Molecular Plant • Volume 6 • Number 5 • Pages 1463–1472 • September 2013 RESEARCH ARTICLE

Genome-Wide Analysis of Histone Modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in *Oryza sativa* L. Japonica

Zhou Du[a,2,3], Hui Li[b,2], Qiang Wei[a,2], Xin Zhao[b], Chunchao Wang[a], Qilin Zhu[b], Xin Yi[a], Wenying Xu[a], X. Shirley Liu[c,1], Weiwei Jin[b,1] and **Zhen Su[a,1]**

a State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100193, China
b National Maize Improvement Center of China, Key Laboratory of Biology and Genetic Improvement of Maize, Ministry of Agriculture, China Agricultural University, Beijing 100193, China
c Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, 450 Brookline Ave, Boston, MA 02215, USA

ABSTRACT While previous studies have shown that histone modifications could influence plant growth and development by regulating gene transcription, knowledge about the relationships between these modifications and gene expression is still limited. This study used chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq), to investigate the genome-wide distribution of four histone modifications: di and trimethylation of H3K4 (H3K4me2 and H3K4me3) and acylation of H3K9 and H3K27 (H3K9ac and H3K27ac) in *Oryza sativa* L. japonica. By analyzing published DNase-Seq data, this study explored DNase-Hypersensitive (DH) sites along the rice genome. The histone marks appeared mainly in generic regions and were enriched around the transcription start sites (TSSs) of genes. This analysis demonstrated that the four histone modifications and the DH sites were all associated with active transcription. Furthermore, the four histone modifications were highly concurrent with transcript regions—a promising feature that was used to predict missing genes in the rice gene annotation. The predictions were further validated by experimentally confirming the transcription of two predicted missing genes. Moreover, a sequence motif analysis was constructed in order to identify the DH sites and many putative transcription factor binding sites.

Key words: bioinformatics; chromatin structure and remodeling; epigenetics; gene regulation; genomics; rice.

INTRODUCTION

Previous studies have demonstrated that histone modifications play an important role in regulating gene transcription during plant growth and development, as well as the plant response to various endogenous and exogenous stimuli (Berger, 2007; Pfluger and Wagner, 2007). For example, in *Arabidopsis*, histone modifications, such as the mono, di, and trimethylation of histone H3 lysine4 (H3K4me1/2/3), displayed dynamic changes in response to dehydration stress (van Dijk et al., 2010). Moreover, gene expression can be fine-tuned by combining histone modification patterns, including modification types and levels (Schones and Zhao, 2008; Wang et al., 2008; Zhou et al., 2010). For example, active promoters contain the histone modifications, H3K4me2, H3K4me3, and H3K9ac, whereas repressed genes are marked by H3K9me2 or H3K27me3 and enhancers are commonly marked by H3K4me1/2 and H3K27ac (Heintzman et al., 2007; Zhou et al., 2010). High-throughput techniques, such as tiling array and next-generation sequencing, enable researchers to construct a global map of histone modification and nucleosome positioning along the genome. Chromatin immunoprecipitation, followed by array hybridization (ChIP-chip) or next-generation sequencing (ChIP-Seq), has been widely used for profiling *in vivo* histone modification and has gained increased popularity due to its improved resolution and sensitivity (Ho et al., 2011). In addition to histone modifications, the occupancy of the nucleosome can orchestrate gene expression by competing for access to regulatory elements

[1] To whom correspondence should be addressed. Z.S. E-mail zhensu@cau.edu.cn, tel. +86-10-62731380, fax +86-10-62731380. W.J. E-mail weiweijin@cau.edu.cn, tel. +86-10-62734909, fax +86-10-62733808. X.S.L. E-mail xsliu@jimmy.harvard.edu, tel. +1-617-632-2472.

[2] These authors contributed equally to the article.

[3] Present address: Department of Bioinformatics, School of Life Sciences and Technology, Tongji University, 1239 Siping Road, Shanghai 200092, China

© The Author 2013. Published by the Molecular Plant Shanghai Editorial Office in association with Oxford University Press on behalf of CSPB and IPPE, SIBS, CAS.

doi: 10.1093/mp/sst018, Advance Access publication 25 January 2013

Received 13 August 2012; accepted 20 January 2013

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

on genome DNA with the transcription complex (Jiang and Pugh, 2009). DNase I digestion followed by high-throughput sequencing (DNase-Seq) is a powerful tool for identifying DNase I hypersensitive (DH) sites, which are typically nucleosome-free regions (NFRs), and putative transcriptional and chromatin regulatory regions, such as promoters, enhancers, and insulators (Pique-Regi et al., 2011; Song et al., 2011).

Rice (*Oryza sativa*), a model plant organism, is one of the world's staple foods, feeding a large proportion of the planet. He et al. (2010) produced a genome-wide map of H3K4me3, H3K9ac, and H3K27me3 in two rice subspecies and their reciprocal hybrids, which produced a considerable amount of valuable data. In this study, a more comprehensive investigation into the four histone marks in *O. sativa* L. japonica, including H3K4me2, H3K4me3, H3K9ac, and H3K27ac, was undertaken using ChIP-Seq and the DH sites were explored by public data. The four histone modifications in this study have been demonstrated to associate with active transcription and are non-transposable-element gene-specific (Li et al., 2008; Zhang et al., 2009; He et al., 2010; van Dijk et al., 2010; Ha et al., 2011), but detailed exploration of the relationships among them is still needed. As far as can be ascertained, the genome-wide study of H3K27ac by the ChIP-Seq method used in this study is the first time this process has been performed on a plant and will broaden understanding of the biological functions exerted by H3K27ac.

RESULTS

Comprehensive Analysis of Rice Sequencing Data

This study investigated four histone modifications (H3K4me2, H3K4me3, H3K9ac, and H3K27ac) in *O. sativa* L. japonica cv Nipponbare seedlings using ChIP-Seq with input DNA (non-ChIP genomic DNA) as the control. The antibodies recognizing these histone modifications were chosen from literatures (Li et al., 2008; Charron et al., 2009; He et al., 2010; Jang et al., 2011). We measured the H3K9ac and H3K27ac in rice seedlings by Western blotting and the total H3 serves as a loading control (Supplemental Figure 1). The tissue samples were also used for an mRNA-Seq experiment. The rice seedling DNase-Seq data were obtained from previous studies (Zhang et al., 2011) and analyzed along with ChIP-Seq from this study, and mRNA-Seq data to gain an integrated map of the epigenome and transcriptome in rice. A computational pipeline was designed, which consisted of published tools and self-developed scripts, for the sequencing data analysis (Supplemental Figure 2). By aligning ChIP-Seq reads to the reference genome (Rice Genome Annotation Project, TIGR/MSU version 6.1) using Burrows-Wheeler Aligner (BWA) software (Li, 2009), it was found that most reads (from 91.7% to 98.9%) were mappable to the genome, which suggested that the sequencing data were of a high quality. Model-based analysis using ChIP-Seq (MACS) software (Zhang et al., 2008) was used to obtain the peaks. Tens of thousands (from 24 324 to 39 971) of histone modification peaks were obtained using the ChIP-Seq data and 27119 DNase-Hypersensitive (DH) sites were obtained using the DNase-Seq data (Supplemental Table 1). Cufflinks software (Trapnell et al., 2010) was used to calculate the gene expression index using this study's mRNA-Seq data and a total of 26 912 genes were identified as being expressed (FPKM > 0). To evaluate the reliability of the ChIP-Seq results obtained in this study, ChIP-Seq data of the histone modifications (H3K4me3, H3K9ac, and H3K27me3) generated by He et al. (2010) were downloaded from the Gene Expression Omnibus database (accession number GSE19602) and analyzed using the same pipeline. The distributions of two of the histone modifications (H3K4me3 and H3K9ac) were consistent between this study's sequencing data with the published data set, confirming the accuracy of the ChIP-Seq data obtained in this study (Supplemental Figure 3). To visualize the sequencing data and analysis results, a customized, publically accessible rice UCSC genome browser was constructed (Kent et al., 2002) using relevant data to facilitate further exploration by the researchers in the plant community (http://structuralbiology.cau.edu.cn/cgi-bin/hgTracks).

Characterizing the Distribution of Histone Modifications and DH Sites in the Rice Genome

The distribution of peaks identified in the ChIP-Seq and DNase-Seq data along the rice genome were characterized using CEAS software (Shin et al., 2009). The rice genome was characterized into six classes that included five classes of genic regions (promoter, 5′ untranslated region (UTR), 3′ UTR, coding exon and intron) and intergenic regions. This study calculated the distribution of histone modifications and DH sites associated with the different genome categories (Figure 1 and Supplemental Figure 4). It was found that the four histone modifications were predominantly enriched in the generic regions, especially in coding exon, intron, and UTR regions. The intergenic and promoter regions occupied a major part of the DH sites (42.4% and 37.3%), which was consistent with previous reports (Zhang et al., 2011). It was also noticed that the H3K27ac, which was similar to other histone marks, accumulated mainly in generic regions (61.9%), and was most similar to another acylated modification, H3K9ac. Another two histone-methylated modifications (H3K4me2 and H3K4me3) also showed a similar pattern, implying that the same modifications at different places on the histone shared similar genome-wide distributions. Furthermore, it was found that two acylated modifications were more likely to be found within the intergenic regions but less likely to be found within the coding regions compared with the two histone-methylated modifications. The results showed different distribution preferences between acylated and methylated histone marks, which suggested that there were different regulation patterns for these two kinds of modifications.

Meta-gene profiles of the four histone modifications and the DH sites were generated along the generic regions for

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

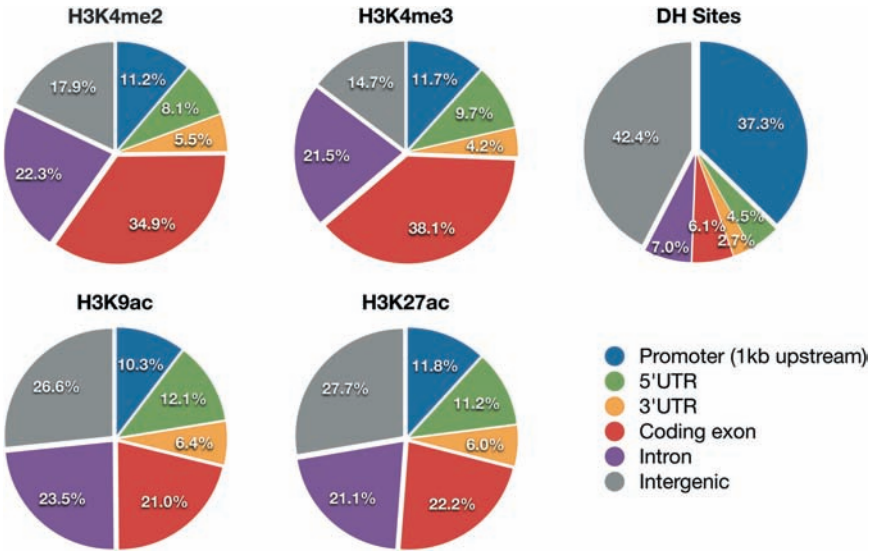

Figure 1. The Distribution of Histone Modifications and DH Sites within Different Regions of the Rice Genome.

12 295 transposable-element (TE) genes and 45 329 non-TE genes (Supplemental Figure 5 and Figure 2A). It was found that the histone modifications and the DH sites in this study accumulated mainly in non-TE genes, but were seldom found in TE genes. The ratio of genes associated with histone modification peaks and DH sites was calculated and uncovered a substantial number of non-TE genes that were associated with histone marks (Figure 2B and Supplemental Table 2). While approximately 65% and 43% of all non-TE genes were associated with the four histone modifications, only a small proportion of TE genes (6.5%) were identified as being correlated with the histone marks. One explanation for this phenomenon is that non-TE genes play an important role in maintaining biological activity and thus demand a comprehensive regulation mechanism, in which the histone modifications play an important role. In contrast, the TE genes, which are mostly non-functional, are repressed and did not contain many histone marks compared with non-TE genes. By exploring non-TE genes, it was discovered that histone modifications were significantly enriched around transcription start sites (TSSs) but seldom near transcription terminal sites (TTSs). The enriched regions of the four histone modifications (H3K4me2/3, H3K9ac, and H3K27ac) and the DH sites were mainly located within the gene promoter regions. Interestingly, the summits of the four histone modification peaks were in the gene body region, downstream of the TSSs, while the peak summits of the DH sites were upstream of the TSSs. One reason for this situation might be the existence of a NFR upstream of the TSSs. A representative region of genes associated with histone modification and DH sites from the UCSC genome browser is represented in Figure 2C.

Histone Modifications and DH Sites Are Associated with Active Transcription

The histone modifications and DNase sites around the TSSs of 25 940 expressed (FPKM > 0) non-TE genes associated with at least one histone modification peak or DH site were interrogated. It was found that the four histone modifications and the DH sites showed positive correlations with gene activity (Figure 3). By calculating the Pearson Correlation Coefficient (r) between the sequencing density of the histone modifications and the DH sites with gene expression, it was found that the correlations with H3K9ac and H3K27ac (both r = 0.55) were higher than the correlations with H3K4me2 and H3K4me3 (r = 0.54 and 0.42). The correlation between DH sites and gene expression was 0.78, which suggested that the DH sites have a strong influence on transcription activity (Supplemental Figure 6). One interesting observation was that, besides the increased density, the peak width of the DH sites was broader around genes with higher expressions.

The concurrence of histone marks and DH sites with transcript regions was calculated by checking whether a histone modification or DH site existed, for a given DH site or a histone modification peak (Figure 4). It was discovered that the concurrence frequencies of H3K4me2–H3K4me3 and H3K9ac–H3K27ac were both high (>80%) and the transcript regions were better correlated with H3K4me2 and H3K4me3 than H3K9ac and H3K27ac. Additionally, two histone-methylated modifications (H3K4me2/3) had relatively low concurrence frequencies compared to H3K9ac, H3K27ac, and the DH sites. Unlike histone modifications, DH sites had moderate concurrence frequencies with histone modifications and transcript regions, which may be due to the absence of histones, and therefore histone modifications.

Prediction of Transcription Factor Binding Sites within DH Sites

The DH sites are open chromatin regions where nucleosomes are depleted to facilitate the binding of transcription factors (TF) on the genomic DNA. Different TFs could specifically recognize corresponding motifs within the DH sites in order to

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

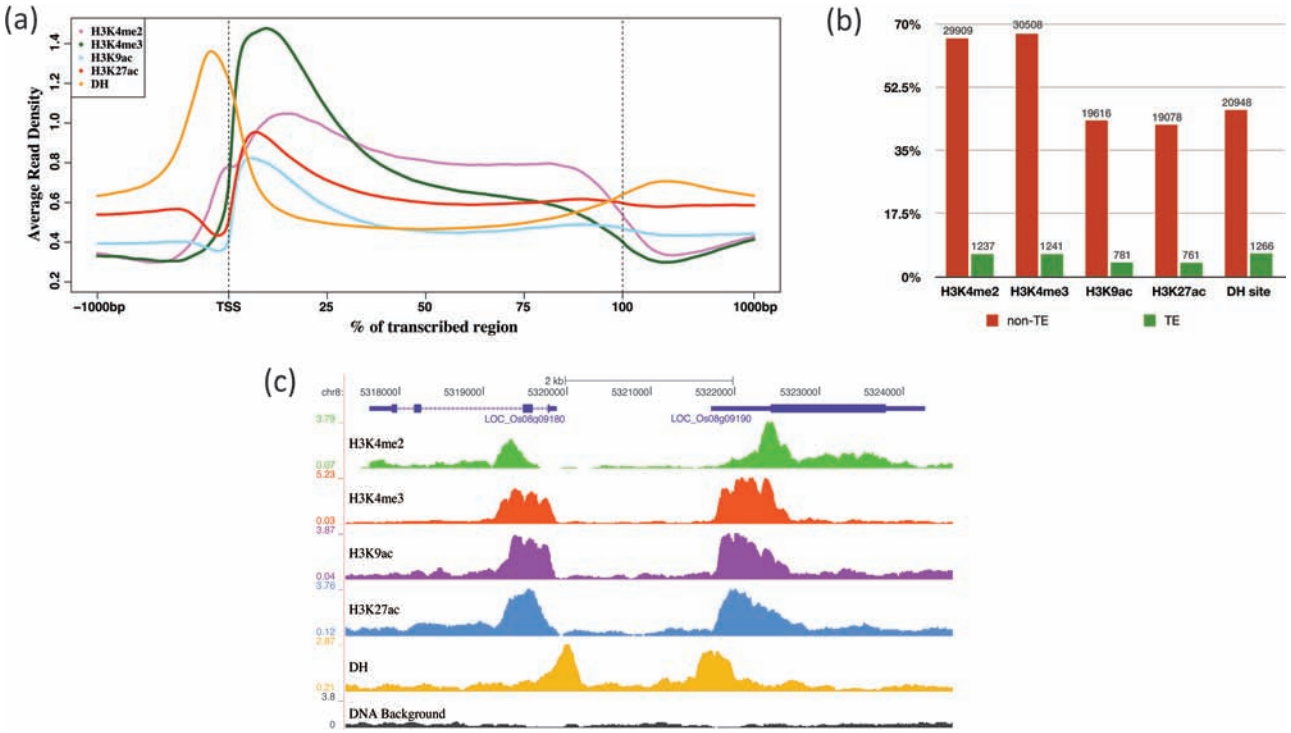

Figure 2. The Association between Histone Modifications and DH Sites with Rice TE and Non-TE Genes.
(A) Distribution of histone marks and DH sites along rice non-TE genes. A meta-gene profile was generated using the normalized sequencing density of the DH sites and four histone modifications: H3K4me3, H3K4me2, H3K9ac, and H3K27me3. The gene body was converted into a percentage in aim to standardize genes of different lengths. The 1-kb upstream and downstream regions of the gene are included.
(B) A histogram displaying the percentage of non-TE and TE genes associated with histone modifications or DH sites in rice. The number of genes in each category is shown at the top of each column.
(C) Representative genes associated with histone modifications and DH sites.

regulate gene transcription. To try and identify the motif hits within DH sites, this study collected 63 plant motifs from the Transfac database (Matys et al., 2003) and screened them with the custom Cistrome platform (Liu et al., 2011). The DH sites were divided into two classes: 15 029 promoter-associated DH sites that overlapped with rice promoter regions and 12 090 non-promoter-associated DH sites. Typically, the hit number within in a DH site ranged from 15 to 40, with a maximum of 25 (Supplemental Figure 7). It was found that the hit number of motifs was higher within promoter-associated DH sites than within non-promoter-associated DH sites (Supplemental File 1). The two motifs (M00948 and M01126) with the most hits in promoter-associated DH sites are shown in Figure 5A. The TFs related to M00948 and M01126 are rice gene PCF2 and *Arabidopsis* gene BPC1, according to previous reports (Kosugi and Ohashi, 2002; Matys et al., 2003). PCF2 is a transcription activator expressed in seedlings and leaves (Kosugi and Ohashi, 2002; Matys et al., 2003) and BPC1 is a regulator of the homeotic *Arabidopsis* gene, SEEDSTICK, which controls ovule identity (Meister et al., 2004). Interestingly, although the total number of promoter-associated DH sites was larger than the non-promoter-associated DH sites, the two motifs (M00503 and M00089), which were recognized by *Arabidopsis* genes, ATHB-5 and ATHB-1 (Sessa et al., 1993; Johannesson et al., 2001), had more hits in the non-promoter-associated DH sites. Previous studies characterized ATHB-5 as a positive regulator of ABA-responsiveness in developing seedlings (Johannesson et al., 2003) and ATHB-1 as a regulator in leaf development (Aoyama et al., 1995).

Identification of Missing Genes in Rice Annotation by Histone Modifications

After discovering the high concurrence frequency between histone modifications and transcript regions, it was proposed that these marks could be used to identify unannotated genes. There are two widely used resources for gene annotation in rice (*Japonica*): MSU/TIGR (Ouyang et al., 2007) and RAP–DB (Itoh et al., 2007). We found 6629 RAP–DB genes whose TSSs did not overlap with any TSS region (500 bp up or downstream) in the MSU/TIGR gene, meaning that these genes were missing in the annotation. Among these unannotated genes, 471 genes had both H3K4me3 and H3K9ac peaks within 400 bp regions downstream of the TSS (Supplemental File 2). These results suggested that these genes existed but were missing in the rice MSU/TIGR annotation. To evaluate this prediction, two RAP–DB genes (Os02g0510400 and Os11g0564800),

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

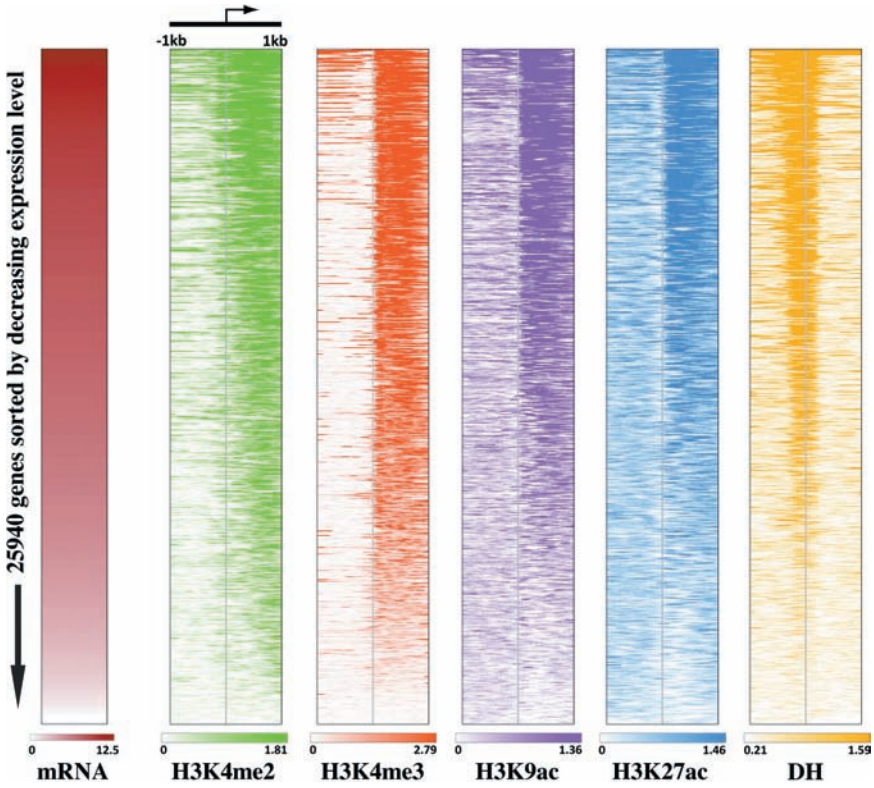

Figure 3. Heatmaps for Transcription Activity, Histone Modifications, and DH Sites around the TSSs of Non-TE Genes.
The genes were sorted according to their expression level, as determined by mRNA-Seq. For each gene, the histone modification intensities are displayed along –1-kb to 1-kb regions around the TSSs. Each color represents one histone modification, DH sites, or transcription activity.

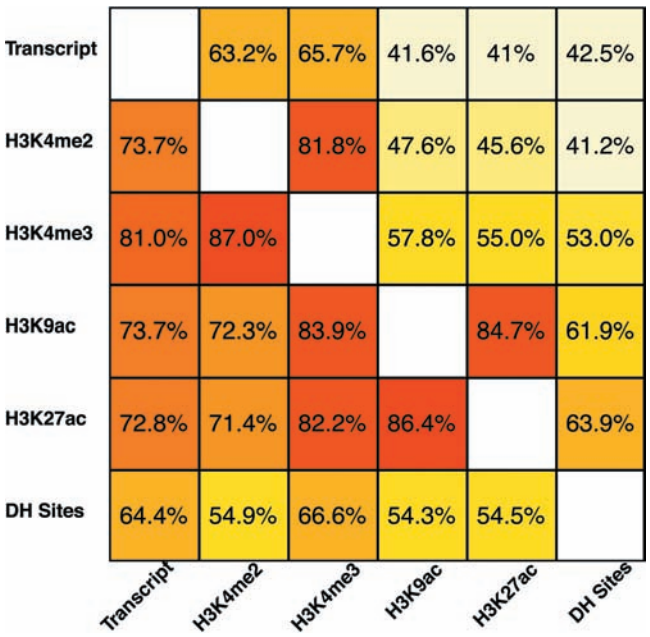

Figure 4. Concurrence Frequencies for Histone Modifications, DH Sites, and Transcript Regions.
The percentage number indicates the possibility that a histone modification peak, DH site, or transcript region on the *x*-axis exists in a given histone modification peak, DH site, or transcript region on the *y*-axis. A transcript region is defined as 1 kb up or downstream of the TSS.

neither of which had a functional annotation, were selected from the list for experimental validation. Analysis using histone modifications showed that Os02g0510400 was located near a MSU/TIGR gene, LOC_Os02g30640, but had no overlap with any other gene (Figure 6A). The PCR experiments proved that Os02g0510400 existed and was transcribed under two different conditions: normal conditions and when treated with methyl viologen (MV) (Figure 6B). In another example, the Os11g0564800 overlapped with the MSU/TIGR annotated gene, LOC_Os11g35840, but was located on the antisense strand (Figure 6C). Intriguingly, both genes were transcribed, but under different conditions (Figure 6D). Os11g0564800 was transcribed after MV treatment and repressed under normal conditions, but LOC_Os11g35840 exhibited the opposite expression pattern. The results demonstrated that histone modifications that have a high concurrence possibility with active transcript loci could be used to predict unannotated genes—a valuable approach for poorly annotated organisms.

DISCUSSION

In this study, a genome-wide examination of four histone marks, H3K4me2, H3K4me3, H3K9ac, and H3K27ac, in rice (*O. sativa* L. japonica) seedlings was performed using ChIP-Seq experiments. By integrating public DNase-Seq data, the distribution patterns of four histone modifications and DH

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

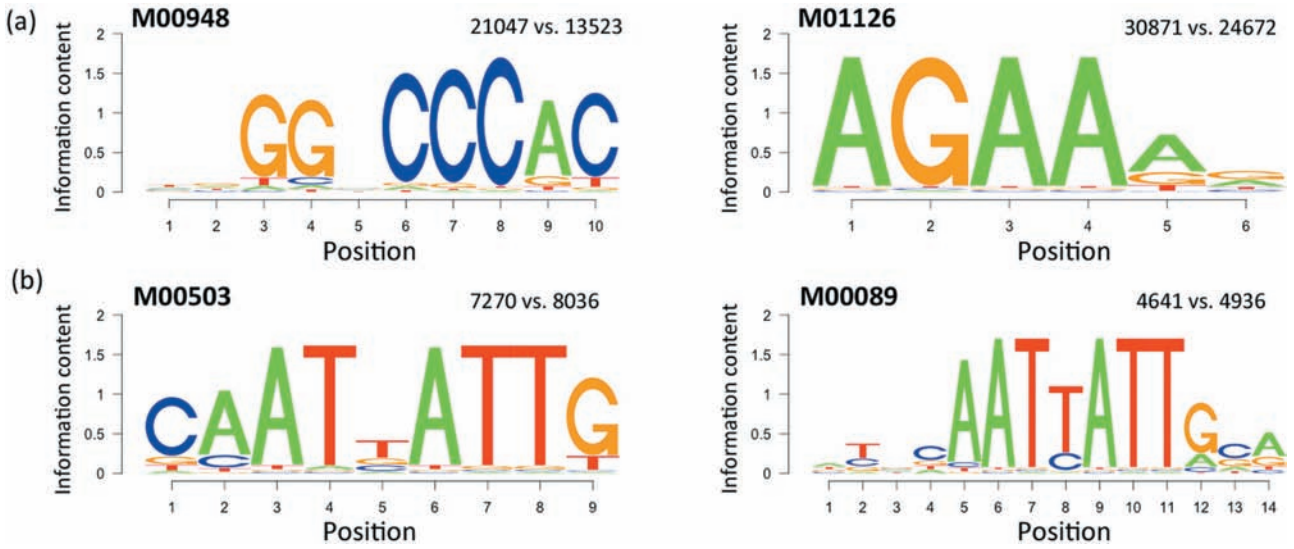

Figure 5. Sequence Logo of Identified Motifs within DH Sites.
(A) The top two motifs have more hits in the promoter-associated DH sites compared to hits in the non-promoter-associated DH sites.
(B) These two motifs have more hits in the non-promoter-associated DH sites than in the promoter-associated DH sites. For each motif, the numbers of motif hits within the promoter-associated DH sites and the non-promoter-associated DH sites are shown on the upper right of each figure.

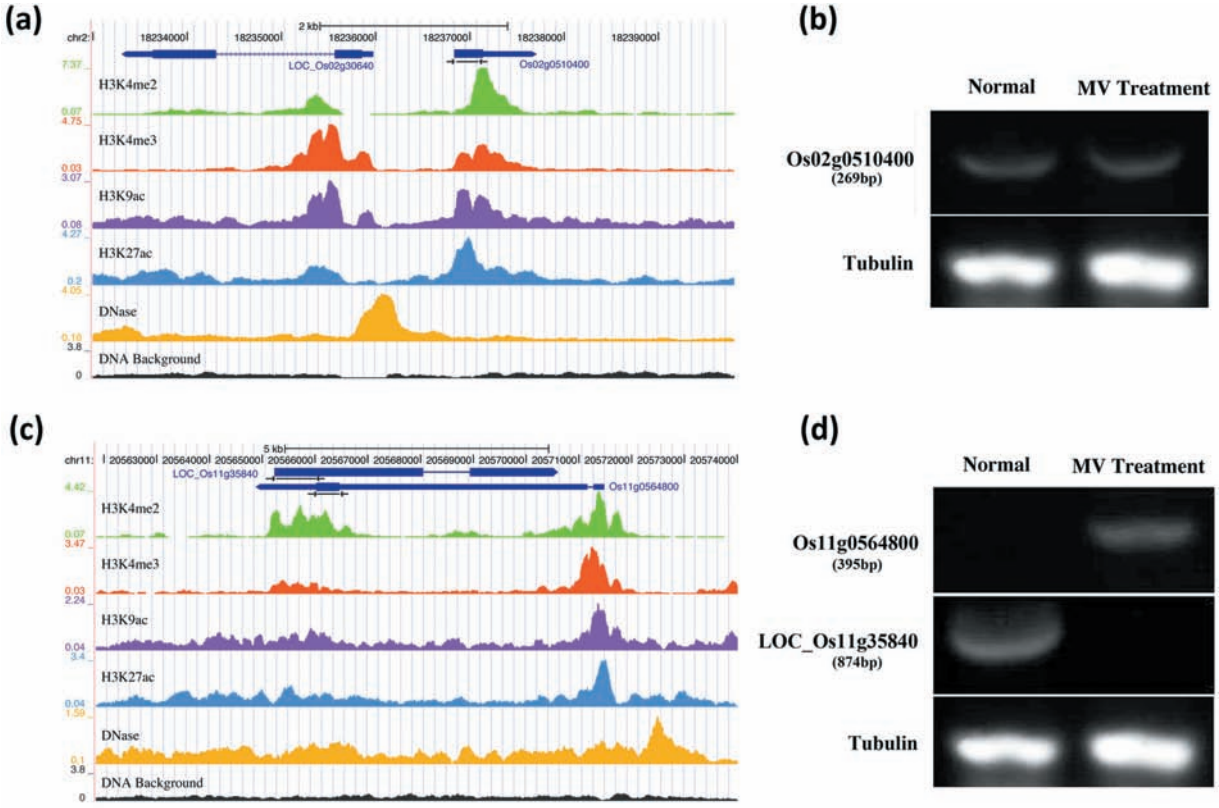

Figure 6. Experimental Validation of Two Unannotated Genes in the TIGR/MSU Annotation.
(A) The gene model for Os02g0510400 **(B)** PCR analysis of Os02g0510400 under normal conditions and under methyl viologen (MV) treatment.
(C) The gene model for Os11g0564800 **(D)** PCR analysis of Os11g0564800 under normal conditions and under MV treatment. Primers designed for the PCR experiments are highlighted by the black arrows.

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

sites along the rice genome were systematically studied. We found that the four histone modifications were mainly located within generic regions of non-TE genes and enriched around TSSs. In the meta-profile of histone modifications in humans, there is an enrichment upstream of TSSs (Barski et al., 2007; Wang et al., 2008, 2009), whereas this study did not find this in rice. Notably, the global characterization of H3K27ac will benefit the plant community, since, as far as can be ascertained, this is the first time that H3K27ac has been investigated using ChIP-Seq in plants. These results indicated that H3K27ac showed a similar distribution pattern to another acylated mark, H3K9ac, in the rice genome. While most identified peaks of H3K27ac were located in generic regions, enrichment was highest in intergenic regions compared with the three other histone modifications. This phenomenon was consistent with a previous report that suggested that H3K27ac may serve as the marker for active enhancers (Creyghton et al., 2010). Additionally, 3442 non-promoter-associated DH sites were found to overlap with H3K27ac peaks, which are important open chromatin regions that play a role in regulating transcription activity.

Nucleosomes compete with TFs for access to genome sequences and play an important part in the gene regulation mechanism. In living cells, nucleosomes associated with histone modifications are more likely to be involved in the regulation of gene transcription because their occupancies are fine-tuned by posttranslational modifications, such as histone modifications. In order to explore the distribution of nucleosomes associated with histone modifications, this study profiled nucleosome distribution around the TSS using the ChIP-Seq data (Supplemental Figure 8) obtained during this study. For H3K4me2, H3K4me3, and H3K9ac, the number of modified nucleosomes was relatively low upstream of the TSS and accumulated in a periodic way downstream of the TSS. The distribution pattern of the nucleosomes associated with H3K27ac was distinct from the other three marks, suggesting a different position for nucleosomes with the H3K27ac modification in rice. Unlike enrichment of H3K4me2/3 upstream of TSSs in humans (Barski et al., 2007), there were few nucleosomes associated with these two histone marks, implying a different mechanism for gene regulation through nucleosome positioning in plants. Finally, both the density and the peak width of DH sites were positively correlated with gene expression, indicating a potential mechanism for gene regulation by nucleosome positioning in rice.

The high concurrence frequency between histone modifications and transcript regions suggested that H3K4me3 and H3K9ac could be used to identify missing genes in the rice MSU/TIGR annotation. One methylation mark (H3K4me3) and one acylation mark (H3K9ac) were selected in order to avoid information redundancy. The strategy was validated using PCR experiments and proved the transcription of both genes. Technology, like mRNA-Seq, could trace missing genes systematically and histone modifications have their advantages. They can be used to identify non-coding genes, such as miRNA precursors, and to investigate genes whose active transcription rely on diverse conditions and are thus out of the reach of a single mRNA-Seq experiment. Histone marks represent a good approach for characterizing missing genes in the genome annotation of organisms that have not been investigated to any great extent.

This study undertook a comprehensive investigation of DH sites and four histone modifications and explored the relationships among them. Compared with our material in the ChIP-Seq experiments, the collected DNase-Seq data (Zhang et al., 2011) used rice seedling in later stage (2 weeks). Although it is possible that some of the DH sites are only active in later seedling stage, we believe the majority of DH sites remain steady when considering the consistence of histone modifications between our ChIP-Seq data and histone modification data in the later seeding stage (4 weeks) from He et al. (2010). Overall, the results of this study should contribute to scientific knowledge by providing a systematic analysis of the regulatory mechanism exerted by histone modifications and DH sites.

METHODS

Plant Materials and Growth Conditions

Seeds of rice (Nipponbare) were germinated in water for 1 d at room temperature and for 1 d at 35°C and then grown in a greenhouse for 4 d (24-h darkness, 28°C). The seedlings of the rice plants, including shoots and roots, were harvested together.

Immunoblotting

One gram of frozen seedlings was grinded to a fine powder in liquid nitrogen. Then, transfer the powder to a new container added 2 ml extraction buffer (0.2 M Tris-HCl, pH 6.8, 8% SDS, 40% glycerol, 20% 2-mercaptoethanol, 0.1% bromophenol blue) and incubate for 5 min at 60°C. The tube was moved on the ice for 5 min and centrifuge for 10 min at 13 000 rpm. The supernatant was loaded into 15% SDS–PAGE, and transferred to Immobilon PVDF membranes. Histone modifications were detected with the antibodies used for ChIP experiments and visualized with film using analkaline phosphatase-conjugated anti-rabbit IgG secondary antibody and the BCIP/NBT solution (Amresco).

Isolation and Immunoprecipitation of Chromatin

A native chromatin immunoprecipitation (ChIP) was performed essentially as described by Nagaki et al. (2003). Approximately 20 g of etiolated seedlings (1 week old) were ground to a fine powder with liquid nitrogen and re-suspended in TBS (10 mMTris, pH 7.5, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.1 mM PMSF, 2/5 tab of complete mini (Roche Applied Science, Indianapolis) with 0.5% Tween 40. The nuclei were purified in a sucrose gradient and digested with 10 units of micrococcal nuclease (MNase: Sigma, St Louis) for 5 min at 37°C. The nucleosome samples were first incubated with pre-immune

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

rabbit serum (1:100 dilution) and then with 4% protein A Sepharose (GH healthcare Bio-Sciences AB, Uppsala) for 4 h, after which they were centrifuged. The histone modification antibodies were same as those used in previous reports (Wang et al., 2008; Shin et al., 2012). The supernatant was incubated with anti-dimethyl histone H3 (Lys 4) (H3K4me2, Upstate (Millipore) 07–030, Temecula), anti-trimethyl-histone H3 (Lys 4) (H3K4me3, Upstate (Millipore) 07–473, Temecula), anti-acetyl-histone H3 (Lys 9) (H3K9Ac, Abcam (Hong Kong) Ltd, ab10812), and anti-acetyl-histone H3 (Lys 27) (H3K27ac, Abcam (Hong Kong) Ltd, ab4729) antibodies at 4°C overnight. An equal amount of pre-immune rabbit serum, which served as a nonspecific binding control in each ChIP experiment, was used in the control experiments. Then the samples were incubated with 25% protein A Sepharose at 4°C for 4 h. After centrifugation, the pellet (bound) fractions were subjected to a series of washes and the immune complexes were eluted from the washed beads using elution buffer (50 mM NaCl, 20 mM Tris-HCl at pH 7.5, 5 mM EDTA, 1% SDS). Immunoprecipitated DNA was extracted using phenol/chloroform and precipitated with ethanol. The ChIPed DNA was re-suspended in 100 µl TE buffer (pH 8.0).

Two biological replicates treated with antibodies (treatment) and two replicates with a mock treatment (control) were used in each ChIP experiment. Chip experiments were undertaken only when two treatment samples both showed significant differences compared with the control samples after real-time PCR. Quantitative real-time PCR analysis of ChIPed DNA was performed on the Applied Biosystems 7500 Real Time PCR System using the SYBR Premix Ex Taq kit (TakaRa Biotechnology (Dalian) Co., Ltd) in order to determine the relative fold enrichment (RFE) of modified histone-associated sequences in the bound fractions. The real-time PCR reactions were performed using the primer, Kinase-F/R, as the positive control, as it shows a higher enrichment of H3K4me2, H3K4me3, H3K9ac, and H3K27ac, and primer sets Os25S-F/R or Cen8.t00719-F/R were used as the negative control because of their lower enrichment (Supplemental Table 3). The PCR cycle threshold (CT) was set at <0.05 for most of the experiments. For each primer set, the ΔCT value was calculated using the comparative CT method (ΔCT = CT (bound DNA) – CT (mock DNA)) according to Wang et al. (2011). Relative enrichment was calculated as $2^{-\Delta\Delta CT}$ ± standard deviation (SD), where ΔΔCT = ΔCT(Kinase) – ΔCT(negative control). The two treatment replicates were then mixed and sent for sequencing.

Computational Analysis of Sequencing Data

Rice seedling ChIP-Seq and mRNA-seq data were sequenced using IlluminaHiSeq™ 2000 followed by standard protocols. The DNase-Seq data was downloaded from the NCBI Gene Expression Omnibus (GEO) (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26610) (Zhang et al., 2011). Sequencing reads were aligned to the reference genome (MSU Rice Genome Annotation Release 6.1) by BWA software using default parameters (Ouyang et al., 2007). MACS software was used to identify peaks (Zhang et al., 2008). The MFOLD parameter (enrichment ratio region for model building) in MACS was set at eight. The CEAS software (Shin et al., 2009) was used to comprehensively analyze the identified peaks and cufflinks software (Trapnell et al., 2010) was used to calculate expression levels using mRNA-Seq data from the default parameters. The gene expression level was normalized by gene length and reads number and then measured using FPKM (Fragments Per Kilobase of transcript per Million mapped reads). The UCSC genome browser (Kent et al., 2002) was customized for the rice genome and hosted the sequencing data and analysis results. In order to identify the TF binding sites, a 300-bp region centered by the peak summit of the DH sites was used for motif screening by the Motif Screen Tool our custom built Cistrome platform (Liu et al., 2011). Lastly, NPS software (Zhang et al., 2008) was used to identify nucleosomes associated with histone marks using the ChIP-Seq data collected in this study and then used the central points of the nucleosomes to chart nucleosome positioning along the TSSs.

RT–PCR Validation of Missing Genes in the MSU/TIGR Rice Annotation

Semi-quantitative RT–PCR was used to validate transcription in both the normal and the MV treatment libraries. Primers were designed according to the rice MSU/TIGR and RAP-DB annotations (Supplemental Table 3). For transcripts amplification, 32 cycles were used and 30 cycles were used for Tubulin amplification.

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online.*

FUNDING

This work was supported by grants from the Natural Science Foundation of China (31071125, 31025018, and 31028011) and the Ministry of Education of China (NCET-09-0735), as well as the National Institutes of Health of US (grant no. GM99409).

ACKNOWLEDGMENTS

We thank Qunlian Zhang for technical support and Yiwen Chen for critical suggestion. No conflict of interest declared.

REFERENCES

Aoyama, T., Dong, C., Wu, Y., and Carabelli, M. (1995). Ectopic expression of the *Arabidopsis* transcriptional activator Athb-1 alters leaf cell fate in tobacco. Plant Cell. **7**, 1773–1785.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007).High-resolution

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

profiling of histone methylations in the human genome. Cell. **129**, 823–837.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature. **447**, 407–412.

Charron, J.-B.F., He, H., Elling, A.A., and Deng, X.W. (2009). Dynamic landscapes of four histone modifications during deetiolation in *Arabidopsis*. Plant Cell. **21**, 3732–3748.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl Acad. Sci. U S A. **107**, 21931–21936.

Ha, M., Ng, D., and Li, W. (2011). Coordinated histone modifications are associated with gene expression variation within and between species. Genome Res. **21**, 590–598.

He, G., Zhu, X., Elling, A.A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., et al. (2010).Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell. **22**, 17–33.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature. **39**, 311–318.

Ho, J.W.K., Bishop, E., Karchenko, P.V., Negre, N., White, K.P., and Park, P.J. (2011). ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. BMC Genomics. **12**, 134.

Itoh, T., Tanaka, T., Barrero, R.A., Yamasaki, C., Fujii, Y., Hilton, P.B., Antonio, B.A., Aono, H., Apweiler, R., Bruskiewich, R., et al. (2007). Curated genome annotation of *Oryza sativa* ssp. japonica and comparative genome analysis with *Arabidopsis thaliana*. Genome Res. **17**, 175–183.

Jang, I.-C., Chung, P.J., Hemmes, H., Jung, C., and Chua, N.-H. (2011). Rapid and reversible light-mediated chromatin modifications of *Arabidopsis* phytochrome A locus. Plant Cell. **23**, 459–470.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. **10**, 161–172.

Johannesson, H., Wang, Y., and Engström, P. (2001). DNA-binding and dimerization preferences of *Arabidopsis* homeodomain–leucine zipper transcription factors *in vitro*. Plant Mol. Biol. **45**, 63–73.

Johannesson, H., Wang, Y., and Hanson, J. (2003). The *Arabidopsis thaliana* homeobox gene ATHB5 is a potential regulator of abscisic acid responsiveness in developing seedlings. Plant Mol. Biol. **51**, 719–729.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. **12**, 996–1006.

Kosugi, S., and Ohashi, Y. (2002). DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J. **30**, 337–348.

Li, H. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. **25**, 1754–1760.

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

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y., et al. (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol. **12**, R83.

Matys, V., Fricke, E., and Geffers, R. (2003). TRANSFAC®: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. **31**, 374–378.

Meister, R., Williams, L., and Monfared, M. (2004). Definition and interactions of a positive regulatory element of the *Arabidopsis* INNER NO OUTER promoter. Plant J. **37**, 426–438.

Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S., and Jiang, J. (2003). Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics. **163**, 1221–1225.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., et al. (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. **35**, D883–D887.

Pfluger, J., and Wagner, D. (2007). Histone modifications and dynamic regulation of genome accessibility in plants. Curr. Opin. Plant Biol. **10**, 645–652.

Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. **21**, 447–455.

Schones, D.E., and Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. Nat. Rev. Genet. **9**, 179–191.

Sessa, G., Morelli, G., and Ruberti, I. (1993). The Athb-1 and –2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. EMBO J. **12**, 3507–3517.

Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: *cis*-regulatory element annotation system. Bioinformatics. **25**, 2605–2606.

Shin, J.H., Li, R.W., Gao, Y., Baldwin, R., and Li, C.J. (2012). Genome-wide ChIP-seq mapping and analysis reveal butyrate-induced acetylation of H3K9 and H3K27 correlated with transcription activity in bovine cells. Funct. Integr. Genomics. **12**, 119–130.

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.K., Sheffield, N.C., Gräf, S., Huss, M., Keefe, D., et al. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. **21**, 1757–1767.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. **28**, 511–515.

van Dijk, K., Ding, Y., Malkaram, S., Riethoven, J.-J.M., Liu, R., Yang, J., Laczko, P., Chen, H., Xia, Y., Ladunga, I., et al. (2010). Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in *Arabidopsis thaliana*. BMC Plant Biol. **10**, 238.

Wang, G., He, Q., Liu, F., Cheng, Z., Talbert, P.B., and Jin, W. (2011).Characterization of CENH3 proteins and

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014

centromere-associated DNA sequences in diploid and allotetraploid Brassica species. Chromosoma. **120**, 353–365.

Wang, Z., Schones, D.E., and Zhao, K. (2009). Characterization of human epigenomes. Curr. Opin. Genet. Dev. **19**, 127–134.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., et al. (2008).Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. **40**, 897–903.

Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J. (2011). High-resolution mapping of open chromatin in the rice genome. Genome Res. **22**, 151–162.

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

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. **9**, R137.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2010). Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. **12**, 7–18.

Downloaded from http://mplant.oxfordjournals.org/ at Harvard University on February 24, 2014