   Petruk S, Sedkov Y, Johnston DM, et al. TrxG and PcG proteins but not methylated

        histones remain associated with DNA through replication. *Cell*. 2012;150(5):922-933.

        doi:10.1016/j.cell.2012.06.046.

16.   Niederhuth CE, Schmitz RJ. Covering your bases: Inheritance of DNA methylation in

        plant genomes. *Mol Plant*. 2014;7(3):472-480. doi:10.1093/mp/sst165.

17.   Law J a, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns

        in plants and animals. *Nat Rev Genet*. 2010;11(3):204-220. doi:10.1038/nrg2719.

18.   Ehrlich M, Gama-Sosa MA, Huang LH, et al. Amount and distribution of 5-

methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res*. 1982;10(8):2709-2721. doi:10.1093/nar/10.8.2709.

19.   Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet*. 2009;10(5):295-304. doi:10.1038/nrg2540.

20.   Cokus SJ, Feng S, Zhang X, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*. 2008;452(7184):215-219. doi:10.1038/nature06745.

21.   Lister R, O'Malley RC, Tonti-Filippini J, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*. 2008;133. doi:10.1016/j.cell.2008.03.029.

22.   Zhang X, Yazaki J, Sundaresan A, et al. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell*. 2006;126(6):1189-1201. doi:10.1016/j.cell.2006.08.003.

23.   Feng S, Cokus SJ, Zhang X, Chen P, Bostick M, Goll MG. Conservation and divergence of methylation patterning in plants and animals. 2010. doi:10.1073/pnas.1002720107.

24.   Stroud H, Do T, Du J, et al. The roles of non-CG methylation in Arabidopsis. *Nat Struct Mol Biol*. 2014;21(1):64-72. doi:10.1126/science.1246980.Common.

25.   Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol*. 2015;16(9):519-532. doi:10.1038/nrm4043.

26.   Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011;21(3):381-395. doi:10.1038/cr.2011.22.

27.   Cosgrove MS, Boeke JD, Wolberger C. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol*. 2004;11(11):1037-1043. doi:10.1038/nsmb851.

28.   Cosgrove MS, Wolberger C. How does the histone code work? *Biochem Cell Biol*. 2005;83:468-476. doi:10.1139/o05-137.

29.   Zhang X. Chromatin Modifications in Plants. Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, eds. *Plant Genome Divers*. 2012;1. doi:10.1007/978-3-7091-1130-7.

30.   Yun M, Wu J, Workman JL, Li B. Readers of histone modifications. *Cell Res*. 2011;21(4):564-578. doi:10.1038/cr.2011.42.

31.   Jacobs S a, Khorasanizadeh S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science*. 2002;295(5562):2080-2083. doi:10.1126/science.1069473.

32.   Lo WS, Trievel RC, Rojas JR, et al. Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Mol Cell*. 2000;5(6):917-926. doi:10.1016/S1097-2765(00)80257-9.

33.   Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol Cell*. 2000;5(6):905-915. doi:10.1016/S1097-2765(00)80256-7.

34.   Hoppmann V, Thorstensen T, Kristiansen PE, et al. The CW domain, a new histone recognition module in chromatin proteins. *EMBO J*. 2011;30(10):1939-1952. doi:10.1038/emboj.2011.108.

35.   Springer NM, Napoli C a, Selinger D a, et al. Comparative analysis of SET domain proteins in maize and Arabidopsis reveals multiple duplications preceding the divergence of monocots and dicots. *Plant Physiol*. 2003;132(2):907-925. doi:10.1104/pp.102.013722.

36.   Johnson L, Mollah S, Garcia BA, et al. Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res*.

2004;32(22):6511-6518. doi:10.1093/nar/gkh992.

37.   Zhang X, Bernatavichute Y V, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome Biol*. 2009;10(6):R62. doi:10.1186/gb-2009-10-6-r62.

38.   Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129(4):823-837. doi:10.1016/j.cell.2007.05.009.

39.   Santos-Rosa H, Schneider R, Bannister AJ, et al. Active genes are tri-methylated at K4 of histone H3. *Nature*. 2002;419(6905):407-411. doi:10.1038/nature01071.Published.

40.   Zhang X, Bernatavichute Y V, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome Biol*. 2009;10. doi:10.1186/gb-2009-10-6-r62.

41.   Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell*. 2003;11(3):709-719. doi:10.1016/S1097-2765(03)00092-3.

42.   Sims RJ, Millhouse S, Chen CF, et al. Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Mol Cell*. 2007;28(4):665-676. doi:10.1016/j.molcel.2007.11.010.

43.   Shi X, Hong T, Walter KL, et al. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature*. 2006;442(7098):96-99. doi:10.1038/nature04835.

44.   Matthews AGW, Kuo AJ, Ramón-Maiques S, et al. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature*. 2007;450(7172):1106-1110. doi:10.1038/nature06431.

45.   Lauberth SM, Nakayama T, Wu X, et al. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell*. 2013;152(5):1021-1036. doi:10.1016/j.cell.2013.01.052.

46.   Hung T, Binda O, Champagne KS, et al. ING4 Mediates Crosstalk between Histone H3 K4 Trimethylation and H3 Acetylation to Attenuate Cellular Transformation. *Mol Cell*. 2009;33(2):248-256. doi:10.1016/j.molcel.2008.12.016.

47.   Alvarez-Venegas R, Pien S, Sadder M, Witmer X, Grossniklaus U, Avramova Z. ATX-1, an Arabidopsis Homolog of Trithorax, Activates Flower Homeotic Genes. *Med Eng Phys*. 2003;13:627-637. doi:10.1016/S.

48.   Guo L, Yu Y, Law JA, Zhang X. SET DOMAIN GROUP2 is the major histone H3 lysine 4 trimethyltransferase in Arabidopsis. *Proc Natl Acad Sci*. 2010;107(43):18557-18562. doi:10.1073/pnas.1010478107.

49.   Barski A, Cuddapah S, Cui K, et al. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*. 2007;129(4):823-837. doi:10.1016/j.cell.2007.05.009.

50.   Roudier F, Teixeira FK, Colot V. Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. *Trends Genet*. 2009;25. doi:10.1016/j.tig.2009.09.013.

51.   Kim SY, He Y, Jacob Y, Noh Y-S, Michaels S, Amasino R. Establishment of the vernalization-responsive, winter-annual habit in Arabidopsis requires a putative histone H3 methyl transferase. *Plant Cell*. 2005;17(12):3301-3310. doi:10.1105/tpc.105.034645.

52.   Xu L, Zhao Z, Dong A, et al. 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*. 2008;28(4):1348-1360. doi:10.1128/mcb.01607-07.

53. Zhao Z, Yu Y, Meyer D, Wu C, Shen W-H. Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. *Nat Cell Biol*. 2005;7(12):1256-1260. doi:10.1038/ncb1329.

54. Xiao T, Hall H, Kizer KO, et al. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev*. 2003;17(5):654-663. doi:10.1101/gad.1055503.

55. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. *Science*. 2010;327(5968):996-1000. doi:10.1126/science.1184208.

56. Guo R, Zheng L, Park JW, et al. BS69/ZMYND11 Reads and Connects Histone H3.3 Lysine 36 Trimethylation-Decorated Chromatin to Regulated Pre-mRNA Processing. *Mol Cell*. 2014:1-13. doi:10.1016/j.molcel.2014.08.022.

57. Wen H, Li Y, Xi Y, et al. ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature*. 2014;508(7495):263-268. doi:10.1038/nature13045.

58. Carrozza MJ, Li B, Florens L, et al. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell*. 2005;123(4):581-592. doi:10.1016/j.cell.2005.10.023.

59. Venkatesh S, Smolle M, Li H, et al. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature*. 2012;489(7416):452-455. doi:10.1038/nature11326.

60. Pan G, Tian S, Nie J, et al. Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell*. 2007;1(3):299-312. doi:10.1016/j.stem.2007.08.003.

61. Zhao XD, Han X, Chew JL, et al. Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. *Cell Stem Cell*. 2007;1(3):286-298. doi:10.1016/j.stem.2007.08.004.

62. Zhang X, Clarenz O, Cokus S, Bernatavichute Y V, Pellegrini M, Goodrich J. Whole-Genome Analysis of Histone H3 Lysine 27 Trimethylation in Arabidopsis. 2007;5(5). doi:10.1371/journal.pbio.0050129.

63. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature*. 2011;469(7330):343-349. doi:10.1038/nature09784.

64. Cao R, Wang L, Wang H, et al. Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science (80-)*. 2002;298(5616):1039-1043. doi:10.1126/science.1084274.

65. Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*. 2002;111(2):185-196. doi:10.1016/S0092-8674(02)00975-3.

66. Kuzmichev A, Nishioka K, Erdjument-bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev*. 2002;16:2893-2905. doi:10.1101/gad.1035902.

67. Muller J, Hart CM, Francis NJ, et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*. 2002;111(2):197-208. doi:10.1016/S0092-8674(02)00976-5.

68. Workman JL, Abmayr SM. *Fundamentals of Chromatin*.

69. Mohn F, Weber M, Rebhan M, et al. Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. *Mol Cell*.

2008;30(6):755-766. doi:10.1016/j.molcel.2008.05.007.

70. Gendall AR, Levy YY, Wilson A, Dean C. The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell*. 2001;107(4):525-535. doi:10.1016/S0092-8674(01)00573-6.

71. Wang X, Elling AA, Li X, et al. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell*. 2009;21. doi:10.1105/tpc.109.065714.

72. Roudier F, Ahmed I, Bérard C, et al. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J*. 2011;30(10):1928-1938. doi:10.1038/emboj.2011.103.

73. Pien S, Grossniklaus U. Polycomb group and trithorax group proteins in Arabidopsis. *Biochim Biophys Acta*. 2007;1769(5-6):375-382. doi:10.1016/j.bbaexp.2007.01.010.

74. Bewick A, Ji JL, Niederhuth CE, et al. On the Origin and Evolutionary Consequences of Gene Body DNA Methylation Adam J. Bewick. *bioRxiv Prepr*. 2016.

75. Du Z, Zhou X, Ling Y, Zhang Z, Su Z. agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res*. 2010;38(SUPPL. 2):64-70. doi:10.1093/nar/gkq310.

76. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. 2009;25(15):1952-1958. doi:10.1093/bioinformatics/btp340.

77. Roudier F, Ahmed I, Bérard C, et al. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J*. 2011;30(10):1928-1938. doi:10.1038/emboj.2011.103.

78. Rosenfeld JA, Wang Z, Schones DE, Zhao K, DeSalle R, Zhang MQ. Determination of

enriched histone modifications in non-genic portions of the human genome. *BMC Genomics*. 2009;10(1):143. doi:10.1186/1471-2164-10-143.

79.   Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev*. 2010;24(3):265-276. doi:10.1101/gad.544410.

80.   Li B, Carey M, Workman JL. The Role of Chromatin during Transcription. *Cell*. 2007;128(4):707-719. doi:10.1016/j.cell.2007.01.015.

81.   Martin C, Zhang Y. The Diverse Functions of Histone Lysine Methylation. *Nat Rev Mol Cell Biol*. 2005;6(November):838-850. doi:10.1038/nrm1761.

82.   Ji X, Dadon DB, Abraham BJ, Ihn T, Jaenisch R, Bradner JE. Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. 2015. doi:10.1073/pnas.1502971112.

83.   Zhang X. The epigenetic landscape of plants. *Science (80-)*. 2008;320(5875):489-492. doi:10.1126/science.1153996.

84.   Neill LPOÕ, Turner BM. Immunoprecipitation of native chromatin : NChIP. 2003;31:76-82. doi:10.1016/S1046-2023(03)00090-2.

85.   Haring M, Offermann S, Danker T, Horst I, Peterhansel C, Stam M. Chromatin immunoprecipitation : optimization , quantitative analysis and data normalization. 2007;16:1-16. doi:10.1186/1746-4811-3-11.

86.   Gilmour DS, Lis JT. In Vivo Interactions of RNA Polymerase II with Genes of Drosophila melanogaster. *Mol Cell Biol*. 1985;5(8):2009-2018.

87.   Gilmour DS, Lis JT. Detecting protein-DNA interactions in vivo : Distribution of RNA polymerase on specific bacterial genes. *Proc Natl Acad Sci*. 1984;81(July):4275-4279.

88.   Solomon, Mark J., Pamela L. Larsen and AV. Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell*. 1988;53(6):937-947.