Genome-wide analysis of histone modifications in heterotic hybrid and its parental lines of Chinese cabbage

HASAN MEHRAJ

(Degree)
博士（農学）

(Date of Degree)
2022-03-25

(Date of Publication)
2023-03-01

(Resource Type)
doctoral thesis

(Report Number)
甲第8369号

(URL)
https://hdl.handle.net/20.500.14094/D1008369

Doctoral Dissertation

Genome-wide analysis of histone modifications in heterotic hybrid and its parental lines of Chinese cabbage

Hasan Mehraj

February 2022

Graduate School of Agricultural Science
Kobe University

Contents

Chapter I

General Introduction

Genus *Brassica* is in the Brassicaceae family, and Brassicaceae consists of more than 330 genera and 3800 species (Bailey et al. 2006, Huang et al. 2016). Plants in the genus Brassica are known as cruciferous or cole crops. It includes many plants of economic importance, which includes vegetables (Chinese cabbage, cabbage, napa cabbage, cauliflower, broccoli, kale, pak choi, mizuna, komatsuna, Brussels sprouts, kohlrabi, rutabaga, turnip), oilseeds (canola), the condiment (mustard), fodder and many wild species grown as weeds. *Brassica* plants get attention for scientific interest due to their genomic pattern and agricultural importance. Among the *Brassica* species, six species are the highest level of scientific interest, and the evolution and genomic relationships of these six *Brassica* species have been described by the "Triangle of U" (UN 1935). The "Triangle of U" described that three ancestral diploid *Brassica* species, *Brassica rapa* L. (AA genome), *B. nigra* L. (BB genome), and *B. oleracea* L. (CC genome), are undergone natural hybridization and developed three allotetraploid species, *B. juncea* L. (AABB genome), *B. napus* L. (AACC genome), and *B. carinata* L. (BBCC genome). Later, molecular studies in the *Brassica* genome proved the "Triangle of U" theory (Chalhoub et al. 2014, Yang J et al. 2016a, Kim et al. 2018, Xue et al. 2020). *Brassica* genome underwent a whole-genome triplication, which is considered as a crucial event for the species diversification and intra-species morphotypes (Cheng et al. 2014, Cheng et al. 2017). It is considered that *Brassica* and the model plant *Arabidopsis thaliana* diverged from a common ancestor 14.5-20.4 million years ago (Blanc et al. 2003, Bowers et al. 2003, Koenig and Weigel, 2015), and belong to the same family. *B. rapa* is a diverse and one of the ancestral species of *B. juncea* and *B. napus*, and it consists of many commercially important leafy vegetables (subsp. *pekinensis* - Chinese cabbage, subsp. *chinensis* - pak choi and narinosa, subsp. *perviridis* - komatsuna, subsp. *nipposinica* - mizuna) and root vegetables (subsp. *rapa* - turnip) (Prakash et al. 2012, Lv et al. 2020). It is difficult to recognize the native of *B. rapa*, while the Mediterranean region is considered as the center of origin of *B. rapa* and subsequently it spread across the world (Dixon 2006, Guo et al. 2014).

Characteristics of any living organisms has been regulated by the gene expression. DNA is the key structure of any living organisms and DNA sequence is the blueprint of all information. DNA sequences are the determinant of gene expression and functions. Alterations in the DNA sequences of gene can alter the gene expressions/ functions. Change in gene expressions without changes in the DNA sequences is known as epigenetics, and it is occurred by different kind of modifications to the DNA that can make an epigenetically modified gene turn on or off. The

epigenetic modification is a natural process and it is reversible. Chromatin, DNA and protein consisting in chromosome, can actively controlled by the epigenetic regulators including DNA methylation, chromatin remodeling, histone posttranslational modifications (PTMs) and non-coding RNAs (ncRNAs). These epigenetic regulators can participate in any kind of biological process at any stage of plant growth and development to regulate the gene expression. The epigenetic regulation depends on the environmental response. In all eukaryotic cells, 147 bp of DNA wraps the histone octamer of the nucleosome in the chromatin in the nucleus, and a histone octamer has two copies of each of H2A, H2B, H3, and H4 core histone proteins (Itabashi et al. 2018, Kim 2021, Talbert and Henikoff 2021). Each histone protein has several amino acid residues such as lysine (K), arginine (R), glycine (G), serine (S), leucine (L), tyrosine (Y), etc. in their N-terminal tail (Li et al. 2007, Bannister and Kouzarides 2011, Black et al. 2012, Zhao et al. 2019). Amino acid residues in the histone tail can be modified differently like methylation (me), acetylation (Ac), phosphorylation (Ph), ubiquitylation (Ub), and sumoylation (Su) to change the gene expression where methylation can be mono-/ di-/ tri-methylated (me1/me2/me3) (Li et al. 2007, Bannister and Kouzarides 2011, Black et al. 2012, Zhao et al. 2019, Demetriadou et al. 2020). Histone modification can regulate the chromatin structure and significantly change the gene expression in plant development and response to stresses (Kim et al. 2015, Meyer 2015). Methylation in few sites of lysine group (lysine 4, K4; lysine 9, K9; lysine 27, K27; and lysine 36, K36) of H3 is consider as the most important epigenetic transcriptional regulators for the many biological functions (Li et al. 2007, Bannister and Kouzarides 2011, Demetriadou et al. 2020).

Genome-wide histone distribution patterns of methylation levels in K4, K9, K27, K36 of histone H3 has already been studied in model plant *A. thaliana* (Bernatavichute et al. 2008, Oh et al. 2008, Turck et al. 2007, Zhang et al. 2007, Zhang et al. 2009, Roudier et al. 2011). In plants, tri-methylation of histone H3 in lysine 4 (H3K4me3) and H3K36me3 activate the transcriptions while H3K9me2 and H3K27me3 repress the transcription (Fujimoto et al. 2012a, Kim et al. 2015, Quadrana and Colot 2016). Histone modifications are very complex and tissue-specific. Besides the independent existence of a histone marks, it is also possible for the synchronous existence of two different histone marks (Kouzarides 2007, Li et al. 2007). Existence of the functionally two opposite histone marks together is known as bivalent histone modifications. Co-existence H3K4me3- and H3K27me3-marks is the most prominent bivalent histone modifications that can result transcriptional activation or repression or poised states, and bivalent chromatin states have been identified in plants which are transcriptionally associated with plant stress-responsive gene expression (Sequeira-Mendes et al. 2014, Qian et al. 2018,

Zeng et al. 2019, Blanco et al. 2020). High-throughput chromatin immunoprecipitation sequencing (ChIP-seq) is used to detect the regions having histone methylation.

The central dogma of molecular biology suggested that the information flow is from the DNA to RNA to Protein (Crick 1970). Initial genome sequence analyses in higher organisms and in *A. thaliana* revealed that only a tiny percentage of the genome are occupied by protein-coding gene (Nowak 1994, Yamada et al. 2003). The remaining transcripts were considered as "junk". However, the so-called "junk" DNA is not trash anymore. The large-scale transcriptomic analyses and high-throughput gene sequencing technology disclosed that a large proportion of the eukaryotic genome is transcribed into RNAs and does not encode proteins. RNAs transcripts lacking protein-coding potential are known as non-protein-coding RNAs or ncRNAs (Collins and Penny 2009, Chen et al. 2021). NcRNAs have a diverse group of transcripts, and considering the regulatory roles, and ncRNAs are mainly categorized into two types. One category is housekeeping ncRNAs and another category is regulatory ncRNAs. Housekeeping ncRNAs such as ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs) are universally expressed in cells and regulate general cellular functions (Collins and Penny 2009**,** Cech and Steitz 2014, Ariel et al. 2015, Zhang et al. 2019). Regulatory ncRNAs are considered as regulators of gene expression at epigenetic, transcriptional, and post-transcriptional levels (Cech and Steitz 2014). According to the length, regulatory ncRNAs have been classified into long ncRNAs (lncRNAs), medium-sized ncRNAs, small ncRNAs (sRNAs). LncRNAs transcripts are longer than 200 nucleotides (nt) in length, medium-sized ncRNAs transcripts are 31-200 nt in length, and sRNAs transcripts are ~18–30 nt in length (Cech and Steitz 2014, Ariel et al. 2015, Chekanova 2015, Liu X et al. 2015, Mattick and Rinn 2015, Karlik et al. 2019, Rai et al. 2019, Song et al. 2021). Based on the position and orientation of transcription, lncRNAs further has been classified as long intergenic noncoding RNAs (lincRNAs), intronic noncoding RNAs (incRNAs), and natural antisense transcripts (NATs) (Cech and Steitz 2014, Ariel et al. 2015, Karlik et al. 2019, Rai et al. 2019, Jha et al. 2020, Song et al. 2021). LincRNAs originate from intergenic regions those are weakly spliced with trans-regulatory function, incRNAs are transcribed from intronic regions, and NATs originate from complementary DNA strands of their associated genes (Wu et al. 2017, Karlik et al. 2019, Rai et al. 2019, Jha et al. 2020). Based on mode of biogenesis and mechanism, there are two broad categories of sRNAs, microRNAs (miRNAs) and small interference RNAs (siRNAs) (Bologna and Voinnet 2014, Liu and Chen 2018). miRNAs are endogenous single stranded RNA (ssRNA), typically 21–22 nt in length, and RNA hairpin structure; while siRNAs are an exogenous double-stranded RNA (dsRNA) that have multiple subclass such as 21-, 22- and 24-nt siRNAs (Liu and Chen 2018, Lunardon et al. 2020). Among the number of siRNA subclasses, 24-nt siRNAs are

considered as the dominant siRNA occurred in the intergenic regions especially at the 5′-flanking regions of protein-coding genes (Lunardon et al. 2020). RNA-sequencing (RNA-seq) allow us to disclose the functional analysis of various genes at their epigenetic level (Li et al. 2018). RNA-seq helps to identify the novel ncRNAs including lncRNAs, and their role in regulating plant growth and development, and responses to biotic and abiotic stress (Yu et al. 2019, Budak et al. 2020, Jha et al. 2020).

Heterosis or hybrid vigor is the superiority of the heterozygous F_1 hybrids for growth, speed of development, fertility, biomass, and resistance compared to the homozygous parental lines (Lippmann and Zamir 2007, Birchler et al. 2010, Chen 2010, Itabashi et al. 2018). Superior phenotype in crossed seedlings than the self-fertilized seedlings in maize was first observed by Charles Darwin in 1876 (Darwin 1876). Shull introduced the heterosis to express the phenotypic superiority (Shull 1908, Shull 1948). Shull has indicated that selfing reduced the growth vigor and yield and the hybrids had uniformly superior growth vigor and yield compared to the inbreeding parents in maize (Shull 1908, Shull 1948). Since then, heterosis is significantly used in plant breeding to increase the production in crop plants (Duvick 2001, Schnable and Springer 2013). Considering the economic importance, heterosis is considered to be one of the breakthrough innovations in agricultural science but molecular mechanism of heterosis is still not completely understood. Understanding the molecular basis of heterosis will help to apply it more efficiently for crop breeding. The genetic basis of heterosis is still under debate. Traditional genetic models such as dominance, overdominance, and epistasis are used to explain the genetic mechanism of heterosis (Birchler et al. 2010, Chen 2010, Fujimoto et al. 2018). Dominance model is considered as the complementation or suppression of the inferior or deleterious alleles from one parent by the superior or dominant alleles from the other parent resulted the heterosis (Davenport 1908, Bruce 1910, Jones 1917, Crow 1998). Heterosis caused by the heterozygosity at a specific locus is known as overdominance (Shull 1908, East 1936, Crow 1998). In epistasis model, interactions between parental derivative non-allelic genes cause heterosis (Richey 1942, Powers 1944, Williams 1959). Heterosis is very complex in nature, and the traditional genetic models cannot explain heterosis completely. Therefore, advance genetic study such as quantitative trait loci (QTL), transcriptomes and proteomes analysis can also use for the molecular mechanisms of heterosis in plants (Fujimoto et al. 2018, Lv et al. 2020, Wu et al. 2021, Yu et al. 2021). However, it cannot be skipped influence of the epigenetic mechanisms for heterosis (Greaves et al. 2015, Ryder et al. 2019). A possible linkage of different epigenetic regulators such as DNA methylation, siRNAs, miRNA, and histone modifications has been noticed with the changes in gene expression of heterotic hybrid phenotypes in crop plants and model plants (Zhao et al. 2007, Ni et al. 2009, He et al. 2010, Law and Jacobson 2010, Nakamura

and Hosaka 2010, Qi et al. 2010, Groszmann et al. 2011, Ng et al. 2011, Shen et al., 2012, Shivaprasad et al. 2012, Chen et al. 2013, Dapp et al. 2015, Fu et al. 2015, Kawanabe et al. 2016a, Shen et al. 2017, Fujimoto et al. 2018, Lauss et al. 2018, Li et al. 2021, Yu et al. 2021). The plant growth and biomass heterosis can be affected by histone modifications alteration associated transcriptional activity (Yu et al. 2021). In the promoter regions of the circadian clock genes (*CCA1, CIRCADIAN CLOCK ASSOCIATED 1* and *LHY, LATE ELONGATED HYPOCOTYL*) of *A. thaliana*, lower levels of H3K9Ac and H3K4me2 enrichment in F_1 hybrid compared with the parental lines changed their transcriptions (Ni et al. 2009), this functional alteration may lead to the vigorous plant growth in F_1 hybrid through the regulation of different biological process (Miller et al. 2012, Shen et al. 2012, Chen 2013, Kim AJ et al. 2017). Global patterns of H3K4me3-, H3K9ac-, and H3K27me3-marks in two subspecies and their subsequent F_1 hybrid of rice by ChIP-Seq were identified where H3K4me3-marks or H3K27me3-marks inversely expressed between hybrids and parents (He et al. 2010). Studies suggests that histone modification might have a linkage to regulate vigourous growth in F_1 hybrid.

The whole-genome sequencing of *B. rapa* is available (Wang et al. 2011). A few researches studied on the epigenetic repressive histone marks (Takahashi et al. 2018, Akter et al. 2019, Payá-Milans et al. 2019), but there is not any report for the active histone marks in *B. rapa*. On the other hand, there is no single evidence for the comparative histone mark profiling study in a heterotic hybrid with the parental lines in *B. rapa*. Therefore, we characterized the active histone marks, H3K4me3 and H3K36me3, using two parental inbreed lines of Chinese cabbage. Thereafter, we compared the histone modifications states of these parental lines with their heterotic F_1 hybrids of Chinese cabbage histone modifications. We also examined association of the H3K4me3-, H3K27me3-, and H3K36me3-marks in the region covering lncRNAs.

Chapter II

Characterization of Histone H3 Lysine 4 and 36 Tri-methylation in *Brassica rapa* L.

Abstract

Covalent modifications of histone proteins act as epigenetic regulators of gene expression. We report the distribution of two active histone marks (H3K4me3 and H3K36me3) in 14-day leaves in two lines of *Brassica rapa* L. by chromatin immunoprecipitation sequencing. Both lines were enriched with H3K4me3 and H3K36me3 marks at the transcription start site, and the transcription level of a gene was associated with the level of H3K4me3 and H3K36me3. H3K4me3- and H3K36me3-marked genes showed low tissue-specific gene expression, and genes with both H3K4me3 and H3K36me3 had a high level of expression and were constitutively expressed. Bivalent active and repressive histone modifications such as H3K4me3 and H3K27me3 marks or antagonistic coexistence of H3K36me3 and H3K27me3 marks were observed in some genes. Expression may be susceptible to changes by abiotic and biotic stresses in genes having both H3K4me3 and H3K27me3 marks. We showed that the presence of H3K36me3 marks was associated with different gene expression levels or tissue specificity between paralogous paired genes, suggesting that H3K36me3 might be involved in subfunctionalization of the subgenomes.

Keywords: histone H3 lysine 4 tri-methylation, histone H3 lysine 36 tri-methylation, epigenetics, subfunctionalization, *brassica*

Introduction

Brassica rapa L. encompasses commercially important cultivars of vegetables, oilseeds, condiments, and fodder and is a crop closely related to *Arabidopsis thaliana* (Cheng et al. 2014, Cheng et al. 2016a, Lv et al. 2020). In addition to its agronomic significance, *B. rapa* is also important for genomic studies, because it has the first complete genome sequence to be determined within the genus *Brassica* (Wang et al. 2011). *B. rapa* (AA genome) is one of the ancestral species of the allotetraploid species, *Brassica nigra* L (AABB) and *Brassica napus* L (AACC) (Chalhoub et al. 2014, Yang J et al. 2016b). The *B. rapa* genome has undergone a whole-genome triplication after speciation between *A. thaliana* and *B. rapa* (Wang et al. 2011). The whole-genome triplication results in multiple copies of paralogous genes and generates three subgenomes, the least fractioned subgenome (LF) and two more fractionated subgenomes (MF1 and MF2) within the *B. rapa* genome (Wang et al. 2011). After the whole genome triplication, subfunctionalization such as different expression levels or DNA methylation levels among three subgenomes or paralogous genes has been observed (Parkin et al. 2014, Chen et al. 2015).

The basic unit of chromatin is the nucleosome, which consists of 147 bp of DNA wrapped around a histone octamer containing two of each of H2A, H2B, H3, and H4 (Li et al. 2007). Chromatin structure is regulated by posttranslational modification of histone proteins such as methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation (Strahl and Allis 2000, Jenuwein and Allis 2001, Li et al. 2007). Specific amino acid residues of the N-terminal tail of histone proteins are targets for posttranslational modifications that can impact gene expression by altering the chromatin structure. In plants, tri-methylation of histone H3 lysine 4 (H3K4me3) and H3K36me3 are often associated with transcriptional activation and that of H3K9me2 and H3K27me3 with transcriptional repression (Fuchs et al. 2006, Li et al. 2007, Xiao et al. 2016). Histone modification is an epigenetic regulatory mechanism, which affects transcriptional activity of chromatin without changing DNA sequence and is crucial for the development and the adaptation of plants to changing environments (Fujimoto et al. 2012a, Quadrana and Colot 2016).

The application of high-throughput sequencing technologies provides an opportunity to identify the genome-wide profiles of histone modification by a combination of chromatin immunoprecipitation (ChIP) and genome-wide sequencing (ChIP-seq). With ChIP-seq, the genome-wide distribution patterns of histone modifications such as H3K4me3, H3K9me2, H3K27me3, and H3K36me3 have been identified in some plant species. In *A. thaliana*, H3K4me3, H3K27me3, and H3K36me3 were found in euchromatin, while H3K9me2 was found

in heterochromatin (Turck et al. 2007, Zhang et al. 2007, Bernatavichute et al. 2008, Zhang et al. 2009, Roudier et al. 2011). In *A. thaliana*, H3K4me3 and H3K36me3 marked highly expressed genes, while H3K27me3 marked lowly expressed genes or genes showing tissue-specific expression (Turck et al. 2007, Zhang et al. 2007, Zhang et al. 2009, Roudier et al. 2011). In *B. rapa*, H3K9me2 was associated with transcriptional repression of genes and was overrepresented in transposable elements (TEs) (Takahashi et al. 2018).

Despite whole-genome epigenome information in *B. rapa* having been obtained for repressive histone marks such as H3K9me2 or H3K27me3 (Takahashi et al. 2018, Akter et al. 2019, Payá-Milans et al. 2019), there is no report of the whole epigenome information for active histone marks. In the present study, we examined the distribution of H3K4me3 and H3K36me3 using two inbred lines of Chinese cabbage, RJKB-T23 and RJKB-T24, and investigated the role of H3K4me3 and H3K36me3 in transcription, tissue-specific gene expression, subfunctionalization of paralogous genes, and species conservation and response to biotic and abiotic challenges.

Materials and Methods

Two Chinese cabbage inbred lines (*B. rapa* var. *pekinensis*), RJKB-T23 (T23) and RJKB-T24 (T24), were used (Kawamura et al. 2016). The genetic distances between the two lines (T23 and T24) and the reference genome were similar, suggesting that the genetic distance between the two lines was moderate among Chinese cabbage accessions (Kawamura et al. 2016, Shea et al. 2018). Seeds were surface-sterilized and grown on agar-solidified Murashige and Skoog (MS) plates with 1% (w/v) sucrose under long-day (LD) conditions (16 h light) at 22°C. First and second leaves were harvested at 14 days after sowing for ChIP analyses, and the plant phenotype at this stage in the two lines was shown in Akter et al. (2019).

ChIP-seq

ChIP experiments were performed as described by Buzas et al. (2011). One gram of first and second leaves of *B. rapa* was used for ChIP analysis, and anti-H3K4me3 (Millipore, 07-473) and H3K36me3 (Abcam, ab9050) antibodies were used. Before ChIP-seq, we validated the enrichment of purified immunoprecipitated DNAs by qPCR using the positive and negative control primer sets of H3K4me3 and H3K36me3 previously developed (Figure II-1) (Kawanabe et al. 2016b). Purified immunoprecipitated DNA and input DNA were sequenced by HiSeq™ 2000 (36 bp single end). These sequence data have been submitted to the DDBJ database (http://www.ddbj.nig.ac.jp) under accession number DRA003120. Low-quality reads or adapter sequences were purged from the ChIP-seq reads using Cutadapt version 1.7.1 and Trim Galore! version 0.3.7. The reads were mapped to the *B. rapa* reference genome v.1.5 (http://brassicadb.cn/) using Bowtie 2 version 2.2.3 (Table II-1). The mapped reads on the interspersed repeat regions (IRRs), such as the TEs detected by RepeatMasker, were examined (Table II-1). For sequential ChIP, experiments were performed as described by Finnegan et al. (2011). Anti- H3K4me3 antibodies were used for first ChIP, and anti- H3K27me3 antibodies (Millipore, 07-449) were used for second ChIP.

Sequential ChIP-qPCR for bivalent enrichment

ChIP-qPCR was performed using a LightCycler Nano (Roche). The immunoprecipitated DNA was amplified using FastStart Essential DNA Green Master (Roche). PCR conditions were 95°C for 10 min followed by 40 cycles of 95°C for 10 s, 60°C for 10 s, and 72°C for 15 s, and melting program (60°C to 95°C at 0.1°C/s). After amplification cycles, each reaction was subjected to melt temperature analysis to confirm single amplified products. Data presented are the average and standard error (s.e.) from three biological and experimental replications.

Enrichment of H3K4me3 or H3K36me3 marks was calculated by comparing the target gene and non-H3K4me3- or non-H3K36me3-marked genes, respectively, by qPCR using immunoprecipitated DNA as a template. Bra011336 that did not have H3K4me3 and H3K27me3 marks was used for reference for sequential ChIP-qPCR for examining enrichment of target genes. The difference between primer pairs was corrected by calculating the difference observed by qPCR amplifying the input DNA as a template. Primer sequences used in this study are shown in Table II-2.

Detection of H3K4me3 and H3K36me3 Peaks by Model-Based Analysis for ChIP-seq

We performed peak calling on alignment results using MACS 2 2.1.0 and identified the regions having H3K4me3 or H3K36me3 peaks. The MACS callpeak was used with the options (effective genome size: 2.30e + 08, band width: 200, model fold: 10–30, tag size: 36) described by Akter et al. (2019). A *p*-value cutoff of 1.00e-05 was used to consider peaks significant. H3K4me3- and H3K36me3-marked genes were defined as genes that had a more than 200-bp-length peak within a genic region (exon-intron) including 200 bp upstream and downstream as described by Akter et al. (2019).

Gene Ontology Analysis

Analysis for enrichment of gene functional ontology terms was completed using the gene ontology (GO) tool agriGO (Du et al. 2010) following the methods described by Shimizu et al. (2014). Statistical tests for enrichment of functional terms used the hypergeometric test and false discovery rate (FDR) correction for multiple testing to a level of 1% FDR.

Results

Identification of H3K4me3- and H3K36me3-marked genes in *B. rapa*

The impact of the active histone marks, H3K4me3 and H3K36me3, on gene expression in *B. rapa*; their interaction with other epigenetic marks; and the diversity and conservation of H3K4me3 and H3K36me3 distribution between *B. rapa* and *A. thaliana* were examined. The presence of H3K4me3 or H3K36me3 marks on the chromatin of 14-day leaves in two inbred lines of Chinese cabbage (RJKB-T23 (T23) and RJKB-T24 (T24)) was mapped by ChIP-seq (Table II-1). Reads mapped in the genic regions were classified into 2 kb upstream, exon, intron, and 2 kb downstream segments. The proportion of reads in exons was slightly higher than in the input DNA (Figure II-2 and Table II-3), suggesting that there is preferential location of H3K4me3 and H3K36me3 in exon regions. H3K4me3 and H3K36me3 were enriched in the transcribed regions in both lines, especially around the transcription start sites (TSS) (Figure II-3 and Figure II-4). The percentage of H3K4me3 and H3K36me3 reads in the IRRs (TEs and repeats) was lower than in the input DNA in both lines (Table II-1). There was no H3K4me3 and H3K36me3 enrichment in IRR sequences or their flanking regions (Figure II-4).

We defined an H3K4me3- or H3K36me3-marked gene as having a peak of more than 200 bp within the genic region, which includes the 200 bp upstream and downstream sequences (see section "Materials and Methods"). In T23 and T24, 18,475 (46.6%) and 19,208 genes (48.5%) had H3K4me3 marks, respectively, and 16,759 genes were common to the two lines and were termed H3K4me3-marked genes (Table II-4). In T23 and T24, 13,395 (33.8%) and 13,771 genes (34.8%) had H3K36me3 marks, respectively, and 11,844 genes were common to the two lines and were termed H3K36me3-marked genes (Table II-4).

Previously obtained SNP data for T23 and T24 were used (Shea et al. 2018, Akter et al. 2019). Genes having H3K4me3 or H3K36me3 marks tended to have a higher SNP number per length in each gene (mutation rate) than genes without H3K4me3 or H3K36me3 marks, respectively, in T23 and T24 (Figure II-5). However, there was no difference between genes having H3K4me3 or H3K36me3 marks and the total genes (Figure II-5).

Identification of genes carrying bivalent or antagonistic active and repressive histone modification

We counted the overlapped genes among three histone marks (H3K4me3 and H3K36me3 in this study and H3K27me3 in Akter et al. (2019). Of the H3K36me3-marked genes, 85.4% (n = 10,119) also had H3K4me3 marks (Table II-4 and Table II-5). The enrichment of

H3K4me3 or H3K36me3 marks was similar between genes having one or two modifications (Figure II-6). These results suggest that these two modifications were in the same region of each gene. Three functional *BrFLC* paralogs (*BrFLC1*, *BrFLC2*, and *BrFLC3*) had H3K4me3 and H3K36me3 marks (Figure II-7).

Of the H3K27me3-marked genes, 6.4% (n = 671) had H3K36me3 (Table II-4); there was antagonistic coexistence of active and repressive histone modifications between H3K36me3 and H3K27me3 marks. Two of the three *BrSOC1* (Bra004928 and Bra039324) were induced by 4 weeks of cold treatment (Shea et al. 2019). All three *BrSOC1* paralogs had the H3K27me3 mark but not the H3K4me3 or H3K36me3 mark (Figure II-7). Previously, we identified genes showing a difference in H3K27me3 levels between 2-day cotyledons and 14-day leaves in both T23 and T24 (Akter et al., 2019). In 903 genes showing a higher H3K27me3 level in 2-day cotyledons than in 14-day leaves, 549 genes (60.8%) had H3K36me3 marks in 14-day leaves (Data is not shown here, Data is available in Supplementary Table S5 of Mehraj et al. 2021). In 395 genes showing a higher H3K27me3 level in 14-day leaves than in 2-day cotyledons, only 10 genes (2.5%) had H3K36me3 marks in 14-day leaves (Data is not shown here, Data is available in Supplementary Table S5 of Mehraj et al. 2021). In 21 genes categorized into "post-embryonic development" with higher H3K27me3 levels in 14-day leaves than in 2-day cotyledons, including *BrFUS3*, *BrDOC1*, and *LEA* genes, none of the 21 genes had H3K36me3 marks in 14-day leaves, suggesting that increasing H3K27me3 may be antagonistic to H3K36me3 accumulation.

Of the H3K27me3-marked genes, 35.4% (n = 3,699) had H3K4me3; there were bivalent active and repressive histone modifications between H3K4me3 and H3K27me3 marks (Table II-4). There are two *BrVIN3* paralogs. *BrVIN3a* (Bra020445) was induced by 4 weeks of cold treatment, while *BrVIN3b* (Bra006824) was not (Shea et al. 2018). *BrVIN3a* had both H3K4me3 and H3K27me3 marks, while *BrVIN3b* had neither H3K4me3 nor H3K27me3 marks (Figure II-7). In genes having both H3K4me3 and H3K27me3 marks, accumulation of the H3K4me3 mark was flat in the genic region, and the H3K4me3 level around the TSS was lower than that in H3K4me3-marked genes (Figure II-6 and Figure II-8). The H3K27me3 level in genes having both H3K4me3 and H3K27me3 marks was lower than that in H3K27me3-marked genes, especially in the middle-to-3' part of the genic region (Figure II-6 and Figure II-8). The level of H3K36me3 mark in the genes having both H3K36me3 and H3K27me3 marks was also lower than that in H3K36me3-marked genes, while accumulation of the H3K27me3 mark in the genes having both H3K36me3 and H3K27me3 marks was similar to that in H3K27me3-marked genes (Figure II-6 and Figure II-9). More than 30 and 80% of genes having both H3K4me3 and

H3K27me3 marks had overlapped peaks with more than 500 and 150 bp in length, respectively (Table II-5), suggesting that different histone modifications tended to detect the same position on the gene. In case of genes having both H3K36me3 and H3K27me3 marks, more than 15 and 55% of genes had overlapped peaks with more than 500 and 150 bp in length (Table II-5). In genes having both H3K4me3 and H3K27me3 marks, the categories related to transcriptional regulation were highly overrepresented including transcription factors such as *LFY*, *WRKY*, *ERF*, *IAA*, and *HFSB* (Figure II-10, Figure II-11 and Table II-6). To examine the simultaneous occupancy of active (H3K4me3) and repressive (H3K27me3) histone modifications, sequential ChIP-qPCR in 14-day leaves in T24 was performed in seven genes having both H3K4me3 and H3K27me3 marks detected by ChIP-seq analysis. Six of the seven genes showed enrichment for the second modification compared with genes having only the H3K4me3 or H3K27me3 mark (Figure II-12). *BrVIN3* showed lower enrichment than the other six genes having both H3K4me3 and H3K27me3 marks (Figure II-12).

H3K36me3 marks were associated with a transcriptionally active state and constitutive expression

The average transcription level of H3K4me3- or H3K36me3- marked genes was higher than that in the total genes, and the average transcription level of H3K4me3-marked genes was lower than that in H3K36me3-marked genes (Figure II-13). The average transcription level of genes having both H3K4me3 and H3K36me3 marks showed higher expression levels than the total genes (Figure II-13). The average transcription level of genes having both H3K4me3 and H3K27me3 marks or having both H3K36me3 and H3K27me3 marks showed a lower expression level than H3K4me3- or H3K36me3-marked genes, respectively, and the decreased average expression level of genes also having H3K27me3 was larger in H3K4me3- marked genes than in H3K36me3-marked genes; genes with bivalent active and repressive histone modifications, H3K4me3 and H3K27me3 marks, showed similar expression levels to those with the repressive histone modification, H3K27me3 (Figure II-13).

Previously, we calculated a tissue specificity index, *T*-value, which interpolates the entire range between 0 for housekeeping genes and 1 for strictly one-tissue-specific genes, using the transcriptome data from six different tissues in *B. rapa* (Tong et al. 2013; Akter et al. 2019). We found that H3K36me3- marked genes and genes having both H3K4me3 and H3K36me3 marks showed significantly lower average *T*-values compared with the total genes (Figure II-14), suggesting that H3K36me3 has a role in constitutive gene expression. H3K4me3-marked genes had lower average *T*-values compared with total genes, but the *T*-value was not as low as

in H3K36me3-marked genes. H3K4me3 or H3K36me3 genes also marked with H3K27me3 showed increased average *T*-values, and this effect was greater in H3K4me3-marked genes than in H3K36me3-marked genes (Figure II-14); bivalent active and repressive histone modifications, H3K4me3 and H3K27me3 marks, showed a higher tissue specificity similar to the H3K27me3-marked genes.

Bivalent active and repressive histone modifications, H3K4me3 and H3K27me3, increased transcriptional sensitivity in stress response

In this study, 42.3, 29.9, and 26.4% of annotated genes were defined as H3K4me3-, H3K36me3-, and H3K27me3-marked genes. Previously, we identified genes in T24 that were differentially expressed following *Fusarium oxysporum* f. sp. *conglutinans* (*Foc*) inoculation or 4 weeks of cold treatment (vernalization) compared with non-treated samples (Miyaji et al. 2017, Shea et al. 2019). We examined whether the genes with H3K4me3, H3K27me3, and H3K36me3 marks showed changed expression caused by these two stress treatments. Of 253 differentially expressed genes following *Foc* inoculation, 131 (51.8%), 45 (17.8%), and 139 (54.9%) genes had H3K4me3, H3K36me3, and H3K27me3 marks, respectively (Figure II-15). Of 1,441 differentially expressed genes resulting from 4 weeks of cold treatment, 729 (50.6%), 356 (24.7%), and 538 (37.7%) genes had H3K4me3, H3K36me3, and H3K27me3 marks, respectively (Figure II-15). These results showed that genes with altered expression in both stress treatments had a significantly lower percentage of H3K36me3- marked (chi-squared test, $p < 10{-}5$) and a higher percentage of H3K27me3-marked genes (chi-squared test, $p < 10{-}10$) than in total genes, suggesting that H3K36me3-marked genes tended to be transcriptionally stable and H3K27me3-marked genes tended to be variable in response to these two stress treatments.

Of the annotated genes, 9.3% had both H3K4me3 and H3K27me3 marks, and 1.7% had both H3K36me3 and H3K27me3 marks. Of 253 genes differentially expressed following *Foc* inoculation, 58 (22.9%) genes had both H3K4me3 and H3K27me3 marks, and five (2.0%) genes had both H3K36me3 and H3K27me3 marks (Figure II-15). Of 1,441 genes differentially expressed following 4 weeks of cold treatment, 295 (20.5%) genes had both H3K4me3 and H3K27me3 marks, and 38 (2.6%) genes had both H3K36me3 and H3K27me3 marks (Figure II-15). These results suggest that genes having both H3K4me3 and H3K27me3 marks tended to have changed levels of transcription associated with the two stress treatments, but genes having both H3K36me3 and H3K27me3 marks did not have changed transcription levels.

Comparison of H3K4me3 and H3K36me3 states between paralogous genes in *B. rapa*

We compared H3K4me3 or H3K36me3 locations between paralogs. Among the 1,675 three-copy sets, 392 had H3K4me3 in all three copies, 317 had H3K4me3 in at least two copies, 364 had H3K4me3 in at least one copy, and 602 sets did not have H3K4me3 in any copies (Figure II-16). In the case of H3K36me3, 235 had H3K36me3 in all three copies, 242 had H3K36me3 in at least two copies, 333 had H3K36me3 in at least one copy, and 865 sets did not have H3K27me3 in any copies (Figure II-16).

We examined whether a difference in H3K4me3 or H3K36me3 states between paralogs was associated with a different level of gene activity. Between paralogous pairs, there was no significant difference of the average expression levels between genes with and without H3K4me3 marks in either T23 or T24 (Figure II-17 and Figure II-18). Between paralogous pairs, the average expression levels of genes with H3K36me3 marks tended to be higher than those without H3K36me3 in both lines, and some comparisons showed significantly different expression levels between paralogous pairs with and without H3K36me3 marks (Figure II-19 and Figure II-20), indicating that the presence of H3K36me3 is associated with a difference of gene expression level between paralogous pairs. We also examined whether a difference in H3K4me3 or H3K36me3 states between paralogs was associated with a different level of *T*-value. *T*-values between paralogs with and without H3K4me3 marks tended to be the same (Figure II-21), while the average *T*-values of genes with H3K36me3 tended to be lower than those without H3K36me3 marks when paralogous pairs were compared (Figure II-22), suggesting an association of H3K36me3 with constitutive gene expression.

Relationship among epigenetic marks

We compared H3K4me3 or H3K36me3 distribution to the other epigenetic marks, H3K9me2, H3K27me3, and DNA methylation (Takahashi et al. 2018, Akter et al. 2019). At the whole-genome level, there was high positive correlation between H3K4me3 and H3K36me3, while there was a negative correlation between H3K4me3 or H3K36me3 and DNA methylation (Figure II-23). In genic regions or IRRs, there was a positive correlation between H3K4me3 and H3K36me3, but the relationship between H3K4me3 or H3K36me3 and DNA methylation was not clear. In IRRs, there was a positive relationship among four histone modifications (r = 0.40–0.79), but there was no relationship between the presence of the four histone modifications and DNA methylation (r = −0.04–0.15; Figure II-23).

The distribution of H3K4me3 and H3K36me3 was similar, with a partial overlap with H3K27me3, but it is antagonistic to the distribution of H3K9me2 and DNA methylation (Figure II-24 and Figure II-25).

The average level of DNA methylation in the region overlapping H3K4me3 or H3K36me3 regions was lower than in the total genome (Figure II-26), indicating that H3K4me3 or H3K36me3 regions were preferentially not DNA methylated.

Half of the genes marked with H3K4me3 or H3K36me3 are shared between *B. rapa* and *A. thaliana*

To gain information about conservation of H3K4me3 and H3K36me3 states beyond *B. rapa*, we compared the genes marked with H3K4me3 and H3K36me3 in *B. rapa* and *A. thaliana*. About 50% of H3K4me3- or H3K36me3-marked genes in *A. thaliana* overlapped with at least one orthologous gene in *B. rapa* (Table II-6). This percentage is higher than that in H3K27me3-marked genes (Table II-6). These overlapped genes obtained using a data set by Engelhorn et al. (2017) were termed species-conserved H3K4me3- or H3K36me3-marked genes.

Discussion

We examined the genomic locations of H3K4me3 and H3K36me3 using two *B. rapa* lines. Since the results in the two lines were quite similar, this study focused on the genes having H3K4me3 or H3K36me3 in both lines as biological replicates, which could suggest a more fundamental epigenetic state in *B. rapa*. About 47 and 34% of total genes had H3K4me3 and H3K36me3 marks, respectively, and these percentages are similar to the case in *A. thaliana* (H3K4me3, 50–70%; H3K36me3, −50%) (Luo et al. 2013, Sequeira-Mendes et al. 2014, Engelhorn et al. 2017), suggesting that our analysis does not overestimate the number of genes having these histone marks. Both H3K4me3 and H3K36me3 marks were enriched close to the TSS, with a sharper peak in H3K4me3 than in H3K36me3. These distribution patterns of H3K4me3 and H3K36me3 marks in the genic region are similar to those in *A. thaliana* and rice (Roudier et al. 2011, Liu et al. 2019). Low levels of H3K4me3 and H3K36me3 marks were observed in IRRs, especially in the body region.

H3K4me3 and H3K36me3 are known as active histone marks, and H3K4me3- and H3K36me3-marked genes showed higher expression levels on average than the total gene expression levels. With RNA-seq data from six tissues, a tissue-specific index, *T*-value, has been developed, and *T*-values indicate a range between 0 for housekeeping or constitutive genes and 1 for genes showing tissue-specific expression (Tong et al. 2013). The average *T*-value of H3K4me3- or H3K36me3-marked genes was lower than that of total genes, suggesting that H3K4me3- and H3K36me3-marked genes had a low level of tissue specificity, being constitutively expressed. Of the two active histone modifications, the H3K36me3 mark was more associated with higher expression levels and lower tissue specificity. A similar trend was observed in response to *Foc* inoculation and 4 weeks of cold treatment; H3K36me3-marked genes are more transcriptionally stable than H3K4me3-marked genes in response to these two stress treatments.

We compared H3K4me3 and H3K36me3 to other epigenetic modifications. The H3K4me3 and H3K36me3 marks showed a negative correlation with DNA methylation and H3K9me2 at the whole genome level. DNA methylation levels in the regions having H3K4me3 or H3K36me3 were low. These results indicate that H3K4me3 and H3K36me3 do not physically coexist with the transcriptional repression mark of heterochromatin. A correlation between H3K4me3 or H3K36me3 and DNA methylation or H3K9me2 has also been identified in other plant species (Zhang et al. 2009, He et al. 2010, Roudier et al. 2011). Only 1.7% of the total genes (671 genes) had both H3K36me3 and H3K27me3 marks, suggesting that H3K36me3

marks were not compatible with H3K27me3 marks. Antagonistic roles for H3K36me3 and H3K27me3 were observed in the transcriptional regulation of *FLC* during vernalization (Yang et al. 2014), and in this study, antagonistic modification of H3K4me3/H3K36me3 and H3K27me3 in pre-vernalized plants was observed in three *BrFLC* paralogs. A higher level of H3K4me3 marks and lower level of H3K27me3 marks at *ABA INSENSITIVE 3* (*ABI3*), *LEAFY COTYLEDON 1/2* (*LEC1/2*), and *FUSCA3* (*FUS3*) levels increased their expression to regulate somatic embryo development in the *set domain group 8* (*sdg8*) and *embryonic flower 2* (*emf2*) double mutant in *A. thaliana* seedlings (Tang et al. 2012). In *A. thaliana*, changes of the active histone H3K4me3 mark and the repressive histone H3K27me3 mark have been observed in *DELAY OFGERMINATION1* (*DOG1*) during seed dormancy cycling; in dormant seeds, the H3K4me3 mark accumulated in *DOG1*, while the H3K4me3 mark was decreased and the H3K27me3 mark was increased during loss of dormancy and the germination process (Footitt et al. 2015, Cheng et al. 2020). In 21 genes that showed higher H3K27me3 levels in 14-day leaves than in 2-day cotyledons and categorized into "post-embryonic development" including *BrDOG1* and *BrFUS3* (Akter et al. 2019), no genes had H3K36me3 marks, suggesting that antagonistic H3K36me3 and H3K27me3 modifications might be important for the regulation of embryogenic expression of these genes.

The distribution of H3K4me3 or H3K36me3 partially overlapped with H3K27me3, and we identified genes with bivalent histone modification. The number of genes having both H3K4me3 and H3K36me3 marks was twice the expected number, suggesting that H3K4me3 and H3K36me3 preferentially co-localized. This is consistent with previous reports in *A. thaliana* (Luo et al. 2013, Sequeira-Mendes et al. 2014, Xi et al. 2020). Coexistence of the H3K4me3 and H3K36me3 marks at *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*) chromatin in *A. thaliana* activates its expression (Berr et al. 2015), suggesting that coexistence of active marks may have an important role in transcriptional regulation during plant development. The numbers of genes having H3K4me3 and H3K27me3 marks or H3K36me3 and H3K27me3 marks were less than 20% or 80% of the expected number, respectively. However, about 9% of the total genes (3,699 genes) had both H3K4me3 and H3K27me3 marks in *B. rapa*. In these genes, a lower accumulation of H3K4me3 and H3K27me3 marks than in genes having either single modification was observed. The genes having H3K36me3 and H3K27me3 marks showed similar characteristics to the H3K36me3-marked genes such as having a high level of expression and a constitutive expression pattern. In contrast, the genes having H3K4me3 and H3K27me3 marks showed similar characteristics to H3K27me3-marked genes such as the low level of gene expression, highly tissue-specific gene expression, and overrepresentation in the category of "regulation of transcription." Coexistence of H3K4me3

and H3K27me3 in the same chromosome fiber has been found in *A. thaliana*, and this was confirmed by sequential ChIP analysis (Sequeira-Mendes et al. 2014). In this study, we showed that some transcription factor genes have a simultaneous presence of H3K4me3 and H3K27me3 marks using sequential ChIP-qPCR. In *A. thaliana*, *VIN3* has a simultaneous presence of H3K4me3 and H3K27me3 marks (Finnegan et al. 2011). ChIP-seq analysis showed that *BrVIN3a* has both H3K4me3 and H3K27me3 marks, but sequential ChIP-qPCR did not show the simultaneous presence of these modifications. There were differences between *BrVIN3a* and *BrVIN3b* in the transcriptional response following vernalization and the histone modification in pre-vernalized states. By examining changes in these histone modifications in response to vernalization, the differing role of *VIN3* between species or between paralogs will be clarified. It has been suggested that the presence of both H3K4me3 and H3K27me3 marks may be alternative states of transcription in different tissues and/or stages in cell differentiation in *A. thaliana* (Sequeira-Mendes et al. 2014). In this study, genes having both H3K4me3 and H3K27me3 marks showed high tissue-specific gene expression and high sensitivity of transcription in response to biotic (*Foc* inoculation) and abiotic (4 weeks of cold treatment) stresses, suggesting that bivalent active and repressive histone modifications, H3K4me3 and H3K27me3, play a role in rapid response of transcription, not only through development but also following stress exposure.

B. rapa has experienced whole-genome triplication and has three subgenomes (LF, MF1, and MF2). There is a biased gene expression, distribution of TEs, and DNA methylation level among the three subgenomes in *B. rapa* (Cheng et al. 2012, Chen et al. 2015, Cheng et al. 2016b). Genes covering LF tended to have a higher expression level than their paralogous genes covering MF1 or MF2, and DNA methylation level in LF was lower than that in MF1 or MF2 (Chen et al. 2015). The difference of expression levels between paralogous paired genes was associated with differences of H3K36me3 levels but not of H3K4me3. Previously, we have shown an association between gene expression levels and H3K27me3 levels between paralogous paired genes (Akter et al. 2019). The variation of the tissue specificity (*T*-values) between paralogous paired genes was associated with a difference of H3K36me3 or H3K27me3 levels (Akter et al. 2019). We also showed that the association between histone modifications and lncRNA expression can be different from that of mRNA. It is difficult to determine whether the relationship between histone modifications and transcription level is a cause or a consequence, but these results suggest that both active and repressive histone modifications, H3K36me3 and H3K27me3, play a role in the variation of gene expression between paralogous paired genes, which may be involved in subfunctionalization. Further analysis will be needed to verify this possibility.

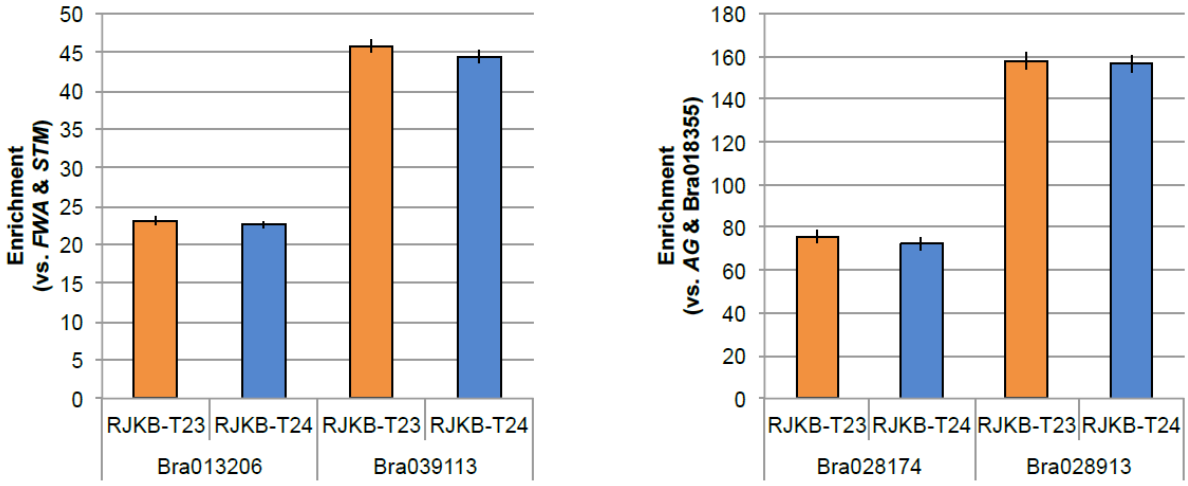


Figure II-1. Validation of the enrichment of purified immunoprecipitated (IP) DNAs by qPCR. Bra013206 and Bra039113 are H3K4me3-marked genes (positive control), and *FWA* and *STM* are low level H3K4me3-marked genes (negative control) (Kawanabe et al. 2016). Bra028174 and Bra028913 are H3K36me3-marked genes (positive control), and *AG* and Bra018355 are low level H3K36me3-marked genes (negative control) (Kawanabe et al. 2016). Values are means ± standard error (s.e.) (three biological and technical replicates) of relative H3K4me3 or H3K36me3 levels.

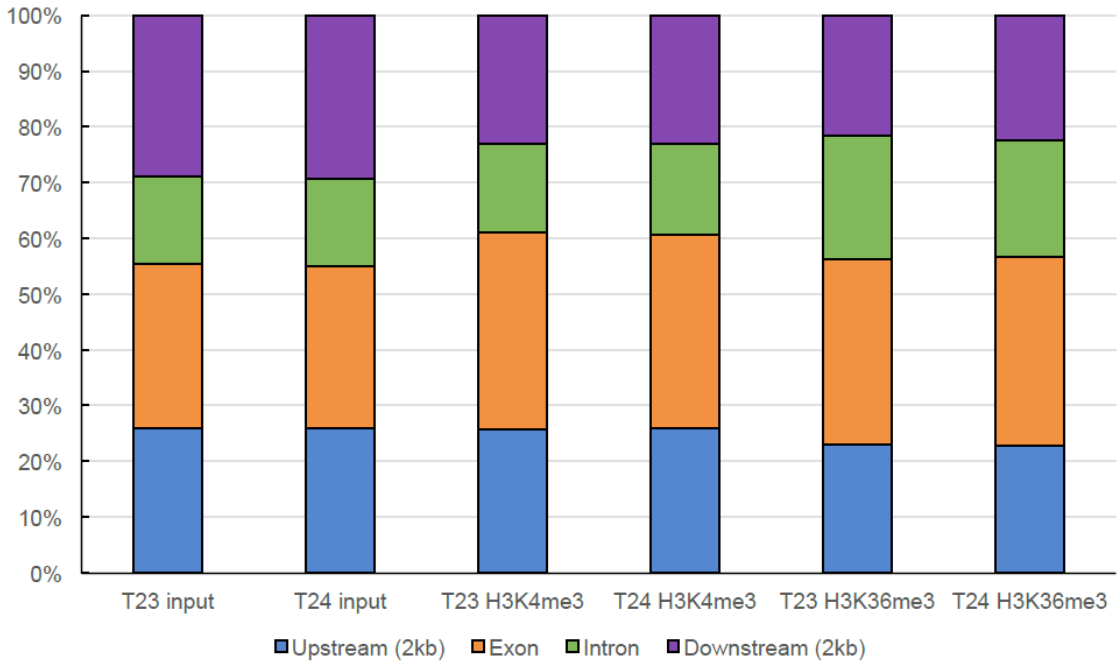


Figure II-2. Mapped reads of input and immunoprecipitated DNA using anti-H3K4me3 or anti-H3K36me3 antibodies in genic regions classified into 2kb-upstream, exon, intron, and 2kb downstream positions.

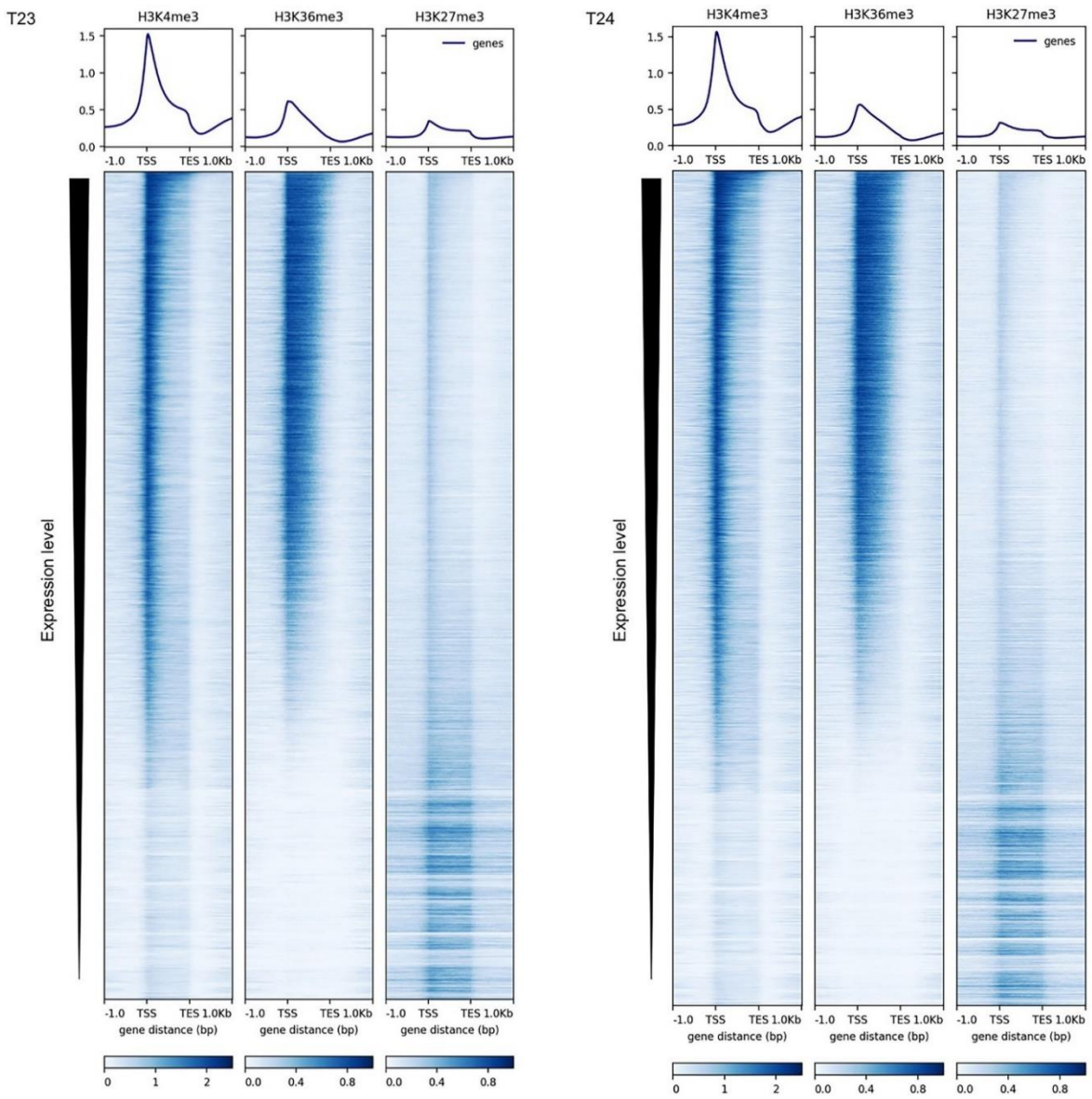


Figure II-3. Distribution of H3K4me3, H3K36me3, and H3K27me3 marks in the genic regions in RJKB-T23 (T23) and RJKB-T24 (T24). The deepTools version 3.1.3 is used for visualization (https://deeptools.readthedocs.io/en/develop/).

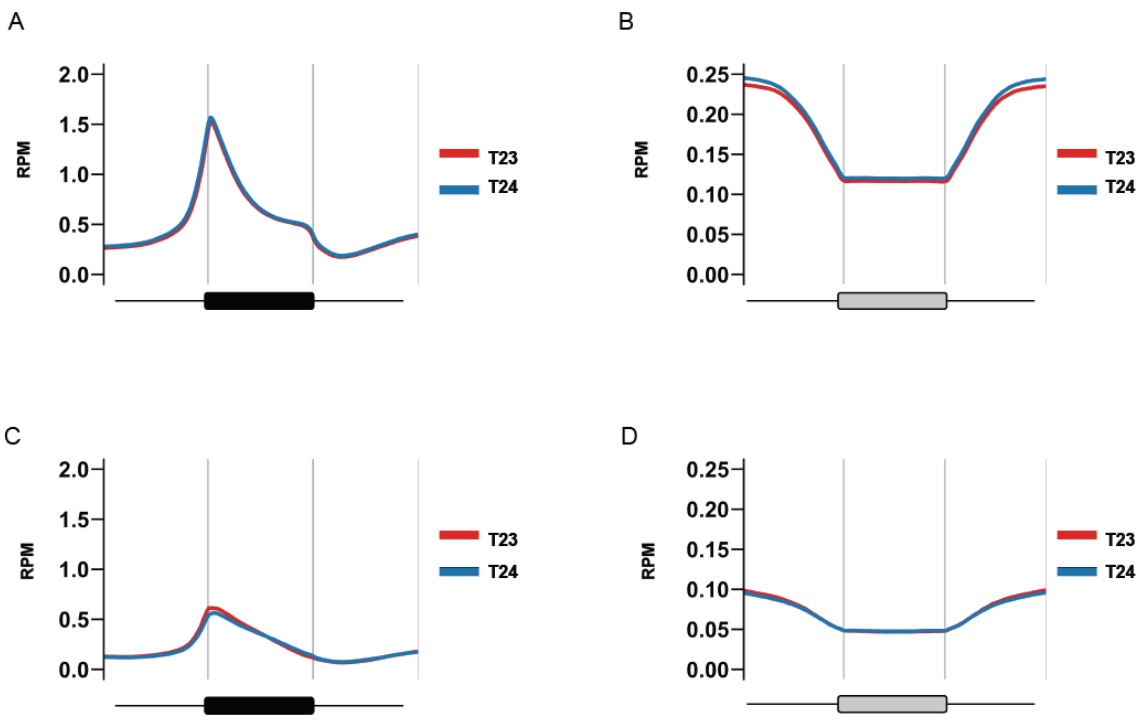


Figure II-4. Mapped reads of immunoprecipitated DNA using anti-H3K4me3 (A, C) or anti-H3K36me3 (B, D) antibodies in genic regions (black boxes) and IRR regions (gray boxes).

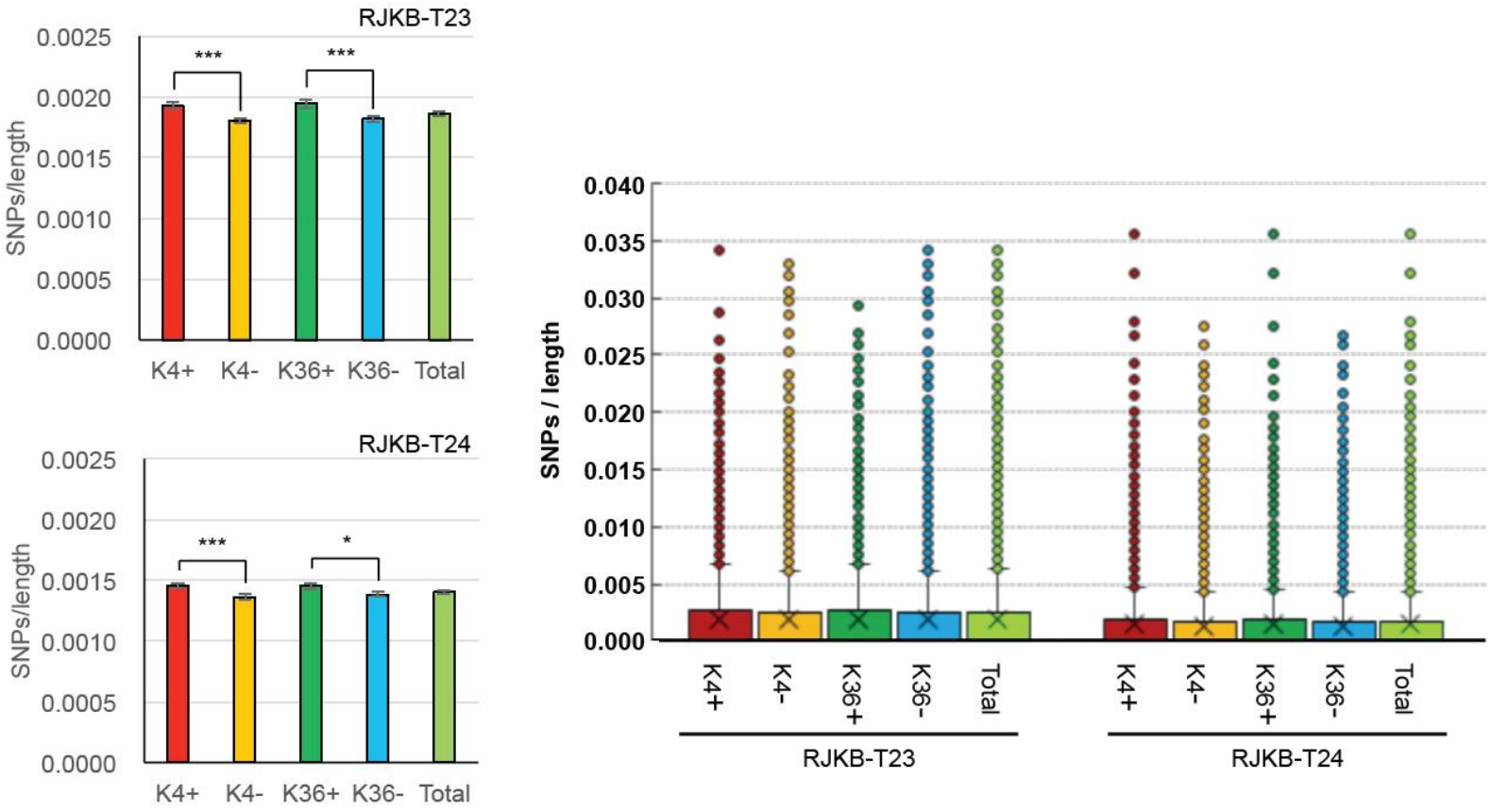


Figure II-5. Comparison of the average mutation rate (SNPs per length) in each gene having an H3K4me3 or H3K36me3 mark in RJKB-T23 and RJKB-T24. Values are means ± standard error (s.e.) of SNP numbers per length. *, $p < 0.05$, ***, $p < 0.001$ (Student *t*-test). '+' and '-' represent the presence and absence of H3K4me3/H3K36me3 marks in genes, respectively.

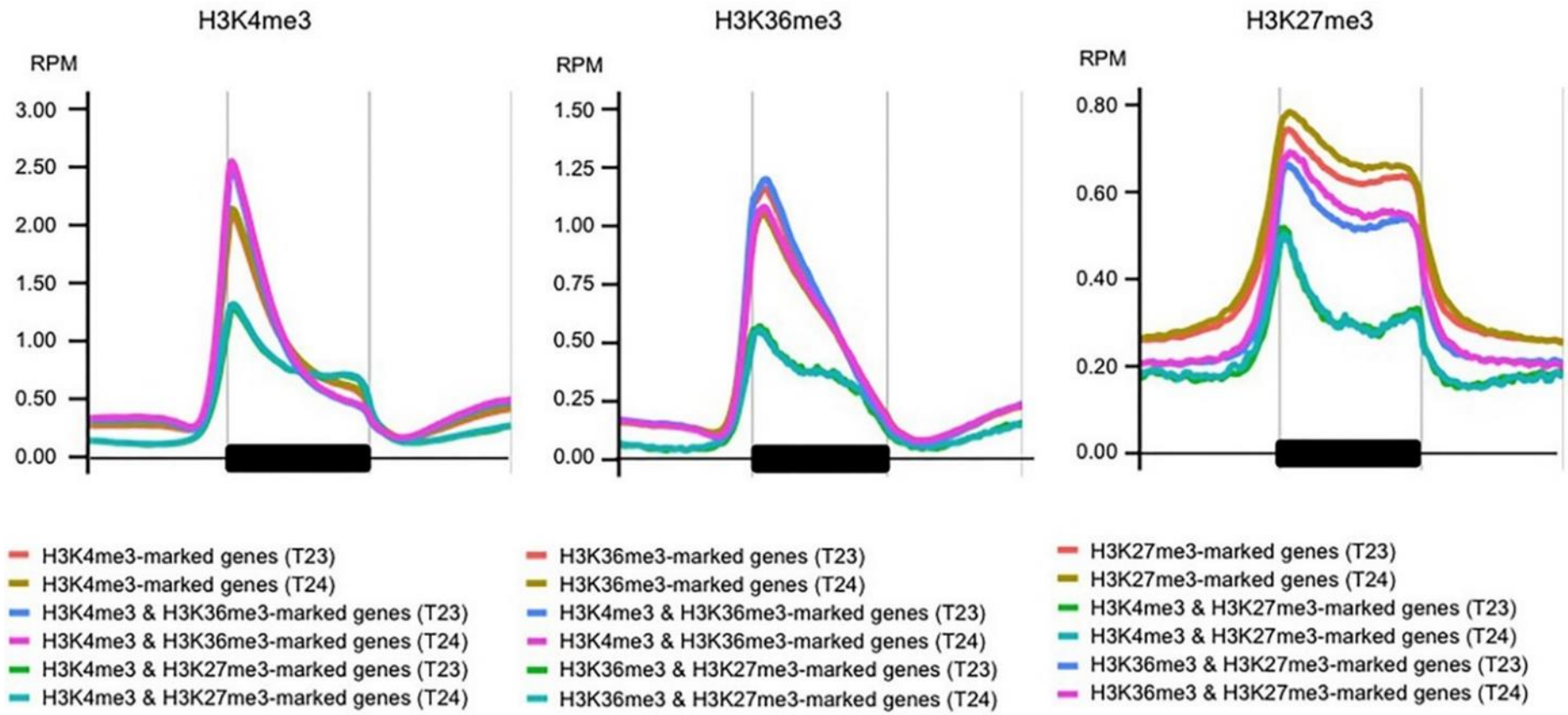


Figure II-6. Metagene plots of H3K4me3, H3K36me3, and H3K27me3 in genic regions. H3K4me3 (left panel), H3K36me3 (middle panel), and H3K27me3 (right panel) levels at the genic region with 1 kb upstream and 1 kb downstream are shown using genes having one or two histone modifications.

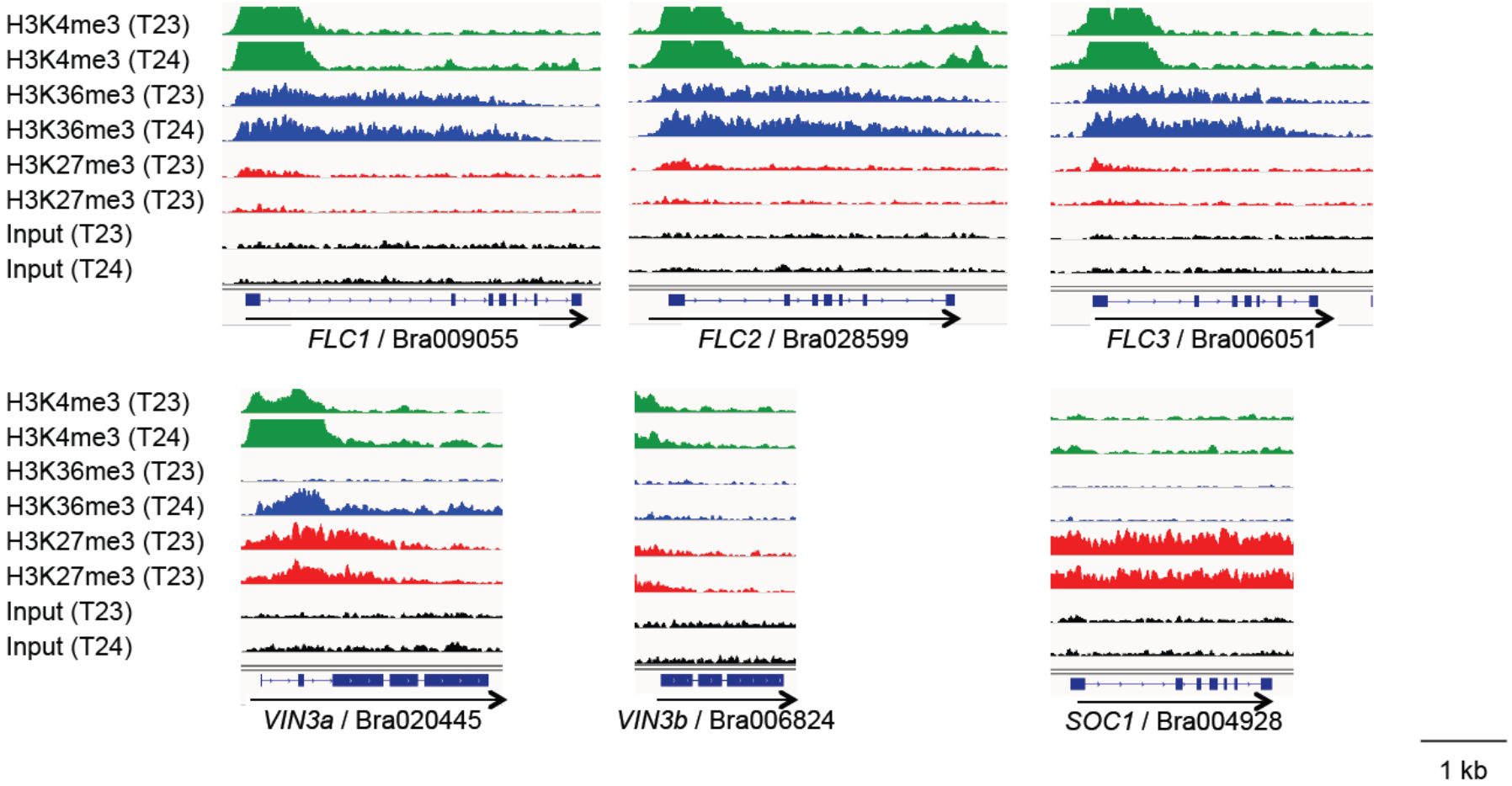


Figure II-7. Visualization of H3K4me3, H3K36me3, and H3K27me3 peaks by Integrative Genomics Viewer (IGV) by ChIP-seq.

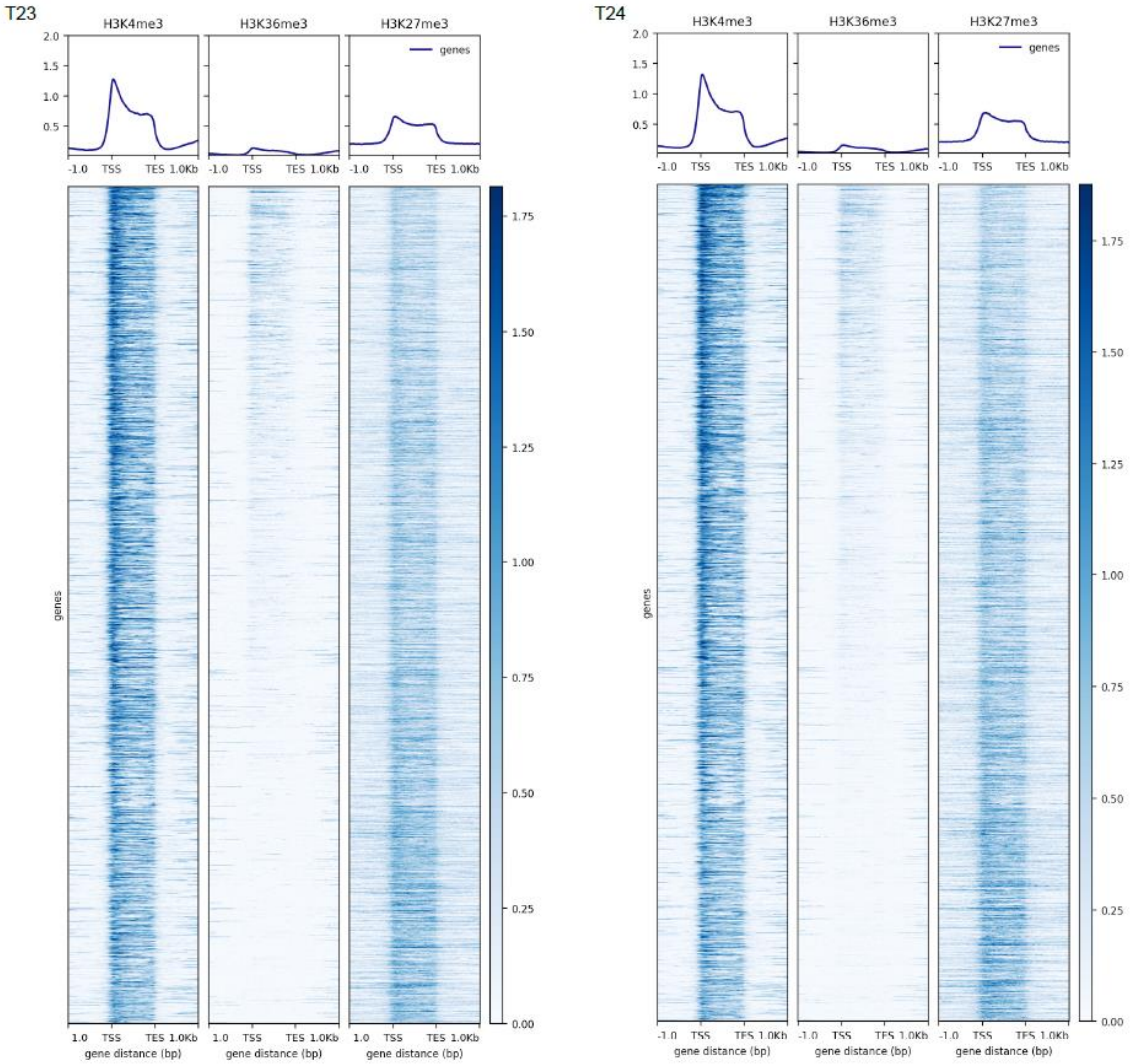


Figure II-8. Distribution of H3K4me3, H3K36me3, and H3K27me3 marks in the genic regions in RJKB-T23 (T23) and RJKB-T24 (T24). The genes having both H3K4me3 and H3K27me3 marks were used. The deeptools version 3.1.3 is used for visualization (https://deeptools.readthedocs.io/en/develop/).

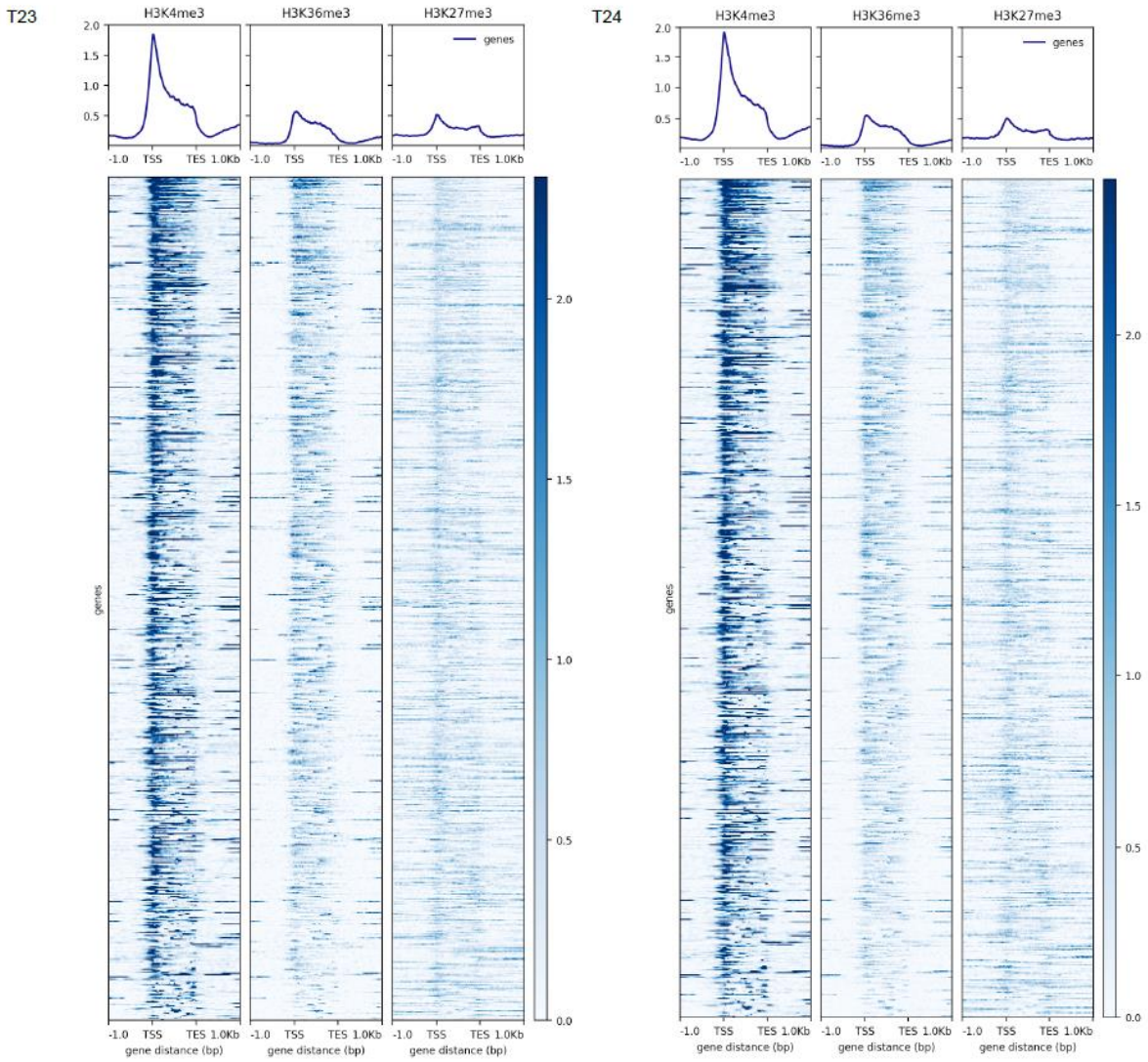


Figure II-9. Distribution of H3K4me3, H3K36me3, and H3K27me3 marks in the genic regions in RJKB-T23 (T23) and RJKB-T24 (T24). The genes having both H3K36me3 and H3K27me3 marks were used. The deeptools version 3.1.3 is used for visualization (https://deeptools.readthedocs.io/en/develop/).

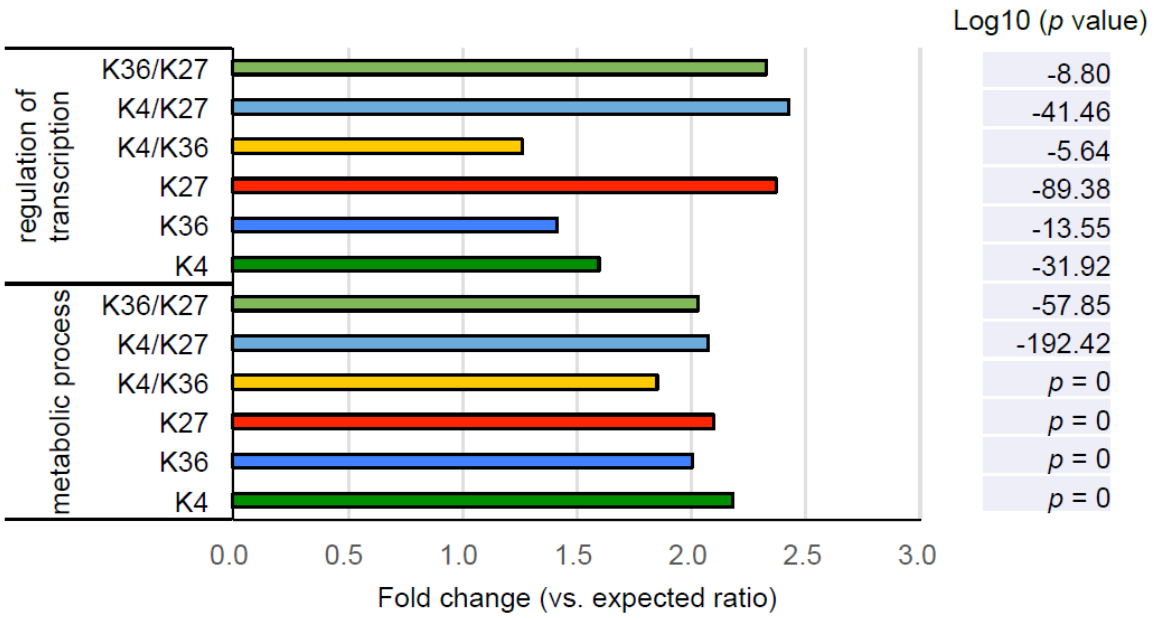


Figure II-10. GO classification in the category of 'regulation of transcription' and 'metabolic process'. The ratio was calculated by the percentage of genes having histone marks dividing the percentage of all annotated genes.

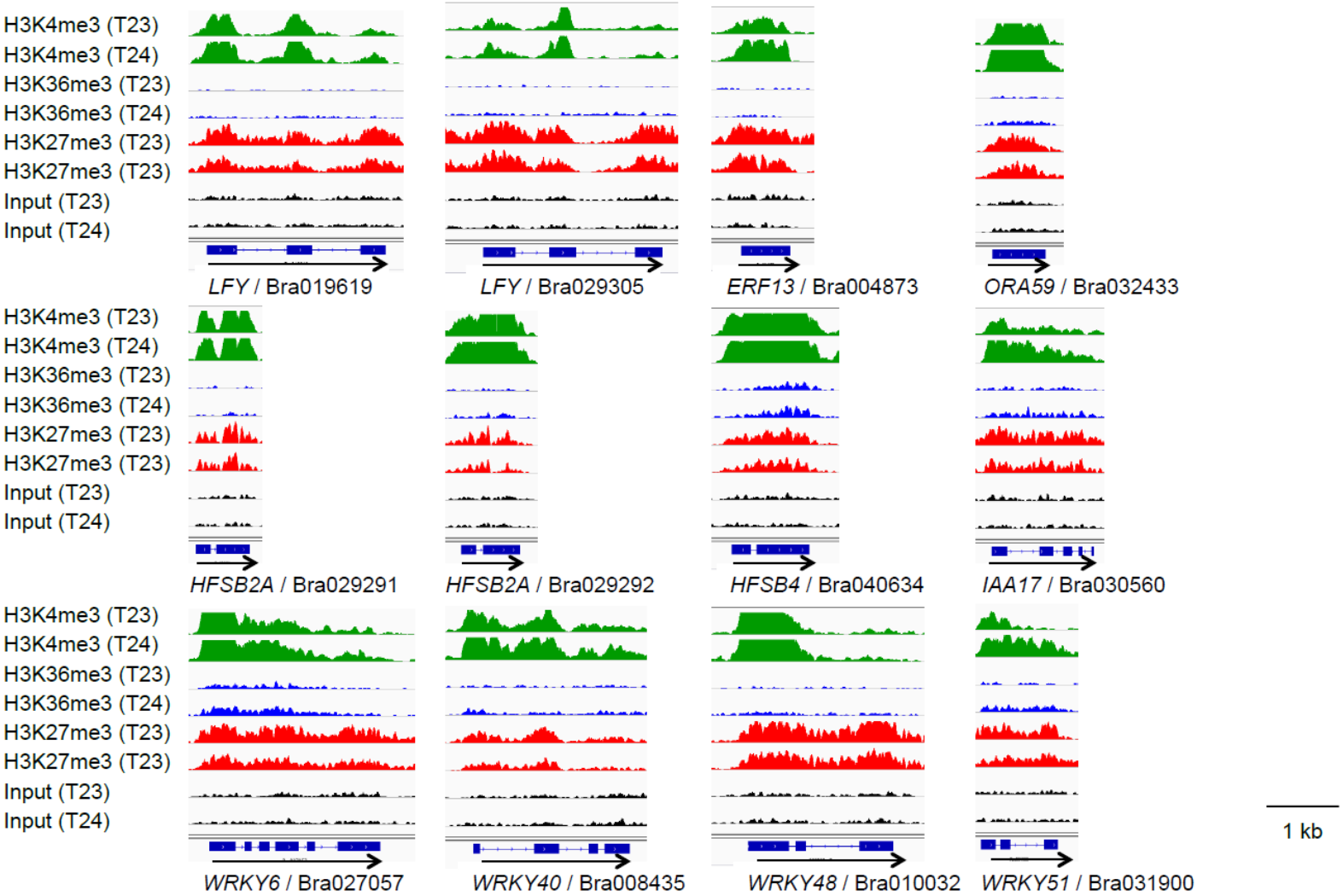


Figure II-11. Visualization of H3K4me3, H3K36me3, and H3K27me3 peaks in genes having H3K4me3 and H3K27me3 marks by Integrative Genomics Viewer (IGV) by ChIP-seq.

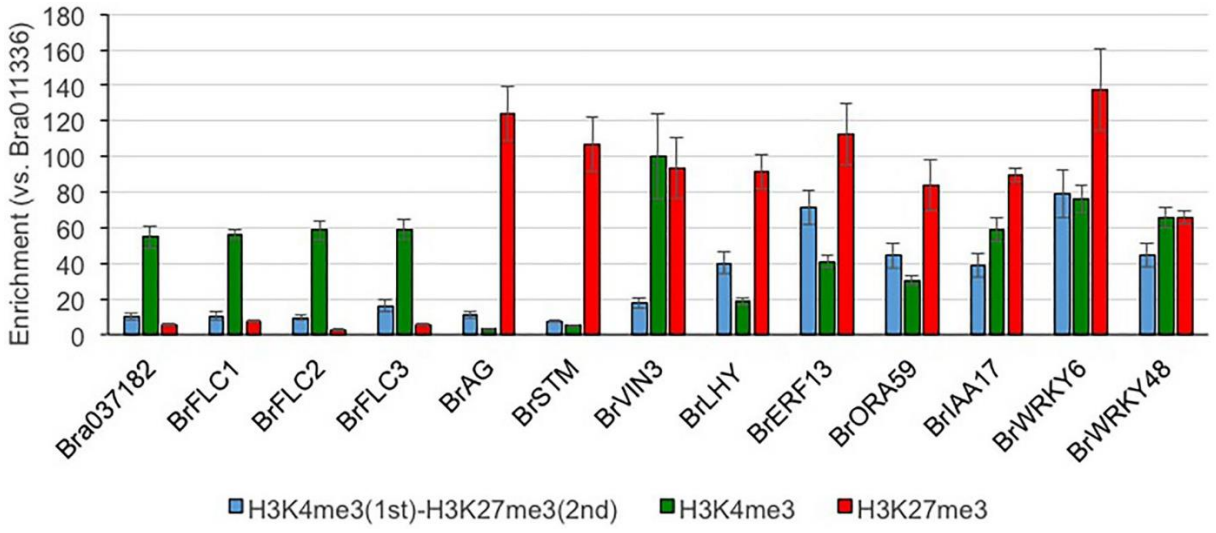


Figure II-12. Sequential ChIP-qPCR in RJKB-T24. Bra011336 that does not have H3K4me3 and H3K27me3 is used as reference gene for qPCR. Bra037182 and three *BrFLC* paralogs are H3K4me3-marked genes, and *BrAG* and *BrSTM* are H3K27me3-marked genes. *BrVIN3*, *BrLHY*, *BrERF13*, *BrORA59*, *BrIAA17*, *BrWRKY6*, and *BrWRKY48* had both H3K4me3 and H3K27me3 marks by ChIP-seq, and distribution of these modification in each gene was shown in Figure II-6 and Figure II-11. Values are means ± standard error (s.e.; three biological and technical replicates) of relative H3K4me3, H3K27me3, or H3K4me3/ H3K27me3 levels.

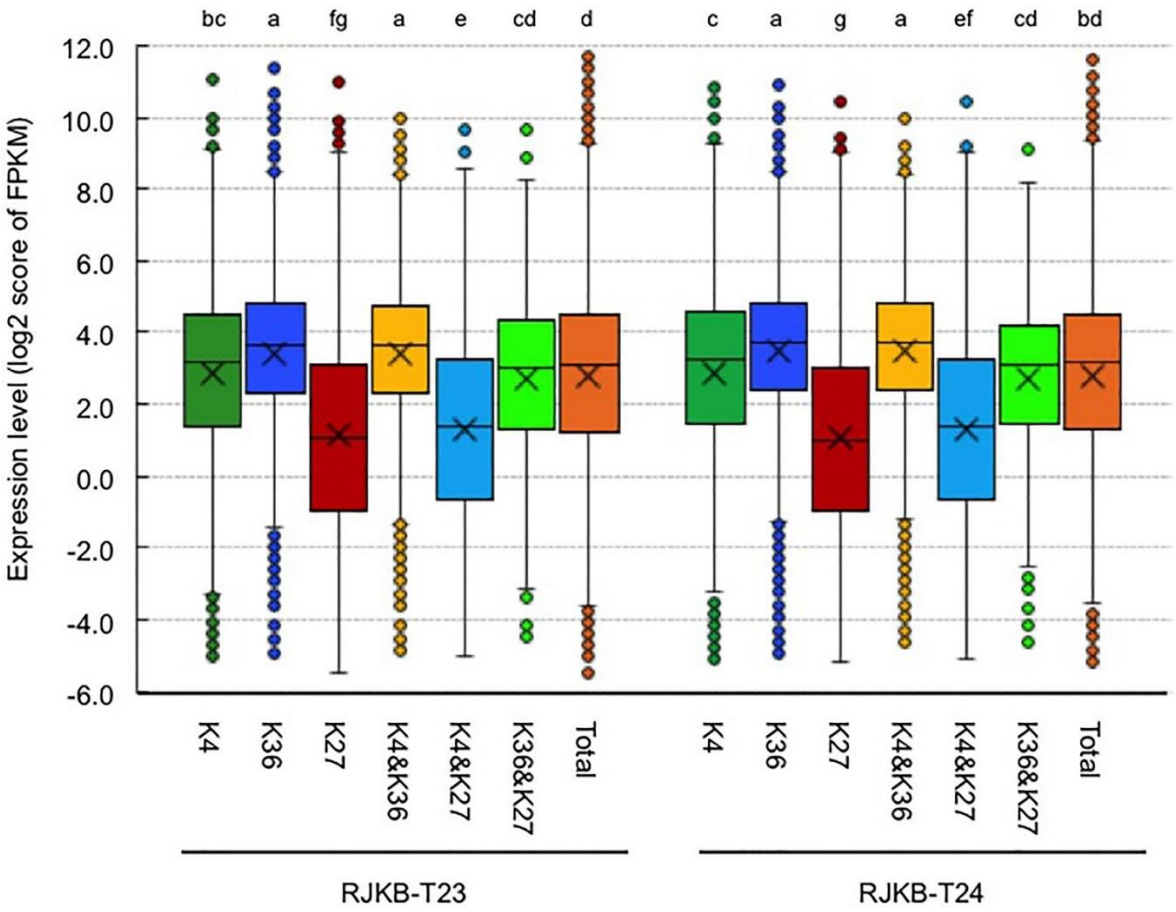


Figure II-13. Box plots of the gene expression levels of the log 2 score of FPKM with H3K4me3 (K4), H3K36me3 (K36), and H3K27me3 (K27) in genic regions of RJKB-T23 and RJKB-T24. "&" represents genes having two different histone marks. "Total" indicates the log 2 score of FPKM in all genes (FPKM < 0.01). FPKM: fragments per kilobase of transcript per million mapped reads. Different letters indicate significant difference (Tukey's HSD test, $p < 0.05$).

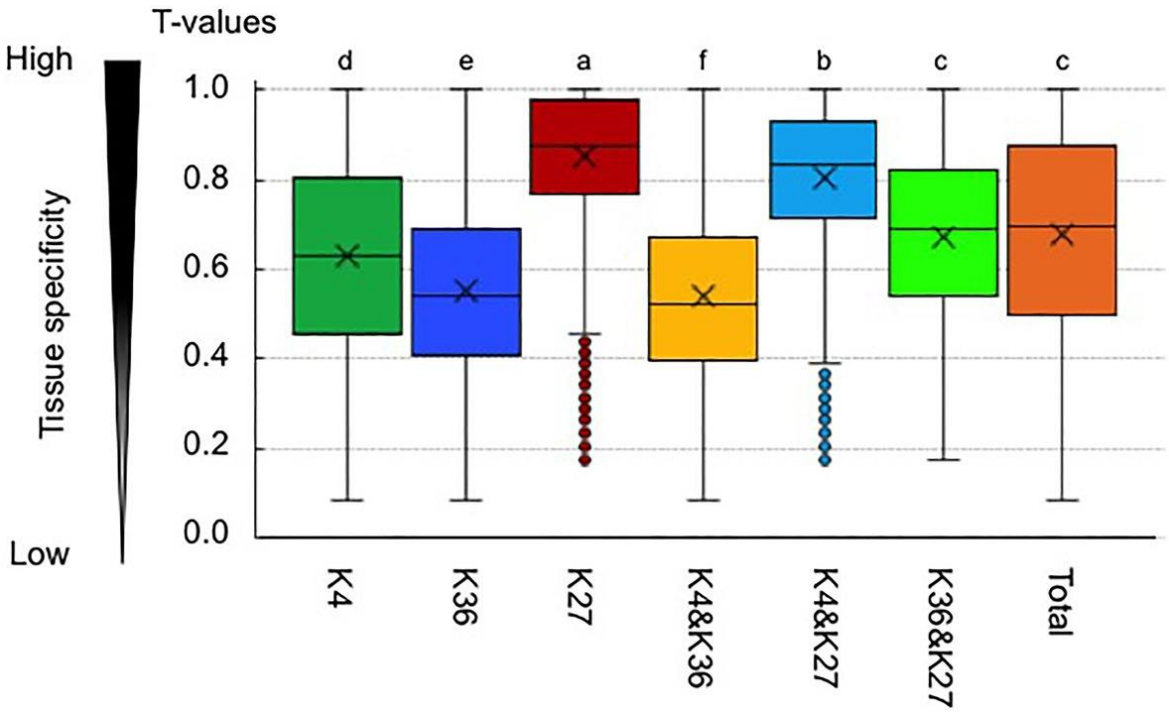


Figure II-14. Tissue specificity of expression in genes having an H3K4me3 (K4) or H3K36me3 (K36) mark. A tissue specificity index, *T*-value, which interpolates the entire range between 0 for housekeeping genes and 1 for strictly one-tissue-specific genes, was calculated using the transcriptome data in six different tissues in *B. rapa*. K27 represents the H3K27me3-marked genes. "&" represents genes having two different histone marks. "Total" indicates *T*-value in all genes. Different letters indicate significant difference (Tukey's HSD test, $p < 0.05$).

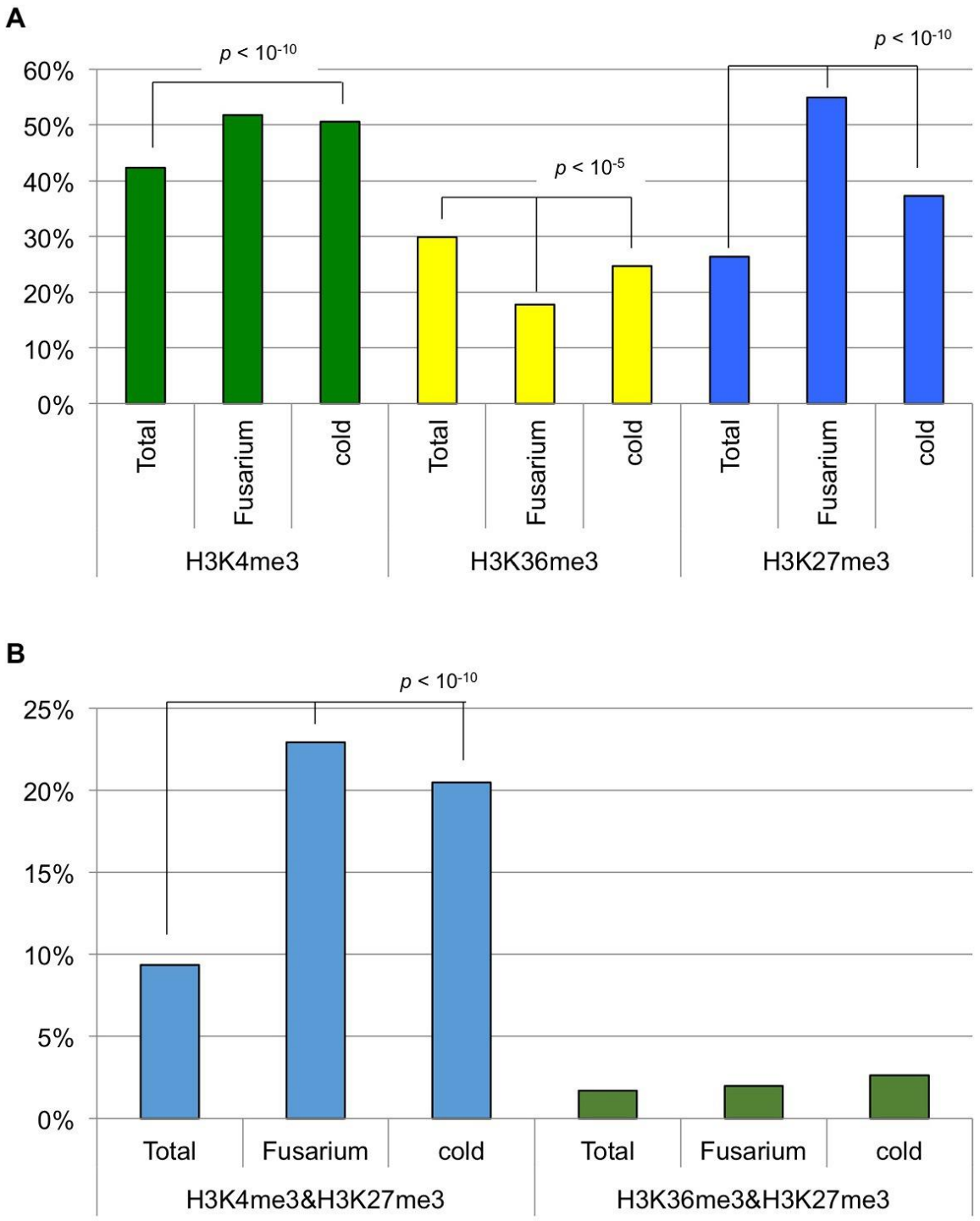


Figure II-15. The percentage of genes having histone modifications. (A) Genes having H3K4me3, H3K36me3, or H3K27me3 marks were used. (B) Genes having both H3K4me3 and H3K27me3 marks or both H3K36me3 and H3K27me3 marks were used. "Total" represents all the annotated genes in *B. rapa*. "Fusarium" and "cold" represent genes differentially expressed following *Fusarium oxysporum* f. sp. *conglutinans* inoculation and 4 weeks of cold treatment (vernalization), respectively.

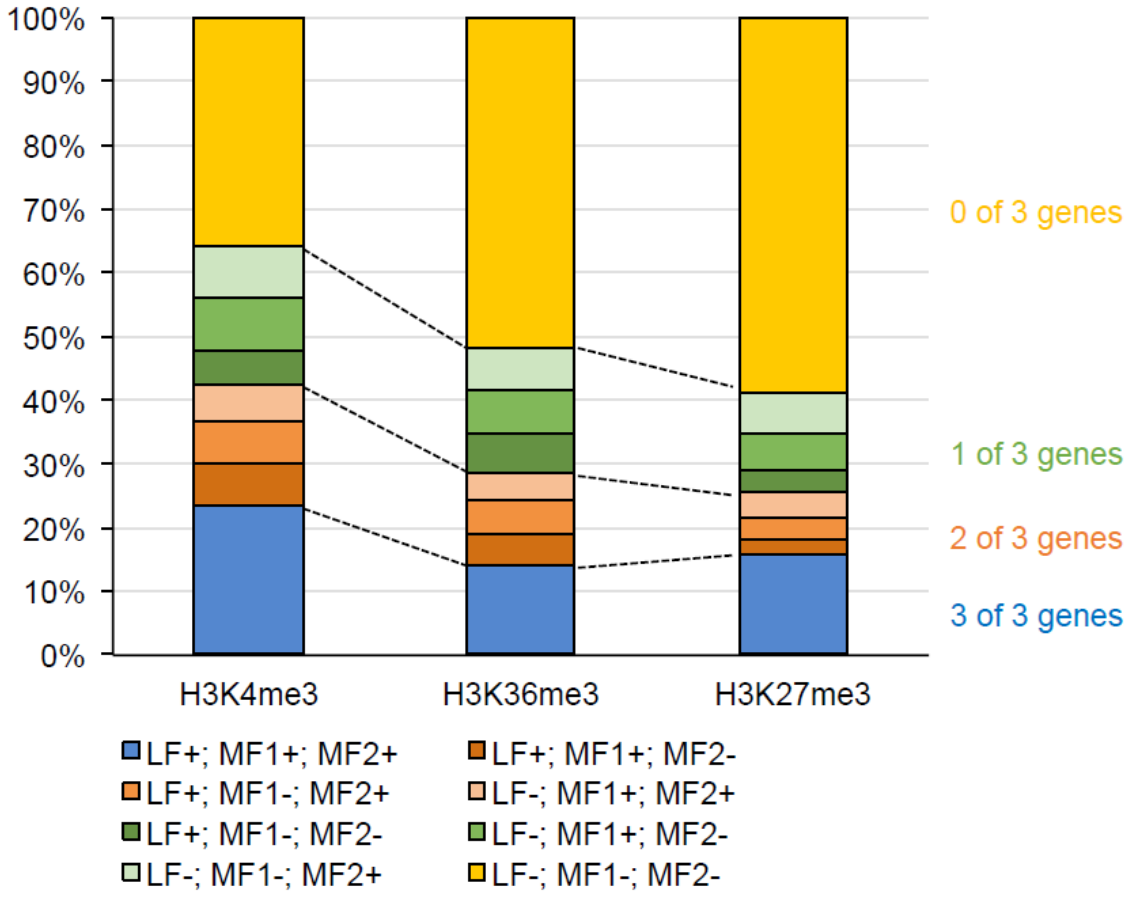


Figure II-16. Number of genes having H3K4me3 and H3K36me3 marks among three paralogous genes. '+' and '-' represent presence and absence of histone marks, respectively.

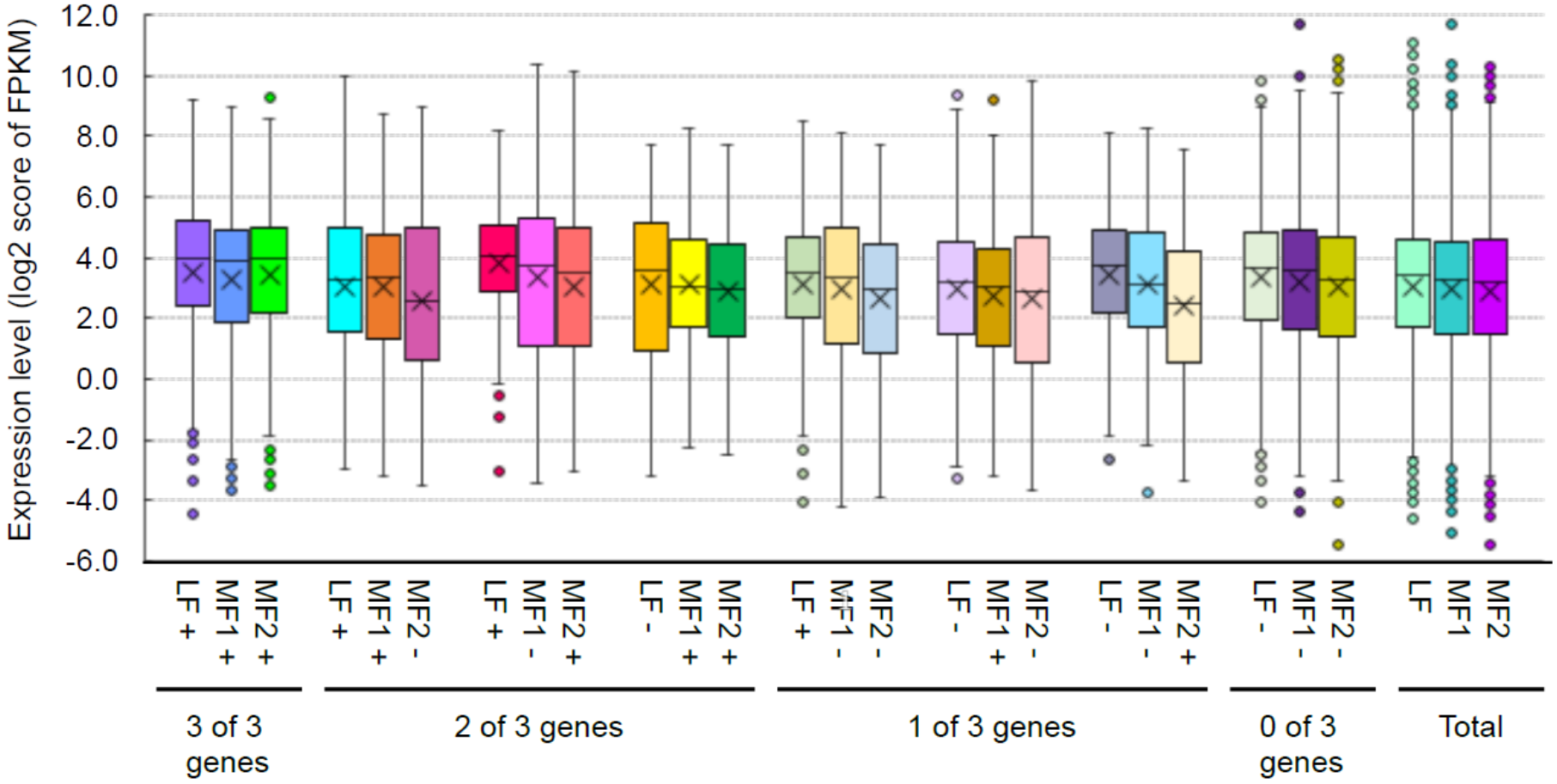


Figure II-17. Comparison of the expression level (FPKM) between paralogous pairs with and without H3K4me3 marks in RJKB-T23. Values are means ± standard error (s.e.) of FPKM. . '+' and '-' represent presence and absence of H3K4me3 marks, respectively.

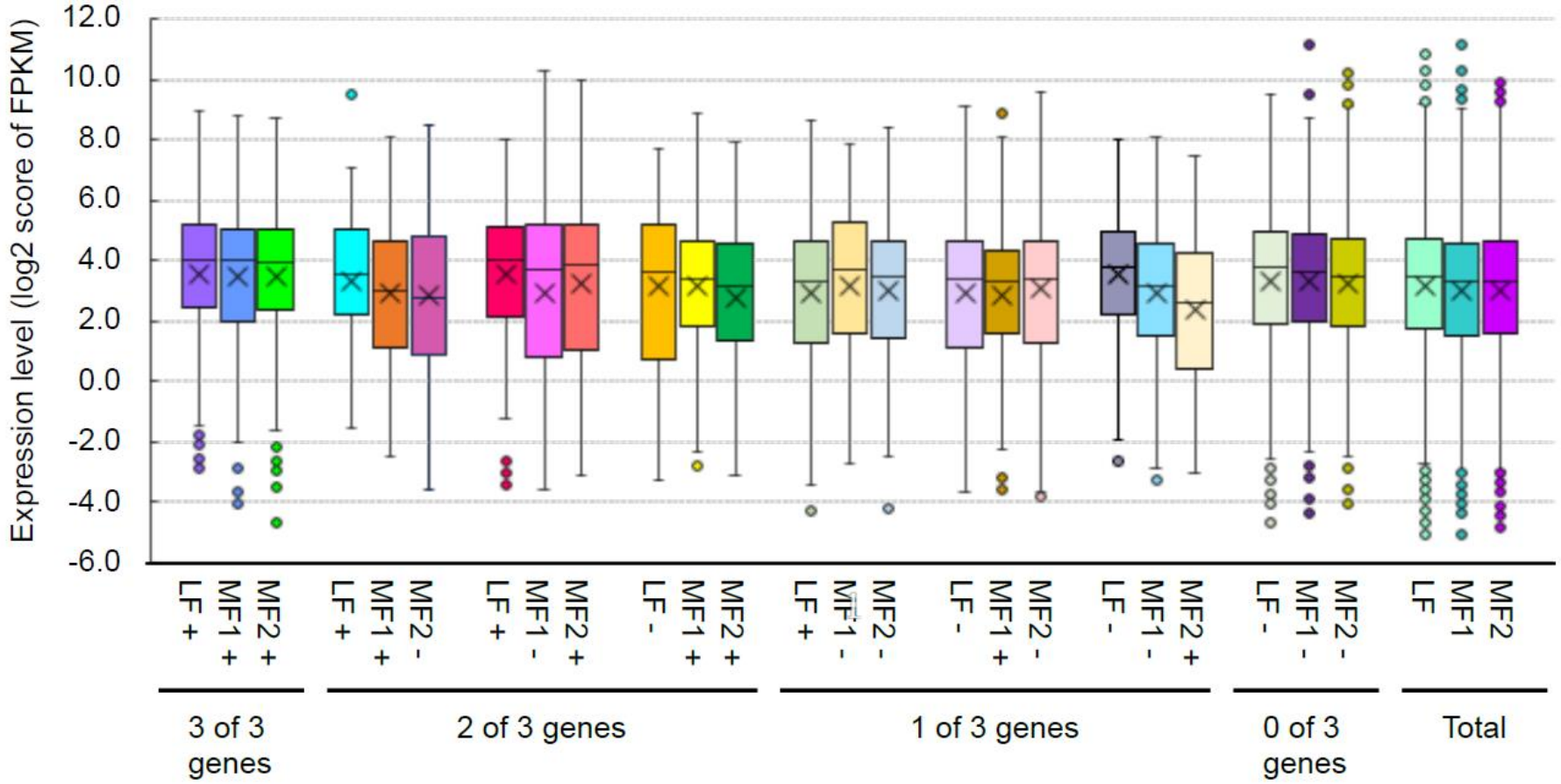


Figure II-18. Comparison of the expression level (FPKM) between paralogous pairs with and without H3K4me3 marks in RJKB-T24. Values are means ± standard error (s.e.) of FPKM. . '+' and '-' represent presence and absence of H3K4me3 marks, respectively.

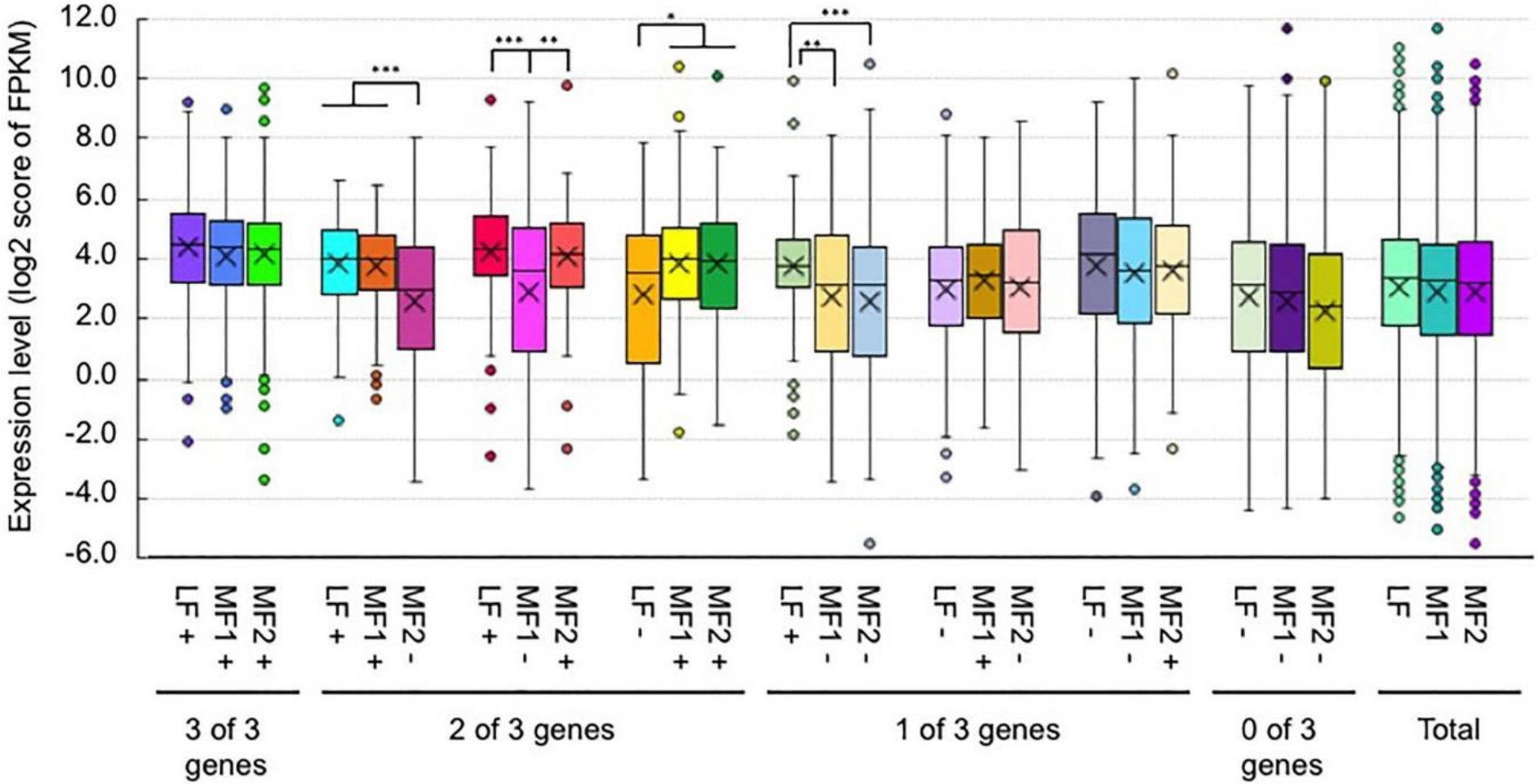


Figure II-19. Comparison of the expression level (FPKM) between paralogous pairs with and without H3K36me3 marks in RJKB-T23. Values are means ± standard error (s.e.) of FPKM. "+" and "−" represent the presence and absence of H3K36me3 marks, respectively. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (Student *t*-test).

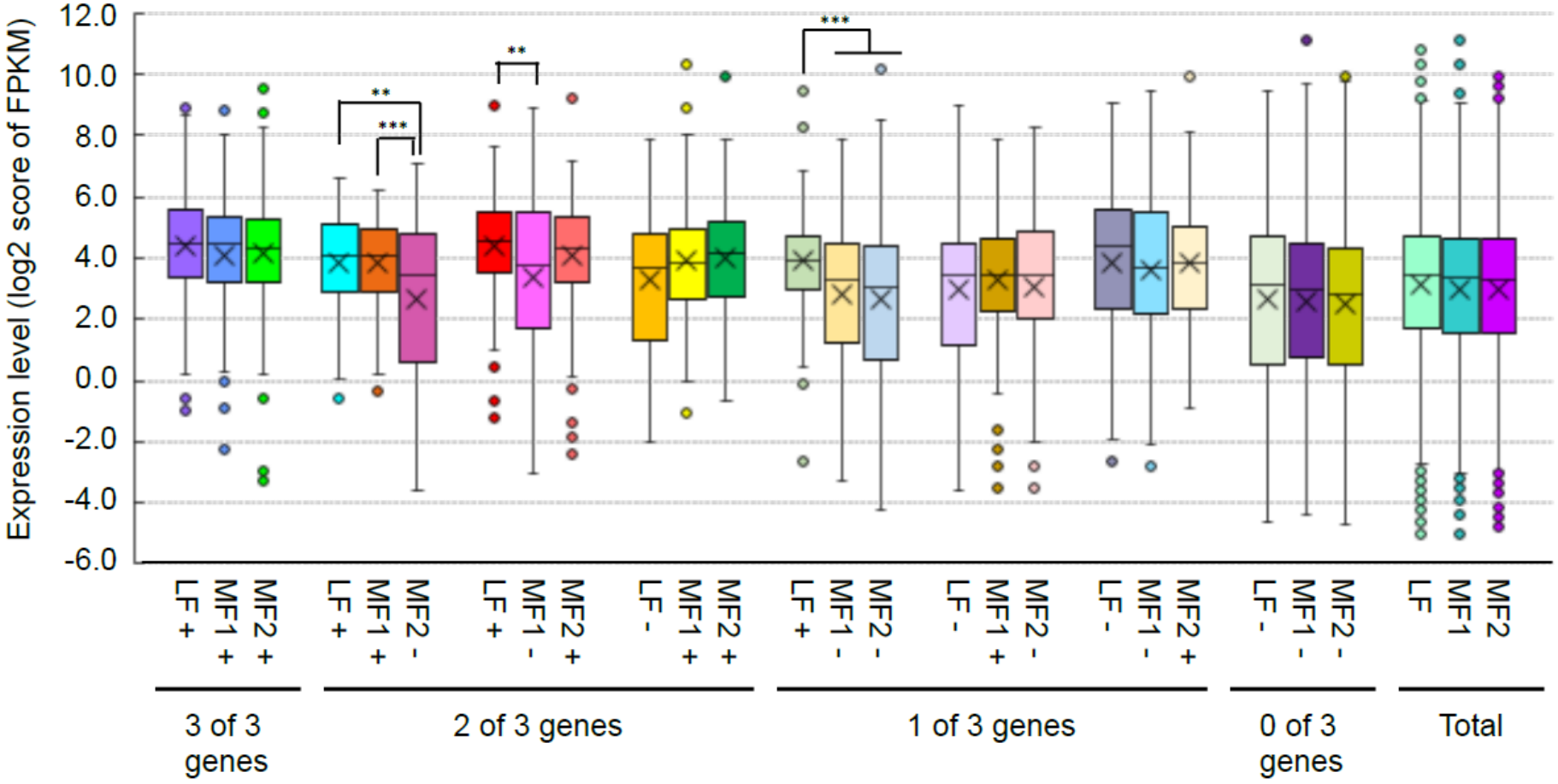


Figure II-20. Comparison of the expression level (FPKM) between paralogous pairs with and without H3K36me3 marks in RJKB-T24. Values are means ± standard error (s.e.) of FPKM. . '+' and '-' represent presence and absence of H3K36me3 marks, respectively. **, $p < 0.01$; ***, $p < 0.001$ (Student *t*-test).

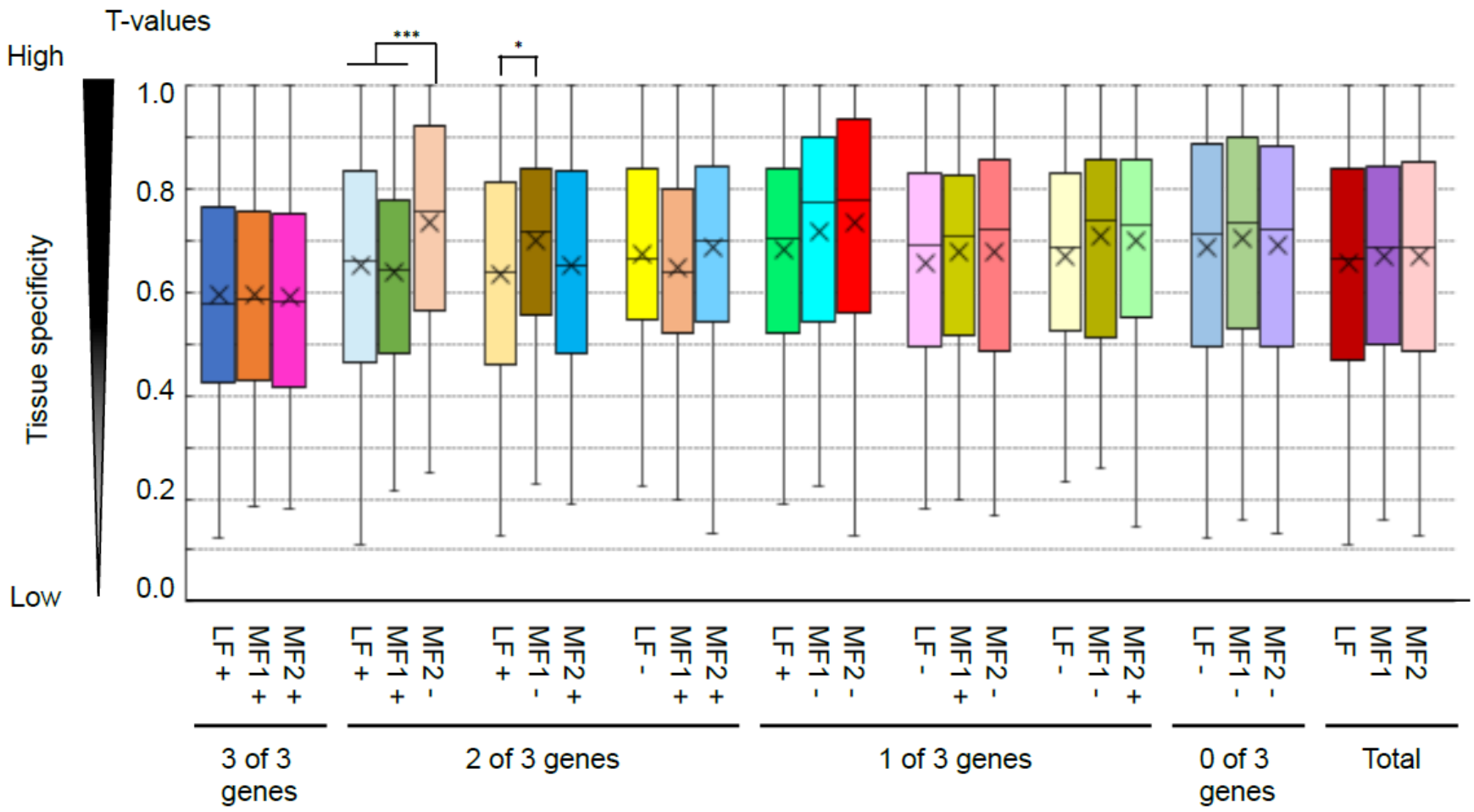


Figure II-21. Comparison of the tissue specificity of expression (a tissue specificity index, T-value) between paralogous pairs with and without H3K4me3 marks. Values are means ± standard error (s.e.) of FPKM. '+' and '-' represent presence and absence of H3K4me3 marks, respectively. *, $p < 0.05$; ***, $p < 0.001$ (Student *t*-test).

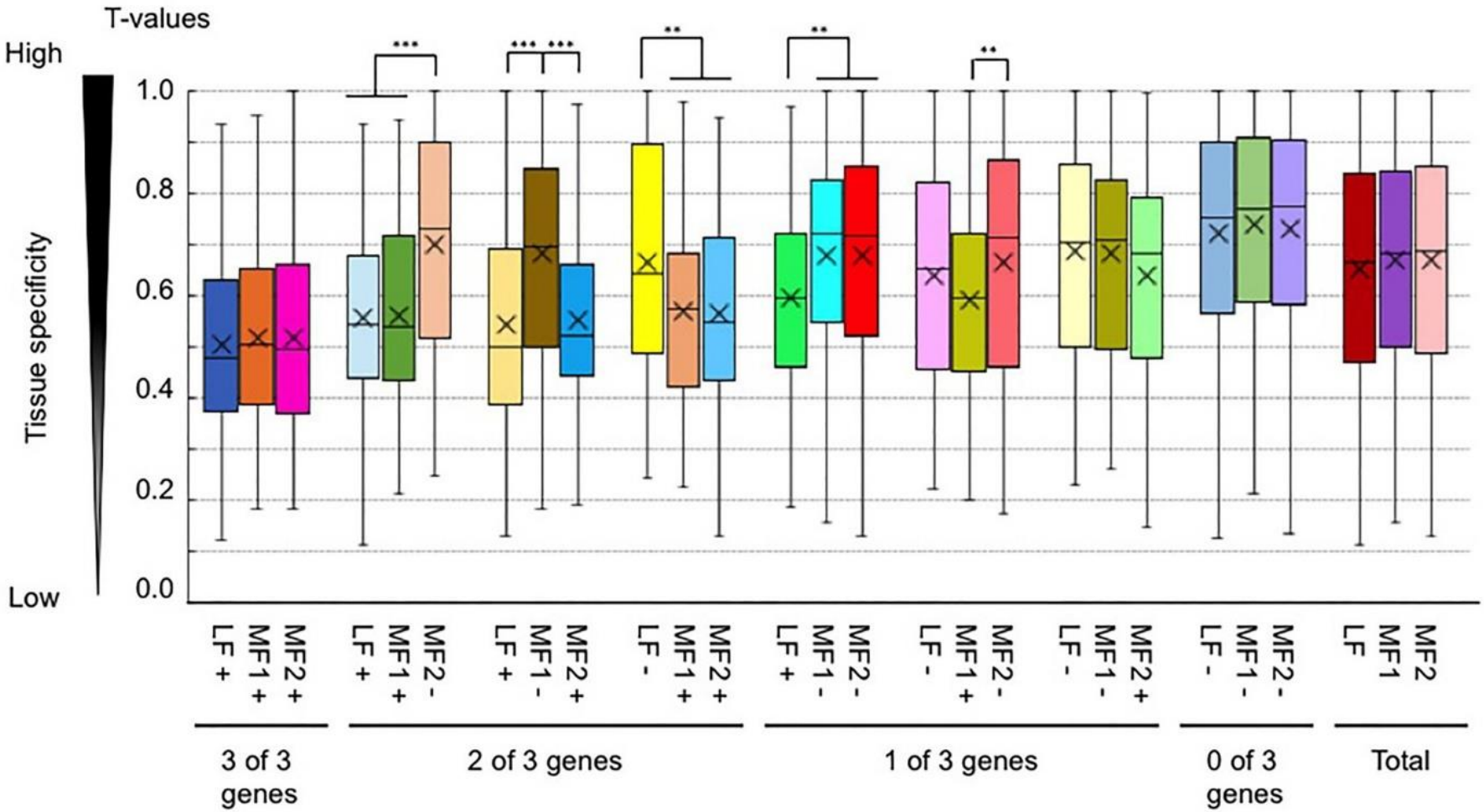


Figure II-22. Comparison of the tissue specificity of expression (a tissue specificity index, *T*-value) between paralogous pairs with and without H3K36me3 marks. Values are means ± standard error (s.e.) of FPKM. "+" and "–" represent the presence and absence of H3K4me3 marks, respectively. $^{**}p < 0.01$; $^{***}p < 0.001$ (Student *t*-test).

Whole genome

	H3K4me3	H3K36me3	H3K9me2	H3K27me3	CG	CHG	CHH
H3K4me3		0.85	-0.23	0.28	-0.73	-0.60	-0.41
H3K36me3	0.87		-0.20	0.12	-0.62	-0.59	-0.45
H3K9me2	-0.25	-0.19		-0.15	0.48	0.35	0.21
H3K27me3	0.25	0.13	-0.17		-0.44	-0.28	-0.16
CG	-0.74	-0.64	0.47	-0.46		0.85	0.63
CHG	-0.61	-0.62	0.33	-0.30	0.85		0.86
CHH	-0.42	-0.48	0.19	-0.20	0.65	0.87	

Genic region

	H3K4me3	H3K36me3	H3K9me2	H3K27me3	CG	CHG	CHH
H3K4me3		0.64	0.30	-0.04	-0.21	-0.26	-0.18
H3K36me3	0.65		0.08	-0.11	0.00	-0.18	-0.15
H3K9me2	0.27	0.21		0.16	0.22	0.26	0.23
H3K27me3	-0.13	-0.18	0.14		-0.11	-0.05	0.00
CG	-0.19	0.04	0.24	-0.15		0.79	0.61
CHG	-0.26	-0.18	0.24	-0.06	0.75		0.79
CHH	-0.18	-0.15	0.21	-0.03	0.60	0.80	

IRRs

	H3K4me3	H3K36me3	H3K9me2	H3K27me3	CG	CHG	CHH
H3K4me3		0.79	0.51	0.56	-0.01	-0.02	-0.04
H3K36me3	0.78		0.53	0.45	0.03	0.00	-0.02
H3K9me2	0.40	0.56		0.57	0.15	0.12	0.04
H3K27me3	0.46	0.43	0.49		0.09	0.07	0.02
CG	-0.02	0.05	0.15	0.05		0.70	0.46
CHG	-0.03	0.02	0.11	0.04	0.71		0.62
CHH	-0.04	-0.01	0.04	0.01	0.47	0.63	

Figure II-23. The comparison between epigenetic states in RJKB-T23 (upper right) and RJKB-24 (lower left). The correlation coefficient of histone modifications quantified by reads per kilobase million (RPKM) and DNA methylation levels in each sliding window per 100 kb at the whole genome level. IRRs, interspersed repeats regions.

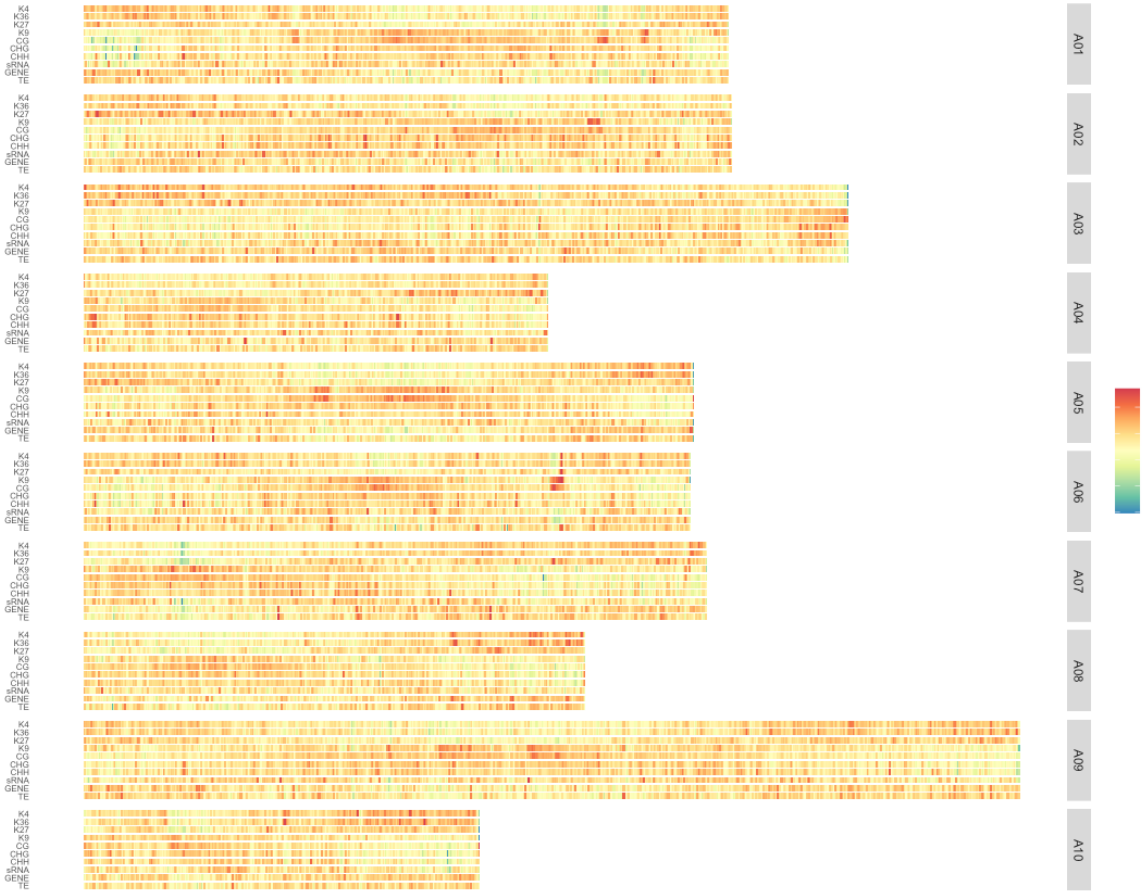


Figure II-24. The distribution of H3K4me3 and H3K36me3 and other epigenetic marks in RJKB-T23. Heatmap is visualized using ggplot2 package version 3.3.2 (https://ggplot2.tidyverse.org/) in R version 3.6.1 (https://www.r-project.org/). GENE and TE represent their expression levels.

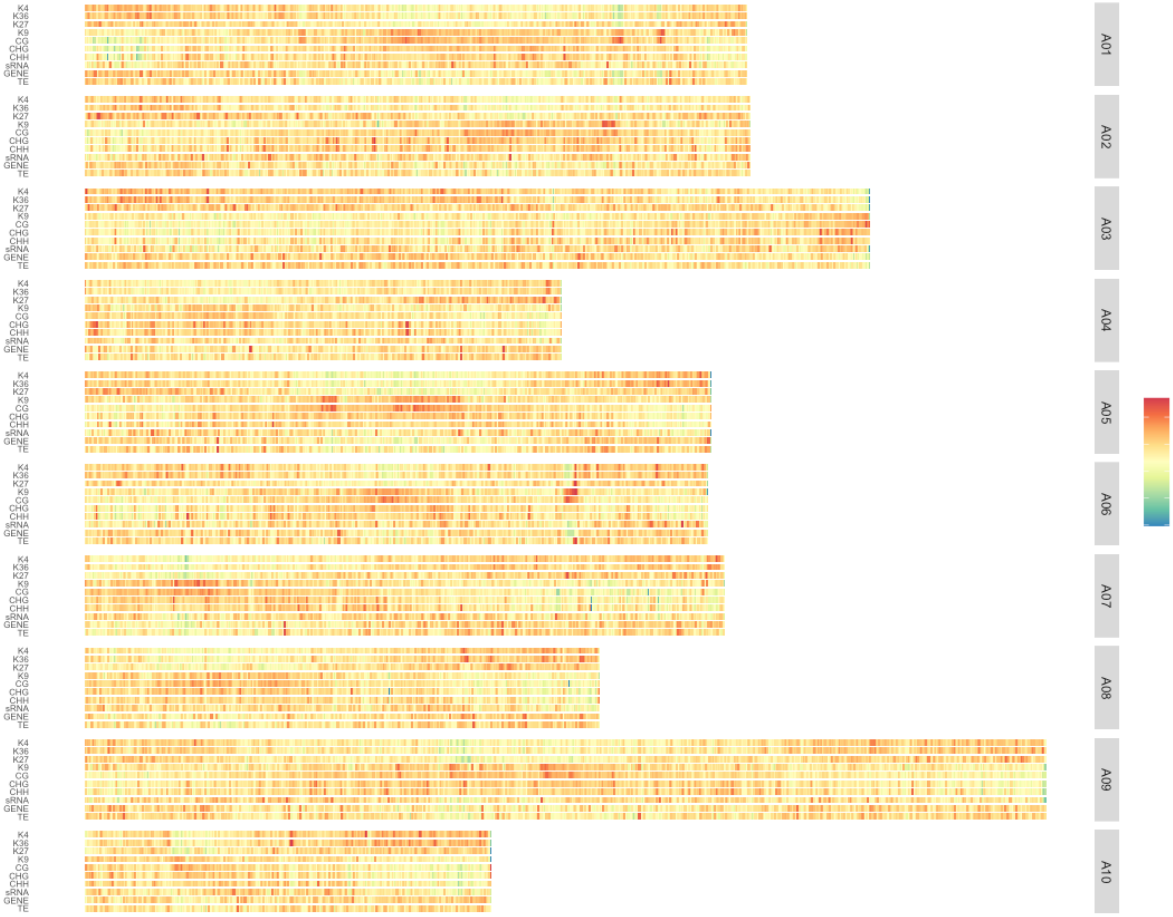


Figure II-25. The distribution of H3K4me3 and H3K36me3 and other epigenetic marks in RJKB-T24. Heatmap is visualized using ggplot2 package version 3.3.2 (https://ggplot2.tidyverse.org/) in R version 3.6.1 (https://www.r-project.org/). GENE and TE represent their expression levels.

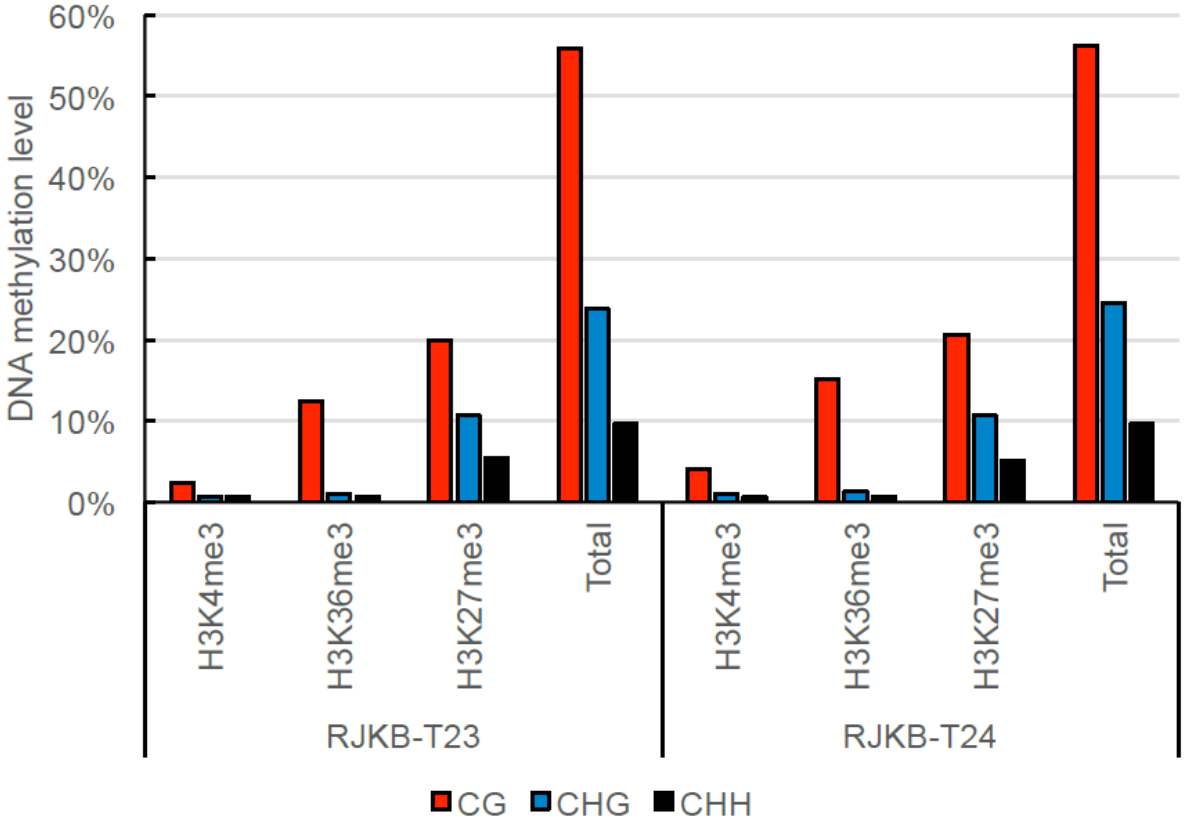


Figure II-26. DNA methylation level in the regions having H3K4me3 or H3K36me3 marks. 'Total' represents the average of DNA methylation levels in all regions of the genome.

Table II-1. Sequence data of ChIP-seq in RJKB-T23 and RJKB-T24

Sample	a. Total number of reads	b. Mapped reads	b/a (%)	c. Mapped reads (Unique aligned)	c/a (%)	d. Mapped in A01-A10	d/c (%)	e. Mapped in IRRs	e/c (%)
Input									
RJKB-T23 rep1*	30,674,566	27,017,433	88.1%	14,575,541	47.5%	12,405,527	85.1%	6,294,140	43.2%
RJKB-T23 rep2*	41,092,509	36,173,019	88.0%	18,713,543	45.5%	16,174,659	86.4%	7,683,962	41.1%
Total (rep1 + rep2)	71,767,075	63,190,452	88.0%	33,289,084	46.4%	28,580,186	85.9%	13,978,102	42.0%
RJKB-T24 rep1*	27,342,223	24,058,760	88.0%	12,247,035	44.8%	10,320,865	84.3%	5,369,066	43.8%
RJKB-T24 rep2*	37,826,546	33,372,769	88.2%	18,221,789	48.2%	15,809,899	86.8%	7,130,500	39.1%
Total (rep1 + rep2)	65,168,769	57,431,529	88.1%	30,468,824	46.8%	26,130,764	85.8%	12,499,566	41.0%
H3K4me3									
RJKB-T23	30,596,386	28,639,370	93.6%	17,041,548	55.7%	16,317,538	95.8%	3,188,886	18.7%
RJKB-T24	41,092,059	38,531,342	93.8%	19,753,824	48.1%	18,974,076	96.1%	3,513,416	17.8%
H3K36me3									
RJKB-T23	21,907,233	19,828,887	90.5%	12,669,101	57.8%	12,121,794	95.7%	1,676,741	13.2%
RJKB-T24	38,707,655	35,231,079	91.0%	19,778,151	51.1%	18,909,961	95.6%	2,878,052	14.6%

* Data from Akter et al. (2019)

Table II-2. Sequences of primers used for ChIP-qPCR

Name	Primer sequences (5'-3')	
ChIP-qPCR (H3K4me3)		
Positive control		
Bra013206	GACGAGCACAAGAGTGGTGA	TAATCGCTGTCGCTGTCACT
Bra039113	CTCCTCTCTCCGGCTTCTTC	AAGGGGTGATGATAGGAGCA
Negative control		
FWA	CGGCATATGATTCGTTTGTG	CCTGGTTGTGTAGCATGTGG
STM	TGGAGAGTGGTTCCAACAGCACTTC	GGAGCTACTTTGTTGGTGGTGTGAC
ChIP-qPCR (H3K36me3)		
Positive control		
Bra028174	ACGGGTTTGTTTACCGTGAG	CTGATGCACTGGACTTGACG
Bra028913	CCCTGGGAGCAACTCTGTTA	GTGGGAGCAATCCTGATGAC
Negative control		
AG	AAATGAGAGGAACAATCCAAGTATG	ACACTAACTGAAGAGCGGTTTGGTC
Bra018355	TCAACTTGTGGAACCGTCAA	CCACCATCTCCTTCCATGTT
Sequential ChIP-qPCR (H3K4me3 and H3K27me3)		
VIN3	AGCGTGAGTACATGTTTTACATGC	CTTGAAATGAAGAAGGATCTATGTC
LHY	ATGGATCCTGAAGGTTTCACGAG	TCCAAACGCTGCCGTCTGCGGTG
ERF13	CAATAACGGTGTTAACCCACCCG	CGCGGAGAGTGTTGTAGATGGCC
ORA59	GAGGGAGAAGACACCGTGGTGGC	GTCGAACGTCCCGAGCCACACTC
IAA17	GGCAGTGTTGGGCTGAATCTGAG	TAGCACTCACGACGTCGTGGCTC

WRKY6	GGGTTGGTCTGGTCTCACTCTTG	CCTGAATTCACTGCCGTCGATGG
WRKY48	CACAGAGACCAAGATAGATCACC	TTCCCAGCTAGCTCCAACTGATG
Bra037182	CAGTGCCATTGGTGACTTTG	CCTCTCCGCTTTCGTTAGTG
FLC-1	TGGGGAGGAAGAAACTTGAA	CACCGGAGGAGAAGCTGTAG
FLC-2	CGACAAGTCACCTTCTCCAA	AGAGGAACGGAAGCGAAAAG
FLC-3	TTGAGAACAAAAGTAGCCGACA	GGCTAATAAAGGAAGGCACAGA
For reference		
Bra011336	GGGAATCAGCTTTTGTGGTG	AAATGACCCGATCAGCAAAG

Table II-3. Mapped reads in genic region using Input-DNA-seq and ChIP-seq data of RJKB-T23 and RJKB-T24

Sample	a. Mapped reads (Unique aligned)	b. Upstream (2kb)	b/a	c. Exon	c/a	d. Intron	d/a	e. Downstream (2kb)	e/a
Input									
RJKB-T23 rep1*	14,575,541	3,104,668	21.3%	3,617,693	24.8%	1,815,451	12.5%	3,410,205	23.4%
RJKB-T23 rep2*	18,713,543	4,185,124	22.4%	4,588,848	24.5%	2,498,728	13.4%	4,712,180	25.2%
RJKB-T24 rep1*	12,247,035	2,549,482	20.8%	2,971,301	24.3%	1,474,871	12.0%	2,787,697	22.8%
RJKB-T24 rep2*	18,221,789	4,192,231	23.0%	4,433,169	24.3%	2,573,304	14.1%	4,778,009	26.2%
H3K27me3									
H3K4me3									
RJKB-T23	17,041,548	6,730,625	39.5%	9,233,871	54.2%	4,271,496	25.1%	5,984,367	35.1%
RJKB-T24	19,753,824	8,076,045	40.9%	10,619,642	53.8%	5,073,689	25.7%	7,080,689	35.8%
H3K36me3									
RJKB-T23	12,669,101	4,310,656	34.0%	6,304,469	49.8%	4,114,773	32.5%	4,054,592	32.0%
RJKB-T24	19,778,151	6,560,514	33.2%	9,776,037	49.4%	6,005,833	30.4%	6,439,203	32.6%

* Data from Akter et al. (2019)

Table II-4. Comparison of the H3K4me3, H3K36me3, and H3K27me3-marked genes

	Total number of genes	1	2	3
1. H3K4me3 (T23&T24)	16,759		10,119	3,699
2. H3K36me3 (T23&T24)	11,844			671
3. H3K27me3 (T23&T24) *	10,456			

* Data from Akter et al. (2019)

Table II-5. Number of genes having overlapped peaks of two histone modification

Number of genes overlapping peaks of different histone modifications

	Total gene number	Number of genes having overlapped peaks					
		> 500bp		> 300bp		>150bp	
H3K4me3 & H3K36me3	10,119	8,556	84.6%	9,670	95.6%	9,970	98.5%
H3K4me3 & H3K27me3	3,699	1,179	31.9%	2,464	66.6%	3,111	84.1%
H3K36me3 & H3K27me3	671	107	15.9%	229	34.1%	373	55.6%

Table II-6. GO function term overrepresented in genes having H3K4me3 and H3K27me3 marks

GO accession	Term type	Term	*p* value	FDR
GO:0005488	F	binding	4.00E-239	6.00E-236
GO:0008152	P	metabolic process	3.80E-193	1.80E-189
GO:0016021	C	integral to membrane	1.60E-166	7.30E-164
GO:0031224	C	intrinsic to membrane	8.60E-109	1.90E-106
GO:0003824	F	catalytic activity	9.60E-107	7.30E-104
GO:0005515	F	protein binding	1.10E-99	5.60E-97
GO:0032774	P	RNA biosynthetic process	3.30E-97	7.60E-94
GO:0006351	P	transcription, DNA-dependent	1.00E-96	1.60E-93
GO:0046872	F	metal ion binding	1.30E-93	5.00E-91
GO:0006355	P	regulation of transcription, DNA-dependent	7.90E-94	9.20E-91
GO:0051252	P	regulation of RNA metabolic process	1.40E-93	1.30E-90
GO:0016705	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	3.10E-92	9.30E-90
GO:0005506	F	iron ion binding	3.40E-89	8.50E-87
GO:0043169	F	cation binding	4.70E-89	8.80E-87
GO:0043167	F	ion binding	4.70E-89	8.80E-87
GO:0044425	C	membrane part	4.60E-78	6.90E-76
GO:0046983	F	protein dimerization activity	5.30E-77	8.90E-75
GO:0016070	P	RNA metabolic process	2.00E-75	1.40E-72
GO:0055085	P	transmembrane transport	1.90E-75	1.40E-72
GO:0044238	P	primary metabolic process	2.60E-72	1.50E-69
GO:0046906	F	tetrapyrrole binding	1.10E-68	1.70E-66
GO:0004497	F	monooxygenase activity	1.20E-60	1.70E-58
GO:0009987	P	cellular process	1.20E-58	6.20E-56
GO:0001882	F	nucleoside binding	3.30E-55	4.20E-53
GO:0001883	F	purine nucleoside binding	9.30E-55	1.00E-52
GO:0030554	F	adenyl nucleotide binding	9.30E-55	1.00E-52
GO:0044237	P	cellular metabolic process	6.60E-53	3.10E-50
GO:0016491	F	oxidoreductase activity	8.40E-52	8.50E-50

GO:0017076	F	purine nucleotide binding	4.10E-50	3.90E-48
GO:0006810	P	transport	6.10E-48	2.60E-45
GO:0051234	P	establishment of localization	1.00E-47	3.90E-45
GO:0051179	P	localization	2.70E-47	9.60E-45
GO:0065007	P	biological regulation	6.20E-47	2.10E-44
GO:0009889	P	regulation of biosynthetic process	1.30E-42	3.80E-40
GO:0031326	P	regulation of cellular biosynthetic process	1.30E-42	3.80E-40
GO:0000166	F	nucleotide binding	8.10E-42	7.20E-40
GO:0045449	P	regulation of transcription	3.50E-42	9.60E-40
GO:0005976	P	polysaccharide metabolic process	4.50E-42	1.20E-39
GO:0006350	P	transcription	5.50E-42	1.30E-39
GO:0010556	P	regulation of macromolecule biosynthetic process	5.90E-42	1.40E-39
GO:0010468	P	regulation of gene expression	9.10E-42	2.00E-39
GO:0032559	F	adenyl ribonucleotide binding	4.10E-41	3.40E-39
GO:0019219	P	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	4.70E-41	1.00E-38
GO:0006807	P	nitrogen compound metabolic process	5.60E-41	1.10E-38
GO:0031323	P	regulation of cellular metabolic process	9.50E-41	1.80E-38
GO:0019222	P	regulation of metabolic process	1.80E-40	3.40E-38
GO:0060255	P	regulation of macromolecule metabolic process	3.50E-40	6.30E-38
GO:0009055	F	electron carrier activity	8.20E-40	6.50E-38
GO:0080090	P	regulation of primary metabolic process	4.20E-40	7.30E-38
GO:0006139	P	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	6.40E-40	1.10E-37
GO:0046914	F	transition metal ion binding	1.80E-39	1.30E-37
GO:0051171	P	regulation of nitrogen compound metabolic process	9.30E-40	1.50E-37
GO:0003677	F	DNA binding	1.40E-38	1.00E-36
GO:0032555	F	purine ribonucleotide binding	2.00E-37	1.30E-35
GO:0032553	F	ribonucleotide binding	2.00E-37	1.30E-35
GO:0043170	P	macromolecule metabolic process	9.50E-38	1.50E-35

GO:0005524	F	ATP binding	1.10E-36	6.90E-35
GO:0005975	P	carbohydrate metabolic process	5.60E-36	8.30E-34
GO:0050790	P	regulation of catalytic activity	9.70E-36	1.40E-33
GO:0065009	P	regulation of molecular function	2.60E-35	3.70E-33
GO:0044264	P	cellular polysaccharide metabolic process	1.30E-33	1.80E-31
GO:0044248	P	cellular catabolic process	3.40E-32	4.50E-30
GO:0016052	P	carbohydrate catabolic process	5.50E-32	7.10E-30
GO:0006811	P	ion transport	9.20E-30	1.20E-27
GO:0044262	P	cellular carbohydrate metabolic process	5.70E-29	7.00E-27
GO:0044260	P	cellular macromolecule metabolic process	3.00E-28	3.50E-26
GO:0050794	P	regulation of cellular process	4.60E-28	5.30E-26
GO:0050789	P	regulation of biological process	8.70E-28	9.90E-26
GO:0009199	P	ribonucleoside triphosphate metabolic process	9.20E-26	1.00E-23
GO:0009141	P	nucleoside triphosphate metabolic process	9.90E-26	1.10E-23
GO:0009058	P	biosynthetic process	1.70E-25	1.80E-23
GO:0009205	P	purine ribonucleoside triphosphate metabolic process	5.90E-25	6.10E-23
GO:0009144	P	purine nucleoside triphosphate metabolic process	1.50E-24	1.50E-22
GO:0016020	C	membrane	1.80E-24	2.00E-22
GO:0016787	F	hydrolase activity	3.50E-24	2.10E-22
GO:0006796	P	phosphate metabolic process	1.10E-22	1.10E-20
GO:0016740	F	transferase activity	2.00E-22	1.10E-20
GO:0006793	P	phosphorus metabolic process	1.20E-22	1.20E-20
GO:0044275	P	cellular carbohydrate catabolic process	1.30E-22	1.20E-20
GO:0043687	P	post-translational protein modification	1.80E-22	1.70E-20
GO:0044249	P	cellular biosynthetic process	2.50E-22	2.30E-20
GO:0009150	P	purine ribonucleotide metabolic process	1.90E-21	1.70E-19
GO:0009259	P	ribonucleotide metabolic process	3.90E-21	3.50E-19
GO:0006163	P	purine nucleotide metabolic process	2.60E-20	2.20E-18
GO:0048037	F	cofactor binding	7.70E-20	4.30E-18

GO:0006464	P	protein modification process	1.10E-19	9.10E-18
GO:0005215	F	transporter activity	2.60E-19	1.40E-17
GO:0016706	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	3.40E-19	1.80E-17
GO:0006812	P	cation transport	2.60E-19	2.20E-17
GO:0032259	P	methylation	1.60E-18	1.30E-16
GO:0006629	P	lipid metabolic process	5.00E-18	4.00E-16
GO:0032787	P	monocarboxylic acid metabolic process	5.30E-18	4.20E-16
GO:0003676	F	nucleic acid binding	1.00E-17	5.20E-16
GO:0030001	P	metal ion transport	6.80E-18	5.30E-16
GO:0006730	P	one-carbon metabolic process	8.60E-18	6.50E-16
GO:0009056	P	catabolic process	1.10E-17	7.90E-16
GO:0042545	P	cell wall modification	1.20E-17	8.50E-16
GO:0004674	F	protein serine/threonine kinase activity	2.10E-17	1.00E-15
GO:0006855	P	multidrug transport	1.80E-17	1.30E-15
GO:0050660	F	FAD binding	2.70E-17	1.30E-15
GO:0030599	F	pectinesterase activity	3.00E-17	1.40E-15
GO:0043412	P	macromolecule modification	2.40E-17	1.70E-15
GO:0022857	F	transmembrane transporter activity	5.30E-17	2.30E-15
GO:0015893	P	drug transport	4.90E-17	3.50E-15
GO:0042493	P	response to drug	6.90E-17	4.70E-15
GO:0015238	F	drug transmembrane transporter activity	1.30E-16	5.40E-15
GO:0046034	P	ATP metabolic process	8.40E-17	5.70E-15
GO:0016310	P	phosphorylation	1.90E-16	1.30E-14
GO:0004857	F	enzyme inhibitor activity	3.20E-16	1.30E-14
GO:0006753	P	nucleoside phosphate metabolic process	3.60E-16	2.40E-14
GO:0009117	P	nucleotide metabolic process	3.60E-16	2.40E-14
GO:0055086	P	nucleobase, nucleoside and nucleotide metabolic process	4.90E-16	3.20E-14
GO:0043436	P	oxoacid metabolic process	5.30E-16	3.30E-14
GO:0019752	P	carboxylic acid metabolic process	5.30E-16	3.30E-14
GO:0006082	P	organic acid metabolic process	5.70E-16	3.50E-14

GO:0016758	F	transferase activity, transferring hexosyl groups	9.30E-16	3.80E-14
GO:0006468	P	protein amino acid phosphorylation	7.20E-16	4.40E-14
GO:0042180	P	cellular ketone metabolic process	7.70E-16	4.60E-14
GO:0034645	P	cellular macromolecule biosynthetic process	1.50E-15	9.10E-14
GO:0004672	F	protein kinase activity	2.50E-15	1.00E-13
GO:0009059	P	macromolecule biosynthetic process	2.50E-15	1.50E-13
GO:0006631	P	fatty acid metabolic process	1.00E-14	5.80E-13
GO:0015698	P	inorganic anion transport	1.10E-14	6.50E-13
GO:0016746	F	transferase activity, transferring acyl groups	4.90E-14	1.90E-12
GO:0005634	C	nucleus	2.80E-14	2.50E-12
GO:0016773	F	phosphotransferase activity, alcohol group as acceptor	1.30E-13	4.80E-12
GO:0050662	F	coenzyme binding	3.50E-13	1.30E-11
GO:0055044	C	symplast	6.10E-13	3.90E-11
GO:0009506	C	plasmodesma	6.10E-13	3.90E-11
GO:0015674	P	di-, tri-valent inorganic cation transport	8.50E-13	4.80E-11
GO:0005911	C	cell-cell junction	2.00E-12	1.10E-10
GO:0046483	P	heterocycle metabolic process	3.60E-12	2.00E-10
GO:0006820	P	anion transport	3.70E-12	2.00E-10
GO:0030054	C	cell junction	5.90E-12	2.90E-10
GO:0005576	C	extracellular region	6.80E-12	3.10E-10
GO:0010467	P	gene expression	7.30E-12	4.00E-10
GO:0046394	P	carboxylic acid biosynthetic process	8.10E-12	4.30E-10
GO:0016053	P	organic acid biosynthetic process	8.10E-12	4.30E-10
GO:0006633	P	fatty acid biosynthetic process	2.10E-11	1.10E-09
GO:0016747	F	transferase activity, transferring acyl groups other than amino-acyl groups	3.30E-11	1.20E-09
GO:0015297	F	antiporter activity	3.30E-11	1.20E-09
GO:0044265	P	cellular macromolecule catabolic process	3.00E-11	1.60E-09
GO:0030243	P	cellulose metabolic process	7.50E-11	3.90E-09
GO:0004175	F	endopeptidase activity	1.90E-10	6.60E-09

GO:0016757	F	transferase activity, transferring glycosyl groups	2.70E-10	8.90E-09
GO:0004091	F	carboxylesterase activity	3.50E-10	1.10E-08
GO:0044042	P	glucan metabolic process	2.50E-10	1.30E-08
GO:0004553	F	hydrolase activity, hydrolyzing O-glycosyl compounds	3.90E-10	1.30E-08
GO:0019842	F	vitamin binding	8.30E-10	2.60E-08
GO:0010167	P	response to nitrate	6.40E-10	3.20E-08
GO:0080135	P	regulation of cellular response to stress	6.90E-10	3.50E-08

C, Cellular component; F, Molecular function; P, Biological process

Table II-7. Number of genes having H3K4me3 or H3K36me3 in both *B. rapa* and *A. thaliana*

	ChIP-chip (Chr. 4) 10-day seedlings (Col) Roudier et al. (2011)			ChIP-seq aerial tissue of 2-week-old (Col) Luo et al. (2013)			ChIP-seq Leaf (L*er*) Engelhorn et al. (2017)		
H3K4me3	1,196 / 2,136	56.0%	a	9,003 / 16,123	55.8%	a	7,272 / 14,205	51.2%	b
H3K36me3	1,202 / 2,432	49.4%	b	N.D.			7,332 / 13,446	54.5%	a
H3K27me3	497 / 1,190	41.8%	c	3,334 / 7,026	47.5%	b	3,054 / 6,309	48.4%	c

Different letters indicate significant difference (Fisher's exact test, $P < 10^{-5}$)

Chapter III: High parent specific inheritance of the H3K4me3-, H3K9me2-, H3K27me3- and H3K36me3-marks in a heterotic hybrid of Chinese cabbage

Abstract

Heterosis, superiority of the F_1 hybrid compared with its parental lines, is important for the crop improvement. The molecular mechanism of heterosis involves the different chromatin marks that might be inherited from the parental lines. Chinese cabbage (*B. rapa* var. *pekinensis*) is an economically important leafy vegetable. We analyzed the heterotic phenotype of an F_1 hybrid and its parental lines at early- and harvesting stages of Chinese cabbage. F_1 hybrid showed a best parent heterosis (BPH) on its 4-days cotyledon and that was maintained thereafter in cotyledon (BPH, 18.2-38%), first- and second-leaves (BPH, 29.5-56.7%), final yield (76.9% BPH for harvested weight, and 46.2% BPH for total biomass). In this study, we analyzed H3K4me3-, H3K9me2-, H3K27me3- and H3K36me3-marks through the genome of a heterotic F_1 hybrid and its parental lines (RJKB-T24, female parent and RJKB-T23, male parent) of Chinese cabbage by chromatin immunoprecipitation (ChIP) sequencing. We identified the differentially modified histone marked genes (DMGs) for the all four histone marks and we found a differentiation of between parental lines. A small number of genes showed heterotic F_1 hybrid specific histone modification in Chinese cabbage. The most DMGs of all four histone marks between F_1 hybrid and each parental line showed high-parent pattern, while a small subset of DMGs showed intermediate or low parent pattern. High parent pattern of the histone marks might be associated with the heterotic growth and development of Chinese cabbage.

Keywords: Chinese cabbage; heterosis; hybrid; chromatin immunoprecipitation; histone marks; and epigenetic inheritance

Introduction

The tendency of the superiority for growth, biomass, and resistance of the heterozygous F_1 hybrids compared to the homozygous parental lines are known as heterosis or hybrid vigor (Lippmann and Zamir 2007, Itabashi et al. 2018). It has great agricultural importance to increase the production in crop plants (Duvick 2001, Schnable and Springer 2013). It is considered multi-genetic traits that are involved in controlling various complex traits (Baranwal et al. 2012). The genetic mechanism of heterosis is still not clear though it has been discovered over a century (Shull 1908). The classical genetic models such as dominance, overdominance, and epistasis are used to explain the heterosis (Lippmann and Zamir 2007, Fujimoto et al. 2018, Itabashi et al. 2018, Lv et al. 2020). Recently, quantitative trait loci (QTL) analysis is applied to explain the genetic basis of heterosis in various plants (Lv et al. 2020, Fujimoto et al. 2018, Yu et al. 2021) while transcriptomes and proteomes analyses are also used to explain molecular mechanisms of heterosis in plants (Fujimoto et al. 2018, Lv et al. 2020, Wu et al. 2021, Yu et al. 2021). Parental genetic distance is also considered to be a predictor for the heterosis in the F_1 hybrid (Itabashi et al. 2018, Wu et al. 2021). Studies in *A. thaliana* and the *B. rapa* did not show any strong positive relationship between genetic distance of parental lines and heterosis (Barth et al. 2003, Meyer et al. 2004, Kawamura et al. 2016). The epigenome of an intraspecific hybrid can provide high levels of heterosis in any complex traits even a high level of genetic similarity with its parental lines. Epigenetic mechanisms can substantially contribute to explaining the molecular mechanism of heterosis (Kawanabe et al. 2016a). Heterosis studies in model plants suggested that heterosis levels can be influenced by changes in gene expression through different epigenetic regulations such as DNA methylation, histone modification, and non-coding RNA (Greaves et al. 2015, Ryder et al. 2019).

Histone modifications can influence the plant biomass heterosis by affecting the transcriptional activity of genes to change the phenotype (Yu et al. 2021). Biological process such as starch biosynthesis and the growth rate has been controlled by the circadian clock in plants (Shen et al 2015, Yu et al. 2021), and when plants match the environment with their internal circadian rhythm then more vigorous plant growth occurs (Ko et al. 2016, Kim AJ et al. 2017, Yu et al. 2021). Histone modification can interplay with the circadian rhythm to change its transcription (Chen and Mas, 2019). The alteration of H3K9Ac and H3K4Me2 marks in F_1 hybrid compared with parental lines in the promoter region of circadian clock genes (*CCA1* and *LHY*) changes the transcriptional response in *A. thaliana* (Ni et al. 2009) which led to the heterotic growth in F_1 hybrids (Miller et al. 2012, Shen et al. 2012, Chen 2013). In the F_1 hybrid of *A. thaliana*, the *CCA1* gene has been found rhythmic histone activation and repression at

different time points of the day under the pathogen attack that leads to the heterosis for the disease resistance (Yang et al. 2021). Global patterns of H3K4me3, H3K9ac, and H3K27me3 marks in two subspecies and their subsequent F_1 hybrid of rice by ChIP-seq were identified where H3K4me3 or H3K27me3 marks inversely expressed between hybrids and parents (He et al. 2010). Global pattern of histone marks in F_1 hybrids might be similar to their parents throughout the genome while it might be altered at localized regions of the genome (Moghaddam et al. 2011, Dong et al. 2012, Zhu A et al. 2017, Zhu W et al. 2017). In Arabidopsis, genome-wide distribution of the H3K4me3, H3K9ac, H3K27me3, H3K9me2 marks by ChIP-seq showed hundreds of genes having altered histone modification in hybrids while some of the genes including floral repressor *FLC* showed a consistent change in histone modification and gene expression. (Zhu A et al. 2017). Alteration of the gene expression in F_1 hybrids has been observed by the histone modifications compared with its parental lines in some plants such as rice, maize, Arabidopsis (Greaves et al. 2015, Wu et al. 2021).

Most of the researchers mainly focused on the final harvesting stage for the heterosis study of the crop plants. Seedling establishment is one of the most important factors for crop development which ultimately affect the final establishment and crop yield. Plants including *B. rapa* shows heterotic growth at the very early developmental stages where F_1 hybrids show increased cotyledon size, leaf size, rosette diameter, and biomass compared with parental lines (Basunanda et al. 2010, Ma et al. 2011, Fujimoto et al. 2012b, Meyer et al. 2012, Groszmann et al. 2014, Ko et al. 2016, Yang et al., 2017, Wang L et al. 2019, Li et al. 2021). In contrast, little is known about the roles of histone modification in heterosis, and studies in Arabidopsis and rice suggest that histone modification possibly regulates some features of heterosis. The whole-genome sequence of *B. rapa* is available (Wang et al. 2011) while the ChIP-seq offers an opportunity for histone marks profiling. Though few studies have been documented the different histone modifications states independently in *B. rapa* inbred lines by ChIP analysis for specific treatment conditions (Kawanabe et al. 2016, Shen et al. 2019). However, there is not any comparative study between parents and their F_1 hybrids in *B. rapa* which is not even enough for the model plants. In our previous study, we have analyzed the genome-wide distribution of the active histone marks (H3K4me3 and H3K36me3) (Chapter II) and repressive histone marks (H3K27me3) in the parental lines (Akter et al. 2019). In this study, we have compared the genome-wide H3K4me3, HK9me2, H3K27me3, and H3K36me3 marks between parents and their F_1 hybrids in *B. rapa* to understand histone modifications states that might be inherited from parental lines to the heterotic F_1 hybrid.

Materials and methods

Plant materials and growth conditions

A Chinese cabbage F_1 hybrid was developed from the crossing between two inbred parental lines, RJKB-T24 (♀) and RJKB-T23 (♂). F_1 hybrid and its parental lines were used in this study. Phenotypes were evaluated at two different developmental stages. For the early developmental stage, seeds of F_1 and its parental lines were surface sterilized and grown on Murashige and Skoog (MS) agar plates with 1% (w/v) sucrose under long day (LD) condition (16h light / 8h dark) at 21˚C. On the 7th day, seedlings were transferred into soil and grow under long day (LD) conditions (16h light / 8h dark) at 24˚C. The area and cell number of the cotyledon at 2-, 4-, and 6-days cotyledon and 10-,12-, and 14-days first and second leaves were respectively measured and counted by ImageJ (Schneider et al. 2012). Plants for the F_1 hybrid and its parental lines were grown in an open field for the harvesting stage. Phenotypic parameters such as plant height, circumference, harvested biomass, and total biomass were measured for the confirmation of the heterosis. Plant height is the length, and circumference is the perimeter of the headed Chinese cabbage without external outer leaves. Harvested biomass is the weight of the leafy head without external leaves and total biomass is the weight of the leafy head with external leaves. Mid-parent heterosis (MPH) and better-parent heterosis (BPH) were calculated and expressed in percentage using the equation $MPH = (((F_1 - MP)/MP) \times 100)$ and $BPH = (((F_1 - BP)/BP) \times 100)$ respectively, where F_1 represents the value of F_1 hybrid, MP is the mean value of two parental lines, and BP is the value of the better-performing parental line.

Genome-wide histone marks detections by ChIP-seq

One gram of first and second leaves of the female line (RJKB-T24; T24), male line (RJKB-T23; T23) and F_1 hybrid of Chinese cabbage (*B. rapa*) were collected from 14-days old plant for ChIP analyses. Anti-H3K4me3 (Millipore, 07-473), anti-H3K9me2 (ACTIVE MOTIF, 39753), anti-H3K27me3 (Millipore, 07-449), and H3K36me3 (Abcam, ab9050) antibodies were used. A detail of the ChIP experimental and data analysis procedures described in the materials and methods section of Chapter II.

Results

F_1 hybrid shows heterosis at early developmental stage and harvesting stage

The Chinese cabbage F_1 hybrid showed strong heterosis for cotyledon and leaf size during the early developmental stage and harvesting stage compared with the parental lines (Figure III-1, Figure III-2). The cotyledon size of the F_1 hybrid was statistically identical with the RJKB-T23 but statistically different from T24 at 2 days after sowing (DAS), however, the F_1 hybrid showed significantly larger cotyledon size than its parental lines at 4 DAS that was maintained at 6 DAS (Figure III-1A). The first and second leaf sizes of the F_1 hybrid showed increased leaf area compared to the parental lines at 10, 12, and 14 DAS (Figure III-1B). The cell numbers per 250 μm^2 area of cotyledon were not varied significantly among the F_1 hybrid and parental lines at 2, 4, and 6 DAS while the cell numbers of F_1 hybrid was statistically identical with one of the parental lines at 10, 12, and 14 DAS of the first and second leaf (Figure III-2). The results suggested that an increase of cotyledon size and first and second leaf size in F_1 hybrid were due to the increase of cell size. At the final harvest stage, we measured the plant height, circumference, harvested weight, and total biomass. F_1 hybrid showed a significantly increased level of plant height, circumference, harvested weight, and total biomass compared to the parental lines (Figure III-3).

MPH and BPH were calculated for the cotyledon size, first and second leaf size, plant height, circumference, harvested weight, and total biomass (Figure III-4). We did not observe best parent heterosis for the cotyledon size at 2 DAS, however, both best parent and mid parent heterosis were first observed in cotyledon size at 4 DAS that was maintained in cotyledon size at 6 DAS (Figure III-4A). In the first and second leaf sizes, we found both the mid parent and best parent heterosis at 10, 12, and 14 DAS (Figure III-4A). At the harvesting stage, we found strong positive mid-parent heterosis and best parent heterosis for the harvested weight (MPH- 68.1%, BPH- 76.9%) and total biomass (MPH- 48.4%, BPH- 46.2%) (Figure III-4B). These results suggested the F_1 hybrid had strong heterosis at the final harvest stage that was obvious at the early developmental stage.

Identification of differentially histone modified genes between parental lines and their F_1 hybrid

Our ChIP-sequencing of the parental lines and F_1 hybrids of Chinese cabbage generated a genome-wide landscape for the H3K4me3, H3K9me2, H3K27me3, and H3K36me3 marks. We identified the differentially modified genes (DMGs) for each of the studied histone

modifications. We used false discovery rate (FDR) correction at a 5% level to identify DMGs between female parent (T24), male parent (T23), and F_1 hybrid for each type of histone modifications. A total of 1836 (T24vsT23), 505 (F_1vsT24), and 359 (F_1vsT23) DMGs for H3K4me3; 1524 (T24vsT23), 340 (F_1vsT24), and 300 (F_1vsT23) DMGs for H3K9me2; 5058 (T24vsT23), 543 (F_1vsT24), and 969 (F_1vsT23) DMGs for H3K27me3; 2452 (T24vsT23), 411 (F_1vsT24), and 633 (F_1vsT23) DMGs for H3K36me2 were identified (Figure III-5A).

For the H3K4me3, 802 of 1836 (43.7%) DMGs showed higher H3K4me3 level in T24 while 1034 (56.3%) were in T23. 486 of 505 (96.2%) DMGs had higher H3K4me3 level in F_1 than in T24, and 340 of 359 (94.7%) DMGs had higher H3K4me3 level in F_1 than in T23 (Figure III-5B and Figure III-5C).

For the H3K9me2, 625 of 1524 (41.0%) DMGs showed higher H3K9me2 level in T24 while 899 (59.0%) genes were in T23. 298 of 340 (87.6%) DMGs had higher H3K9me2 level in F_1 than in T24, and 263 of 300 (87.7%) DMGs had higher H3K9me2 level in F_1 compared than in T23 (Figure III-5B and Figure III-5C).

For the H3K27me3, 3173 of 5058 (62.7%) DMGs were enriched in T24 while 1885 (37.3%) were enriched in T23. 521 of 543 (95.9%) DMGs had higher H3K27me3 level in F_1 than in T24, and 768 of 969 (79.3%) DMGs had higher H3K27me3 level in F_1 than in T23 (Figure III-5B and Figure III-5C).

For the H3K36me3, 1483 of 2452 (60.5%) DMGs were enriched in T24 while 969 (39.5%) were enriched in T23, 403 of 411 (98.1%) DMGs had higher H3K36me3 level in F_1 than in T24, 606 of 633 (95.5%) DMGs had higher H3K36me3 level in F_1 than in T23 (Figure III-5B and Figure III-5C).

In the comparison between parental lines, female parents (T24) showed a higher level in H3K27me3 and H3K36me3 marks than in T23, while male parents showed a higher level in H3K4me3 and H3K9me2 marks than in T24 (Figure III-5C). The comparison of the F_1 hybrid with parental lines showed that most of the DMGs showed higher level of all four histone marks in the F_1 hybrid (around 87~98% in F_1vsT24, and around 87~95% in F_1vsT23) (Figure III-5C).

For all four histone marks, we constructed the Venn-diagram showing using DMGs of three combinations (T24vsT23, F_1vsT24, and F_1vsT23) (Figure III-6). A small number of DMGs were overlapped between three combinations and the number of the overlapped DMGs were 16, 22, 47, and 29 for the H3K4me3, H3K9me2, H3K27me3, and H3K36me3 marks, respectively (Figure III-6). A major number of DMGs between F_1 and each parental line overlapped with

DMGs between parental lines (Figure III-6). A few DMGs between F_1 and each parental line were not different histone modification between parental lines, and the number of each histone mark was 8, 68, 72, and 42 for the H3K4me3, H3K9me2, H3K27me3, and H3K36me3 marks, respectively (Figure III-6).

High parent-specific inheritance of the differentially modified histone marks in F_1 hybrid

We classified the DMGs in F_1 hybrid into eight classes in total, and these were one class as overdominance (OD, Class I), two classes as the high parent (HP, Class II and III), two classes as the intermediate in parents (IP; Class IV and V), two classes as the low parent (LP, Class VI and VII), and one class underdominance (UD, Class VIII) using the enrichment levels in parental lines and their F_1 hybrid (Figure III-7). Genes that were not fit within the mentioned eight classes were categorized into 'Unclassified'. We did not find any genes under the underdominance class for all four histone marks. We did not find overdominance for the H3K4me3, H3K9me2, and H3K27me3 marks, however, we found only 12 overdominance genes for H3K36me3 marks (Figure III-8A). For all four histone marks, a low number of genes were in Class IV to Class VII while most of the genes were in Class II and Class III (Figure III-8A).

For H3K4me3, 804 of 864 DMGs between F_1vsT24 and F_1vsT23 were high parent that was 2.2% of the total genes (Figure III-8A and Figure III-8B). For H3K9me2, 519 of 640 between F_1vsT24 and F_1vsT23 were high parent that was 1.3% of the total genes (Figure III-8A and Figure III-8B). For H3K27me3, 1239 of 1512 DMGs between F_1vsT24 and F_1vsT23 were high parent that was 3.1% of the total genes (Figure III-8A and Figure III-8B). For H3K36me3, 907 of 1144 DMGs between F_1vsT24 and F_1vsT23were high parent that was 2.3% of the total genes (Figure III-8A and Figure III-8B). These results suggest the high parent was major class in H3K4me3, H3K9me2, H3K27me3, and H3K36me3 marks.

Discussion

Heterosis at the early developmental stage of Chinese cabbage

The interspecific hybrid of Chinese cabbage showed strong heterosis at their early developmental stage and final harvest. At the early developmental stage, larger cotyledon area and leaf area of Chinese cabbage F_1 hybrid compared to parental lines instead of the identical cell numbers in per unit area suggests that the leaf or cotyledon size of genetic material used in this study has the increased cell size. Heterosis can result from the increasing either of cell number or cell size or both (Fujimoto et al. 2012b, Groszmann et al. 2014). Heterosis at the early developmental stage has a significant influence to lead the heterotic growth, development, and yield of the plant. It is the fundamental knowledge that an increase of either cell number or cell size is important for the generations of more photosynthetic area, and the more photosynthetic area will lead to increase plant size and biomass. Cellular organ-specific yield heterosis has also been confirmed in some plants (Flint-Garcia et al. 2009, Shi et al. 2011). The larger cotyledon and leaves of F_1 hybrid at the early developmental stage are expected that facilitate more carbon fixation by photosynthesis in F_1 hybrid compared to the parental lines (Fujimoto et al. 2012b, Meyer et al. 2012, Groszmann et al. 2014) which might influence the heterosis in yield traits of the F_1 hybrid. Like our result, obvious heterosis was also observed in the F_1 hybrid of Chinese cabbage at the early developmental stage (Saeki et al. 2016, Li et al. 2021).

Inheritance of the differentially modified histone marks from parental alleles into the F_1 hybrid

The molecular mechanism of heterosis is a mystery and it cannot be explained by only genetic mechanisms. The phenotypic variation can be occurred by both genetic and epigenetic variation (Schmitz and Ecker 2012, Turck and Coupland 2014). The overall histone modifications in F_1 hybrid might be equal to the average levels of the parental lines (He et al. 2010; Moghaddam et al. 2011; Yang M et al. 2016). Some researchers studied genetic mechanisms of heterosis in Chinese cabbage, however, the inheritance pattern of histone marks into the F_1 hybrid is still not studied. Histone marks are known as the epigenetic regulators; however, many histone marks are heritable (Bonasio et al. 2010). In this study, we analyzed the distribution of H3K4me3, HeK9me2, H3K27me3 and H3K36me3 marks in F_1 hybrid compared with parental lines to understand the epigenetic changes in F_1. We have identified the DMGs by the comparative analysis of the ChIP-seq data of the parental lines and its F_1 hybrid. Parental alleles showed variation for all four histone marks, and the variations for the over-presentation of these histone marks in genes could be associated with the natural phenotypic variation.

Thereafter we have identified the over-presented histone marks in the F_1 hybrids inherited from the parental alleles. In Arabidopsis, numbers of genes were found in altered levels of histone modifications at the localized genomic regions (Moghaddam et al. 2011, Dong et al. 2012, Zhu A et al. 2017, Zhu W et al. 2017). More numbers of the H3K27me3- and H3K36me-marked genes were over-presented in the female allele while more numbers of the H3K4me3- and H3K9me2-marked genes were over-presented in male allele. Excluding the H3K27me3-marked genes the variation for the number of genes between parental alleles were not noticeable. Less than half of the H3K4me3-, H3K9me2- and H3K36me3-marked DMGs while less than one third of the H3K27me3-marked showed additive in F_1 hybrid. Alteration of the histone marks in a heterotic F_1 hybrid compared to it parental lines can changes the gene expression in F_1 hybrids (Greaves et al. 2015, Wu et al. 2021), thus we should check whether additive histone modification is associated with transcriptional level in F_1 hybrids.

High parental allele specific inheritance of the differentially modified histone marks into F_1 hybrid

We categorized the inherited DMGs in F_1 hybrid into overdominance, high parent specific, intermediate between parents', and low parent specific modifications for all four histone marks. Our allele specific inheritance of the over-presented histone marked gene analysis resulted that all four marks in F_1 hybrid were inherited irrespective to the parental alleles. We did not find any noticeable number of genes under the overdominance, intermediate between parents' and low parent specific modifications in F_1 hybrid. Most of the inherited genes were belonged to the high parent specific histone modification for all four histone marks. In *A. thaliana*, histone modifications states in hybrid are independent on the difference for the modifications between the parents at the early developmental stage (Zhu A et al. 2017). Our results suggest the modifications of H3K4me3, HeK9me2, H3K27me3 and H3K36me3 marks are not dependent on a specific parent in Chinese cabbage, and we found that most of the inherited histone marked genes were high parent specific for all four histone marks and that was ranged about 80-95% of the total inherited DMGs. Allele-specific alteration of the H3K4me3, H3K27me3 and H3K36me3 marks had also been identified in rice and Arabidopsis (He et al. 2010, Guo et al. 2015, Moreno-Romero et al. 2016). In a reciprocal hybrid of *A. thaliana*, a small subset of the inherited H3K4me3-, H3K9me2- and H3K27me3-marked DMGs has been correlated with the gene expression levels (Zhu A et al. 2017). The allele specific changes of the H3K36me3 marks were significantly regulate the allele specific gene expression in heterotic hybrid rice (Guo et al. 2015). However, we did not co-investigate transcriptome analysis in this study, therefore, regulation of the gene expression might be or might not be associated with the

high parent specific alteration of the H3K4me3, H3K9me2, H3K27me3, and H3K36me3 marks in the F_1 hybrid. A further investigation is need to know the association of these inherited DMGs with transcriptional response, and it will help us to elucidate the role of inherited H3K4me3-, H3K9me2-, H3K27me3-, and H3K36me3-marked genes for the epigenetic regulation in heterosis in Chinese cabbage.

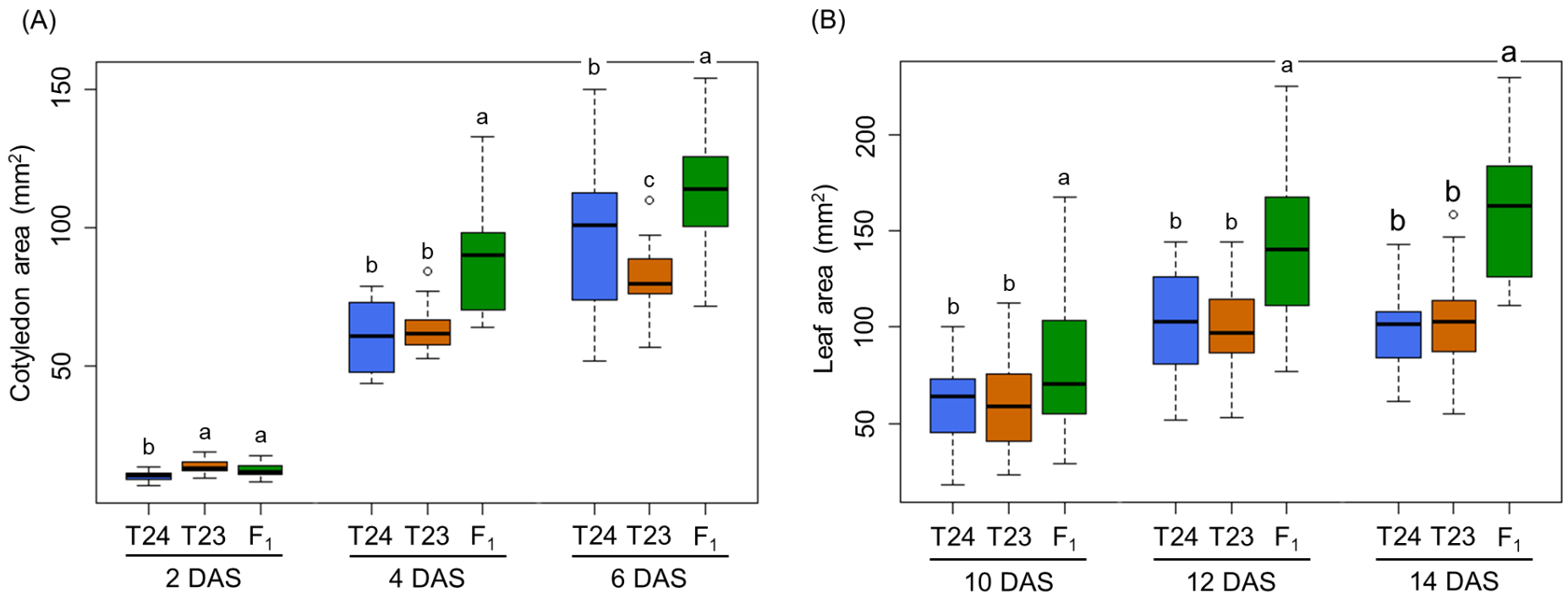


Figure III-1. Cotyledon area (A) and first and second leaf area (B) of the parental lines and F_1 hybrid at different days after sowing (DAS). "T24" is the female line (RJKB-T24) and "T23" is the male line (RJKB-T23). Boxplots having similar letter are the statistically identical at Tukey's HSD test ($p < 0.01$).

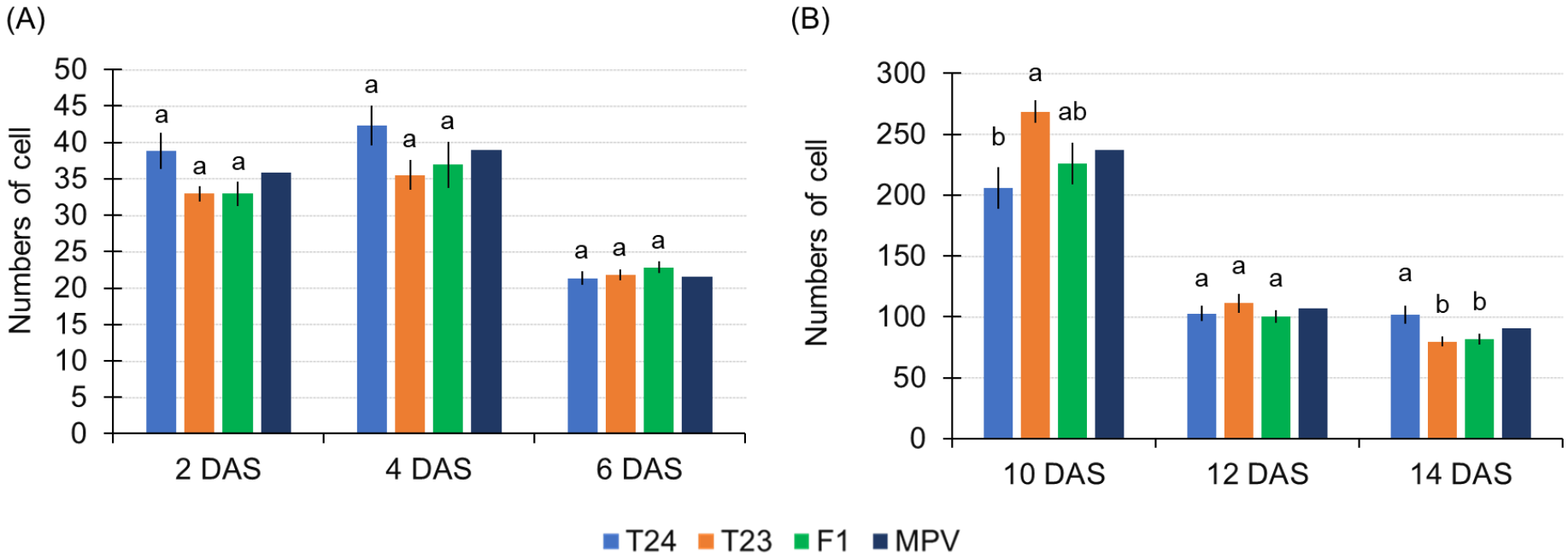


Figure III-2. Numbers of cell per unit area in cotyledon (A), and first and second leaf (B) of the parental lines and F_1 hybrid at different days after sowing (DAS). Bars having similar letter are the statistically identical at Tukey's HSD test ($p < 0.05$). "T24" is the female line (RJKB-T24) and "T23" is the male line (RJKB-T23).

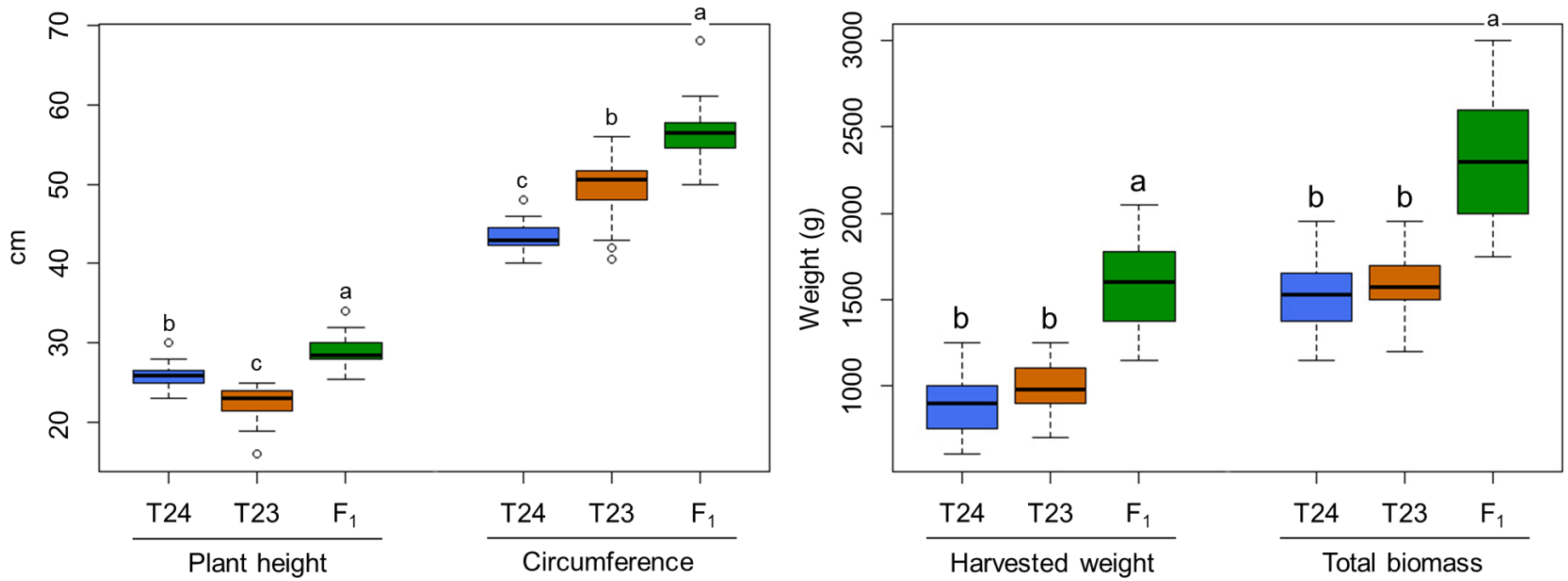


Figure III-3. Plant growth at harvesting stage (A) and yield (B) of Chinese cabbage of the parental lines and F_1 hybrid. "T24" is the female line (RJKB-T24) and "T23" is the male line (RJKB-T23). Boxplots having similar letter are the statistically identical at Tukey's HSD test ($p < 0.01$).

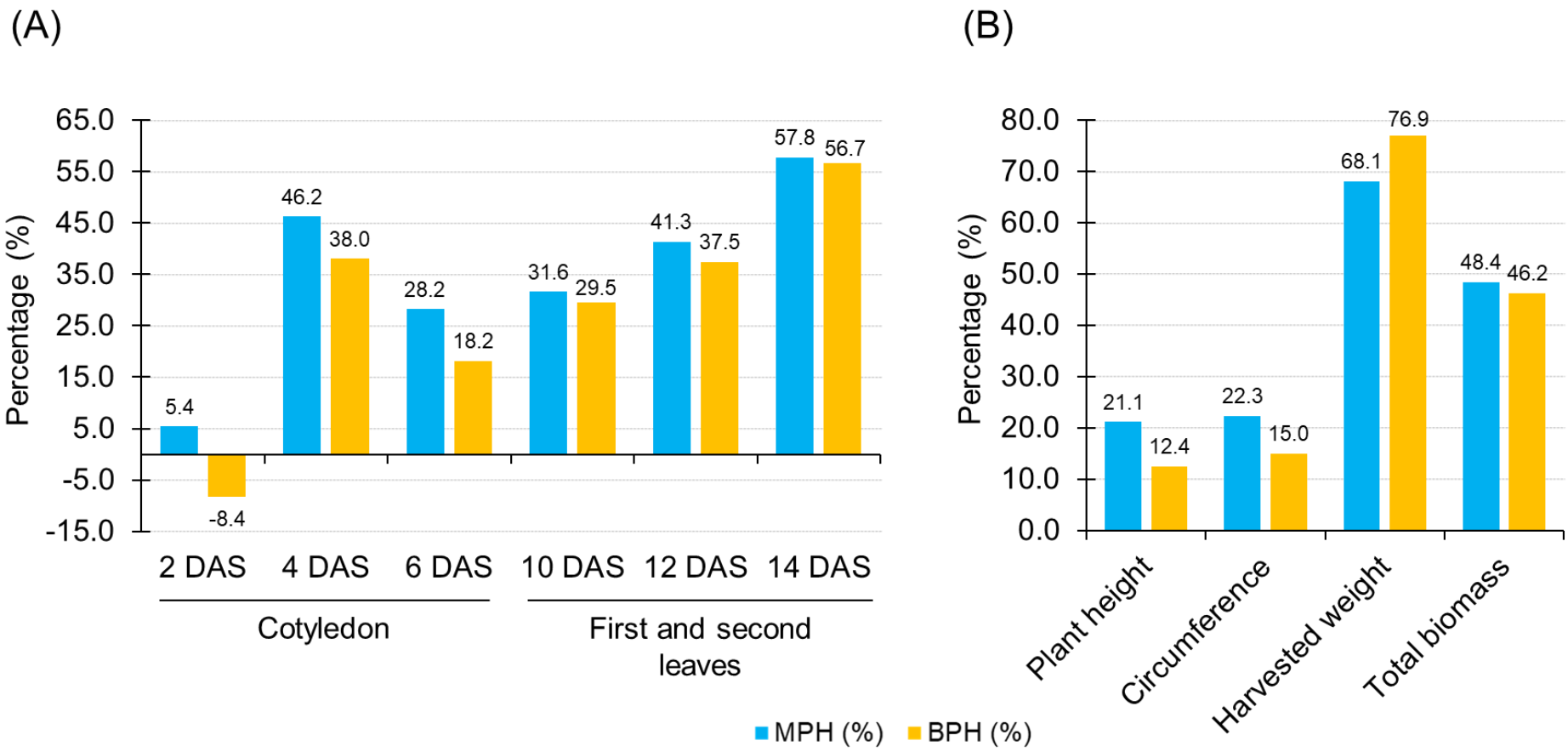


Figure III-4. Heterosis levels for the vegetative growth at early development stage (A) and yield characters (B). MPH is the mid parent heterosis and BPH is the best parent heterosis.

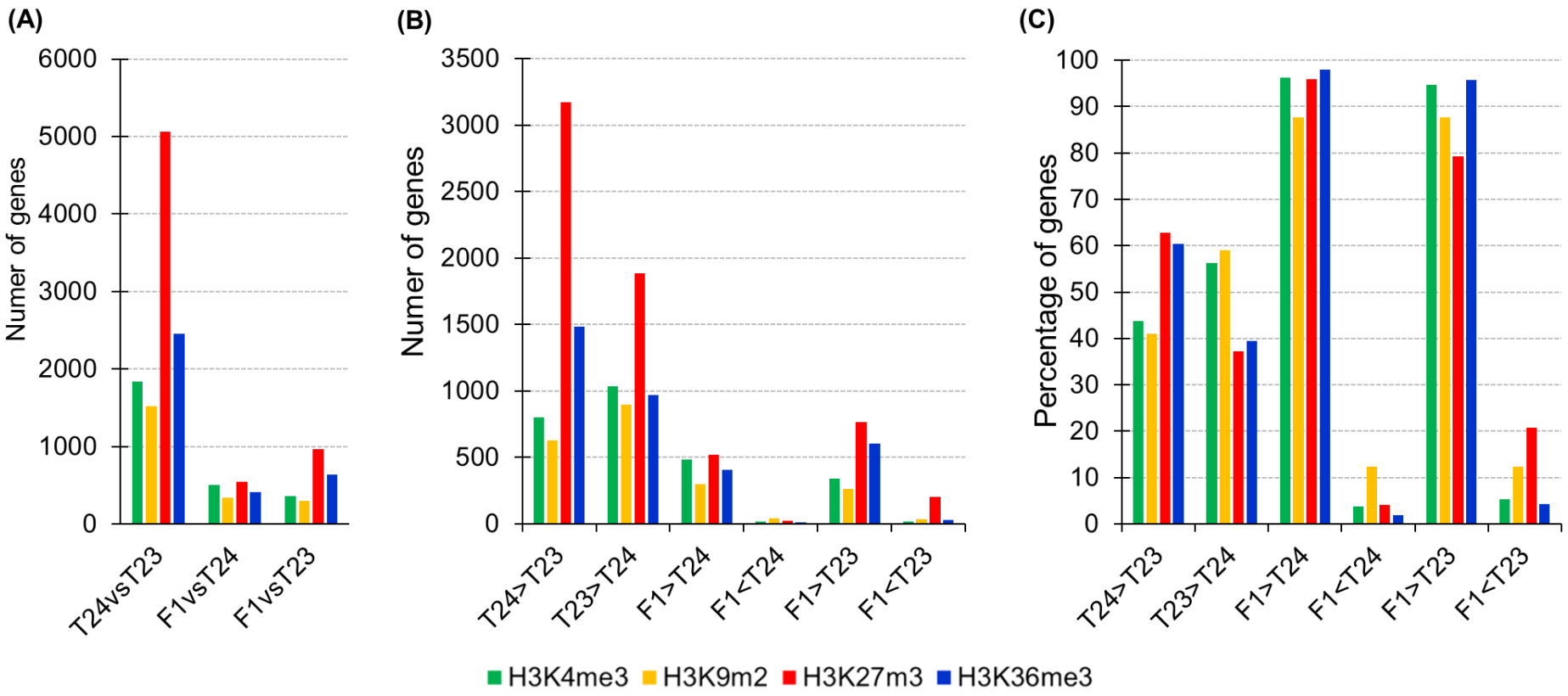


Figure III-5. Variation of the differentially modified histone marked genes between parental alleles and F_1 hybrid (A). The number (B) and percentage (C) of the over-presented histone marked gene in parental lines and F_1 hybrid. "T24" represents the female parent (RJKB-T24) and "T23" represents the male parent (RJKB-T23).

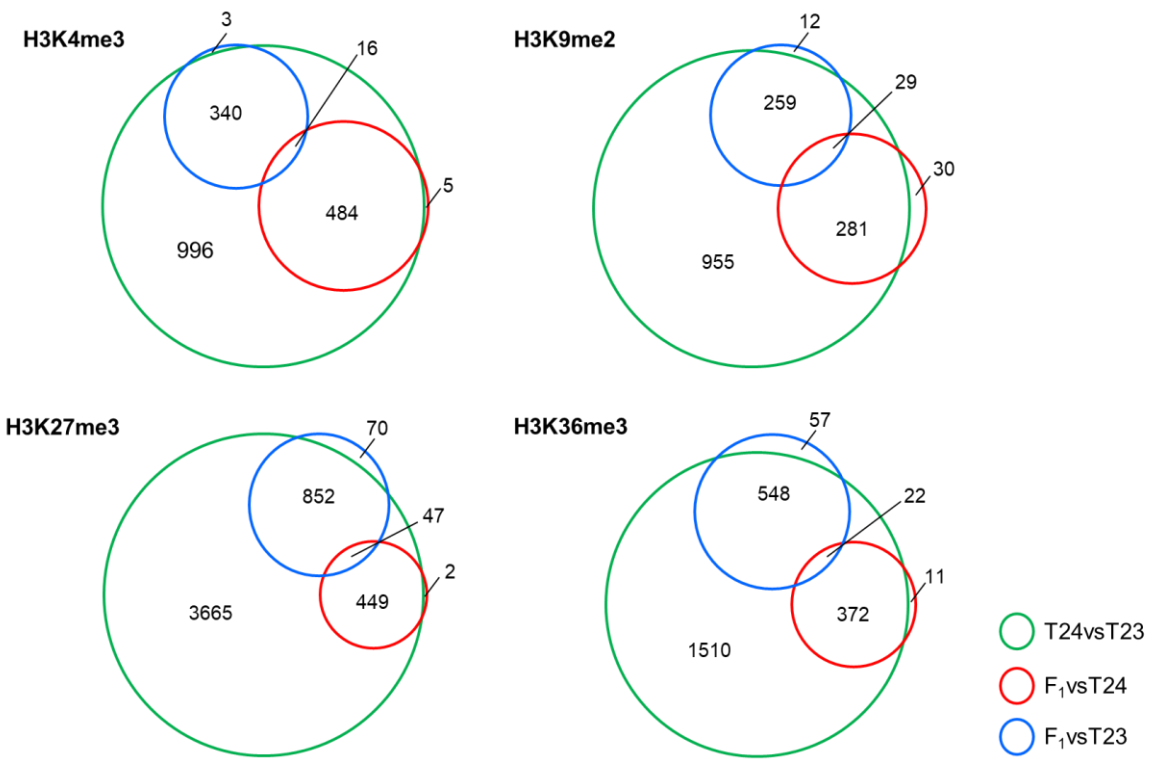


Figure III-6. Venn diagrams between differentially modified histone marked genes (DMGs) between parental lines and F_1 hybrid. Upper left panel represents for the H3K4me3 marked genes, upper right panel represents for the H3K36me3 marked genes, lower left panel represents for the H3K27me3 marked genes, and lower right panel represents for the H3K9me2 marked genes. "T24" represents the female parent (RJKB-T24) and "T23" represents the male parent (RJKB-T23).

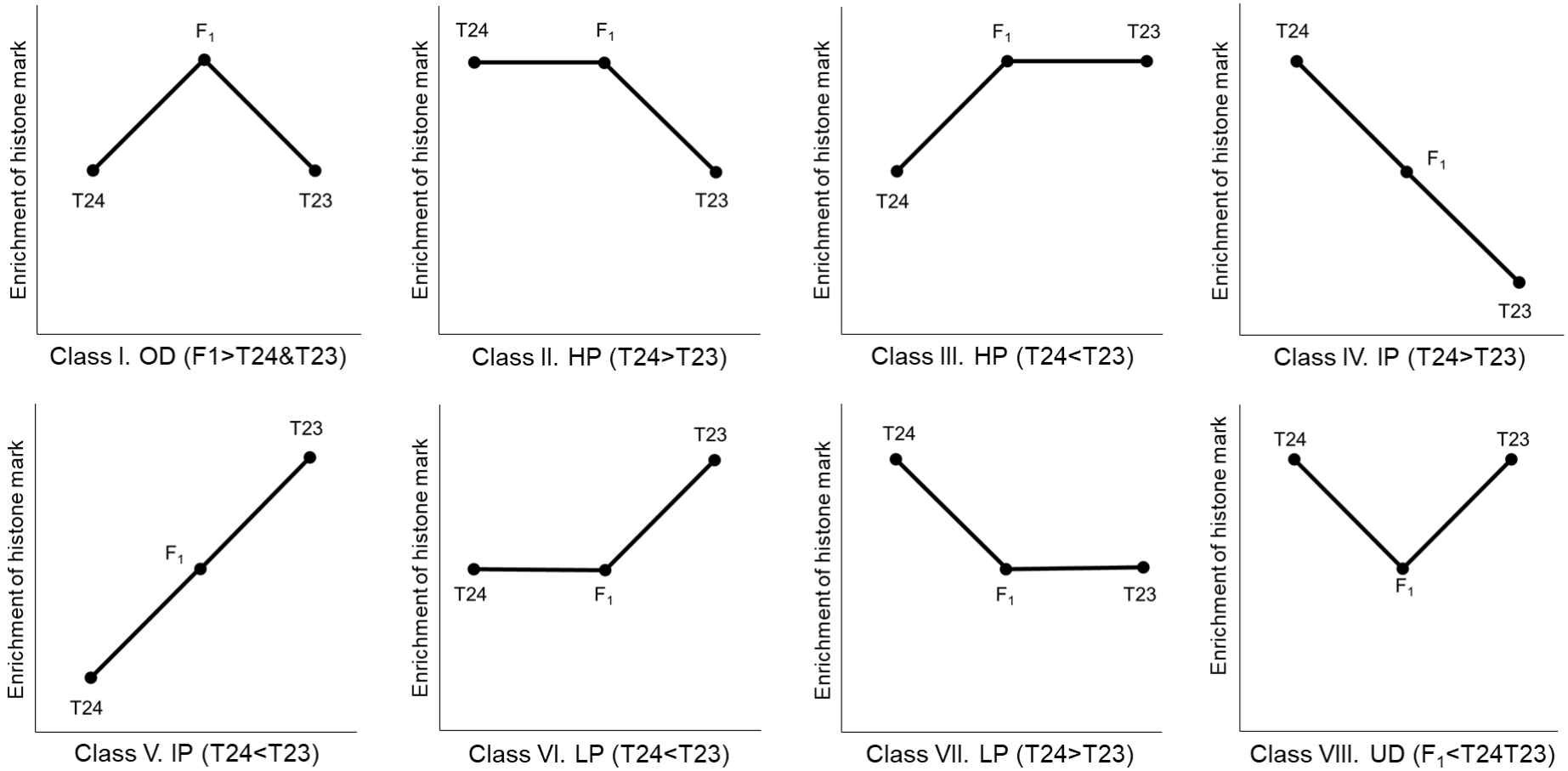


Figure III-7. Enrichment patterns for the differentially modified histone marked genes between parental lines and F_1 hybrid. “OD”, “HP”, “IP”, “LP” and, “UD” respectively represents the “over-dominance”, “high parent”, “intermediate between parents”, “low parent”, and “underdominance”. “T24” represents the female parent (RJKB-T24) and “T23” represents the male parent (RJKB-T23).

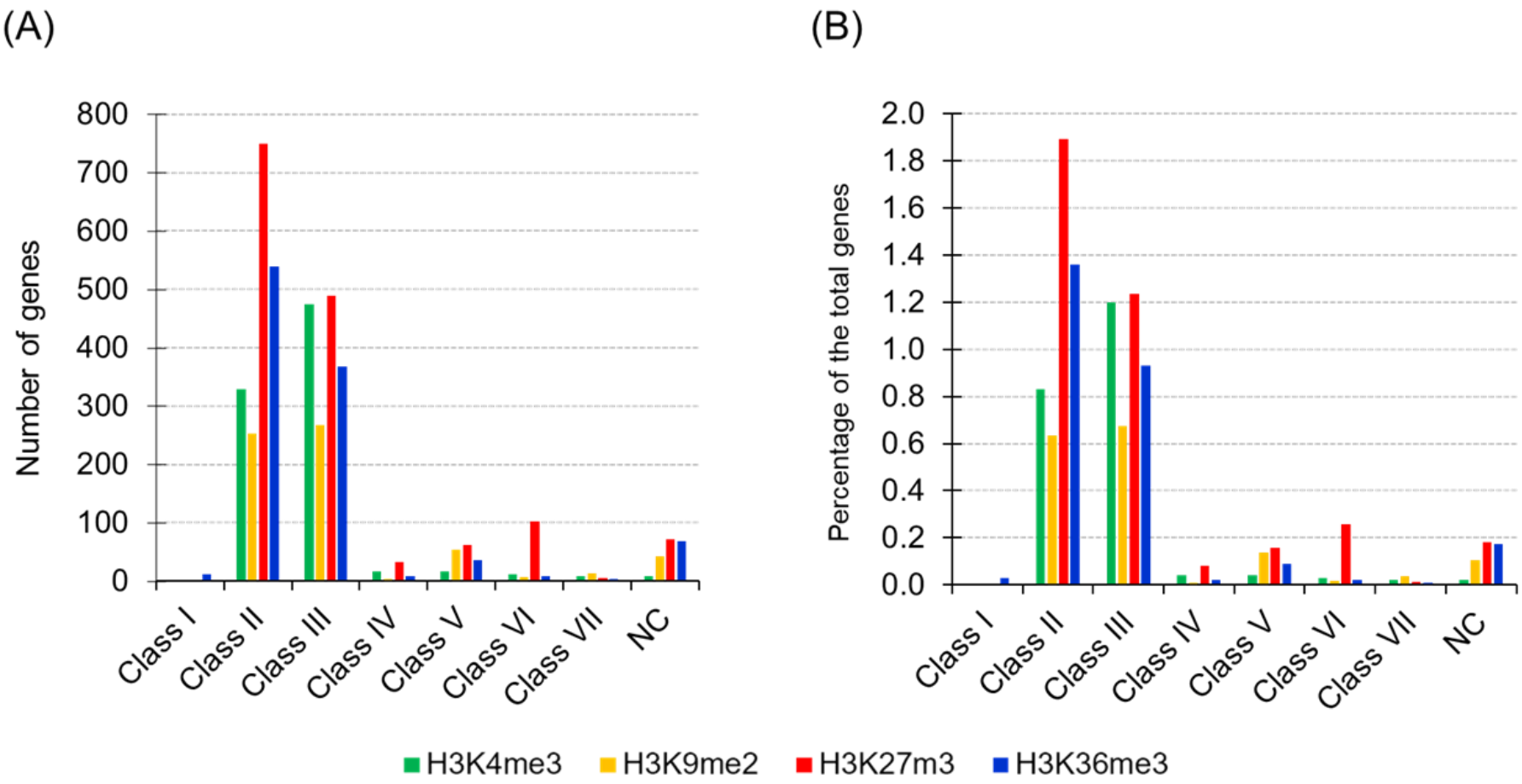


Figure III-8. Classification of the DMGs in F_1 hybrids. The number of genes (A) and the percentage of the total genes (B) under each class for each type of histone marks.

Chapter IV

Genome-wide analysis of long noncoding RNAs, 24-nt siRNAs, DNA methylation and H3K27me3 marks in *Brassica rapa*

Abstract

Long noncoding RNAs (lncRNAs) are RNA fragments that generally do not code for a protein but are involved in epigenetic gene regulation. In this study, lncRNAs of *Brassica rapa* were classified into long intergenic noncoding RNAs, natural antisense RNAs, and intronic noncoding RNAs and their expression analyzed in relation to genome-wide 24-nt small interfering RNAs (siRNAs), DNA methylation, and histone H3 lysine 27 trimethylation marks (H3K27me3). More than 65% of the lncRNAs analyzed consisted of one exon, and more than 55% overlapped with inverted repeat regions (IRRs). Overlap of lncRNAs with IRRs or genomic regions encoding for 24-nt siRNAs resulted in increased DNA methylation levels when both were present. LncRNA did not overlap greatly with H3K27me3 marks, but the expression level of intronic noncoding RNAs that did coincide with H3K27me3 marks was higher than without H3K27me3 marks. The Brassica genus comprises important vegetables and oil seed crops grown across the world. *B. rapa* is a diploid (AA genome) thought to be one of the ancestral species of both *B. juncea* (AABB genome) and *B. napus* (AACC) through genome merging (allotetrapolyploidization). Complex genome restructuring and epigenetic alterations are thought to be involved in these allotetrapolyploidization events. Comparison of lncRNAs between *B. rapa* and *B. nigra*, *B. oleracea*, *B. juncea*, and *B. napus* showed the highest conservation with *B. oleracea*. This study presents a comprehensive analysis of the epigenome structure of *B. rapa* at multi-epigenetic levels (siRNAs, DNA methylation, H3K27me3, and lncRNAs) and identified a suite of candidate lncRNAs that may be epigenetically regulated in the Brassica genus. Both H3K4me3 and H3K36me3 marks were enriched in chromatin regions encoding lncRNAs, especially around the transcription start site. The transcription level of long intergenic noncoding RNAs was positively associated with the level of H3K4me3 and H3K36me3, while this association was not observed in natural antisense RNAs (NATs) and intronic noncoding RNAs.

Keywords: LncRNAs, 24-nt siRNAs, H3K4me3, H3K36me3, H3K27me3, DNA methylation

Introduction

The Brassica genus comprises vegetable and oil seed crops. The “Triangle of U” proposed the genomic relationship among six major species of the Brassica genus. Three allotetraploid species, each of which contains two complete diploid genomes derived from two different parental species, *Brassica juncea* L. (AABB genome, 2n = 4x = 36), *B. napus* L. (AACC, 2n = 4x = 38), and *B. carinata* L. (BBCC, 2n = 4x = 34) are derived from the natural hybridization of the diploid species, *B. rapa* L. (AA, 2n = 2x = 20), *B. nigra* L. (BB, 2n = 2x = 16), and *B. oleracea* L. (CC, 2n = 2x = 18) (UN 1935). Some species in the Brassica genus show morphological divergence (termed morphotype). *B. rapa* includes leafy vegetables such as Chinese cabbage (var. *pekinensis*), pak choi (var. *chinensis*), and komatsuna (var. *perviridis*), root vegetables including turnip (var. *rapa*), and oilseed crops (var. *oleifera*) (Lv et al. 2020). The first whole genome sequence determined in the genus Brassica was that of *B. rapa* (Wang et al. 2011). Later, whole genome sequences of *B. oleracea*, *B. nigra*, *B. napus*, and *B. juncea* were determined (Chalhoub et al. 2014, Liu et al. 2014, Parkin et al. 2014, Yang J et al. 2016b).

Plant transcriptome analyses have revealed RNAs devoid of protein-coding potential, which are called noncoding RNAs (ncRNAs). Some ncRNAs contain exons, which potentially code for short proteins or peptides, however, experimental investigation is required to validate their functions. Two families of ncRNAs are known; long noncoding RNAs (lncRNAs) that are longer than 200 nucleotides (nt), and small RNAs (sRNAs) that are ~18–30 nt in length (Chekanova 2015, Liu J et al. 2015, Mattick and Rinn 2015, Karlik et al. 2019, Rai et al. 2019). LncRNAs are classified by their position and orientation of transcription; long intergenic noncoding RNAs (lincRNAs), intronic noncoding RNAs (incRNAs) derived from introns, and natural antisense transcripts (NATs) transcribed from the complementary DNA strand of their associated genes (Ponting et al. 2009, Chekanova 2015, Liu J et al. 2015, Mattick and Rinn 2015, Karlik et al. 2019, Rai et al. 2019). Recent studies have shown that lncRNAs play crucial roles in various physiological processes such as vernalization (Swiezewski et al. 2009, Heo and Sung 2011, Kim and Sung 2017), photoperiod-sensitive male sterility (Ding et al. 2012), red-light-mediated seedling photomorphogenesis (Wang Y et al. 2014), seed dormancy (Fedak et al. 2016), and the transcriptional regulation of plant innate immunity (Seo et al. 2017).

Epigenetics has been defined as ‘‘the study of changes in genome expression that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence” (Wu and Morris, 2001). DNA methylation and histone modification are well-known epigenetic modifications, and lncRNAs are considered to be involved in epigenetic regulation (Wang C et

al. 2017). In plants, DNA methylation is established through RNA-directed DNA methylation (RdDM) (Fujimoto et al. 2012a, Matzke et al. 2015). Plant-specific RNA POLYMERASE IV (Pol IV) and Pol V are involved in RdDM. Pol IV-transcribed ncRNAs are cleaved into 24-nucleotide small interfering RNAs (24-nt siRNAs) by DICER--LIKE 3 (DCL3), and 24-nt siRNAs are loaded into ARGONAUTE 4 (AGO4). Pol V-transcribed lncRNAs act as scaffold molecules of AGO-siRNA complexes, which recruit DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) for catalyzing *de novo* DNA methylation. In *Arabidopsis thaliana* two cold-induced lncRNAs (COLDAIR and COLDWRAP) regulate histone methylation by recruiting polycomb repressive complex 2 (PRC2), which catalyzes the tri-methylation of histone H3 lysine 27 (H3K27me3), to the chromatin region of *FLOWERING LOCUS C* (*FLC*) (Kim and Sung 2017, Kim DH et al. 2017, Itabashi et al. 2018, Liu et al. 2020). H3K27me3 also showed an association with transcriptional repression of genes and a role in tissue- or developmental stage-specific transcriptional regulation (Akter et al. 2019, Payá-Milans et al. 2019, Shen et al. 2019). However, only a limited number of studies have examined the relationship between lncRNA expression and histone modifications in plants (Hung et al. 2020). We also examined the relationship between lncRNA expression level and active histone modifications (H3K4me3 and H3K36me3)

Previous studies have identified lncRNAs associated with different lines, tissues or environmental changes in Brassica (Yu X et al. 2013, Liu X et al. 2015, Song et al. 2016, Huang et al. 2018, Zhang et al. 2018, Shea et al. 2019, Wang A et al. 2019). However, there are few reports of the association between lncRNAs and epigenetic modifications in *B. rapa*. In this study, we examined the DNA methylation levels and H3K27me3 levels in the region covering lncRNAs and found an association between lncRNAs and inverted repeat regions (IRRs) or 24-nt siRNAs, and association between genes overlapping with lncRNAs and different epigenetic marks (DNA methylation and H3K27me3).

Materials and methods

Plant materials and growth conditions

Six lines of *B. rapa*, inbred lines of RJKB-T24 (var. *pekinensis*) (Kawamura et al. 2016) and Yellow sarson (var. *trilocularis*), doubled haploid lines of BRA2209 (var. *rapa*), Homei (var. *pekinensis*), and Osome (var. *perviridis*), and a commercial F_1 hybrid cultivar, 'Harunosaiten' (var. *pekinensis*) (Watanabe Seed Co., Ltd.) were used. The six lines of *B. rapa* were selected to compare between different varieties, and the three *B. oleracea* lines were included to compare between species. RJKB-24 is breeding material used to produce a commercial F_1 hybrid cultivar and expresses the common agricultural traits desired for Chinese cabbage such as appropriate heading date and vernalization requirement (Miyaji et al 2017, Shea et al. 2018, Shea et al. 2019). Three *B. oleracea* cabbage (var. *capitata*) F_1 hybrid cultivars, 'Reiho', 'Matsunami' (Ishii Seed Growers CO., LTD), and 'Kinkei 201' (Sakata Seed Co., Ltd.) were used.

Seeds were surface sterilized and grown on agar solidified Murashige and Skoog (MS) plates with 1% (w/v) sucrose under long day (LD) condition (16h light / 8h dark) at 21˚C. Fourteen day first and second leaves of *B. rapa* and 19-day first and second leaves of *B. oleracea* were harvested for isolation of genomic DNA or total RNA. RJKB-T24 was used for RNA-sequencing (RNA-seq) for detection of lncRNAs. Briefly, after performing QC of the sequenced reads, putative mRNAs were identified by aligning the sequence reads to the *B. rapa* reference genome v1.5 using HISAT2 and then assembling transcripts with Stringtie. Assembled transcripts with a mapping code of 'u', indicating they are intergenic but not part of the annotated reference genome, were then compared to the SwissProt database using blastx (e-value 1e-10; (Kong et al. 2007)). Transcripts with hits were classified as putative mRNAs, while transcripts with no hits were classified as putative lincRNAs. A more detailed description of this method can be found in Shea et al. (2019).

DNA extraction and PCR

Genomic DNA was isolated by the Cetyl trimethyl ammonium bromide (CTAB) method (Murray and Thompson 1980). The PCR reaction was performed using the following conditions; 1 cycle of 94˚C for 3 min, 35 cycles of 94˚C for 30 s, 55˚C for 30 s, and 72˚C for 1 min, and final extension at 72˚C for 3 min. The PCR products were electrophoresed on 1.0% agarose gel. Primer sequences used in this study are shown in Table IV-1.

RNA extraction and RT-PCR

Total RNA from the first and second leaves were isolated by SV Total RNA Isolation System (Promega Co., WI, USA). To analyze lncRNA expression, cDNA was synthesized from 500 ng total RNA using PrimeScript RT reagent Kit (Takara Bio., Shiga, JAPAN). The absence of genomic DNA contamination was confirmed by PCR using a control without reverse transcriptase. The PCR conditions were 94˚C for 2 min followed by 35 cycles of 94˚C for 30 s, 55˚C for 30 s, and 68˚C for 30 s. The primers used in this study are listed in Table IV-1.

Detection of epigenetic states in lncRNA regions

To examine the epigenetic states (DNA methylation levels, H3K27me3, and 24-nt siRNA levels) of lncRNA coding regions in *B. rapa*, we used previous sequence data of whole genome bisulfite sequencing (WGBS) (Takahashi et al. 2018), chromatin immunoprecipitation sequencing (ChIP-seq) (Akter et al. 2019), and small RNA-sequencing (sRNA-seq) data (Takahashi et al. 2018), which were generated using samples from the same line, tissue, and developmental stages but harvested independently. Two biological replicates were used for all analyses.

The reads of WGBS were mapped to the *B. rapa* reference genome v.1.5 using Bowtie2 version 2.2.3 and Bismark v0.14.3, and data covering genomic regions encoding lncRNA in chromosomes A01 to A10 were extracted. In order to estimate the methylation levels of CG, CHG, and CHH (H is A, C, or T) contexts, the numbers of methylated and unmethylated reads were extracted for each cytosine position using bismark methylation extractor script with the paired-end parameter. The methylation level at each cytosine position was calculated by dividing the number of methylated cytosines (mC) reads by the total number of reads.

The reads of sRNA-seq were mapped to the *B. rapa* reference genome v.1.5 using Bowtie2 version 2.2.3. We classified the alignment reads by length, and the 24-nt aligned reads covering the lncRNA regions in chromosomes A01 to A10 were extracted.

The reads of ChIP-seq using anti-H3K27me3 (Millipore, 07–449) antibodies were mapped to the *B. rapa* reference genome v.1.5 using Bowtie2 version 2.2.3, and data covering genomic regions encoding for lncRNA in chromosomes A01 to A10 were extracted.

Comparison of putative mRNA and lncRNAs in *B. rapa* to other related species of *Brassica*

The putative mRNAs and lncRNAs were first compared to the reference genomes of *B. nigra*, *B. oleracea*, *B. juncea*, and *B. napus* by best-hit blastn (e-value 1e-10) to identify homologous regions in closely related *Brassica* species (Yu J et al. 2013, Chalhoub et al. 2014, Yang et al. 2016). In order to parse the local High-scoring segment pair (HSP) alignments produced by blastn, genBlastA was used to produce a representative putative gene that is homologous to the query (She et al. 2009). The parsed local alignments were then analyzed by a custom python script (available at http://www.github.com/danshea/lncRNA) to examine the overall alignment length and computed coverage of the aligned homologous region to the putative mRNA or lncRNA query transcript sequence. These results were then imported into R and plotted to assess overall relative coverage of the sequences among the *Brassica* species using Simple plot and ggbio in the Bioconductor package (Gentleman et al. 2004).

Gene ontology (GO) analysis

The genes covering incRNAs or NATs were used for GO analysis. The incRNAs and NATs having DNA methylation, siRNAs, or H3K27me3 were identified and their corresponding genes were also used for GO analysis using agriGO (Du et al. 2010) following the methods described by Shimizu et al. (2014). Statistical tests for enrichment of functional terms used the hypergeometric test and false discovery rate (FDR) correction for multiple testing to a level of 1% FDR.

Sequential ChIP-qPCR for bivalent enrichment of the targeted mRNA-NAT pairs

The enrichment of bivalent H3K4me3-H3K27me3-marks of the targeted mRNA-NAT pairs (Bra016382-MSTRG.19710 and Bra033594-MSTRG.1355) was calculated by comparing the target mRNAs/NATs and non H3K4me3 and H3K27me3 marks (Bra011336) by qPCR using immunoprecipitated DNA as a template. The difference between primer pairs was corrected by calculating the difference observed by qPCR amplifying the input DNA as a template. Primer sequences used in this study are shown in Table IV-1.

Results

Characterization of lncRNA in *B. rapa*

We identified 1,444 lincRNAs, 551 NATs, and 93 incRNAs using RNA-seq data of 14-day first and second leaves with and without four weeks of cold treatment (Shea et al. 2019). In order to investigate the relationship between lncRNAs and epigenetic modifications or the species specificity of lncRNAs, we analyzed in more detail the RNA-seq data of the 14-day first and second leaves without cold treatment. There was no strong bias in the chromosomal distribution of expressed lncRNAs (Figure IV-1). More than 65% of lncRNAs contained one exon (lincRNAs, 65.7%; NATs, 72.4%; incRNAs, 71.0%), whereas the proportion of mRNAs containing only one exon is 15.1% (Figure IV-2A). The mean transcript lengths of lincRNAs (725 nt) and incRNAs (779 nt) were shorter than that of NATs (1,271 nt) and mRNAs (1,305 nt) (Figure IV-2B). About 40% of lincRNAs were located within 2 kb of the genic regions, and about 10% of lincRNAs were located more than 20kb from the genic region (Figure IV-2C). In this study, we focused on lncRNAs mapped to the chromosomes A01 to A10 as previous WGBS, ChIP-seq, and sRNA-seq have omitted the placed scaffolds for their analyses (Takahashi et al. 2018, Akter et al. 2019). 763 of 1,173 (65.0%) lincRNAs, 291 of 529 (55.0%) NATs, and 66 of 92 (71.7%) incRNAs overlapped with IRRs such as transposable elements (TEs) detected by RepeatMasker, suggesting that IRRs are the source of more than half of lncRNAs in *B. rapa* (Figure IV-3).

Relationship between lncRNAs and siRNA or DNA methylation

We have performed sRNA-seq using 14-day first and second leaves, which are identical developmental stages and tissues to previously analyzed samples (Takahashi et al. 2018), but independently harvested for this study. We identified the lncRNAs having perfect sequence identity to genomic regions encoding 24-nt siRNAs. 219 of 1,173 (18.7%) lincRNAs, 74 of 529 (14.0%) NATs, and 16 of 92 (17.4%) incRNAs in A01-A10 overlapped with unique-mapped genomic regions encoding 24-nt siRNAs, and more than 80% of the lncRNAs overlapping with genomic regions encoding 24-nt siRNAs were from regions that harbored IRRs (Figure IV-3). 24-nt siRNAs were mapped in a similar way to lncRNA and its 5’ and 3’ flanking regions; this mapping pattern is different from those of genic regions or IRRs (Figure IV-4).

We examined the whole genome DNA methylation state by WGBS using the same 14-day first and second leaves (Takahashi et al. 2018). The average DNA methylation levels in regions covering lncRNAs were similar to those of the whole genome (Figure IV-5). DNA

methylation level in regions producing NATs was lower than those producing any of the three types of lncRNAs. Overlap of lncRNAs with IRRs or genomic regions encoding for 24-nt siRNAs was associated with increased DNA methylation levels and overlapping with both caused further increases in DNA methylation levels (Figure IV-6).

The DNA methylation pattern was similar in lincRNA and NAT coding regions, and DNA methylation levels in the 5' and 3' flanking regions were higher than in the body regions encoding lincRNAs and NATs (Figure IV-7). The DNA methylation levels over the incRNA coding regions and the flanking regions did not change (Figure IV-7). Overall, mapping of DNA methylation levels to lncRNA coding regions were different from those of genic regions or IRRs (Figure IV-7).

We examined the influence on the lncRNA expression level when IRRs or siRNA clusters on the genome spanned the lncRNA regions. IRRs did not affect the expression levels of lncRNAs (Figure IV-8). The expression levels of lncRNAs covered with siRNA clusters were higher than those not covered by siRNA clusters (Figure IV-8).

Relationship between lncRNAs and H3K27me3

We have examined the H3K27me3 distribution using the 14-day first and second leaf samples (Akter et al. 2019) and identified the genomic regions encoding for lncRNAs that have H3K27me3 marks. 127 of 1,173 (10.8%) lincRNAs, 83 of 529 (15.7%) NATs, and 15 of 92 (16.3%) incRNA coding genomic regions had H3K27me3 marks (Figure IV-9). H3K27me3 was enriched in the transcribed region of lncRNAs, similar to the genic regions, but H3K27me3 levels in lncRNA coding genomic regions were lower than in regions coding for mRNAs (Figure IV-10). The expression level of incRNAs was higher when the encoding regions had H3K27me3 than without H3K27me3 (Figure IV-8). In NATs and lincRNAs, there was no difference of expression level with and without H3K27me3 on their encoding regions (Figure IV-8).

Characterization of un-annotated genes

In 2,052 lncRNAs mapped to intergenic regions of the genome, 608 transcripts had hits (e-value < 1.0e-10) against the Swissport database using BLASTX, indicating they could be unannotated genes. The expression levels of the putative mRNAs from these regions were similar to that of mRNA (Shea et al. 2019). Putative mRNAs tended to have fewer exons compared with annotated mRNAs (Figure IV-2A). The mean transcript length of putative mRNA genes was similar to that of mRNAs (Figure IV-2B).

314 of 490 (64.1%) putative mRNAs in A01-A10 overlapped with IRRs (Figure IV-3) but the expression level of putative mRNAs overlapping with IRRs was similar to those not-overlapping with IRRs (Figure IV-8). Mapping of 24-nt siRNAs and DNA methylation levels to genomic regions encoding putative mRNAs were similar to that of genic regions (Figure IV-4, Figure IV-7). The average expression level of putative mRNAs with overlapping 24-nt siRNAs was higher than that without any overlap (Figure IV-8). The pattern of the average of DNA methylation level in the genomic region encoding putative mRNAs and their flanking regions was similar to that to the average of the genic regions (Figure IV-7). Overlapping IRRs or 24-nt siRNAs resulted in an increase in DNA methylation levels and overlapping with both causes further increase in DNA methylation levels (Figure IV-6).

53 of 490 (10.8%) genomic regions corresponding to putative mRNAs had H3K27me3. H3K27me3 was enriched in the genomic regions encoding for lncRNAs, especially around the transcription start site, and this was similar to the pattern in the genic region (Figure IV-10). The expression levels of putative mRNAs with corresponding genomic H3K27me3 were lower than those without H3K27me3, but there was no significant difference (Figure IV-8).

GO analysis of incRNAs and NATs and their relationship with epigenetic states

The paired genes overlapping with incRNA and NATs were subjected to GO analysis to assess whether there is any relationship between the gene ontology and the different epigenetic marks (Figure IV-11). GO category “transcription and DNA-dependent” and “metabolic process” were overrepresented for the incRNA data. For NATs, GO category “oxidation reduction”, “transcription and DNA-dependent”, and “carbohydrate metabolic process” were overrepresented.

In relation to the paired genes that overlap with incRNA and DNA methylation, siRNA or H3K27me3 marks, DNA methylation was overrepresented for “transcription and DNA dependent”. This indicates that the expression of these genes that overlap with incRNAs may be regulated through DNA methylation, and involved in transcription. Similarly, for NATs, H3K27me3 was overrepresented for “oxidation reduction” and “transcription and DNA dependent”, genes overlapping with NATs and each epigenetic mark were overrepresented for “primary metabolic process” and “metabolic process”, whereas siRNA was absent for “carbohydrate metabolic process”. A detailed list of each GO analysis is presented in Table IV-2 to Table IV-7.

Examination of the conservation of lncRNAs among the Brassica genus

Using the sequences of the 2,088 lncRNAs of *B. rapa*, a best-blast hit search against the *B. nigra*, *B. oleracea*, *B. juncea*, and *B. napus* reference genomes was conducted using GenBlastA (e-value = 1e-10) to examine the conservation of lncRNAs. *B. rapa* lncRNAs were most conserved in *B. oleracea* and moderate conservation was observed in *B. nigra* and *B. napus*.

LncRNAs were less conserved in *B. juncea*, especially in the incRNAs (Figure IV-12). We selected twelve lncRNAs that showed high sequence similarity with the *B. oleracea* reference genome. We tested whether these lncRNA coding genomic sequences were conserved in three commercial cultivars of cabbage (*B. oleracea*) by PCR using genomic DNA as template. PCR amplification of genomic regions corresponding to all twelve lncRNAs was confirmed in all three *B. oleracea* lines. Next, we tested by RT-PCR whether these putative lncRNAs in *B. oleracea* are expressed in the first and second leaves (Figure IV-13). Six of twelve lncRNAs were expressed in all three cultivars. One lncRNA was expressed in two of three lines, and one lncRNA was expressed in one of three lines. The remaining four lncRNAs were not expressed or were slightly expressed in all three lines (Table IV-8). We also examined the variation within *B. rapa* species using six lines. Seven of the twelve lncRNAs were expressed in all six lines, and the remaining five lncRNA showed line-specificity (Table IV-8). The sequences of the twelve lncRNA are listed in (Table IV-9).

Discussion

We analyzed 2,088 lncRNAs in leaves of *B. rapa*; their characteristics such as number of exons, length of transcripts (except for NATs), and lower expression levels were similar to those reported in other plant species (Liu et al. 2012, Joshi et al. 2016, Wang H et al. 2014, Shen et al. 2018, Wang et al. 2018, Chen et al. 2019). The total number of lncRNAs seems to be similar among species (Kopp and Mendell 2018, Kung et al. 2013), but there tends to be low sequence conservation between different species (Vandivier et al. 2016). Lower conservation of lncRNAs than mRNAs was observed between *B. napus* and its two ancestral species, *B. rapa* and *B. oleracea* (Zhang et al. 2018, Shen et al. 2018), even though the hybridization was relatively recent event, ~1,910-7, 180 years ago (Lu et al. 2019). We also found a lower sequence conservation of *B. rapa* lncRNAs among other species of the genus Brassica (*B. nigra*, *B. oleracea*, *B. juncea*, and *B. napus*). In 12 selected highly conserved lncRNAs, we found conservation of the lncRNAs at both the sequence and transcriptional level between *B. rapa* and *B. oleracea*, but there was transcriptional variation within species agreeing with the IRR diversification of each species after the allotetraploidization of *B. napus* (Sampath et al. 2014). The analysis of sequence homology of lncRNAs in *B. rapa* with other members of the *Brassica* genus may depends the understanding of the evolutionary dynamics of lncRNAs in the genus *Brassica*.

We found over 55% of each lncRNAs (65.0% lincRNAs, 55.0% NATs, and 71.7% incRNAs) overlapped with IRRs including TEs throughout the *B. rapa* genome. TEs are considered a source of siRNAs (Cho 2018), and lincRNAs identified in *A. thaliana*, rice, and maize, were associated with 22.2%, 49.7%, and 51.5% of TEs, respectively (Wang D et al 2017). In this study, we found 18.7% lincRNAs, 14.0% NATs, and 17.4% of incRNAs covered unique-mapped 24-nt siRNAs, and over 80% of lncRNAs covering the genomic regions encoding for 24-nt siRNAs overlapped with IRRs with perfect sequence match, suggesting that lncRNAs covering IRRs are also the likely source of siRNAs in *B. rapa*. For mRNA, the average expression levels of lncRNAs having 24-nt siRNAs was higher than that of lncRNAs not having 24-nt siRNAs (Takahashi et al. 2018). LncRNAs and 24-nt siRNAs are known to be involved in the increase of DNA methylation, which agrees with reports of increased gene expression when DNA methylation overlaps with an exonic region (Bewick et al. 2017). The detailed mechanism is not clear, but our results also agrees that lncRNAs may be involved in this gene regulation mechanism.

The epigenetic functions of different lncRNAs have been identified in diverse organisms (Wang C et al. 2017, Matzke et al. 2015), but there is limited information in *B. rapa*. LncRNAs show a close association with DNA methylation of the genomic region encoding the lncRNA (Di Ruscio et al. 2013), but studies have focused on the model plant species, *A. thaliana* (Chekanova et al. 2015, Au et al. 2017). In this study, we found a similar level of DNA methylation in regions encoding lncRNA compared to previously described DNA methylation in the *B. rapa* genome (Takahashi et al. 2018). Levels of DNA methylation can be positively regulated by siRNAs through the RdDM pathway (Yan et al. 2018), and Pol IV is important for biogenesis of 24-nt siRNAs (Zhai et al. 2015). Identification of long Pol IV-dependent transcripts is difficult because these transcripts can be rapidly processed to produce 24-nt siRNAs. There are several reports that detected Pol Ivdependent siRNA-precursor transcripts in *A. thaliana* of different lengths (Blevins et al. 2015, Li et al. 2015, Zhai et al. 2015). Most Pol IV-dependent transcript regions overlapped with Pol IV-dependent siRNA loci, and CHH methylation depends on the production of Pol IV-dependent transcripts (Li et al. 2015). However, if Pol IV-dependent transcripts are short as 30–40 nucleotides (Zhai et al. 2015, Blevins et al. 2015), the lncRNAs that we identified are much longer and not likely to be Pol-IV dependent transcripts that act as siRNA precursors.

Pol V transcripts accumulate at very low levels and have been difficult to identify. However, using RNA immunoprecipitation, Pol V-dependent lncRNAs were identified, and CHH methylation and 24-nt siRNA accumulation were shown to be restricted to Pol V transcribed regions (Böhmdorfer et al. 2016). In this study, about 13% of lncRNAs overlapped with genomic regions encoding 24-nt siRNAs, and the DNA methylation level of those regions was higher than the average. It has been reported that the median length of Pol V transcripts was 689 nt (Böhmdorfer et al. 2016). The length of the lncRNAs overlapping genomic regions encoding for 24-nt siRNAs identified in this study resembles the length of the Pol V transcript precursors.

The level of H3K27me3 plays a role in tissue specific gene expression (Makarevitch et al. 2013). LncRNAs are also considered to be involved in regulating histone modifications. Binding of lncRNAs to PRC2 has been observed in human/animal (Tu et al. 2017, Gaiti et al. 2018), and COLDAIR and COLDWRAP have been shown to recruit the PRC2 complex to *FLC* during vernalization in *A. thaliana* (Heo and Sung 2011, Kim and Sung 2017, Tian et al. 2019, Yu et al. 2019). EMF2B is a component of the PRC2 complex and in rice, a mutant of *emf2b* was reported to lose H3K27me3 and derepress some of the lincRNAs, suggesting that expression of these lincRNAs is regulated by PRC2-mediated histone methylation of the region encoding

the lncRNA (Johnson et al. 2018). We found that about 10–16% of lncRNAs overlapped with H3K27me3 regions, which is lower than the overlap with mRNAs. The pattern of H3K27me3 in lncRNA overlapping regions was similar to mRNAs; H3K27me3 accumulated in the body region of transcripts, especially around the transcription start site. However, there was no negative relationship between the presence of H3K27me3 and lncRNA expression levels, and incRNA coding regions having H3K27me3 marks showed higher expression levels than incRNA coding regions without H3K27me3 marks. H3K27me3 over lncRNA coding regions was also found in different tissues of maize, which was responsible for the regulation of tissue-specific lncRNAs expression (Li et al. 2014). Coincidentally, DNA methylation in intronic regions is known to influence splicing patterns (Regulski et al. 2013), which may reflect epigenetic tissue specific gene regulation involving RdDM and H3K27me3. However, as our study focused on the analyses in a single tissue/developmental stage, we could not identify the lncRNAs that are transcriptionally silenced by H3K27me3, leading to the underestimation of lncRNAs with H3K27me3 marks. Further study will be required to examine the transcriptional regulation of lncRNAs by H3K27me3.

GO analysis of the genes that overlap with incRNA and NATs loci revealed that both were overrepresented in the “transcription and DNA-dependent” category. For incRNAs, all five genes were annotated as transcription factors/regulators but the function of these genes remains unknown without further analyses. However, for genes overlapping with NATs and DNA methylated loci, three of the six identified genes were annotated as beta-galactosidase that are known to be involved in pollen development and fertilization (Jakobsen et al. 2005, Wang et al. 2008). Interestingly, annotated genes enriched for the “transcription and DNA-dependent” category in “NATs” contained WRKY27 and WRKY62 genes that are related to pathogen defense (Mukhtar et al. 2008, 78. Fukushima et al. 2016). Furthermore, a gene (MAF4) known to be involved in flowering regulation that is regulated by NATs in Arabidopsis (Zhao et al. 2018) was also identified, indicating that this NAT may be regulated by DNA methylation and involved in flowering regulation in Brassica. However, further analyses such as deletion of the incRNA/NAT locus or manipulating their expression will be needed to understand the function of these lncRNAs in Brassica.

This study revealed that a small proportion of lncRNAs in *B. rapa* are conserved with other *Brassica* species. The majority of lncRNAs in *B. rapa* overlap with IRRs and there is some overlap with DNA methylation and 24-nt siRNAs, hinting at regulation of lncRNA expression through the RdDM pathway. Interestingly, some of the lncRNAs that overlapped with the genomic regions having H3K27me3 marks were more highly expressed, which was unexpected

because of the known gene suppression activity of H3K27me3. This may indicate an unknown regulatory mechanism of lncRNAs by H3K27me3. However, the current study focuses on the genome-wide quantification of lncRNAs, 24-nt siRNAs, DNA methylation, and H3K27me3, and further quantitative analysis at the locus specific level may reveal a clearer relationship between these marks.

Epigenetic state of the genome (including DNA methylation and H3K27me3, expression of lncRNAs and 24-nt siRNAs) can dynamically change depending on the stage of development, biotic and abiotic stress conditions, and the expression of lncRNAs can also change depending on the tissue or environment and influence gene expression (Chen et al. 2020). This study focused on the association of lncRNAs and epigenetic marks, but to understand the association of lncRNAs with agronomical traits, further exploration of lncRNAs under different tissues, environmental conditions, and cultivars (or varieties) will be needed.

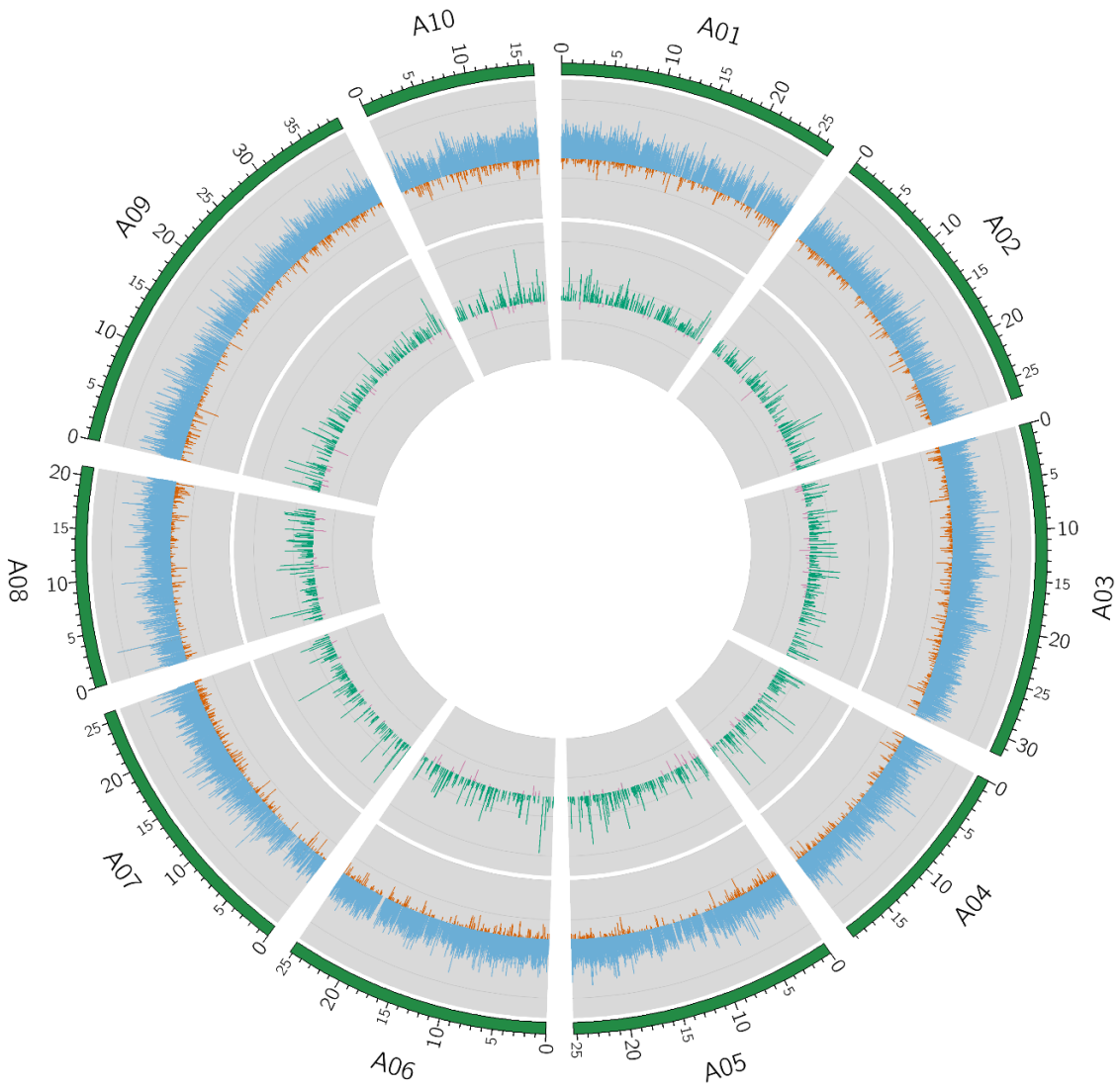


Figure IV-1. Chromosomal (A01-A10) distribution of lncRNAs and mRNAs. Log2 FPKM of mRNAs represented in blue (positive values) and orange (negative values), and lncRNAs in green (positive values) and pink (negative values).

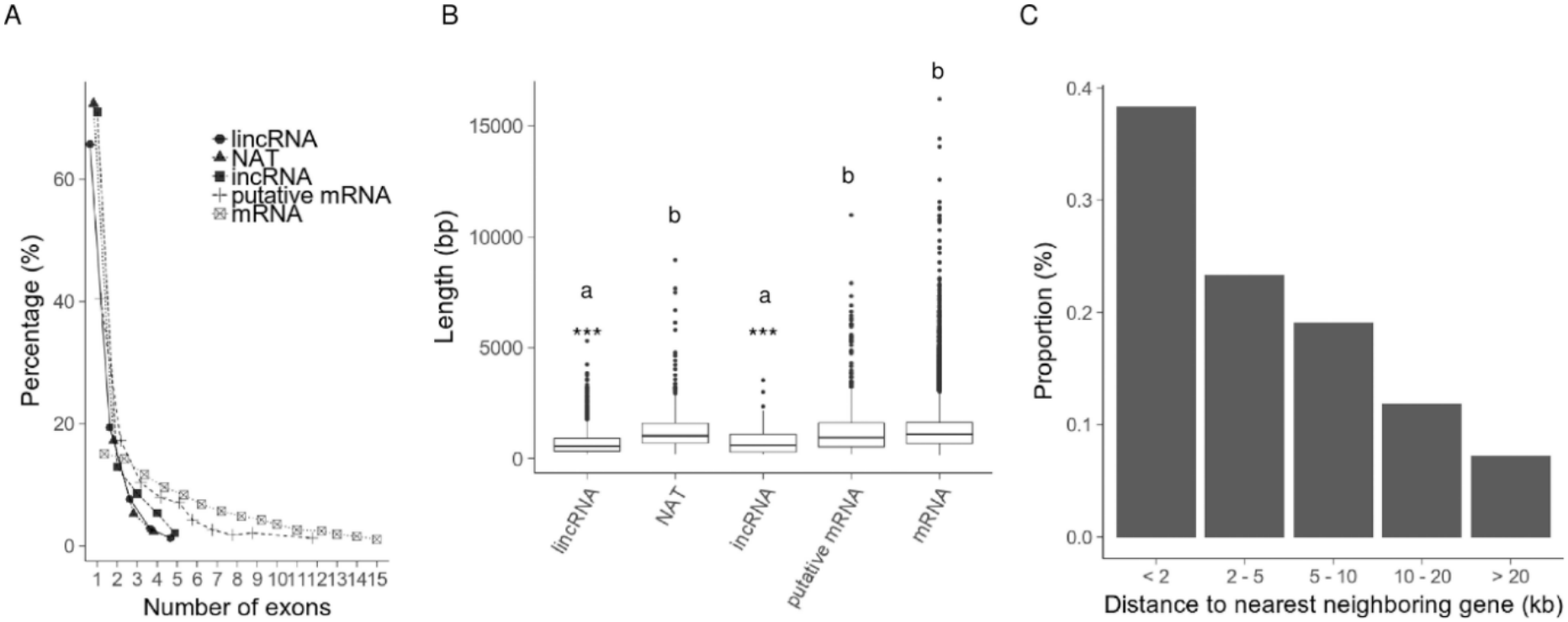
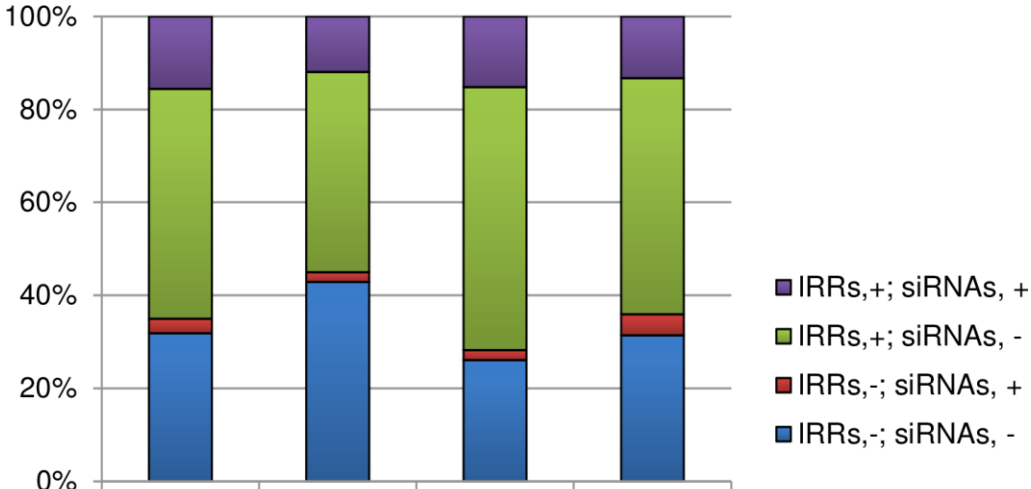


Figure IV-2. Analysis of lncRNAs of 14-day first and second leaves of *B. rapa*. (A) Number of exons in lncRNA. (B) Nucleotide length of lincRNA, NAT, incRNA, putative mRNA and mRNA. "a" and "b" represent significant differences by one-way ANOVA test ("*", $p<0.05$; "***", $p<0.001$). (C) The proportion lncRNA distances to the nearest gene.



	lincRNA	NAT	incRNA	putative mRNA
IRRs,+; siRNAs, +	15.6%	11.9%	15.2%	13.3%
IRRs,+; siRNAs, -	49.4%	43.1%	56.5%	50.8%
IRRs,-; siRNAs, +	3.1%	2.1%	2.2%	4.5%
IRRs,-; siRNAs, -	31.9%	42.9%	26.1%	31.4%

Figure IV-3. Proportion of each type of lncRNA and putative mRNAs that overlap with IRRs and siRNA. "+" indicates with and "-" indicated without overlapping IRRs or genomic regions encoding siRNAs.

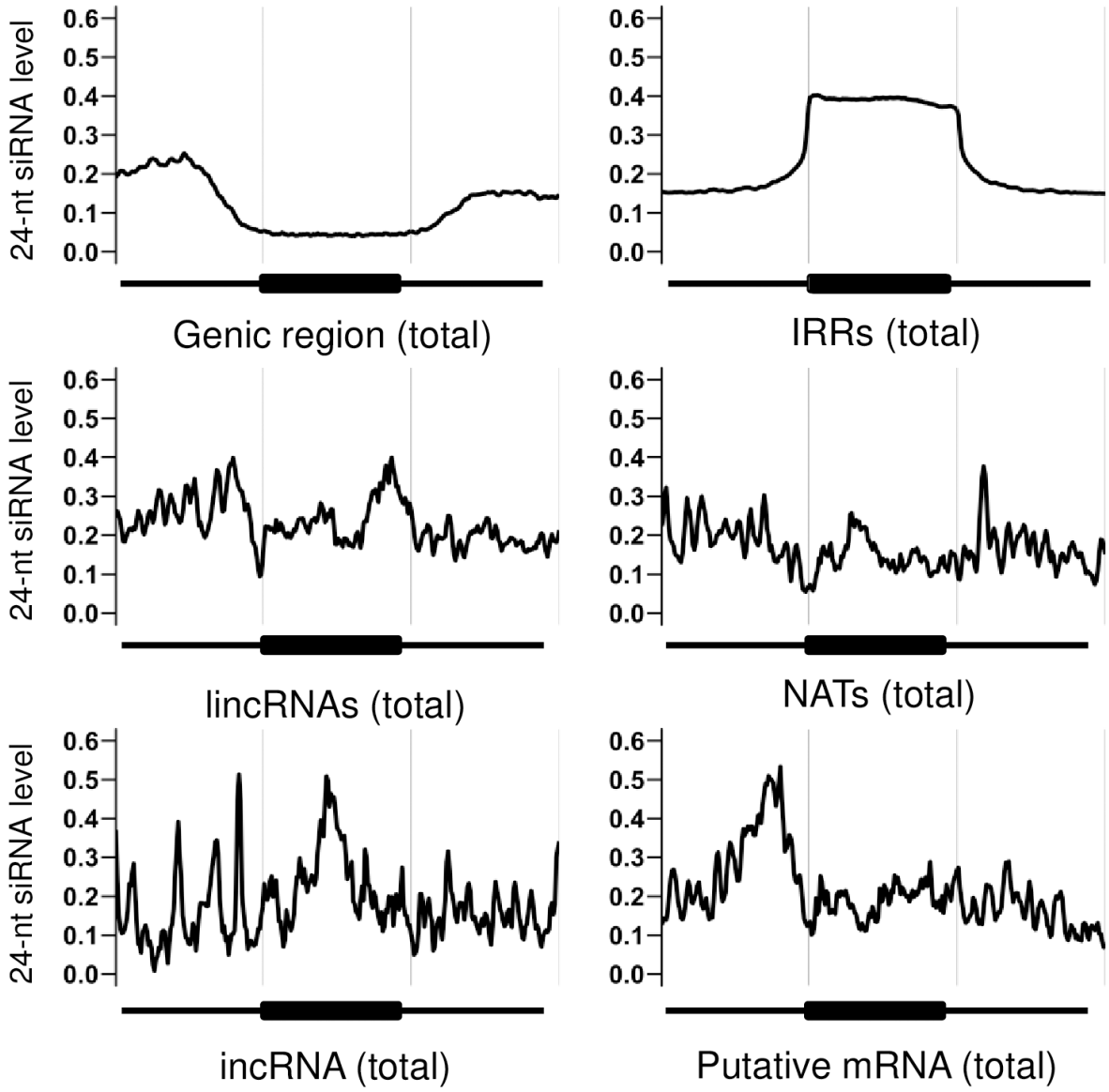


Figure IV-4. 24nt-siRNAs mapped to genic region, IRRs, putative mRNAs and each type of lncRNAs. The x-axis represents the target region (genic region, IRRs, putative mRNAs, or lncRNAs) and the flanking 5' and 3' regions. The y-axis represents the reads per million (RPM) value of 24-nt siRNA that overlaps with the corresponding target region.

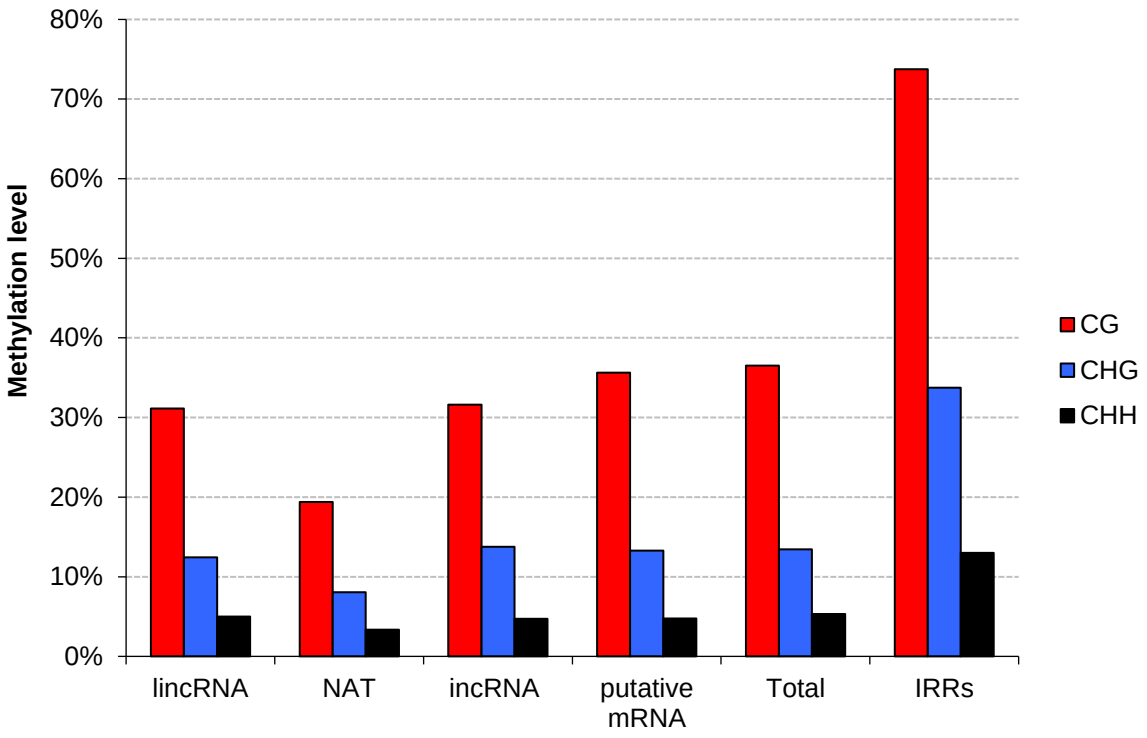


Figure IV-5. DNA methylation level of each type of lncRNAs, putative mRNA, total, and IRRs. "Total" represents the DNA methylation level of all regions.

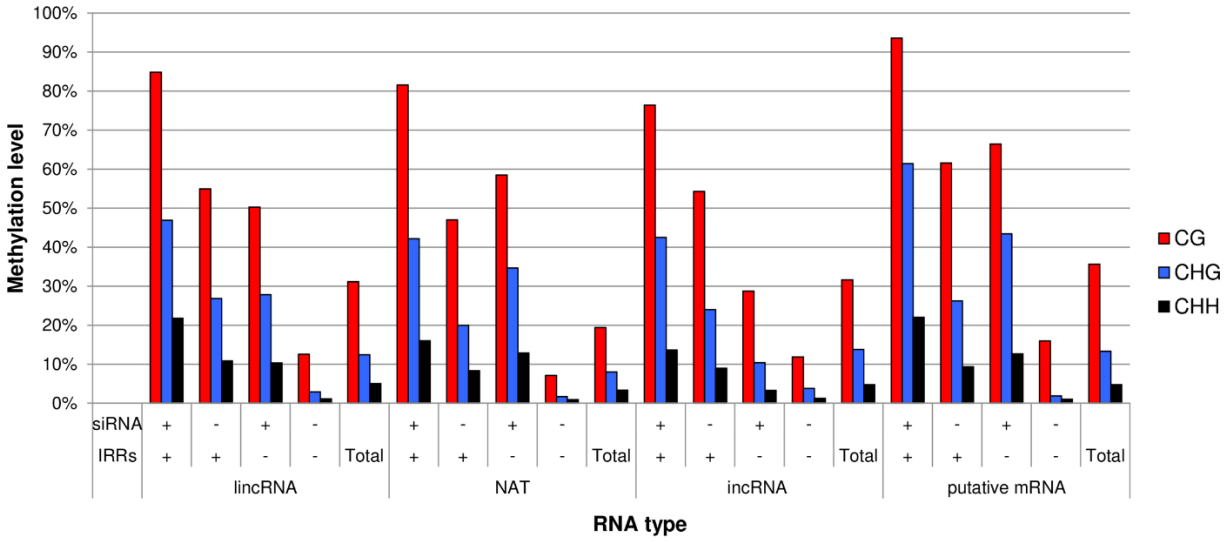


Figure IV-6. DNA methylation level of genomic regions encoding each RNA type with siRNAs or IRRs. "+" indicates with and "-" indicates without overlapping genomic regions encoding for siRNAs or IRRs.

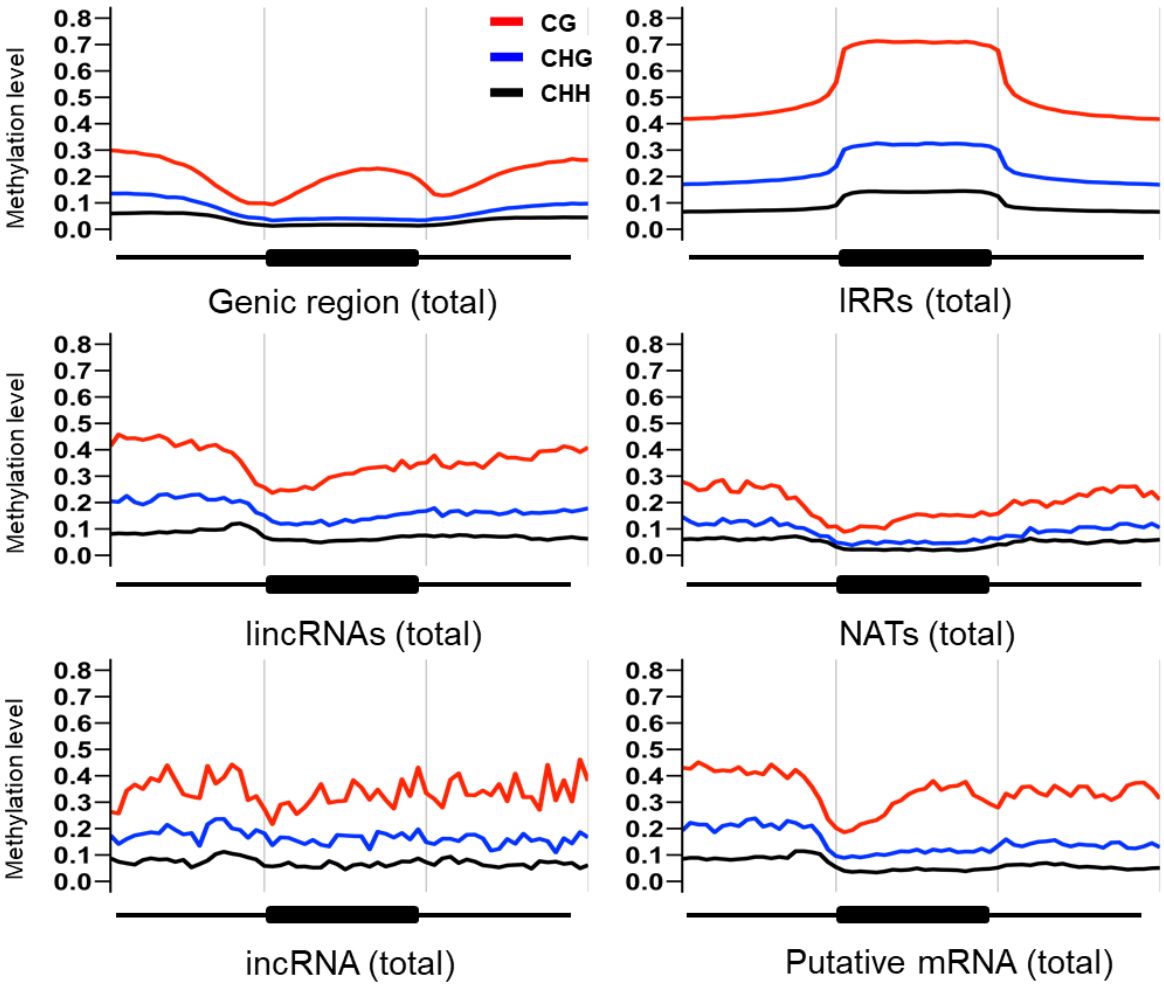


Figure IV-7. DNA methylation of the genic region, IRRs, or regions coding for putative mRNA and lncRNAs.

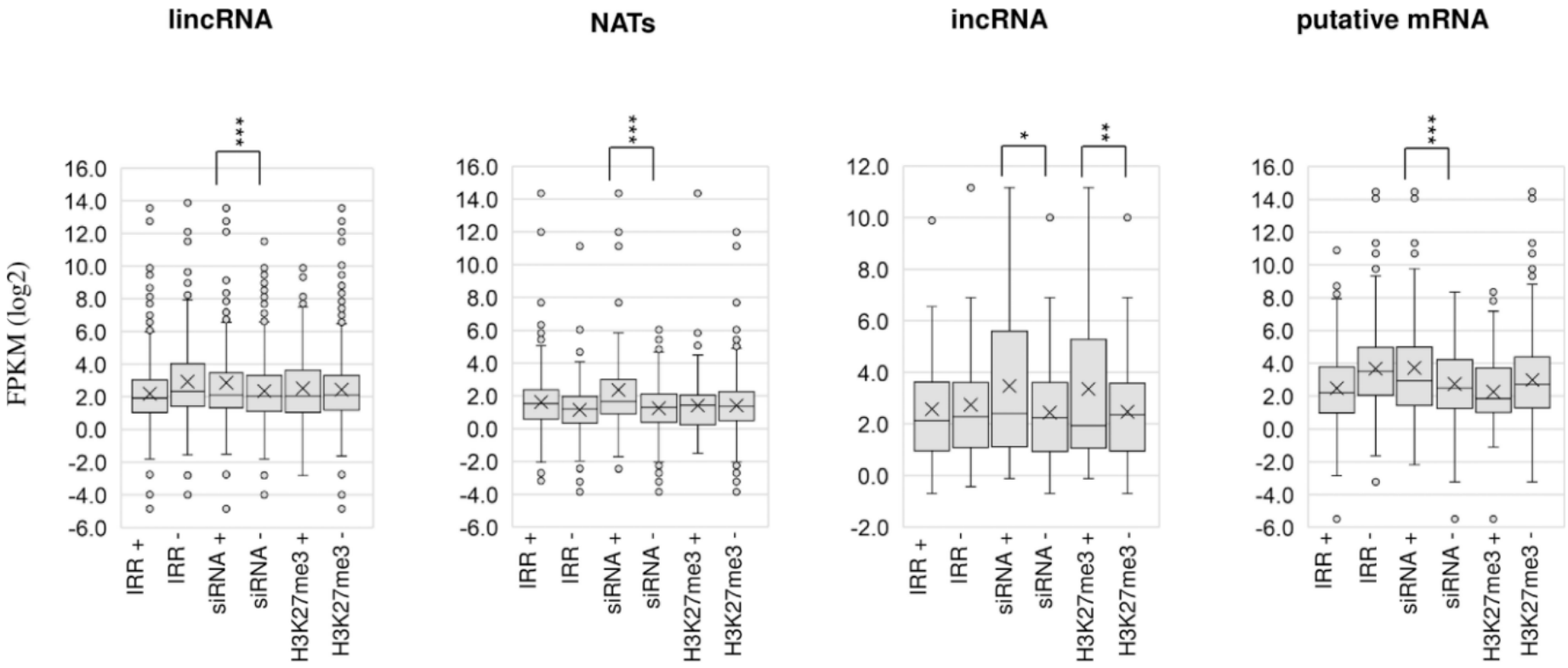


Figure IV-8. Expression level of each type of RNAs with (+) or without (-) overlapping IRRs, 24nt-siRNAs, or H3K27me3. *, *p<0.05*; **, *p<0.01*; ***, *p<0.001* (Student t-test).

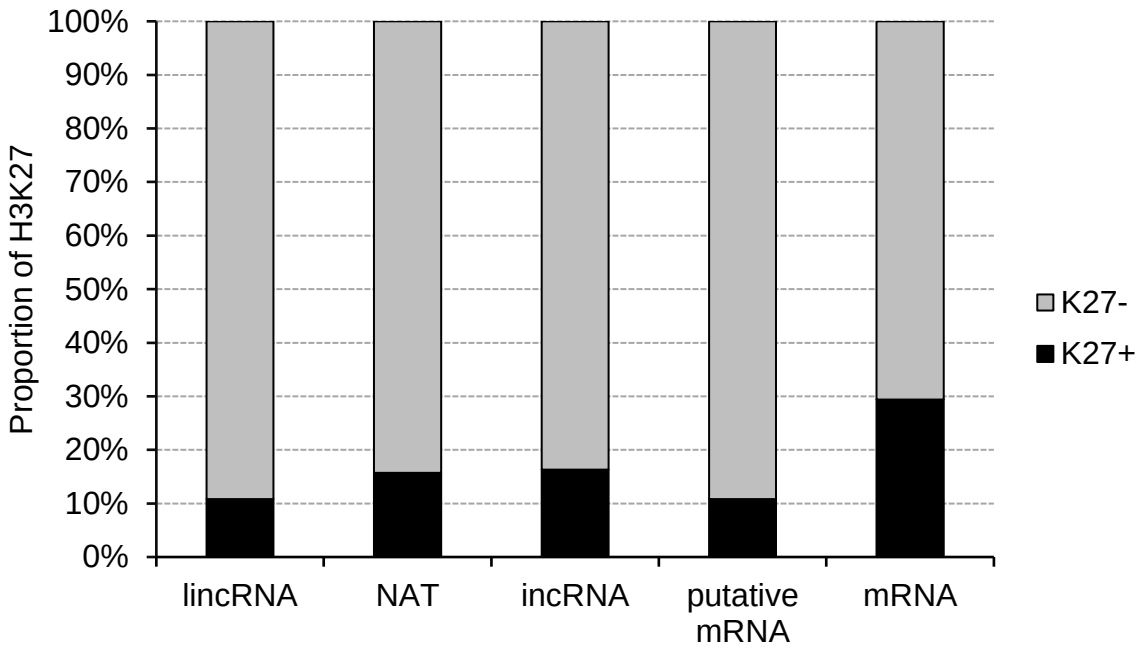


Figure IV-9. Proportion of lncRNAs, putative mRNA, mRNA with (K27+) or without (K27-) H3K27me3 on coding regions.

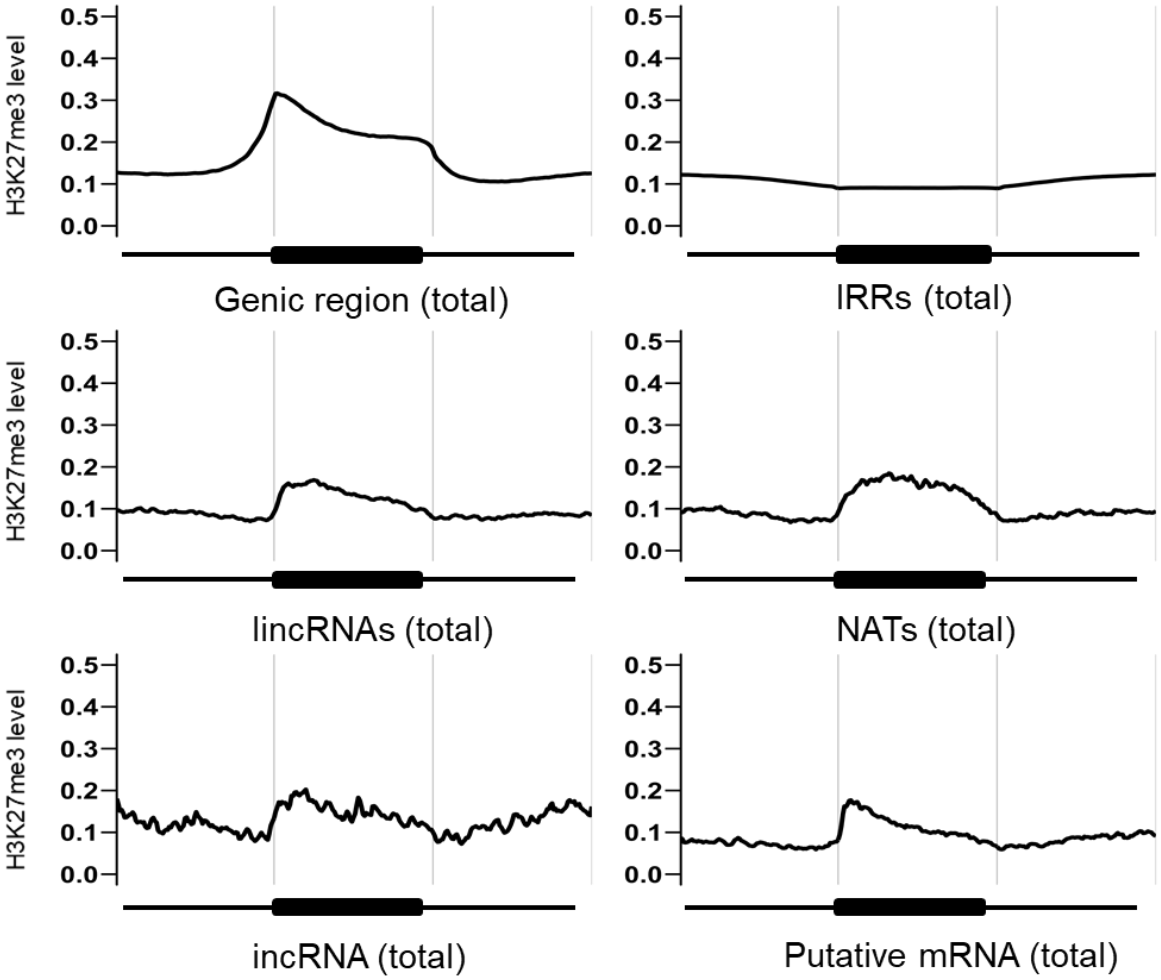


Figure IV-10. H3K27me3 mapped to the genic region, IRRs, or coding regions of putative mRNA or lncRNAs.

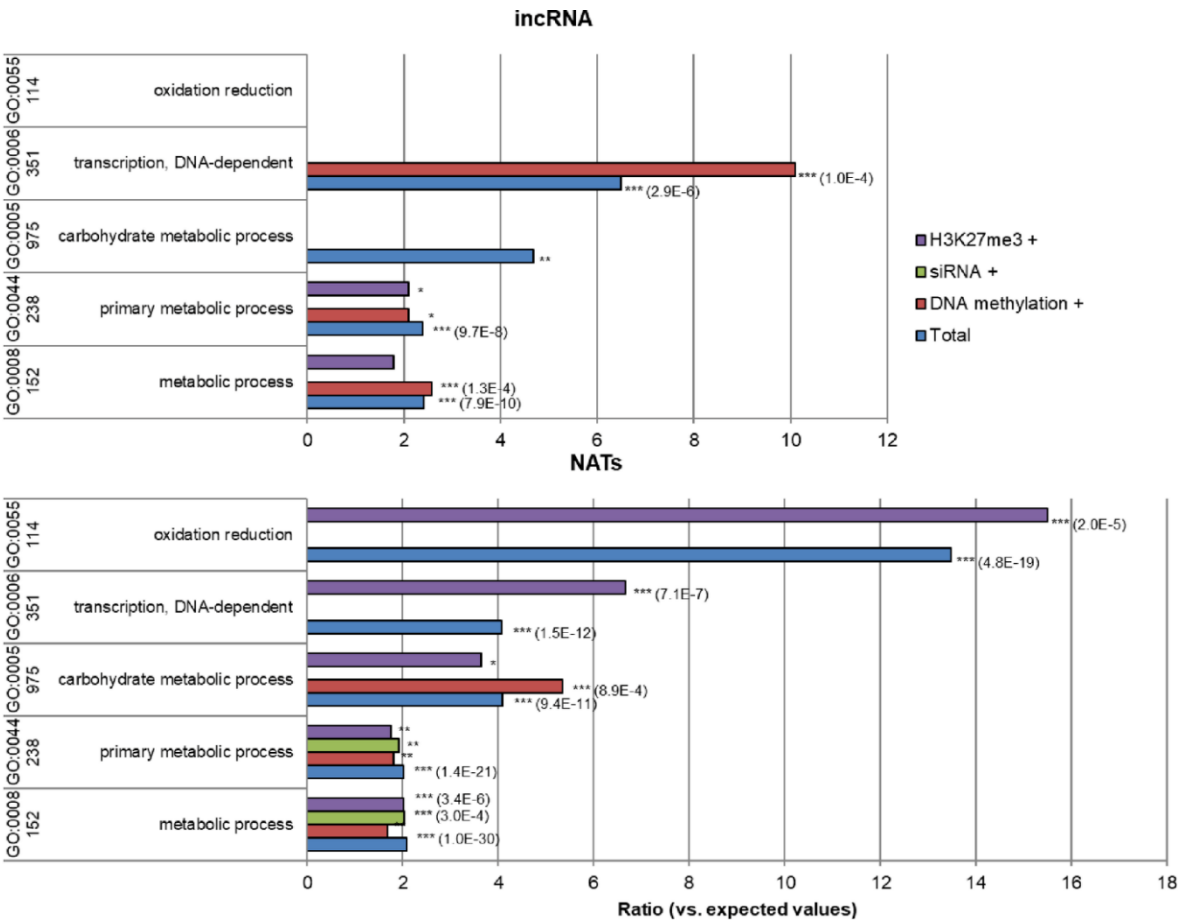


Figure IV-11. GO analysis of incRNA and NATs and different epigenetic marks (DNA methylation, siRNA, H3K27me3). *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$. Overrepresentation of the identified paired genes between incRNA or NATs to DNA methylation, siRNAs or H3K27me3 loci for each GO category is shown.

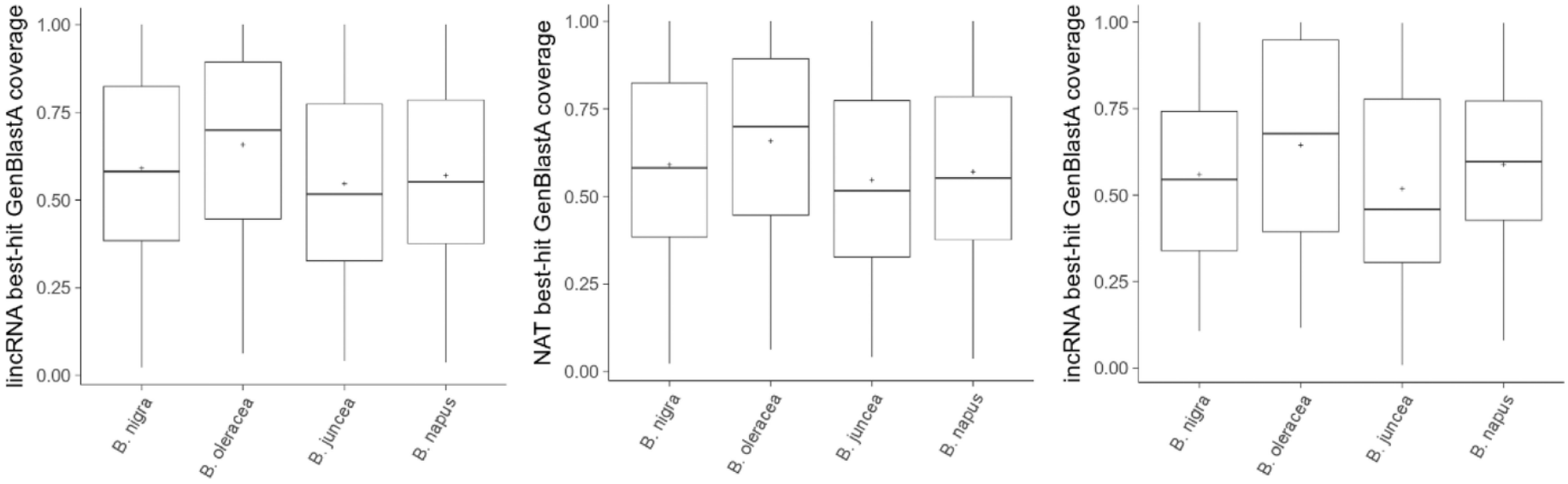


Figure IV-12. BLAST search of each type of lncRNA from *B. rapa* against *B. nigra*, *B. oleracea*, *B. juncea*, and *B. napus*.

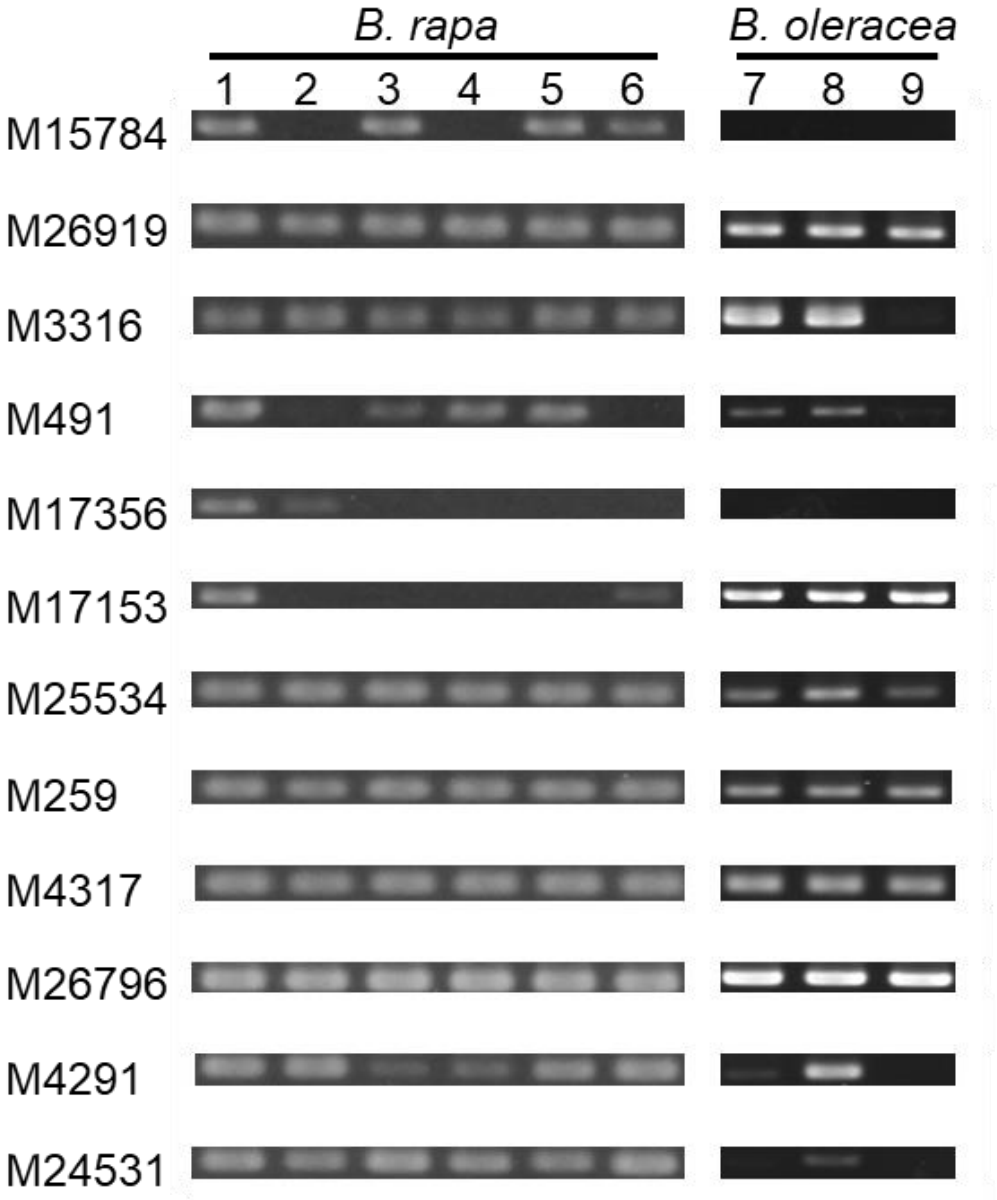


Figure IV-13. RT-PCR of each identified lncRNAs in *B. rapa* and *B. oleracea* lines. The number represents each line as follows: For *B. rapa* lines, 1 = RJKB-T24, 2 = Homei, 3 = Harunosaiten, 4 = BRA2209, 5 = Osome, 6 = Yellow Sarson. For *B. oleracea* lines, 7 = Reiho, 8 = Matsunami, 9 = Kinkei 201.

Table IV-1. Sequence of primers used in this study

Name	Forward primer (5'-3')	Reverse primer (5'-3')
M15784	CCAAAATCCTTCCATCTCTC	CCATATAAAACATGAACTTGG
M26919	AGAAAGCATGGGAATGCAAG	ACTTGCAGCAAACCCAAAAC
M3316	AGGGAGCAGCCACGGGCACA	CGCGGCGGCGCGTTCCTGTG
M491	ACCGCTGAGCAAGAAACCCG	TCATACATATCAGAGAGAAC
M17356	TAGGGGTTGCTTTACTTGAG	TTGACCATTGACCAGTTGAA
M17153	GCCGCAATATTAGACCTCTC	GTTGACAAAAGAGTTGATGG
M25534	TGGAGTTTGAACTACTTTAG	TTTTCCAGCAAATCCTCTGC
M259	CTCGATCAAGATTCATTTTC	TTAAGCTTGGAGAGTTTCAG
M4317	GCCTAGCAATATCCTTTTCC	TGTTAAGCAAAACGACAGTA
M26796	CGGTGGATACCTAGGCACCC	TACTTCGCTATCGGTCACCC
M4921	CGCCAACGCGGCCAAGCTTG	CTTCTCCATGCAGCGGCTCA
M24531	GGATGGCTCTAAATTCTTAA	AATGGTAATACGGTTTTCAC
For reference		
Bra011336	GGGAATCAGCTTTTGTGGTG	AAATGACCCGATCAGCAAAG

Table IV-2. GO function term overrepresented in genes overlapping with incRNA expressing loci (p<0.001)

GO accession	Term type	Term	*p* value	FDR
GO:0008152	P	metabolic process	7.90E-10	1.80E-07
GO:0044238	P	primary metabolic process	9.70E-08	1.10E-05
GO:0005488	F	binding	1.00E-07	1.00E-05
GO:0032774	P	RNA biosynthetic process	2.90E-06	1.30E-04
GO:0006351	P	transcription, DNA-dependent	2.90E-06	1.30E-04
GO:0044237	P	cellular metabolic process	3.00E-06	1.30E-04
GO:0044260	P	cellular macromolecule metabolic process	5.60E-06	1.80E-04
GO:0046483	P	heterocycle metabolic process	5.80E-06	1.80E-04
GO:0043170	P	macromolecule metabolic process	9.40E-06	2.60E-04
GO:0005515	F	protein binding	1.80E-05	7.90E-04
GO:0003824	F	catalytic activity	2.40E-05	7.90E-04
GO:0006807	P	nitrogen compound metabolic process	2.50E-05	6.40E-04
GO:0016070	P	RNA metabolic process	2.80E-05	6.40E-04
GO:0006139	P	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	4.70E-05	9.70E-04
GO:0006464	P	protein modification process	5.80E-05	1.10E-03
GO:0043687	P	post-translational protein modification	8.90E-05	1.50E-03
GO:0043412	P	macromolecule modification	1.40E-04	2.20E-03
GO:0005524	F	ATP binding	1.90E-04	3.70E-03
GO:0032559	F	adenyl ribonucleotide binding	1.90E-04	3.70E-03
GO:0009058	P	biosynthetic process	2.90E-04	4.40E-03
GO:0001883	F	purine nucleoside binding	2.90E-04	3.70E-03
GO:0001882	F	nucleoside binding	2.90E-04	3.70E-03
GO:0030554	F	adenyl nucleotide binding	2.90E-04	3.70E-03
GO:0006725	P	cellular aromatic compound metabolic process	3.30E-04	4.70E-03
GO:0009987	P	cellular process	3.60E-04	4.80E-03
GO:0006350	P	transcription	4.90E-04	6.10E-03
GO:0044249	P	cellular biosynthetic process	5.70E-04	6.70E-03
GO:0006355	P	regulation of transcription, DNA-dependent	6.40E-04	7.20E-03
GO:0032555	F	purine ribonucleotide binding	6.40E-04	6.40E-03

GO:0032553	F	ribonucleotide binding	6.40E-04	6.40E-03
GO:0051252	P	regulation of RNA metabolic process	6.80E-04	7.20E-03
GO:0017076	F	purine nucleotide binding	9.00E-04	8.20E-03

F, Molecular function; P, Biological process

Table IV-3. GO function term overrepresented in genes overlapping with incRNA expressing loci and DNA methylation ($p<0.001$)

GO accession	Term type	Term	*p* value	FDR
GO:0032774	P	RNA biosynthetic process	1.00E-04	2.50E-03
GO:0006351	P	transcription, DNA-dependent	1.00E-04	2.50E-03
GO:0008152	P	metabolic process	1.30E-04	2.50E-03
GO:0016070	P	RNA metabolic process	8.70E-04	1.30E-02

P, Biological process

Table IV-4. GO function term overrepresented in genes overlapping with NAT expressing loci (p<0.0001)

GO accession	Term type	Term	*p* value	FDR
GO:0005488	F	binding	2.60E-39	9.50E-37
GO:0008152	P	metabolic process	1.00E-30	1.00E-27
GO:0044238	P	primary metabolic process	1.40E-21	7.30E-19
GO:0046872	F	metal ion binding	1.10E-19	2.10E-17
GO:0055114	P	oxidation reduction	4.80E-19	1.60E-16
GO:0043169	F	cation binding	9.90E-19	9.10E-17
GO:0043167	F	ion binding	9.90E-19	9.10E-17
GO:0009987	P	cellular process	1.30E-18	3.20E-16
GO:0044237	P	cellular metabolic process	2.20E-18	4.50E-16
GO:0005515	F	protein binding	5.70E-18	4.20E-16
GO:0003824	F	catalytic activity	3.10E-17	1.90E-15
GO:0016021	C	integral to membrane	1.10E-15	1.20E-13
GO:0046914	F	transition metal ion binding	9.10E-14	4.80E-12
GO:0006351	P	transcription, DNA-dependent	1.50E-12	2.20E-10
GO:0032774	P	RNA biosynthetic process	1.50E-12	2.20E-10
GO:0006355	P	regulation of transcription, DNA-dependent	6.90E-12	8.70E-10
GO:0051252	P	regulation of RNA metabolic process	8.80E-12	9.90E-10
GO:0030554	F	adenyl nucleotide binding	9.40E-12	3.50E-10
GO:0001883	F	purine nucleoside binding	9.40E-12	3.50E-10
GO:0001882	F	nucleoside binding	9.40E-12	3.50E-10
GO:0006807	P	nitrogen compound metabolic process	1.20E-11	1.20E-09
GO:0020037	F	heme binding	1.60E-11	5.30E-10
GO:0031224	C	intrinsic to membrane	1.70E-11	9.30E-10
GO:0043565	F	sequence-specific DNA binding	2.00E-11	6.10E-10
GO:0005506	F	iron ion binding	5.00E-11	1.40E-09
GO:0046906	F	tetrapyrrole binding	6.60E-11	1.70E-09
GO:0032559	F	adenyl ribonucleotide binding	8.40E-11	2.10E-09
GO:0005975	P	carbohydrate metabolic process	9.40E-11	8.60E-09
GO:0043170	P	macromolecule metabolic process	2.20E-10	1.90E-08
GO:0017076	F	purine nucleotide binding	3.40E-10	7.60E-09
GO:0000166	F	nucleotide binding	3.50E-10	7.60E-09

GO:0016740	F	transferase activity	3.80E-10	7.80E-09
GO:0016070	P	RNA metabolic process	4.00E-10	3.10E-08
GO:0016758	F	transferase activity, transferring hexosyl groups	5.10E-10	9.90E-09
GO:0006139	P	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1.50E-09	1.10E-07
GO:0044262	P	cellular carbohydrate metabolic process	2.20E-09	1.50E-07
GO:0032555	F	purine ribonucleotide binding	2.60E-09	4.60E-08
GO:0032553	F	ribonucleotide binding	2.60E-09	4.60E-08
GO:0055086	P	nucleobase, nucleoside and nucleotide metabolic process	3.00E-09	1.90E-07
GO:0005524	F	ATP binding	3.10E-09	5.10E-08
GO:0044260	P	cellular macromolecule metabolic process	3.50E-09	2.10E-07
GO:0016757	F	transferase activity, transferring glycosyl groups	6.80E-09	1.10E-07
GO:0044425	C	membrane part	7.40E-09	2.70E-07
GO:0016705	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	4.90E-08	7.50E-07
GO:0043412	P	macromolecule modification	7.80E-08	4.40E-06
GO:0008270	F	zinc ion binding	8.30E-08	1.20E-06
GO:0046700	P	heterocycle catabolic process	1.20E-07	6.40E-06
GO:0046983	F	protein dimerization activity	1.30E-07	1.80E-06
GO:0006753	P	nucleoside phosphate metabolic process	1.40E-07	6.70E-06
GO:0009117	P	nucleotide metabolic process	1.40E-07	6.70E-06
GO:0006464	P	protein modification process	1.60E-07	7.30E-06
GO:0034641	P	cellular nitrogen compound metabolic process	2.20E-07	9.80E-06
GO:0043687	P	post-translational protein modification	3.40E-07	1.40E-05
GO:0030246	F	carbohydrate binding	3.90E-07	5.30E-06
GO:0044249	P	cellular biosynthetic process	6.40E-07	2.60E-05
GO:0009058	P	biosynthetic process	7.10E-07	2.80E-05
GO:0080090	P	regulation of primary metabolic process	2.10E-06	7.80E-05
GO:0055085	P	transmembrane transport	2.90E-06	1.00E-04
GO:0006793	P	phosphorus metabolic process	2.90E-06	1.00E-04

GO:0004497	F	monooxygenase activity	3.20E-06	4.20E-05
GO:0050794	P	regulation of cellular process	3.30E-06	1.10E-04
GO:0010556	P	regulation of macromolecule biosynthetic process	3.50E-06	1.10E-04
GO:0019219	P	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.70E-06	1.20E-04
GO:0003677	F	DNA binding	4.10E-06	5.10E-05
GO:0065007	P	biological regulation	4.30E-06	1.30E-04
GO:0009889	P	regulation of biosynthetic process	5.50E-06	1.50E-04
GO:0031326	P	regulation of cellular biosynthetic process	5.50E-06	1.50E-04
GO:0031323	P	regulation of cellular metabolic process	5.50E-06	1.50E-04
GO:0051171	P	regulation of nitrogen compound metabolic process	5.90E-06	1.60E-04
GO:0006796	P	phosphate metabolic process	8.30E-06	2.20E-04
GO:0006350	P	transcription	8.90E-06	2.30E-04
GO:0050789	P	regulation of biological process	1.20E-05	2.90E-04
GO:0045449	P	regulation of transcription	1.30E-05	3.20E-04
GO:0019222	P	regulation of metabolic process	1.50E-05	3.50E-04
GO:0016310	P	phosphorylation	1.50E-05	3.50E-04
GO:0000096	P	sulfur amino acid metabolic process	1.50E-05	3.50E-04
GO:0006468	P	protein amino acid phosphorylation	1.60E-05	3.60E-04
GO:0060255	P	regulation of macromolecule metabolic process	1.70E-05	3.70E-04
GO:0004672	F	protein kinase activity	1.80E-05	2.20E-04
GO:0032259	P	methylation	2.30E-05	5.00E-04
GO:0044264	P	cellular polysaccharide metabolic process	2.50E-05	5.40E-04
GO:0003676	F	nucleic acid binding	3.40E-05	4.00E-04
GO:0016773	F	phosphotransferase activity, alcohol group as acceptor	4.40E-05	5.10E-04
GO:0044248	P	cellular catabolic process	4.90E-05	1.00E-03
GO:0006810	P	transport	4.90E-05	1.00E-03
GO:0051234	P	establishment of localization	5.20E-05	1.00E-03
GO:0046483	P	heterocycle metabolic process	8.10E-05	1.50E-03
GO:0006730	P	one-carbon metabolic process	8.10E-05	1.50E-03

GO:0005976	P	polysaccharide metabolic process	8.50E-05	1.60E-03

C, Cellular component; F, Molecular function; P, Biological process

Table IV-5. GO function term overrepresented in genes overlapping with NAT expressing loci and DNA methylation ($p<0.0001$)

GO accession	Term type	Term	*p* value	FDR
GO:0005488	F	binding	3.80E-10	3.80E-08
GO:0017076	F	purine nucleotide binding	1.50E-06	4.60E-05
GO:0001883	F	purine nucleoside binding	2.30E-06	4.60E-05
GO:0001882	F	nucleoside binding	2.30E-06	4.60E-05
GO:0030554	F	adenyl nucleotide binding	2.30E-06	4.60E-05
GO:0005515	F	protein binding	3.00E-06	4.90E-05
GO:0000166	F	nucleotide binding	4.50E-06	6.50E-05
GO:0032555	F	purine ribonucleotide binding	6.40E-06	7.10E-05
GO:0032553	F	ribonucleotide binding	6.40E-06	7.10E-05
GO:0032559	F	adenyl ribonucleotide binding	9.90E-06	9.20E-05
GO:0003824	F	catalytic activity	1.00E-05	9.20E-05
GO:0005524	F	ATP binding	6.40E-05	5.30E-04
GO:0016787	F	hydrolase activity	9.30E-05	7.20E-04

F, Molecular function

Table IV-6. GO function term overrepresented in genes overlapping with NAT expressing loci and siRNAs (p<0.0001)

GO accession	Term type	Term	*p* value	FDR
GO:0046872	F	metal ion binding	1.20E-06	1.40E-05
GO:0046914	F	transition metal ion binding	1.50E-06	1.40E-05
GO:0043167	F	ion binding	1.80E-06	1.40E-05
GO:0043169	F	cation binding	1.80E-06	1.40E-05
GO:0008270	F	zinc ion binding	2.20E-06	1.40E-05
GO:0005488	F	binding	5.30E-05	2.90E-04

F, Molecular function

Table IV-7. GO function term overrepresented in genes overlapping with NAT expressing loci and H3K27me3 (p<0.0001)

GO accession	Term type	Term	*p* value	FDR
GO:0005488	F	binding	4.30E-09	3.50E-07
GO:0043565	F	sequence-specific DNA binding	6.90E-07	2.80E-05
GO:0032774	P	RNA biosynthetic process	7.10E-07	6.70E-05
GO:0006351	P	transcription, DNA-dependent	7.10E-07	6.70E-05
GO:0046872	F	metal ion binding	2.40E-06	6.40E-05
GO:0008152	P	metabolic process	3.40E-06	1.40E-04
GO:0006355	P	regulation of transcription, DNA-dependent	3.50E-06	1.40E-04
GO:0051252	P	regulation of RNA metabolic process	3.80E-06	1.40E-04
GO:0043169	F	cation binding	3.90E-06	6.40E-05
GO:0043167	F	ion binding	3.90E-06	6.40E-05
GO:0016758	F	transferase activity, transferring hexosyl groups	1.90E-05	2.50E-04
GO:0055114	P	oxidation reduction	2.00E-05	6.40E-04
GO:0016757	F	transferase activity, transferring glycosyl groups	2.70E-05	3.10E-04
GO:0046914	F	transition metal ion binding	4.50E-05	4.50E-04
GO:0016070	P	RNA metabolic process	5.50E-05	1.50E-03

F, Molecular function; P, Biological process

Table IV-8. Conservation and diversity of the expression of lncRNAs within *B. rapa* species or between species

	B. rapa						*B. oleracea*		
lncRNAs	**RJKB-T24**	**Homei**	**Harunosaiten**	**BRA2209**	**Osome**	**Yellow Sarson**	**Reiho**	**Matsunami**	**Kinkei 201**
M15784	+++	-	+++	-	+++	+++	-	-	-
M26919	+++	+++	+++	+++	+++	+++	+++	+++	+++
M3316	+++	+++	+++	+++	+++	+++	+++	+++	-
M491	+++	-	+++	+++	+++	-	+	+	-
M17356	+++	+	-	-	-	-	-	-	-
M17153	+++	-	-	-	-	+	+++	+++	+++
M25534	+++	+++	+++	+++	+++	+++	+++	+++	+++
M259	+++	+++	+++	+++	+++	+++	+++	+++	+++
M4317	+++	+++	+++	+++	+++	+++	+++	+++	+++
M26796	+++	+++	+++	+++	+++	+++	+++	+++	+++
M4921	+++	+++	+	+	+++	+++	-	+++	-
M24531	+++	+++	+++	+++	+++	+++	-	+	-

RT-PCR results of each lncRNAs "+++" indicates strong PCR band, "+" indicates weak PCR band, and "-" indicates no PCR band.

Table IV-9. Sequences of the identified 12 lncRNAs

Name	Type	Sequence
M491	lincRNA	CAAAGAGAAAACAAGAGAAAAAGAGAGTGAGCAAGAGA GAAAAAAAAAAAAATCTTTCTTGTGAGAAAATCTTTCTCT ATAAGAAATTAAGTGTAATTTCTTGAGTGTATTTGGGGTC GGGTGTGAGTTGAGAGCTTGTAAAAATTTCTCGGGTTGAT AATAAAAGATCTGTAGCAGGTCCGGAGACGTAGGCAAGA TTGGCCGAACTCCGTTAACAATTGTGTGTGTGTTTTTCCGC TCCCATAAATTCGGTTTTCCGCAAAAATTTGACCGCATAT TTCCTAACAACTGGTATCAGAGCTTGAGTTACGGTGATCG GTCAAAAAAAATTTTGCGGAGTTACTATTCACGTGAACAG TAATTCGTGAATAGTAAATTTTGTCGGTAACATCGGTTTG TCGACGTAGGTGTTCTAATACACGGTGGTTCCAGAAACGA TGGGAGGCGAAGACGGTTCGGCGCACAGTATCGGGAAAT TTGAGGTACAGATTATGCATTTTGGAGAATGCAAATTGAA GATTATCTGTACGGAAAGAAGCTTCATCAACCGCTGAGCA AGAAACCCGAGAAGATGGATCAGGATGAGTGGGAGCTCC TTGATAGACAAGTTCTGGGTGTTATAAGGTTAACACTGTC AAAAAACGTTGCTCACAACGTTGCGAAGGAGAAGACCAC AGAAGGGCTCATGAAAGTTCTCTCTGATATGTATGAGAGA AGAGGGACGTCCGTAGATAAAGGACCTATAACACTGAAT GTCGTTGGATTCAAGCTTTTTGCAATCTAAATTAATTTGA AAGATGATATGTACTCATGTGTACGTGTAAACTTATATTA TACCCAAAACTTTACTTTACTATTTTTATCATCAAAATATT ATAGACTTCGCTAGTTCTTGAGATTTGCTTTCTATATTAAA CATGGGGAGTG
M15784	lincRNA	CCAAAATCCTTCCATCTCTCTTTCTCTCCTGTGTCTCTTCTT CAAAATCTATTGCCCAGATAATTAATCTTCTTCTCTTTCTC TCTCTTATATCTCTTTAAAATCCCTAAACTTATTTGTTTTTC CATTTCTGCAATGGATTCTTCAATTCTTTTTTGTTTTCATTG ATGTGATTATCTTTTCATCTAAGTGTATTTCTTTCAACACA GTGTTTAAATCGAAAGAGTATTATCCTCTCGATGAGTTCA TCTACCATAGCTTCTCTGTCAATTTGTTTGAGGTCCACCAC GATATAAAAGTTCTAAACAATTTTATGACGCATAAAGGG

		GAGATATCTTCGACTCACATGCTGGTAAAGGCACATGGCT GTATAACAAGAAAAAGATTTCTGTTAACTGGTAAATTTTG ATTAGATGTTACCAAATAATTCTCGTGAGCATATACAAGT TTGGTTAGTTGTTAAAGCTTTATGTTAAATTTCTTCGTTTC AAAATAGTTATGGTTATATGATACATGGCTATAAATTTCT ATAGAATATGTTTCTCTTATCTGATACTTGTTTGGTTAGAT GCTACAAAATATATTGTCGATTTAAAATAGTTGATAAGAT GCTAAAAATAACATCGATAGCTGATATGTGGCTTGGTAGA TGATACCAAAAAGGTTACCATGCTAACTTTACCATTATCT ATTCAACTTTTATATGCTTAACCATTAATAATTATTATACT ATATATAAACCAATATTACATCCAAGTTCATGTTTTATAT GGAAATATCATAATA
M24531	lincRNA	GCTTCTCTAGTAGATTTTACCGTAAAATCCATTAAACTAA TATTTAAACAAGAAAGTAAATTTAAGCATTTGAATCGAGC CTTATTTATATCCTTTGGGGATGGCTCTAAATTCTTAAGTA TGAATGGAGATTCATATCCAACATATATTCCCAACTCATT AGAAAACTCATCTTAAAGCTATATGGTGGAGCAACAAAA ATTGCGTATTGCACATTTGAAATTCTTCGATGGGAATGTT TGGTTTCTGGCCCAAAACCAGTGATGGGAGAACTATTTAT AACTTATTGGCTTAACGCGAATTATGTTATTCAAATGTTC AATACTTGTTTCAGTGAACATGTCATATACTTAAGACATG AATTCACCAATATTATAAAGATGCATAGTGGGTGATCAAT CCACCAACATCTCTTCTATTTATGTCAATAATTTCATTTAG ATGGTGAAAACCGTATTACCATTTTATCTTTTAAGCAATG AACATATG
M26919	lincRNA	GGAGAAAGCATGGGAATGCAAGGAAGAACGGCTCTACAG TAGAGAGAAGAGATCGGGTTAGATCATATTGGGAGTCTA CAAGCAAAGAAATGAAAAAAGAATTACTTAGGATTAAGG TTAGTGATCTTAAGAGTCACTTTAGTGCATCCAAGGATGG CGATGCAAATGATATAATAACTGAAGCTTTGTCCTTTTGT GAAGCTAATAAGACTTGGCGGTTTTGGGTTTGCTGCAAGT GTAGTGAAAAGTTCAAGGATTCTGAGTCTTATATGCAGCA TATTGTAGGGGAGCATATGGGTAA

M259	incRNA	CTCCTCTTCTCTCTTCGATAATTTCGCCTTTCTTCTTCTTCT TCTATCTCATTGTTTCTTCCTCGATCAAGATTCATTTTCAA AGATCTCATATCTTGTGGGATTAATCAATAAAAATAGTTA GAAAAAAACATTGATGCTTAACCAGCAGTTCTTGGTCTTT GAGACACATGTCTCCGATTATACTCAGTGAGATCTTCCTC TCTGGGTTTATGCTAAACTCCACAGTCAGGCGCAGGACCC ATCTCGTTCAATCTTTCTCTGTTGTCTTCCTTTACTGGCTTT ACTACGTCTCATAGTTTCTGAAACTCTCCAAGCTTAATTTA TCTTTCATATA
M4317	incRNA	GCCTAGCAATATCCTTTTCCGTTCAGTCTTCGAGTCCTACA ACCTTAAGCCAAAATTGTTCTCTTCAGTTCACTTGTACTTT GTACGTCATTTCTGGTCTGTAATTAAGCTCACTTGTCCTTT GGTTGGCTGTTTTTCTCTTTGCGTATCAACATTTTCGTACC ACCACATTTTTGTGGCTGCCTTCAGTGTATTTATATACTGT CGTTTTGCTTAACA
M25534	incRNA	TGGAGTTTGAACTACTTTAGTACCCGGTTACATGGAATTG AAAGATGAACAAAAAATAACCAGCAGAAGTGACATAGA AAAATGGTTTCGAAATGGTCTTTCTGGATGAACATTGATG ACAATATTTCACATAAGGACTGGAGGAAAAAACGAGAAA GTATATTTTGGATAGATGATGACTGACCAATTAGAAAACG ACCGAACATGAAAATTAGCAGAGGATTTGCTGGAAAATA AATTTGAATAATCGGAAATTGGGAATCATTACATTCATGA ACATCTAAGTTGAGAATATTTGAAAGAGGAGAAAACTAG AGCTAAGAGTGTTAAAGACCGTCTTTATGATCTTCGGAA
M3316	NAT	AGGGAGCAGCCACGGGCACAAACCGAGCTCGTCGGCGTT TCCGATCACCGGAACGGCGGTTAAGTCGACGGTGGTGGA ATCGCCGAGGACTCCAAACGTGGAAGGTGAGATTTTCGC GACGGCTGAGTCGAAGAAAGACTCGACGGGGACTCGTTC ATGGCGGCTAGCGAGAGGATTAGCCGAGTGGATGTCGGA GTCGCAGGTGACGCAGAGTGAGGCGGCGTCGGCTTTGCA GGTGACGGCGGCGGGAGCTTGTTCGCAGACTTCACAGAG GTACACGCGCTCGTGACGCGTGAAGGAGTGGATGCTTGT GTCGCAAGTGATGCATAAGAAGGCTGAGTCGAATCGGCA GTACACGGCGGCGGAAGCTGATTTACAAGCGTCACAGGA ACGCGCCGCCGCG

M4921	NAT	TTTTAACGCCAACGCGGCCAAGCTTGGTCATAGCTTTAAC AAAGGCGGCGTTGAAAGCAGCTGAGTTCTTGGCCCAAGA GTCAACGGTGGGCCTAGAGCGACGATTGGTGTAGAGAAC TTGGTCGGAAGTGAAGAGTCCTTTGCCTTGTTGGAGATTC CTGAAGTAAATGTTGTCTAACTTCTTTGGAGTGATTGGGT CAATGTTAATAGCGATTCTTGGGTCAACGTTCTTGGGACA AGCCAATCGAAGGTCCTTAGCGTAGCCTTTATTTAGAGTT GGATCAATGATATGTGTGCGGTTGAAGTTGTAGATTCTGT TGAATACTTTGCTACAATGGGCGAATCCGAGGGTGTGAGC CGCTGCATGGAGAAGTTTGAAAAGGTTTAAATTAAATGTA TTTTGAT
M17153	NAT	GTTTATCAGACGTGTATATAAAAGAAAATATTATACAAAA CCACTTAGGTGAAAAGATTATTGTTAATGTCCATAGGTGA AGATGTTTTGTTCAACACAATCTCAAACTCACTTATCTCAT TGTCTCCTGCGCCCTCTTCTTCTTCATCTTCGTAATCATCC TCCCCGCTCTCTCCTTCTTTATTATCAACTTTGTCCCTCTG ATTAGTCACTCAGCCTTCGTTTCTTGTGAAAGAAGAAACT GTCTCGGTTAACCAATAGAGGCAATATTGGTAGTGGCCCA GTTTTGTTCCTCTCGTTATACAACGCTTCTAGCTGGTGAAA ACAAGGGCATGTCTTGGAATCAAGGGGTCGTCTCTTGTTG CTCTCTTTGACTTTCTTGAAGTACTTGTTGATGTTTTCCCA CTTCTCTTTGCATCGTTTCGTGCTTCGATTGTATCCGAATC TTCTTATCTCTGCAGAGATCTCTTCCCATAACGGCCCCTTA GTACCGTTTTCTAGATAGTTAGCTTCAAGATTCTTTCTTAT TCTTATTAAGGCCTCGACCTCAGTTTTAGGCCATCTTGAA GAACTAGGAGATGCCATTGGAAGTCCCTATCTTCATTATT GTTGTATCTAGTACAACCTGACTTGTCTCTTCGCTCTCAAA TGTAATTGAATGTTGTTCACTTTGATATTGTTTCCTTTGTG ATACTTTATAATTCTGCTGATGTGGTTGTTGTTGTCCTCCT GATATTTTTTGTAAGAATGATATGATTGCAACGTCTTTAG CCGCAATATTAGACCTCTCATGGACTAGTGTATTGTGTTC ACTATTGATTCTCTCTACCTCTTGGACCCTCCAAGCTTCTT CTCTTGAGATTCTCTCTCTCTCGCCAGTCTCCAAAGCTTTC AAGAACCTTTTATGCATCTTCTCTTGTTTCTCCATCAACTC TTTTGTCAACTTCCTCTTCTTCCTCG

M17356	NAT	AAGAAAACAAAAACCTAAGAGATTGGAAAGAGAAAGAG TCAACGAGAGGTCGTGATGGGATTAGGTTGAAGCATGGA AGGAAGAACAAGTGGGGACAATAGAGGGATTGGATCTGT GCCGTTTTGACAGAGCAGAGACGCAGACGCGGAAGGAGC TACGGTGGAAGGAGGAGGTCTTGGTGGTACTAAGAGAGG ACAAGGAGCTCGTCTCTGAAGACGGCAGCTTAAAGATCC CCGGAATACACAGTCGTCGTTGTGAACACTCATCTCAAAA CCTGGAAATGGAACTCCTTTACGCCCTGTTTCCTCCATTTT CTCATCAATTCTTGATTTTGTGAACTGGTGTTTTGGATTTG CAATGTTAGTGTTGGTGCTTGTTTTGTTCTATTCAGCTTCT TTGAAGGGTTTAAATATGATTTCACAATGGACGATTATTA GGAAACTAGGGGTTGCTTTACTTGAGGATTTTAATGGAGT AATTAATTGAACGGATAACAATACCAACTAAGAAAATGG CTCTAAGTCTAATAAAAGAAAAAGAAGTTGGGTGAAAGT TTGGCGGTTGTGTGGTGCGTGTATTGATTGACTCAATATG GCAGAAAAAAACTATTTCAACTGGTCAATGGTCAAAAGA TTGGATAGGTATTAG
M26796	NAT	GAGGAAAGGCTTGCGGTGGATACCTAGGCACCCAGAGAC GAGGAAGGGCGTAGTAAGCGACGAAAAGCTTCGGGGAGT TGAAAATAAGCATAGATCCGGAGATTCCCAAATAGGTCA ACCTTTTAAACTGCCTGCTGAATCCATGAGCAGGCAAGAG ACAACCTGGCGAACTGAAACATCTTAGTAGCCAGAGGAA AAGAAAGCAAAAGCGATTCCCGTAGTAGCGGCGAGCGAA ATGGGAGCAGCCTAAACCGTGAAAACGGGGTTGTCTGAC CCGAGTAGCATGGGGCACGTGGAATCCCGTGTGAATCAG CAAGGACCACCTTGCAAGGCTAAATACTCCTGGGTGACC GATAGCGAAGTAGTACCGTGAGGGAAAGGTGAAAAGAAC CCCCA

Chapter V

Transcriptional association between long non-coding RNAs and active histone marks in *Brassica rapa* L.

Abstract

The present study examined the relationship between lncRNA expression level and two active histone modifications (H3K4me3 and H3K36me3) in *Brassica rapa*. Both histone marks were enriched in the chromatin regions encoding lncRNAs, especially around the transcription start site. The transcription level of long intergenic noncoding RNAs was positively associated with the level of H3K4me3 and H3K36me3, while this association was not observed in natural antisense RNAs (NATs) and intronic noncoding RNAs. We have identified 12 mRNA and NAT pairs in 14-day leaves in RJKB-T24 using previous datasets of *Foc* inoculation (24 and 72 HAI in RJKB-T23 and RJKB-T24) and paired genes overlapping lncRNAs (in RJKB-T24) and examined the modification states of active marks (H3K4me3 and H3K36me3) and repressive mark (H3K27me3). Three pairs did not have any histone modifications, one pair had only H3K4me3 or H3K27me3 and four pairs had H3K4me3 and H3K36me3. Two pairs (Bra016382/MSTRG.19710 and Bra033549/MSTRG.1355) had bivalent active and repressive histone modifications H3K4me3 and H3K27me3. This result suggests that the bivalent modification might have some role in the transcriptional regulation in biotic stress.

Keywords: *Brassica rapa*; lncRNAs; bivalent histone marks; transcription; gene expression

Introduction

Brassica rapa L. includes leafy vegetables such as Chinese cabbage (var. *pekinensis*), pak choi (var. *chinensis*), and komatsuna (var. *perviridis*), as well as root vegetables such as turnip (var. *rapa*) (Lv et al. 2020). These vegetables provide nutrition, vitamins, minerals, dietary fiber, and health-promoting substances. Most of the modern cultivars of these vegetables are F_1 hybrids (Fujimoto et al. 2018). In all eukaryotic cells, approximately 145-147 bp of DNA wraps the histone octamer of the nucleosome in the chromatin in the nucleus. A histone octamer has two copies of each of four core histone proteins, H2A, H2B, H3, and H4 (Kim 2013, Itabashi et al. 2018, Talbert and Henikoff 2021). The histone proteins have an N-terminal tail, and modification of amino acid residues in the histone tail can change gene expression. Histone modification includes methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation (Li et al. 2007, Bannister and Kouzarides 2011, Zhao et al. 2019). Methylated lysine in histone tails can be found in mono (me1)-, di (me2)-, or tri (me3)-methylated states at lysine 4 (K4), 9 (K9), 27 (K27), and 36 (K36) of histone H3 (Li et al. 2007, Bannister and Kouzarides 2011, Demetriadou et al. 2020). Histone modification has an important role in plant development and response to stresses through regulating chromatin structure and gene expression (Kim et al. 2015, Meyer et al. 2015). In plants, tri-methylation of histone H3 lysine 4 (H3K4me3) and H3K36me3 are histone marks that are associated with transcriptional activation, while H3K9me2 and H3K27me3 are associated with transcriptional repression (Kim et al. 2015, Fujimoto et al. 2012a, Quadrana and Colot 2016). Bivalent chromatin states where active and repressive histone marks co-exist have been identified (Sequeira-Mendes et al. 2014, Qian et al. 2018, Blanco et al. 2020), and this chromatin state may trigger plant stress-responsive gene expression (Zeng et al. 2019, Chapter II). LncRNAs are longer than 200 nt in length. LncRNAs are classified into three major groups: long intergenic noncoding RNAs (lincRNAs), natural antisense RNAs (NATs) transcribed from the complementary DNA strand of their associated genes, and intronic noncoding RNAs (incRNAs) derived from introns (Ponting et al 2009, Cech and Steitz 2014, Ariel et al. 2015, Chekanova 2015, Rai et al. 2019). Chromatin immunoprecipitation sequencing (ChIP-seq) allows us to identify the distribution of the histone modifications. The association between histone modification and expression levels in genes has been analyzed in some plant species including *B. rapa* (Chapter II, Asensi-Fabado et al. 2017, Akter et al. 2019, Kim 2021). However, only a limited number of studies have examined the relationship between lncRNA expression and histone modifications in plants (Hung et al. 2020, Chapter IV). We examined the relationship between lncRNA expression level and active histone modifications (H3K4me3 and H3K36me3) in *B. rapa*.

Materials and Methods

Plant materials and growth conditions

After surface sterilization, the seeds were placed on Murashige and Skoog (MS) agar supplemented with 1% (w/v) sucrose and grown under long day (LD) condition (16 h light/8 h dark) at 21 °C. First and second leaves were collected from 14-day-old plants of RJKB-T24.

RNA extraction and lncRNA identification

Total RNAs from first and second leaves were extracted by SV Total RNA Isolation System (Promega) in RJKB-T24 for RNA-sequencing (RNA-seq) for detection of lncRNAs. A more detailed description of lncRNA identification can be found in Shea et al. (2019).

Detection of epigenetic states in lncRNA coding regions

Chromatin immunoprecipitation sequencing (ChIP-seq) data were previously produced using the same line, tissue, and developmental stages but were harvested independently (Chapter II). We used these data to investigate the epigenetic states (H3K4me3 and H3K36me3) of lncRNA coding regions in *B. rapa*. Data covering genomic regions encoding for lncRNA in chromosomes A01 to A10 were extracted from the ChIP-seq reads using anti-H3K4me3 (Millipore, 07-473) and H3K36me3 (Abcam, ab9050) antibodies that were mapped to the *B. rapa* reference genome v.1.5 using Bowtie2 version 2.2.3.

Sequential ChIP-qPCR

For sequential ChIP analysis, one gram of first and second leaves of RJKB-T24 was used. Anti-H3K4me3 antibodies (Millipore, 07-473) and anti-H3K27me3 antibodies (Millipore, 07-449) were used for the first and second ChIP, respectively. Experiments were performed as described by (Chapter II, Finnegan et al. 2011). The average and standard error (s.e.) of three biological and more than three technical replicates are presented. Amplification of target mRNAs/NATs and non H3K4me3 and H3K27me3 marks (Bra011336) were compared by qPCR using immunoprecipitated DNA as a template to calculate the enrichment of bivalent H3K4me3- and H3K27me3-marks of the targeted mRNA-NATs pairs (Bra016382-MSTRG.19710 and Bra033594-MSTRG.1355). The difference in the qPCR amplification of the input DNA as a template was used to correct the difference between the primer pairs. The primer sets used in this study are listed in Table V-1.

Results

Comparison of the H3K4me3 and H3K36me3 states in the lncRNA coding region

We identified overlapped genes between two data sets, DEGs following *Foc* inoculation (24 and 72 HAI in RJKB-T23 and RJKB-T24) and paired genes overlapping lncRNAs (in RJKB-T24); twelve mRNA and NAT pairs were identified (Miyaji et al. 2017). We identified 1,444 long intergenic noncoding RNAs (lincRNAs), 551 natural antisense transcripts (NATs), and 93 intronic noncoding RNAs (incRNAs) in 14-day first and second leaves in RJKB-T24 (Shea et al. 2019). In the present study, the genomic regions encoding lncRNAs that have H3K4me3 and H3K36me3 marks were examined. 704 of 1,173 (60.0 %) lincRNAs, 458 of 529 (86.6 %) NATs, and 47 of 92 (51.1 %) incRNA coding genomic regions had H3K4me3 marks (Figure V-1). 574 of 1,173 (48.9 %) lincRNAs, 380 of 529 (71.8 %) NATs, and 38 of 92 (41.3 %) incRNA coding genomic regions had H3K36me3 marks (Figure V-1). H3K4me3 and H3K36me3 were enriched in the transcribed region of lncRNAs, especially around the transcription start site, similar to the genic regions (Figure V-2). The expression level of lincRNAs was higher when the encoding regions had H3K4me3 or H3K36me3 marks than without H3K4me3 or H3K36me3 marks (Figure V-3). In NATs and incRNAs, there was no difference of expression level between with and without H3K4me3 or H3K36me3 marks on their encoding regions (Figure V-3).

The histone modification states (active marks, H3K4me3 and H3K36me3; repressive mark, H3K27me3) of the genomic regions encoding these 12 mRNA and NAT pairs in 14-day leaves in RJKB-T24 were examined (Table V-2). Three pairs did not have any histone modifications (Table IV-10). One pair had only H3K4me3 or H3K27me3 and four pairs had H3K4me3 and H3K36me3 (Table V-1). Two pairs had H3K4me3 and H3K27me3, indicating bivalent active and repressive histone modifications (Figure V-4, Table V-2). One pair had all three histone modifications (Table V-2). To examine the simultaneous occupancy of active (H3K4me3) and repressive (H3K27me3) histone modifications, sequential ChIP-qPCR in 14-day leaves in RJKB-T24 was performed in two mRNA and NAT pairs (Bra016382/MSTRG.19710 and Bra033549/MSTRG.1355) that have both H3K4me3 and H3K27me3 marks (Akter et al. 2019). The genomic region encoding Bra016382/MSTRG.19710 and Bra033549/MSTRG.1355 showed enrichment for the second modification similar to *BrWRKY48* that has been shown to have bivalent active and repressive histone modifications by sequential ChIP-qPCR (Figure V-4).

Discussion

Histone modification plays a role in the regulation of mRNA expression (Kim et al. 2015, Meyer et al. 2015), but it remains uncertain whether the role of histone modifications in the expression of lncRNAs is the same as that of mRNAs. Unlike mRNAs there was no negative association between the presence of H3K27me3 marks and lncRNA expression levels (Akter et al. 2019), we have also examined H3K27me3 distribution in the genomic regions encoding lncRNAs in *B. rapa*. incRNAs with encoding regions having H3K27me3 marks showed higher expression levels than those without H3K27me3 marks; this trend is the opposite of mRNA. There was a lower proportion having H3K27me3 marks in the region encoding lncRNAs than that in mRNAs. In this study, we identified lncRNAs with active histone modification (H3K4me3- and H3K36me3-marks) in their encoding regions and examined the relationship of histone marks to expression level. The genomic region encoding lincRNAs and incRNAs had H3K4me3 or H3K36me3 marks in a similar proportion as mRNAs, whereas regions encoding for NATs were highly enriched. Like genic regions, both active histone marks were highly enriched around the transcription start site of lincRNAs, incRNAs, and NATs. In lincRNAs, the enrichment of active histone marks either H3K4me3 or H3K36me3 in their encoding regions resulted in a higher expression level like in mRNAs, while this association was not found in incRNAs and NATs. In *A. thaliana*, H3K4me3 and H3K36me3 marks were positively correlated with the expression levels of lncRNAs, while H3K9me2 and H3K27me3 marks were less correlated (Hung et al. 2020). These results suggest that in *B. rapa* lncRNAs do not follow the same rules as mRNAs regarding histone modification states.

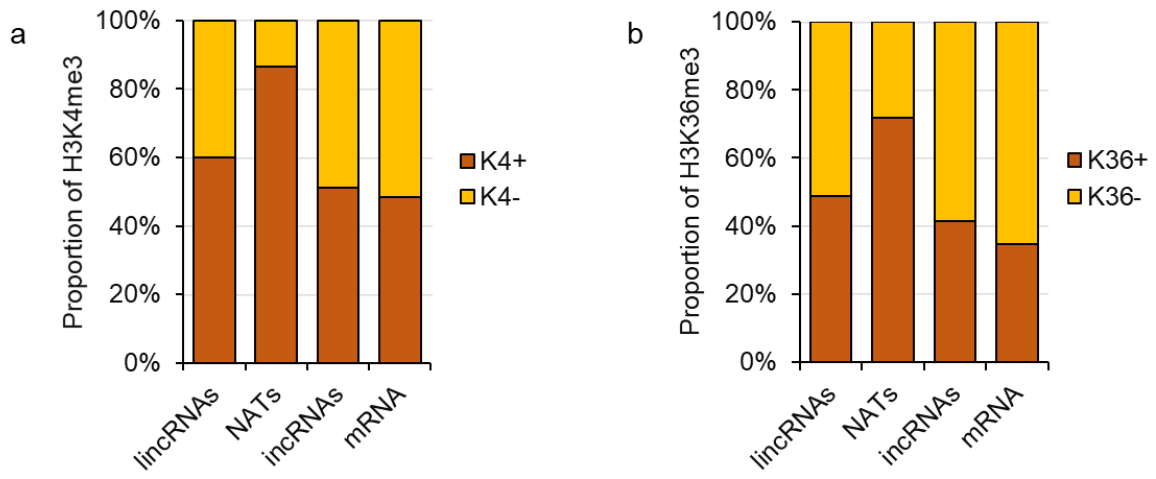


Figure V-1. Proportion of each type of lncRNAs and mRNAs with (+) or without (-) H3K4me3 (A) and H3K36me3 (B) marks on the coding region in RJKB-T24. K4 and K36 represent H3K4me3 and H3K36me3, respectively.

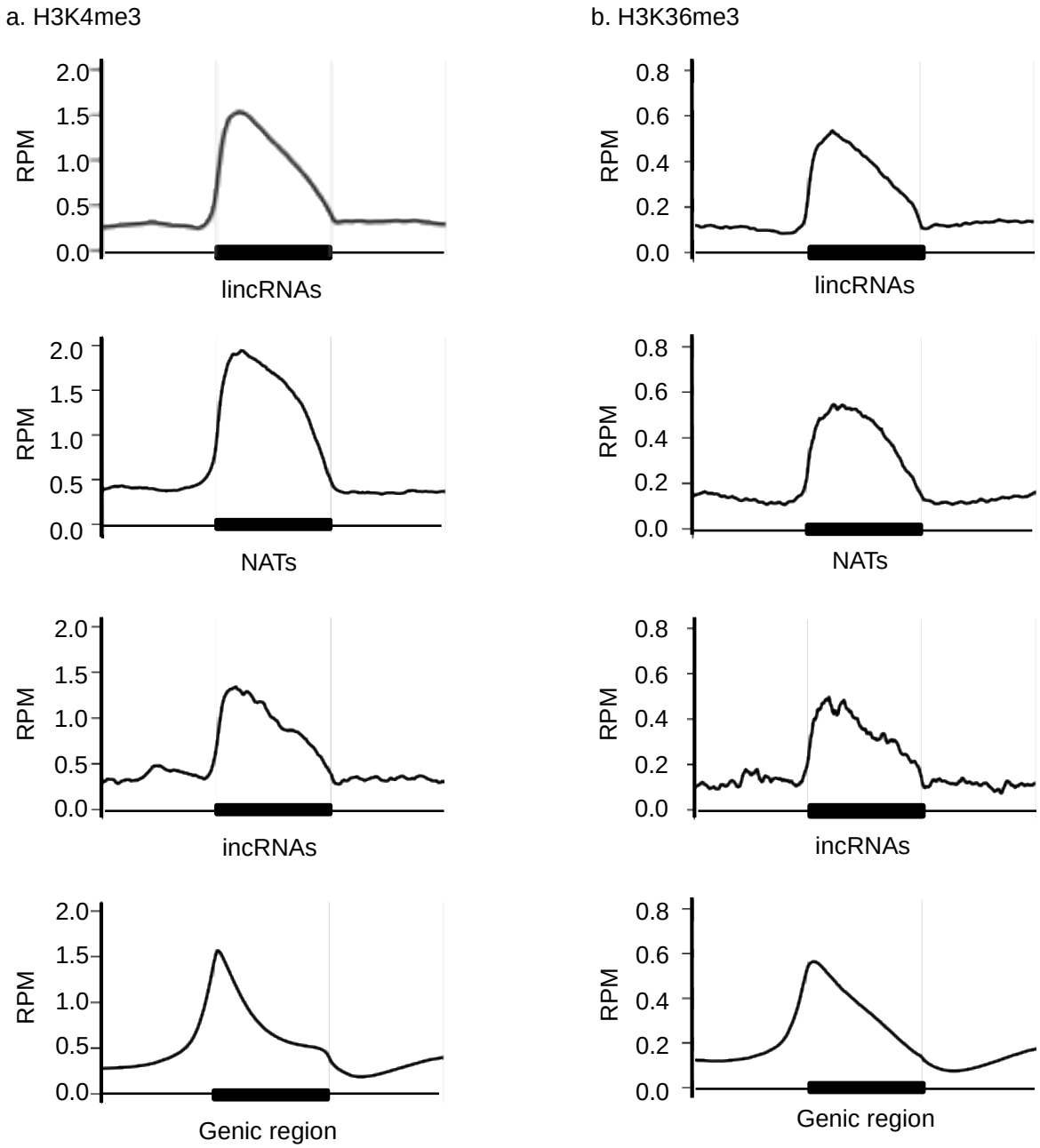


Figure V-2. Enrichment of the H3K4me3 (a) and H3K36me3 (b) in the lncRNAs and genic region. The X-axis represents the region with 1 kb upstream and 1 kb downstream. The Y-axis represents the reads per million (RPM).

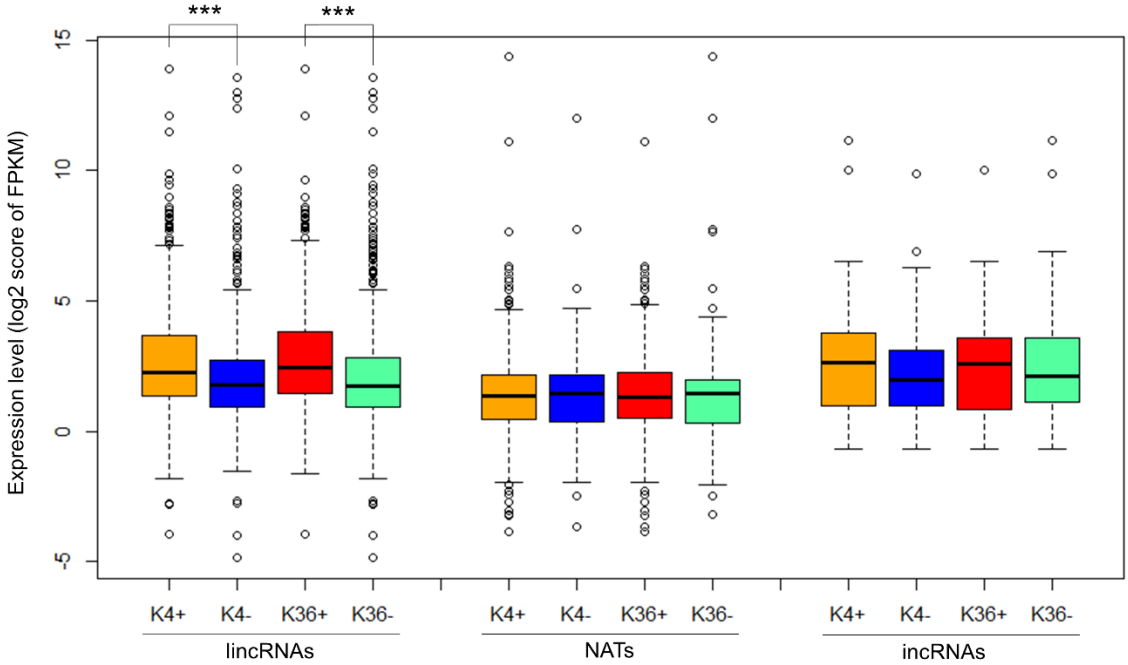


Figure V-3. Boxplots of the expression level of each type of lncRNAs with (+) or without (-) H3K4me3 (K4) and H3K36me3 (K36) on the coding region of RJKB-T24. ***, $p < 0.001$ (Student *t*-test).

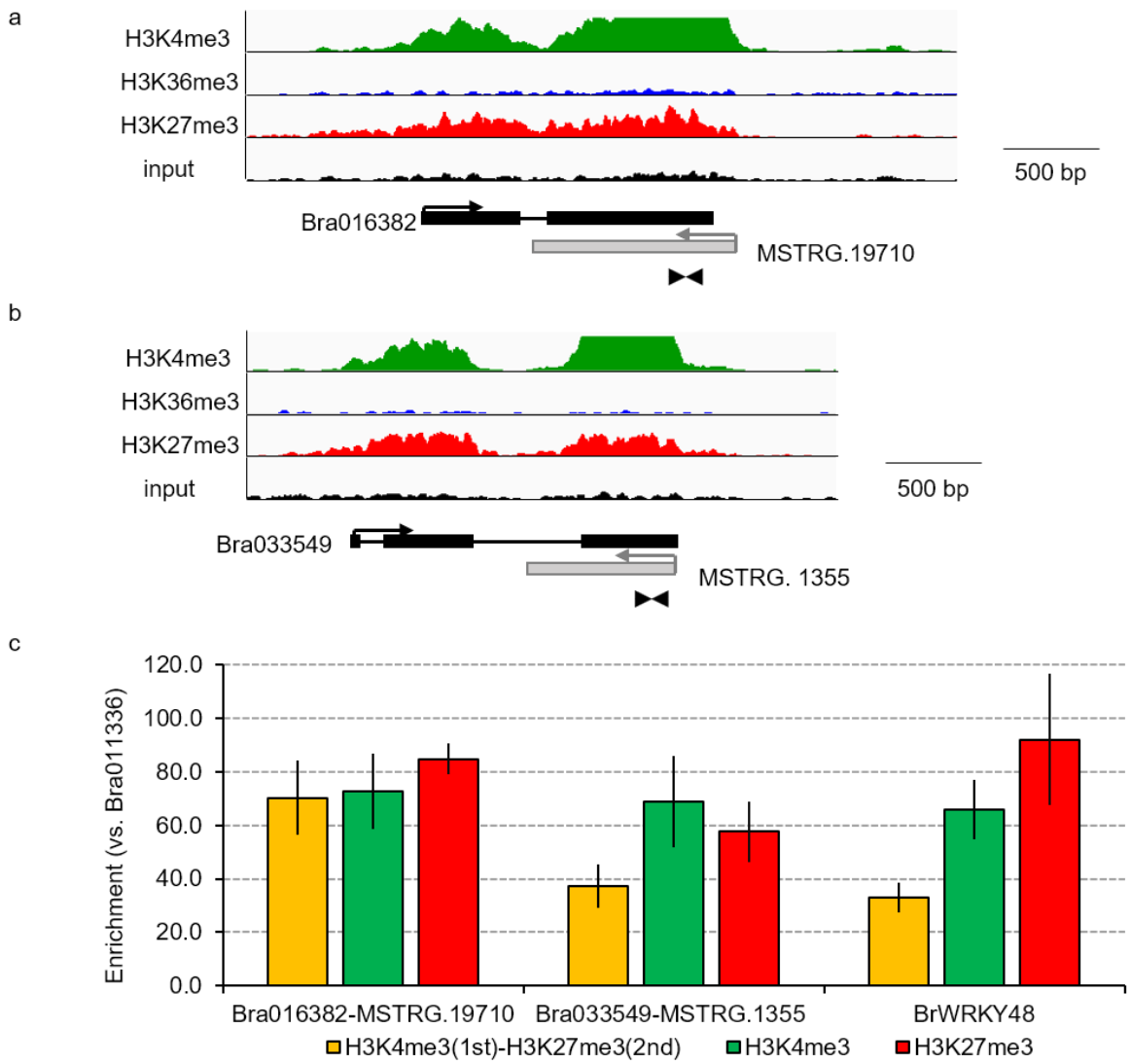


Figure V-4. Bivalent histone modification in the genic regions and their paired NATs. (a,b) Visualization of H3K4me3, H3K36me3, and H3K27me3 peaks by Integrative Genomic Viewer (IGV) by ChIP-seq. Black and gray boxes represent the exon regions of genes or NATs, respectively. Arrows represent the direction of transcription. Arrowheads represent the position of primer sets for sequential ChIP-qPCR (c) Sequential ChIP-qPCR analysis in 14-day leaves in RJKB-T24. Bra011336 that does not have H3K4me3 and H3K27me3 is used as reference gene for qPCR. The previous sequential ChIP-qPCR showed that *BrWRKY48* has both H3K4me3 and H3K27me3 marks (Figure II-12). Values are means ± standard error (s.e.; three biological and technical replicates) of relative H3K4me3, H3K27me3, or H3K4me3/H3K27me3 levels.

Table IV-1. Sequence of primers used in this study

Name	Forward primer (5'-3')	Reverse primer (5'-3')
Sequential ChIP-qPCR (H3K4me3 and H3K27me3) of the targeted mRNA-NAT pairs		
Bra033549.MSTRG.1355	CAAACGGCGGAAACGCTGCGAG	CCGAAGTTCTTCTCGATCTCAGG
Bra016382.MSTRG.19710	CGAACTGTAAGTTCTGTCGCG	GAACCACGCTTAGGCTCAGTCG
For reference		
Bra011336	GGGAATCAGCTTTTGTGGTG	AAATGACCCGATCAGCAAAG

Table V-2. Histone modification states in the selected 12 mRNA and their paired NATs

mRNA and NAT pair	H3K4me3	H3K36me3	H3K27me3
Bra009234 MSTRG.25721	YES	YES	NO
Bra025668 MSTRG.13790	YES	YES	NO
Bra029414 MSTRG.4734	NO	NO	NO
Bra033549 MSTRG.1355	YES	NO	YES
Bra034404 MSTRG.11674	YES	YES	NO
Bra035320 MSTRG.26084	NO	NO	NO
Bra039006 MSTRG.15696	NO	NO	NO
Bra028523 MSTRG.16709	YES	YES	NO
Bra029946 MSTRG.1546	YES	YES	YES
Bra020438 MSTRG.3318	NO	NO	YES
Bra003511 MSTRG.17011	YES	NO	NO
Bra016382 MSTRG.19710	YES	NO	YES

Chapter VI

General discussion

In this dissertation, we analyzed the H3K4me3-, H3K9me2-, H3K27me3-, and H3K36me3-marks using two inbred lines of Chinese cabbage (*B. rapa*) and its heterotic hybrid using ChIP-seq. At first, we have characterized two active histone marks H3K4me3 and H3K36me3 using the parental lines at 14-days old first and second leaves. We added H3K27me3 data from the previous datasets (Akter et al. 2019). We found respectively 47% and 34% of total genes having H3K4me3- and H3K36me3-marks, and both marks were enriched in the genic regions specially TSS. The average expression levels of the genes having H3K4me3- and H3K36me3-marks were higher than the average expression levels of all genes. We categorized the tissue specificity in each gene by RNA-seq from six different tissues of Chinese cabbage. Both the H3K4me3- and H3K36me3-marked genes showed low tissue specificity, while H3K27me3 was highly tissue specific. H3K36me3-marked genes were much lower than that of H3K4me3-marked genes for tissue specificity which suggests H3K36me3-marked genes are more constitutive in gene expression. The presence of the H3K36me3 marks at the subgenome levels (LF, MF1, and MF2) increased the average expression and tissue specificity between paralogous pairs, while H3K4me3 was not associated gene expression level between paralogous pairs (Chapter II).

The coexistence of H3K36me3- and H3K27me3-marks showed a similar pattern to the H3K36me3-marked genes for gene expression and tissue specificity. The bivalent active (H3K4me3) and repressive (H3K27me3) histone marks were identified and confirmed by the sequential ChIP-qPCR. About 9% of the total genes were enriched with bivalent H3K4me3- and H3K27me3-marks. The expression levels and tissue specificity of the bivalent modification showed in a similar way to the repressive histone marks, H3K27me3. More than 20% genes were bivalently modified under these two stress conditions *Fusarium oxysporum* f. sp. *conglutinans* (*Foc*) inoculated and 4 weeks of cold treatment (stress datasets from Miyaji et al. 2017 and Shea et al. 2019). Our results suggest that bivalent active and repressive histone modifications might have a role in the transcriptional response under *Foc* inoculation and cold stress (Chapter II).

We analyzed the genome-wide distribution of H3K4me3-, H3K9me2-, H3K27me3- and H3K36me3-marks in heterotic hybrid of Chinese cabbage. Parental alleles had a variation in for all four histone marks. We found that less than half of the H3K4me3-, H3K9me2- and H3K36me3-marked DMGs were inherited into the F_1 hybrid while the number of inherited DMGs were one third for H3K27me3 marks. The inheritance of all four histone marked genes

was not specific to any allele, most of the inherited DMGs were showed high parental allele. We found only twelve DMGs for H3K36me3 marks that have higher enrichment than that of both parental lines. A low percentage (5-20%) of the inherited DMGs showed their enrichment levels between the parental lines (intermediate parents) or similar levels of enrichment with the low parent (low parent) (Chapter III).

We analyzed the transcriptional responses of different lncRNAs (lincRNAs, incRNAs, and NATs) by the association of the different histone marks (previous datasets for lncRNA, Shea et al. 2019). H3K4me3-, H3K27me3-, and H3K36me3-marks enrichment in lncRNAs were followed the similar trend to mRNAs and they were highly enriched in the body region of transcripts specially around the TSS. The genomic region encoded with each of three types of lncRNAs had a lower proportion of H3K27me3-marks compared to the mRNA. H3K4me3- and H3K36me3-marks did not have the similar trend to H3K27me3-marks for accumulation in the genomic regions encoded with the lncRNAs. Both active histone marks accumulated with an equal proportion of mRNAs for their genomic regions covering lincRNAs and incRNAs while NATs showed opposite pattern of H3K27me3-marks. H3K27me3 marks repress expression level in mRNA, however, the lncRNA did not show this negative relationship with the expression levels. The enrichment of H3K27me3 did not involve in the transcriptional regulation for the lincRNAs and NATs. The incRNAs having H3K27me3 increase the gene expression which is completely opposite to the mRNA (Chapter IV).

H3K4me3- and H3K36me3-marks enriched the lincRNAs increases the expression levels like to the mRNA, while these two active marks did not have any impact on the gene expression levels in incRNAs and NATs. We identified two bivalently modified mRNAs and its paired NATs (Bra033549/ MSTRG.1355, and Bra016382/ MSTRG.19710) using differentially expressed genes and their paired lncRNAs following *Foc* inoculation (*Foc* inoculation datasets from Miyaji et al. 2017). This bivalent modification might have some role in the transcriptional regulation in *Foc* stress, however, we need further study for the confirmation. (Chapter V).

In conclusion, we showed the genome-wide distribution of H3K4me3- and H3K36me3-marks in Chinese cabbage and their role in transcriptional regulation. Genes having bivalent histone marks might have some role in biotic and abiotic stress response. Genome-wide inheritance of the four major histone marks from parents to F_1 hybrid was high parent specific. The regions covering histone marks in lncRNAs were not followed the similar pattern to the mRNAs for the enrichments and transcriptional regulation. However, we studied only in the 14-days old first and second leaves, and this research can be expanded using different tissues developmental stages for more clear information.

Publications List

Original research articles

1. Hasan Mehraj, Satoshi Takahashi, Naomi Miyaji, Ayasha Akter, Yutaka Suzuki, Motoaki Seki, Elizabeth S. Dennis, Ryo Fujimoto. Characterization of histone H3 lysine 4 and 36 tri-methylation in *Brassica rapa* L. Frontiers in Plant Science, 12, 659634. [June 2021] [doi] 10.3389/fpls.2021.659634

2. Hasan Mehraj, Daniel J. Shea, Satoshi Takahashi, Naomi Miyaji, Ayasha Akter, Motoki Seki, Elizabeth S. Dennis, Ryo Fujimoto and Kenji Osabe. Genome-wide analysis of long noncoding RNAs, 24-nt siRNAs, DNA methylation and H3K27me3 marks in *Brassica rapa* and related species. PLoS ONE, 16(3), e0242530. [March 2021] [doi] 10.1371/journal.pone.0242530

Review article

1. Hasan Mehraj, Ayasha Akter, Naomi Miyaji, Junji Miyazaki, Daniel J Shea, Ryo Fujimoto and Md. Asad-ud Doullah. Genetics of clubroot and Fusarium wilt disease resistance in Brassica vegetables: The application of marker assisted breeding for disease resistance, Plants, 9(6), 726. [2020] [doi] 10.3390/plants9060726

Book chapters

1. Takumi Okamoto, Xiaochun Wei, Hasan Mehraj, Mohammad Rashed Hossain, Ayasha Akter, Naomi Miyaji, Yoshinobu Takada, Jong-In Park, Ryo Fujimoto, Ill-Sup Nou and Masao Watanabe. Chinese cabbage (*Brassica rapa* L. var. *pekinensis*) breeding: Application of molecular technology. In JM Al-Khayri (eds), Advances in Plant Breeding Strategies: Vegetable Crops - Fruits, Leaves, Stems. Springer, Cham. pp 59-94. [doi] 10.1007/978-3-030-66969-0_2

2. Mst Arjina Akter, Hasan Mehraj, Takeru Itabashi, Tomoe Shindo, Masaaki Osaka, Ayasha Akter, Naomi Miyaji, Naoki Chiba, Junji Miyazaki, Ryo Fujimoto. Breeding for disease resistance in brassica vegetables using DNA marker selection. In: AKM Aminul Islam,

Mohammad Anwar Hossain and AKM Mominul Islam (eds), Brassica breeding and biotechnology. IntechOpen, London, pp 1-16. [doi] 10.5772/intechopen.96263

3. Honghao Lv, Naomi Miyaji, Kenji Osabe, Ayasha Akter, Hasan Mehraj, Daniel J. Shea and Ryo Fujimoto. The importance of genetic and epigenetic research in the *Brassica* vegetables in the face of climate change. In: Kole C. (eds) Genomic designing of climate-smart vegetable crops. Springer, Cham. pp 161-255. [March 2020] [doi] 10.1007/978-3-319-97415-6_3

Acknowledgements

I have been privileged by a great deal of support and assistance in my Ph.D. study.

I express my sincere gratitude to my supervisor Dr. Ryo Fujimoto, Associate Professor at Kobe University, who allows me to study in his laboratory. I am lucky to have him as an advisor for my Ph.D. study. I am deeply grateful to him for his continuous support, encouragement to research work, patience, and motivation to research, sharing his immense knowledge, and outstanding supervision throughout my Ph.D. study.

I am indebt to my co-supervisors and dissertation committee members for their comments, suggestions, and encouragement to research work.

I would like to acknowledge "Rotary Yoneyama Memorial Foundation, Japan" for supporting me with the scholarship for the Ph.D. study. I would like to thank my host club (Kobe Higashi Rotary Club - 2680), my counselors, and all Rotarians in Japan.

I want to thank Kobe University for the subsidy of tuition fees, and staffs of the Agriculture faculty for their kind supports.

I would like to express my deepest appreciation to, my co-authors, Dr. Kenji Osabe, Daniel J Shea, Satoshi Takahashi, Dr. Naomi Miyaji, Dr. Ayasha Akter, Dr. Motoaki Seki, Dr. Elizabeth S Dennis, and Yutaka Suzuki.

My deepest appreciation goes to Kiyomi Imamura, Makiko Tosaka, Taiji Kikuchi, Terumi Horiuchi, and Ms. Tomoko Kusumi for their technical assistance and Dr. Jean Finnegan for advice on the experimental technique of sequential ChIP.

Special thanks to my lab members (Mr. Kusuke Uezono, Ms. Namiko Nishida, Mr. Kodai Matsuo, Ms. Terumi Murakami, Ms. Moe Yamashita, Mr. Takahiro Iwaasa, Ms. Kana Yumioka, Ms. Arjina Akter, and others) for their kind assistance in my PhD journey. I am also thankful to my family members for their kind support.

-- Hasan Mehraj

References

Akter A, Takahashi S, Deng W, Shea DJ, Itabashi E, Shimizu M, Miyaji N, Osabe K, Nishida N., Suzuki Y, Helliwell CA, Seki M, Peacock WJ, Dennis ES, Fujimoto R. The histone modification H3 lysine 27 tri-methylation has conserved gene regulatory roles in the triplicated genome of *Brassica rapa* L. *DNA Research* **2019**, 26, 433-443.

Ariel F, Romero-Barrios N, Jégu T, Benhamed M, Crespi M. Battles and hijacks: noncoding transcription in plants. *Trends Plant Science* **2015**, 20, 362-71.

Asensi-Fabado MA, Amtmann A, Perrella G. Plant responses to abiotic stress: The chromatin context of transcriptional regulation. Biochim. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2017**, 1860, 106-122.

Au PCK, Dennis ES, Wang MB. Analysis of argonaute 4-associated long non-coding RNA in *Arabidopsis thaliana* sheds novel insights into gene regulation through RNA-directed DNA methylation. *Genes* **2017**, 8, 198.

Bailey CD, Koch MA, Mayer M, Mummenhoff K, O'Kane SL, Warwick SI, Al-Shehbaz IA. Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution* **2006**, 23, 2142-2160.

Bannister A, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Research* **2011**, 21, 381-395.

Baranwal VK, Mikkilineni V, Zehr UB, Tyagi AK, Kapoor S. Heterosis: emerging ideas about hybrid vigor. *Journal of Experimental Botany* **2012**, 63, 6309-6314.

Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **2012**, 9, 671-675.

Barth S, Busimi AK, Utz HF, Melchinger AE. Heterosis for biomass yield and related traits in five hybrids of *Arabidopsis thaliana* L. Heynh. *Heredity* **2003**, 91, 36-42

Basunanda P, Radoev M, Ecke W, Friedt W, Becker HC, Snowdon RJ. Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theoretical and Applied Genetics* **2010**, 120, 271–281.

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

Berr A, Shafiq S, Pinon V, Dong A, Shen WH. The trxG family histone methyltransferase SET DOMAIN GROUP 26 promotes flowering via a distinctive genetic pathway. *The Plant Journal* **2015**, 81, 316-328.

Bewick AJ, Schmitz RJ. Gene body DNA methylation in plants. *Current Opinion in Plant Biology* **2017**, 36, 103-110.

Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA. Heterosis. *The Plant Cell* 2010, 22, 2105-2112.

Black JC, Van RechemC, Whetstine JR. Histone lysine methylation dynamics: Establishment, regulation, and biological impact. *Molecular* Cell **2012**, 48, 491-507.

Blanc G, Hokamp K, Wolfe KH. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Research* **2003**, 13, 137-144.

Blanco E, González-Ramírez M, Alcaine-Colet A, Aranda S, Croce LD. The bivalent genome: Characterization, structure, and regulation. *Trends in genetics* **2020**, *36*, 118-131.

Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, Tang H, Pikaard CS. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife* **2015**, 4, e09591.

Böhmdorfer G, Sethuraman S, Jordan Rowley M, Krzyszton M, Hafiz Rothi M, Bouzit L, Wierzbicki AT. Long noncoding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *Elife* **2016**, 5, e19092.

Bologna NG, Voinnet O. The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. *Annual Review of Plant Biology* **2014**, 65, 473-503.

Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. *Science* **2010**, 330, 612-616.

Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **2003**, 422, 433-438.

Bruce AB. The mendelian theory of heredity and the augmentation of vigor. *Science* **1910**, 32, 627-628.

Budak H, Kaya SB and Cagirici HB. Long non-coding RNA in plants in the era of reference sequences. *Frontiers in Plant Science* **2020**, 11, 276.

Buzas DM, Robertson M, Finnegan EJ, Helliwell CA. Transcription-dependence of histone H3 lysine 27 trimethylation at the Arabidopsis polycomb target gene *FLC*. *The Plant Journal* **2011**, 65, 872-881.

Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **2014**, 157, 77-94.

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier MC, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VH, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CH, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **2014**, 345, 950-953.

Chekanova JA. Long non-coding RNAs and their functions in plants. *Current Opinion in Plant Biology* **2015**, 27, 207-216.

Chen L, Zhu QH, Kaufmann K. Long non-coding RNAs in plants: emerging modulators of gene activity in development and stress responses. *Planta* **2020**, 252, 92.

Chen Q, Liu K, Yu R, Zhou B, Huang P, Cao Z, Zhou Y, Wang J. From “Dark Matter” to “Star”: Insight into the regulation mechanisms of plant functional long non-coding RNAs. *Frontiers in Plant Science* **2021**, 12, 650926.

Chen R, Li M, Zhang H, Duan L, Sun X, Jiang Q, Zhang H, Hu Z. Continuous salt stress-induced long non-coding RNAs and DNA methylation patterns in soybean roots. *BMC Genomics* **2019**, 20, 730.

Chen ZJ. Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science* **2010**, 15, 57-71.

Chen ZJ. Genomic and epigenetic insights into the molecular bases of heterosis. *Nature Reviews Genetics* **2013**, 14, 471-482

Chen X, Ge X, Wang J, Tan C, King GJ, Liu K. Genome-wide DNA methylation profiling by modified reduced representation bisulfite sequencing in *Brassica rapa* suggests that epigenetic modifications play a key role in polyploid genome evolution. *Frontiers in Plant Science* **2015**, 6, 836.

Chen ZJ, Mas P. Interactive roles of chromatin regulation and circadian clock function in plants. *Genome Biology* **2019**, 20, 62.

Cheng F, Liang J, Cai C, Cai X, Wu J, Wang X. Genome sequencing supports a multi-vertex model for Brassiceae species. *Current Opinion in Plant Biology* **2017**, 36, 79-87.

Cheng F, Sun C, Wu J, Schnable J, Woodhouse MR, Liang J, Cai C, Freeling M, Wang X. Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. *The New Phytologist* **2016b**, 211, 288-299.

Cheng F, Sun R, Hou X, Zheng H, Zhang F, Zhang Y, Liu B, Liang J, Zhuang M, Liu Y, Liu D, Wang X, Li P, Liu Y, Lin K, Bucher J, Zhang N, Wang Y, Wang H, Deng J, Liao Y, Wei K, Zhang X, Fu L, Hu Y, Liu J, Cai C, Zhang S, Zhang S, Li F, Zhang H, Zhang J, Guo N, Liu Z, Liu J, Sun C, Ma Y, Zhang H, Cui Y, Freeling MR, Borm T, Bonnema G, Wu J, Wang X. Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nature Genetics* **2016a**, 48, 1218-1224.

Cheng F, Wu J, Wang X. Genome triplication drove the diversification of *Brassica* plants. *Horticulture Research* **2014**, 1, 14024.

Cheng F, Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang X. Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* **2012**, 7, e36442.

Cheng K, Xu Y, Yang C, Ouellette L, Niu L, Zhou X, Chu L, Zhuang F, Liu J, Wu H, Charron JB, Luo M. Histone tales: lysine methylation, a protagonist in Arabidopsis development. *Journal of Experimental Botany* **2020**, 71, 793-807.

Cho J. Transposon-derived non-coding RNAs and their function in plants. *Frontiers in Plant Science* **2018**, 9, 600.

Collins LJ, Penny D. The RNA infrastructure: dark matter of the eukaryotic cell? *Trends in Genetics* **2009**, 25, 120-128.

Crick F. Central dogma of molecular biology. *Nature* **1970**, 227, 561-563.

Crow JF. Alternative hypothesis of hybrid vigor. *Genetics* **1948**, 33, 477-487.

Dapp M, Reinders J, Bédiée A, Balsera C, Bucher E, Theiler G, Granier C, Paszkowski J. Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids. *Nature Plants* **2015**, 1, 15092.

Darwin CR. The effects of cross and self fertilization in the vegetable kingdom. John Murray: London, UK. **1876**.

Davenport CB. Degeneration, albinism and inbreeding. *Science* **1908**, 28, 454-455.

Demetriadou C, Koufaris C, Kirmizis A. Histone N-alpha terminal modifications: genome regulation at the tip of the tail. *Epigenetics Chromatin* **2020**, 13, 29.

Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, Figueroa ME, De Figueiredo Pontes LL, Alberich-Jorda M, Zhang P, Wu M, D'Alò F, Melnick A, Leone G, Ebralidze KK, Pradhan S, Rinn JL, Tenen DG. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **2013**, 503, 371-376.

Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, Xu C, Li X, Xiao J, Zhang Q. A long noncoding RNA regulates photoperiod sensitive male sterility, an essential component of hybrid rice. *Proceedings of the National Academy of Sciences of USA* **2012**, 109, 2654-2659.

Dixon GR. Vegetable brassicas and related crucifers. In Dixon GR (ed.), Origins and diversity of *Brassica* and its relatives. CABI: Wallingford, UK. **2006**, pp 1-33.

Dong X, Reimer J, Göbel U, Engelhorn J, He F, Schoof H, Turck F. Natural variation of H3K27me3 distribution between two *Arabidopsis* accessions and its association with flanking transposable elements. *Genome Biology* **2012**, 13, R117.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research* **2010**, 38, W64–W70.

Duvick DN. Biotechnology in the 1930s: the development of hybrid maize. *Nature Reviews Genetics* **2001**, 2, 69-74.

East EM. Heterosis. *Genetics* **1936**, 21, 375-397.

Engelhorn J, Blanvillain R, Kröner C, Parrinello H, Rohmer M, Posé D, Ott F, Schmid M, Carles CC. Dynamics of H3K4me3 chromatin marks prevails over H3K27me3 for gene regulation during flower morphogenesis in *Arabidopsis thaliana*. *Epigenomes* **2017**, 1, 8.

Fedak H, Palusinska M, Krzyczmonik K, Brzezniak L, Yatusevich R, Pietras Z, Kaczanowski S, Swiezewski S. Control of seed dormancy in *Arabidopsis* by a *cis*-acting noncoding antisense transcript. *Proceedings of the National Academy of Sciences of USA* **2016**, 113, E7846-E7855.

Finnegan EJ, Bond DM, Buzas DM, Goodrich J, Helliwell CA, Tamada Y, Yun JY, Amasino RM, Dennis ES. Polycomb proteins regulate the quantitative induction of *VERNALIZATION INSENSITIVE 3* in response to low temperatures. *The Plant Journal* **2011**, 65, 382-391.

Flint-Garcia SA, Buckler ES, Tiffin P, Ersoz E, Springer NM. Heterosis is prevalent for multiple traits in diverse maize germplasm. *PLoS One* **2009**, 4, e7433.

Footitt S, Müller K, Kermode AR, Finch-Savage WE. Seed dormancy cycling in Arabidopsis: chromatin remodelling and regulation of DOG1 in response to seasonal environmental signals. *The Plant Journal* **2015**, 81, 413-425.

Fu D, Xiao M, Hayward A, Jiang G, Zhu L, Zhou Q, Li J, Zhang M. What is crop heterosis: new insights into an old topic. *Journal of Applied Genetics* **2015**, 56, 1-13.

Fuchs J, Demidov D, Houben A, Schubert I. Chromosomal histone modification patterns - from conservation to diversity. *Trends in Plant Science* **2006**, 11, 199-208.

Fujimoto R, Sasaki T, Ishikawa R, Osabe K, Kawanabe T, Dennis ES. Molecular mechanisms of epigenetic variation in plants. *International Journal of Molecular Sciences* **2012a**, 13, 9900-9922.

Fujimoto R, Taylor JM, Shirasawa S, Peacock WJ, Dennis ES. Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proceedings of the National Academy of Sciences of USA* **2012b**, 109, 7109-7114.

Fujimoto R, Uezono K, Ishikura S, Osabe K, Peacock WJ, Dennis ES. Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. *Breeding Science* **2018**, 68, 145-158.

Fukushima S, Mori M, Sugano S, Takatsuji H. Transcription factor WRKY62 plays a role in pathogen defense and hypoxia-responsive gene expression in rice. *Plant and Cell Physiology* **2016**, 57, 2541-2551.

Gaiti, F., Degnan, B. M., & Tanurdžić, M. Long non-coding regulatory RNAs in sponges and insights into the origin of animal multicellularity. *RNA Biology* **2018**, 15, 696-702.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **2004**, 5, R80.

Greaves IK, Gonzalez-Bayon R, Wang L, Zhu A, Liu PC, Groszmann M, Peacock WJ, Dennis ES. Epigenetic changes in hybrids. *Plant Physiology* **2015**, 168, 1197-1205.

Groszmann M, Gonzalez-Bayon R, Greaves IK, Wang L, Huen AK, Peacock WJ, Dennis ES. Intraspecific Arabidopsis hybrids show different patterns of heterosis despite the close relatedness of the parental genomes. *Plant Physiology* **2014**, 166, 265-280.

Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES. Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proceedings of the National Academy of Sciences of USA* **2011**, 108, 2617-2622.

Guo Y, Chen S, Li Z, Cowling WA. Center of origin and centers of diversity in an ancient crop, *Brassica rapa* (Turnip Rape), *Journal of Heredity* **2014**, 105, 555-565,

Guo Z, Song G, Liu Z, Qu X, Chen R, Jiang D, Sun Y, Liu C, Zhu Y, Yang D. Global epigenomic analysis indicates that epialleles contribute to Allele-specific expression via Allele-specific histone modifications in hybrid rice. *BMC Genomics* **2015**, 16, 232.

He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F, Qi Y, Chen R, Deng XW. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *The Plant Cell* **2010**, 22, 17-33.

Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **2011**, 331, 76-79.

Huang L, Dong H, Zhou D, Li M, Liu Y, Zhang F, Feng Y, Yu D, Lin S, Cao J. Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa*. *The Plant Journal* **2018**, 96, 203-222.

Huang CH, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP, Pires JC, Tan DY, Zhong Y, Ma H. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molecular Biology and Evolution* **2016**, 33, 394-412.

Hung FY, Chen C, Yen MR, Hsieh JA, Li C, Shih YH, Chen FF, Chen PY, Cui Y, Wu K. The expression of long non-coding RNAs is associated with H3Ac and H3K4me2 changes regulated by the HDA6-LDL1/2 histone modification complex in *Arabidopsis*. *NAR Genomics and Bioinformatics* **2020**, 2, lqaa066.

Itabashi E, Osabe K, Fujimoto R, Kakizaki T. Epigenetic regulation of agronomical traits in Brassicaceae. *Plant Cell Reports* **2018**, 27, 87-101.

Jakobsen MK, Poulsen LR, Schulz A, Fleurat-Lessard P, Møller A, Husted S, Schiøtt M, Amtmann A, Palmgren MG. Pollen development and fertilization in *Arabidopsis* is dependent on the *MALE GAMETOGENESIS IMPAIRED ANTHERS* gene encoding a type V P-type ATPase. *Genes and Development* **2005**, 19, 2757-2769.

Jenuwein T, Allis CD. Translating the histone code. *Science* **2001**, 293, 1074-1080.

Jha UC, Nayyar H, Jha R, Khurshid M, Zhou M, Mantri N, Siddique KHM. Long non-coding RNAs: emerging players regulating plant abiotic stress response and adaptation. *BMC Plant Biology* **2020**, 20, 466.

Johnson C, Conrad LJ, Patel R, Anderson S, Li C, Pereira A, Sundaresan V. Reproductive long intergenic noncoding RNAs exhibit male gamete specificity and polycomb repressive complex 2-mediated repression. *Plant Physiology* **2018**, 177, 1198-1217.

Jones DF. Dominance of linked factors as a means of accounting for heterosis. *Genetics* **1917**, 2, 466-479.

Joshi RK, Megha S, Basu U, Rahman MH, Kav NNV. Genome wide identification and functional prediction of long non-coding RNAs responsive to *Sclerotinia sclerotiorum* infection in *Brassica napus*. *PloS One* **2016**, 11, e0158784.

Karlik E, Ari S, Gozukirmizi N. LncRNAs: genetic and epigenetic effects in plants. *Biotechnology and Biotechnological Equipment* **2019**, 33, 429-439.

Kawamura K, Kawanabe T, Shimizu M, Nagano AJ, Saeki N, Okazaki K, Kaji M, Dennis ES, Osabe K, Fujimoto R. Genetic distance of inbred lines of Chinese cabbage and its relationship to heterosis. *Plant Gene* **2016**, 5, 1-7

Kawanabe T, Osabe K, Itabashi E, Okazaki K, Dennis ES, Fujimoto R. Development of primer sets that can verify the enrichment of histone modifications, and their application to examining vernalization-mediated chromatin changes in *Brassica rapa* L. *Genes Genetic Systems* **2016b**, 91, 1-10.

Kawanabe T, Ishikura S, Miyaji N, Sasaki T, Wu LM, Itabashi E, Takada S, Shimizu M, Takasaki-Yasuda T, Osabe K, Peacock WJ, Dennis ES, Fujimoto R. Role of DNA methylation in hybrid vigor in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of USA* **2016a**, 113, E6704-E6711.

Kim DH, Sung S. Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. *Developmental Cell* **2017**, 40, 302-312.

Kim DH, Xi Y, Sung S. Modular function of long noncoding RNA, COLDAIR, in the vernalization response. *PLoS Genetics* **2017**, 13, e1006939.

Kim JA, Kim HS, Choi SH, Jang JY, Jeong MJ, Lee SI. The Importance of the Circadian Clock in Regulating Plant Metabolism. *International Journal of Molecular Sciences* **2017**, 18, 2680.

Kim CK, Seol YJ, Perumal S, Lee J, Waminal NE, Jayakodi M, Lee SC, Jin S, Choi BS, Yu Y, Ko HC, Choi JW, Ryu KY, Sohn SH, Parkin I, Yang TJ. Re-exploration of U's Triangle *Brassica* species based on chloroplast genomes and 45S nrDNA sequences. *Scientific Reports* **2018**, 8, 7353.

Kim JH. Multifaceted chromatin structure and transcription changes in plant stress response. *International Journal of Molecular Sciences* **2021**, 22, 2013. 427

Kim JM, Sasaki T, Ueda M, Sako K, Seki M. Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. *Frontiers in Plant Science* **2015**, 6, 114.

Ko DK, Rohozinski D, Song Q, Taylor SH, Juenger TE, Harmon FG, Chen ZJ. Temporal shift of circadian-mediated gene expression and carbon fixation contributes to biomass heterosis in maize hybrids. *PLoS Genetics* **2016**, 12, e1006197.

Koenig D, Weigel D. Beyond the thale: comparative genomics and genetics of *Arabidopsis* relatives. *Nature Reviews Genetics* **2015**, 16, 285-298.

Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research* **2007**, 35, W345-W349.

Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **2018**, 172, 393-407.

Kouzarides T. Chromatin modifications and their function. *Cell* **2007**, 128, 693-705.

Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: Past, present, and future. *Genetics* 2013, 193, 651-669.

Lauss K, Wardenaar R, Oka R, van Hulten MHA, Guryev V, Keurentjes JJB, Stam M, Johannes F. Parental DNA methylation states are associated with heterosis in epigenetic hybrids. *Plant Physiology* **2018**, 176, 1627-1645.

Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **2010**, 11, 204-220.

Li JR, Liu CC, Sun CH, Chen YT. Plant stress RNA-seq Nexus: a stress-specific transcriptome database in plant cells. *BMC Genomics*, **2018**, 19, 966.

Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE, Evans MM, Scanlon MJ, Yu J, Schnable PS, Timmermans MC, Springer NM, Muehlbauer GJ. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biology* **2014**, 15, R40.

Li S, Vandivier LE, Tu B, Gao L, Won SY, Li S, Zheng B, Gregory BD, Chen X. Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in *Arabidopsis* reveals features and regulation of siRNA biogenesis. *Genome Research* **2015**, 25, 235-245.

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

Li P, Su T, Zhang D, Wang W, Xin X, Yu Y, Zhao X, Yu S, Zhang F. Genome-wide analysis of changes in miRNA and target gene expression reveals key roles in heterosis for Chinese cabbage biomass. *Horticulture Research* **2021**, 8, 39.

Lippman ZB, Zamir D. Heterosis: revisiting the magic. *Trends in Genetics* **2007**, 23, 60-66.

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua NH. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* **2012**, 24, 4333-4345.

Liu J, Wang H, Chua NH. Long noncoding RNA transcriptome of plants. *Plant Biotechnology Journal* **2015**, 13, 319-328.

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang TJ, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J,

Jiang L, Batley J, Sharpe AG, Park BS, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee TH, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications* **2014**, 5, 3930.

Liu X, Hao L, Li D, Zhu L, Hu S. Long non-coding RNAs and their biological roles in plants. *Genomics Proteomics and Bioinformatics* **2015**, 13, 137-147.

Liu ZW, Zhao N, Su YN, Chen SS, He XJ. Exogenously overexpressed intronic long noncoding RNAs activate host gene expression by affecting histone modification in Arabidopsis. *Scientifc Reports* **2020**, 10, 3094.

Liu L, Chen X. Intercellular and systemic trafficking of RNAs in plants. *Nature Plants* **2018**, 4, 869-878.

Liu Y, Liu K, Yin L, Yu Y, Qi J, Shen WH, Zhu J, Zhang Y, Dong A. H3K4me2 functions as a repressive epigenetic mark in plants. *Epigenetics and Chromatin* **2019**, 12, 40.

Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H, Cheng F, Zhang K, Du H, Cheng X, Qu C, Qian W, Liu L, Wang R, Zou Q, Ying J, Xu X, Mei J, Liang Y, Chai YR, Tang Z, Wan H, Ni Y, He Y, Lin N, Fan Y, Sun W, Li NN, Zhou G, Zheng H, Wang X, Paterson AH, Li J. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nature Communications* **2019**, 10, 1154.

Lunardon A, Johnson NR, Hagerott E, Phifer T, Polydore S, Coruh C, Axtell MJ. Integrated annotations and analyses of small RNA-producing loci from 47 diverse plants. *Genome Research* **2020**, 30, 497-513.

Luo C, Sidote DJ, Zhang Y, Kerstetter RA, Michael TP, Lam E. Integrative analysis of chromatin states in Arabidopsis identified potential regulatory mechanisms for natural antisense transcript production. *The Plant Journal* **2013**, 73, 77-90.

Lv H, Miyaji N, Osabe K, Akter A, Mehraj H, Shea DJ, Fujimoto R. The importance of genetic and epigenetic research in the *Brassica* vegetables in the face of climate change. In Kole C (ed.), Genomic designing of climate-smart vegetable crops. Springer: Cham, Switzerland. **2020**, pp. 161-255.

Ma Q, Hedden P, Zhang QF. Heterosis in rice seedlings: its relationship to gibberellin content and expression of gibberellin metabolism and signaling genes. *Plant Physiology* **2011**, 156, 1905-1920.

Makarevitch I, Eichten SR, Briskine R, Waters AJ, Danilevskaya ON, Meeley RB, Myers CL, Vaughn MW, Springer NM. Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. *Plant Cell* **2013**, 25, 780-793.

Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. *Nature Structural and Molecular Biology* **2015**, 22, 5-7.

Matzke MA, Kanno T, Matzke AJM. RNA-directed DNA methylation: The evolution of a complex epigenetic pathway in flowering plants. *Annual Review of Plant Biology* **2015**, 66, 243-267.

Mehraj H, Takahashi S, Miyaji N, Akter A, Suzuki Y, Seki M, Dennis ES, Fujimoto R. Characterization of Histone H3 Lysine 4 and 36 Tri-methylation in *Brassica rapa* L. *Frontiers in Plant Science* **2021**, 12, 659634.

Meyer RC, Witucka-Wall H, Becher M, Blacha A, Boudichevskaia A, Dörmann P, Fiehn O, Friedel S, von Korff M, Lisec J, Melzer M, Repsilber D, Schmidt R, Scholz M, Selbig J, Willmitzer L, Altmann T. Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *The Plant Journal* **2012**, 71, 669-683.

Meyer RC, Törjék O, Becher M, Altmann T. Heterosis of biomass production in Arabidopsis. Establishment during early development. *Plant Physiology* **2004**, 134, 1813-1823.

Meyer P. Epigenetic variation and environmental change. *Journal of Experimantal Botany* **2015**, 66, 3541-3548.

Miller M, Zhang C, Chen ZJ. Ploidy and hybridity effects on growth vigor and gene expression in Arabidopsis thaliana hybrids and their parents. *Genes Genomes Genetics* **2012**, 2, 505-513.

Miyaji N, Shimizu M, Miyazaki J, Osabe K, Sato M, Ebe Y, Takada S, Kaji M, Dennis ES, Fujimoto R, Okazaki K. Comparison of transcriptome profiles by *Fusarium oxysporum* inoculation between Fusarium yellows resistant and susceptible lines in *Brassica rapa* L. *Plant Cell Reports* **2017**, 36, 1841-1854.

Moghaddam AM, Roudier F, Seifert M, Bérard C, Magniette ML, Ashtiyani RK, Houben A, Colot V, Mette MF. Additive inheritance of histone modifications in *Arabidopsis thaliana* intra-specific hybrids. *The Plant Journal* **2011**, 67, 691-700.

Moreno-Romero J, Jiang H, Santos-González J, Köhler C. Parental epigenetic asymmetry of *PRC2*-mediated histone modifications in the *Arabidopsis* endosperm. *The EMBO Journal* **2016**, 35, 1298-1311.

Mukhtar MS, Deslandes L, Auriac MC, Marco Y, Somssich IE. The Arabidopsis transcription factor WRKY27 influences wilt disease symptom development caused by *Ralstonia solanacearum*. *The Plant Journal* **2008**, 56, 935-947.

Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* **1980**, 8, 4321-4326.

Nakamura S, Hosaka K. DNA methylation in diploid inbred lines of potatoes and its possible role in the regulation of heterosis. *Theoretical and Applied Genetics* **2010**, 120, 205-214

Ng DW, Zhang C, Miller M, Palmer G, Whiteley M, Tholl D, Chen ZJ. *cis*- and *trans*-Regulation of miR163 and target genes confers natural variation of secondary metabolites in two *Arabidopsis* species and their allopolyploids. *Plant Cell* **2011**, 23, 1729-1740.

Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* **2009**, 457, 327-331.

Nowak R. Mining treasures from "junk DNA". *Science* **1994**, 263, 608-610.

Oh S, Park S, van Nocker S. Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genetics* **2008**, 4, e1000077.

Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, Denoeud F, Belcram H, Links MG, Just J, Clarke C, Bender T, Huebert T, Mason AS, Pires JC, Barker G, Moore J, Walley PG, Manoli S, Batley J, Edwards D, Nelson MN, Wang X, Paterson AH, King G, Bancroft I, Chalhoub B, Sharpe AG. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biology* **2014**, 15, R77.

Payá-Milans M, Poza-Viejo L, Martín-Uriz PS, Lara-Astiaso D, Wilkinson MD, Crevillén P. Genome-wide analysis of the H3K27me3 epigenome and transcriptome in *Brassica rapa*. *GigaScience* **2019**, 8, giz147.

Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* **2009**, 136, 629-41.

Powers L. An expansion of Jones's theory for the explanation of heterosis. *The American Naturalist* **1944**, 78, 275-280.

Prakash S. Wu X. Bhat SR. History, evolution and domestication of *Brassica* crops. *Plant Breeding Reviews* **2012**, 35, 19-84.

Qi X, Li ZH, Jiang LL, Yu XM, Ngezahayo F, Liu B. Grain-yield heterosis in *Zea mays* L. shows positive correlation with parental difference in CHG methylation. *Crop Science* **2010**, 50, 2338-2346

Qian S, Lv X, Scheid RN, Lu L, Yang Z, Chen W, Liu R, Boersma MD, Denu JM, Zhong X, Du J. Dual recognition of H3K4me3 and H3K27me3 by a plant histone reader SHL. *Nature Communications* **2018**, 9, 2425.

Quadrana L, Colot V. Plant transgenerational epigenetics. *Annual Review of Genetics* **2016,** 50, 467-491.

Rai MI, Alam M, Lightfoot DA, Gurha P, Afzal AJ. Classification and experimental identification of plant long non-coding RNAs. *Genomics* **2019**, 111, 997-1005.

Regulski M, Lu Z, Kendall J, Donoghue MTA, Reinders J, Llaca V, et al. The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Research* **2013**, 23, 1651-1662.

Richey FD. Mock-dominance and hybrid vigor. *Science* **1942**, 96, 280-281.

Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, Giraut L, Després B, Drevensek S, Barneche F, Dèrozier S, Brunaud V, Aubourg S, Schnittger A, Bowler C, Martin-Magniette ML, Robin S, Caboche M, Colot V. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *The EMBO Journal* **2011**, 30, 1928-1938.

Ryder P, McKeown PC, Fort A, Spillane C. Epigenetics and heterosis in crop plants. In Alvarez-Venegas R, De-la-Peña C, Casas-Mollano J (eds.), Epigenetics in plants of agronomic importance: Fundamentals and applications. Springer: Cham, Switzerland. **2019**, pp. 129-148.

Saeki N, Kawanabe T, Ying H, Shimizu M, Kojima M, Abe H, Okazaki K, Kaji M, Taylor JM, Sakakibara H, Peacock WJ, Dennis ES, Fujimoto R. Molecular and cellular

characteristics of hybrid vigour in a commercial hybrid of Chinese cabbage. *BMC Plant Biology* **2016**, 16, 45.

Sampath P, Murukarthick J, Izzah NK, Lee J, Choi HI, Shirasawa K, Choi BS, Liu S, Nou IS, Yang TJ. Genome-wide comparative analysis of 20 miniature inverted-repeat transposable element families in *Brassica rapa* and *B. oleracea*. *PLoS One* **2014**, 9, e94499.

Schmitz RJ, Ecker JR. Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends in Plant Science* **2012**, 17, 149-154

Schnable PS, Springer NM. Progress toward understanding heterosis in crop plants. *Annual Review of Plant Biology* **2013**, 64, 71–88.

Seo JS, Sun HX, Park BS, Huang CH, Yeh SD, Jung C, Chua NH. ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in Arabidopsis. *Plant Cell* **2017**, 29, 1024-1038.

Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C. The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **2014**, 26, 2351-2366.

She R, Chu JSC, Wang K, Pei J, Chen N. genBlastA: Enabling BLAST to identify homologous gene sequences. *Genome Research* **2009**, 19, 143-149.

Shea DJ, Nishida N, Takada S, Itabashi E, Takahashi S, Akter A, Miyaji N, Osabe K, Mehraj H, Shimizu M, Seki M, Kakizaki T, Okazaki K, Dennis ES, Fujimoto R. Long noncoding RNAs in *Brassica rapa* L. following vernalization. *Scientific Reports* **2019**, 9, 9302.

Shea DJ, Shimizu M, Itabashi E, Miyaji N, Miyazaki J, Osabe K, Kaji M, Okazaki K, Fujimoto R. Genome re-sequencing, SNP analysis, and genetic mapping of the parental lines of commercial F_1 hybrid cultivar of Chinese cabbage. *Breeding Science* **2018**, 68, 375-380.

Shen E, Zhu X, Hua S, Chen H, Ye C, Zhou L, et al. Genome-wide identification of oil biosynthesis related long non-coding RNAs in allopolyploid *Brassica napus*. *BMC Genomics* **2018**, 19, 745.

Shen G, Hu W, Zhang B, Xing Y. The regulatory network mediated by circadian clock genes is related to heterosis in rice. *Journal of Integrative Plant Biology*, **2015**, 57, 300-312.

Shen H, He H, Li J, Chen W, Wang X, Guo L, Peng Z, He G, Zhong S, Qi Y, Terzaghi W, Deng XW. Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell* **2012**, 24, 875-892.

Shen X, Xu L, Liu Y, Dong H, Zhou D, Zhang Y, Lin S, Cao J, Huang L. Comparative transcriptome analysis and ChIP-sequencing reveals stage-specific gene expression and regulation profiles associated with pollen wall formation in *Brassica rapa*. *BMC Genomics* **2019**, 20, 264.

Shen Y, Sun S, Hua S, Shen E, Ye CY, Cai D, Timko MP, Zhu QH, Fan L. Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs. *The Plant Journal* **2017**, 91, 874-893.

Shi J, Li R, Zou J, Long Y, Meng J. A dynamic and complex network regulates the heterosis of yield-correlated traits in rapeseed (*Brassica napus* L.). *PLoS One* **2011**, 6, e21645.

Shimizu M, Fujimoto R, Ying H, Pu ZJ, Ebe Y, Kawanabe T, Saeki N, Taylor JM, Kaji M, Dennis ES, Okazaki K. Identification of candidate genes for Fusarium yellows resistance in Chinese cabbage by differential expression analysis. *Plant Molecular Biology* **2014**, 85, 247-257.

Shivaprasad PV, Dunn RM, Santos BA, Bassett A, Baulcombe DC. Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *The EMBO Journal* **2012**, 31, 257-266.

Shull GH. The composition of a field of maize. *Journal of Heredity* **1908**, 4, 296-301.

Shull GH. What Is “Heterosis”? *Genetics* **1948**, 33, 439-446.

Song L, Fang Y, Chen L, Wang J, Chen X. Role of non-coding RNAs in plant immunity. *Plant Communications* **2021**, 2, 100180.

Song X, Liu G, Huang Z, Duan W, Tan H, Li Y, Hou X. Temperature expression patterns of genes and their coexpression with LncRNAs revealed by RNA-Seq in non-heading Chinese cabbage. *BMC Genomics* **2016**, 17, 297.

Strahl B, Allis C. The language of covalent histone modifications. *Nature* **2000**, 403, 41-45.

Swiezewski S, Liu F, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* **2009**, 462, 799-802.

Takahashi S, Osabe K, Fukushima N, Takuno S, Miyaji N, Shimizu M, Takasaki-Yasuda T, Suzuki Y, Dennis ES, Seki M, Fujimoto R. Genome-wide characterization of DNA methylation, small RNA expression, and histone H3 lysine nine di-methylation in *Brassica rapa* L. *DNA Research* **2018**, 25, 511-520.

Talbert PB, Henikoff S. Histone variants at a glance. *Journal of Cell Science* **2021**, 134, jcs244749.

Tang X, Lim MH, Pelletier J, Tang M, Nguyen V, Keller WA, Tsang EW, Wang A, Rothstein SJ, Harada JJ, Cui Y. Synergistic repression of the embryonic programme by SET DOMAIN GROUP 8 and EMBRYONIC FLOWER 2 in *Arabidopsis* seedlings. *Journal of Experimantal Botany* **2012**, 63, 1391-1404.

Tian Y, Zheng H, Zhang F, Wang S, Ji X, Xu C, He Y, Ding Y. PRC2 recruitment and H3K27me3 deposition at *FLC* require FCA binding of *COOLAIR*. *Science Advvances* **2019**, 5, eaau7246.

Tong C, Wang X, Yu J, Wu J, Li W, Huang J, Dong C, Hua W, Liu S. Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. *BMC Genomics* **2013**, 14, 689.

Tu S, Yuan GC, Shao Z. The PRC2-binding long non-coding RNAs in human and mouse genomes are associated with predictive sequence features. Scientific Reports **2017**, 7, 41669.

Turck F, Coupland G. Natural variation in epigenetic gene regulation and its effects on plant developmental traits. *Evolution* **2014**, 68, 620-631.

Turck F, Roudier F, Farrona S, Martin-Magniette ML, Guillaume E, Buisine N, Gagnot S, Martienssen RA, Coupland G, Colot V. Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genetics* **2007**, 3, e86.

U N. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **1935**, 7, 389-452.

Vandivier LE, Anderson SJ, Foley SW, Gregory BD. The conservation and function of RNA secondary structure in plants. *Annual Review of Plant Biology* **2016**, 67, 463-488.

Wang A, Hu J, Gao C, Chen G, Wang B, Lin C, Song L, Ding Y, Zhou G. Genome-wide analysis of long non-coding RNAs unveils the regulatory roles in the heat tolerance of Chinese cabbage (*Brassica rapa* ssp. *chinensis*). *Scientific Reports* **2019**, 9, 5002.

Wang C, Wang L, Ding Y, Lu X, Zhang G, Yang J, Zheng H, Wang H, Jiang Y, Xu L. LncRNA structural characteristics in epigenetic regulation. *International Journal of Molecular Sciences* **2017**, 18, 2659.

Wang D, Qu Z, Yang L, Zhang Q, Liu ZH, Do T, Adelson DL, Wang ZY, Searle I, Zhu JK. Transposable elements (TEs) contribute to stress related long intergenic noncoding RNAs in plants. *The Plant Journal* **2017**, 90, 133-146.

Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, Xu J, Chua NH. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Research* **2014**, 24, 444-453.

Wang L, Liu PC, Wu LM, Tan J, Peacock WJ, Dennis ES. Cotyledons contribute to plant growth and hybrid vigor in Arabidopsis. *Planta* **2019**, 249, 1107-1118.

Wang R, Zou J, Meng J, Wang J. Integrative analysis of genome-wide lncRNA and mRNA expression in newly synthesized *Brassica* hexaploids. *Ecology and Evolution* **2018**, 8, 6034-6052.

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **2011**, 43, 1035-1040.

Wang Y, Fan X, Lin F, He G, Terzaghi W, Zhu D, Deng XW. *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light. *Proceedings of the National Academy of Sciences of USA* **2014**, 111, 10359-10364.

Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH. Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in Arabidopsis. *Plant Physiology* **2008**, 148, 1201-1211.

Williams W. Heterosis and the genetics of complex characters. *Nature* **1959**, 184, 527-530.

Wu CT, Morris JR. Genes, genetics, and epigenetics: A correspondence. *Science* **2001**, 293, 1103-1105.

Wu H, Yang L, Chen LL. The diversity of long noncoding RNAs and their generation. *Trends in Genetics* **2017**, 33, 540-52.

Wu X, Liu Y, Zhang Y, Gu R. Advances in research on the mechanism of heterosis in plants. *Frontiers in Plant Science* **2021**, 12, 745726.

Xi Y, Park SR, Kim DH, Kim, ED, Sung S. Transcriptome and epigenome analyses of vernalization in *Arabidopsis thaliana*. *The Plant Journal* 2020*,* 103, 1490-1502.

Xiao J, Lee US, Wagner D. Tug of war: adding and removing histone lysine methylation in *Arabidopsis*. *Current Opinion in Plant Biology* **2016**, 34, 41-53.

Xue JY, Wang Y, Chen M, Dong S, Shao ZQ, Liu Y. Maternal inheritance of U's Triangle and evolutionary process of *Brassica* mitochondrial genomes. *Frontiers in Plant Science* **2020**, 11, 805.

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **2003**, 302, 842-846.

Yan H, Bombarely A, Xu B, Frazier TP, Wang C, Chen P, Chen J, Hasing T, Cui C, Zhang X, Zhao B, Huang L. siRNAs regulate DNA methylation and interfere with gene and lncRNA expression in the heterozygous polyploid switchgrass. *Biotechnology for Biofuels* **2018**, 11, 208.

Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, Hu Z, Chen S, Pental D, Ju Y, Yao P, Li X, Xie K, Zhang J, Wang J, Liu F, Ma W, Shopan J, Zheng H, Mackenzie SA, Zhang M. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nature Genetics* **2016b**, 48, 1225-1232.

Yang J, Liu G, Zhao N, Chen S, Liu D, Ma W, Hu Z, Zhang M. Comparative mitochondrial genome analysis reveals the evolutionary rearrangement mechanism in *Brassica*. *Plant Biology* **2016a**, 18, 527-536.

Yang M, Wang X, Ren D, Huang H, Xu M, He G, Deng XW. Genomic architecture of biomass heterosis in *Arabidopsis*. *Proceedings of the National Academy of Sciences of USA* **2017**, 114, 8101-8106.

Yang H, Howard M, Dean C. Antagonistic roles for H3K36me3 and H3K27me3 in the cold induced epigenetic switch at *Arabidopsis FLC*. *Current Biology* **2014**, 24, 1793-1797.

Yang L, Liu P, Wang X, Jia A, Ren D, Tang Y, Tang Y, Deng XW, He G. A central circadian oscillator confers defense heterosis in hybrids without growth vigor costs. *Nature Communications* **2021**, 12, 2317.

Yang M, Wang X, Huang H, Ren D, Su Y, Zhu P, Zhu D, Fan L, Chen L, He G, Deng XW. Natural variation of H3K27me3 modification in two *Arabidopsis* accessions and their hybrid. *Journal of Integrative Plant Biology* **2016**, 58, 466-474.

Yu D, Gu X, Zhang S, Dong S, Miao H, Gebretsadik K, Bo K. Molecular basis of heterosis and related breeding strategies reveal its importance in vegetable breeding. *Horticulture Research* **2021**, 8, 120.

Yu J, Zhao M, Wang X, Tong C, Huang S, Tehrim S, Liu Y, Hua W, Liu S. Bolbase: A comprehensive genomics database for *Brassica oleracea*. *BMC Genomics* **2013**, 14, 664.

Yu X, Yang J, Li X, Liu X, Sun C, Wu F, He Y. Global analysis of cis-natural antisense transcripts and their heat-responsive nat-siRNAs in *Brassica rapa*. *BMC Plant Biology* **2013**, 13, 208.

Yu Y, Zhang Y, Chen X, Chen Y. Plant noncoding RNAs: hidden players in development and stress responses. *Annual Review of Cell and Developmental Biology* **2019**, 35, 407-31.

Zeng Z, Zhang W, Marand AP, Zhu B, Buell CR, Jiang J. Cold stress induces enhanced chromatin accessibility and bivalent histone modifications H3K4me3 and H3K27me3 of active genes in potato. *Genome Biology* **2019**, 20, 123.

Zhai J, Bischof S, Wang H, Feng S, Lee TF, Teng C, Chen X, Park SY, Liu L, Gallego-Bartolome J, Liu W, Henderson IR, Meyers BC, Ausin I, Jacobsen SE. A one precursor one siRNA model for Pol Iv dependent siRNA biogenesis. *Cell* **2015**, 163, 445-455.

Zhang J, Wei L, Jiang J, Mason AS, Li H, Cui C, Chai L, Zheng B, Zhu Y, Xia Q, Jiang L, Fu D. Genome-wide identification, putative functionality and interactions between lncRNAs and miRNAs in *Brassica* species. *Scientific Reports* **2018**, 8, 4960.

Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biology* **2009**, 10, R62.

Zhang P, Wu W, Chen Q, Chen M. Non-Coding RNAs and their Integrated Networks. *Journal of Integrative Bioinformatics* **2019**, 16, 20190027.

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

Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. Global identification of *Arabidopsis* lncRNAs reveals the regulation of *MAF4* by a natural antisense RNA. *Nature Communications* **2018**, 9, 5056.

Zhao XX, Chai Y, Liu B. Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids. *Plant Science* **2007**, 172, 930-938

Zhao T, Zhan Z, Jiang D. Histone modifications and their regulatory roles in plant development and environmental memory. *Journal of Genetics and Genomics* **2019**, 46, 467-476.

Zhu A, Greaves IK, Dennis ES, Peacock WJ. Genome-wide analyses of four major histone modifications in Arabidopsis hybrids at the germinating seed stage. *BMC Genomics* **2017**, 18, 137.

Zhu W, Hu B, Becker C, Doğan ES, Berendzen KW, Weigel D, Liu C. Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific Arabidopsis hybrid. *Genome Biology* **2017**, 18, 157.