



## High-resolution mapping of open chromatin in the rice genome

Wenli Zhang, Yufeng Wu, James C. Schnable, et al.

*Genome Res.* 2012 22: 151-162 originally published online November 22, 2011  
Access the most recent version at doi:[10.1101/gr.131342.111](https://doi.org/10.1101/gr.131342.111)

---

**References** This article cites 67 articles, 21 of which can be accessed free at:  
<http://genome.cshlp.org/content/22/1/151.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Copyright © 2012 by Cold Spring Harbor Laboratory Press

# High-resolution mapping of open chromatin in the rice genome

Wenli Zhang,<sup>1,4</sup> Yufeng Wu,<sup>1,4</sup> James C. Schnable,<sup>2</sup> Zixian Zeng,<sup>1</sup> Michael Freeling,<sup>2</sup> Gregory E. Crawford,<sup>3</sup> and Jiming Jiang<sup>1,5</sup>

<sup>1</sup>Department of Horticulture, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA; <sup>2</sup>Department of Plant and Microbial Biology, University of California–Berkeley, Berkeley, California 94720, USA; <sup>3</sup>Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina 27708, USA

Gene expression is controlled by the complex interaction of transcription factors binding to promoters and other regulatory DNA elements. One common characteristic of the genomic regions associated with regulatory proteins is a pronounced sensitivity to DNase I digestion. We generated genome-wide high-resolution maps of DNase I hypersensitive (DH) sites from both seedling and callus tissues of rice (*Oryza sativa*). Approximately 25% of the DH sites from both tissues were found in putative promoters, indicating that the vast majority of the gene regulatory elements in rice are not located in promoter regions. We found 58% more DH sites in the callus than in the seedling. For DH sites detected in both the seedling and callus, 31% displayed significantly different levels of DNase I sensitivity within the two tissues. Genes that are differentially expressed in the seedling and callus were frequently associated with DH sites in both tissues. The DNA sequences contained within the DH sites were hypomethylated, consistent with what is known about active gene regulatory elements. Interestingly, tissue-specific DH sites located in the promoters showed a higher level of DNA methylation than the average DNA methylation level of all the DH sites located in the promoters. A distinct elevation of H3K27me3 was associated with intergenic DH sites. These results suggest that epigenetic modifications play a role in the dynamic changes of the numbers and DNase I sensitivity of DH sites during development.

[Supplemental material is available for this article.]

The identification and functional characterization of the regulatory DNA elements is essential for understanding the regulation of gene expression in eukaryotic genomes. Although the genomes of an increasing number of eukaryotic species have been sequenced, genome-wide identification of regulatory DNA elements, such as that being done in the ENCODE project (The ENCODE Project Consortium 2007) and the Epigenomics Roadmap (Bernstein et al. 2010) in humans and in the modENCODE projects in *Caenorhabditis elegans* and *Drosophila melanogaster* (Gerstein et al. 2010; Roy et al. 2010), has been initiated only in few species. Active regulatory DNA elements, such as promoter and enhancers, interact with regulatory proteins. As a result, these regions are either free of nucleosomes or are under dynamic nucleosome modifications or displacements (Henikoff et al. 2009; Jin et al. 2009). Thus, active DNA elements are associated with “open chromatin” in higher eukaryotic genomes. One distinct characteristic of the genomic regions of open chromatin is a pronounced sensitivity to cleavage of endonuclease DNase I (Wu 1980; Keene et al. 1981; McGhee et al. 1981). Almost all active regulatory elements, including promoters, enhancers, suppressors, insulators, and locus control regions, have been shown to be marked by DNase I hypersensitive (DH) sites (Gross and Garrard 1988).

Until recently, mapping of individual DH sites in higher eukaryotes was mostly achieved using the traditional gel-based approach (Nedosposov and Georgiev 1980; Wu 1980; Kodama et al. 2007). However, new techniques have been developed for genome-wide mapping of DH sites using microarray-based platforms

(Crawford et al. 2006a; Sabo et al. 2006) or high-throughput sequencing-based platforms (Crawford et al. 2006b). Genome-wide DH site maps have been generated in *Saccharomyces cerevisiae* (Hesselberth et al. 2009), *D. melanogaster* (Kharchenko et al. 2011), and humans (Boyle et al. 2008b) using these new techniques. In addition, mapping of DH sites has been proven to be an effective approach to identify regulatory elements. For example, the binding sites of several of the best-characterized regulatory proteins in mammalian species, including the insulator protein CTCF in humans and the glucocorticoid receptor in mouse, overlapped well with the DH sites (Boyle et al. 2011; John et al. 2011). In *Drosophila*, the binding patterns of 21 developmental regulator are quantitatively correlated with DNA accessibility in chromatin that can be measured by the DNase I sensitivity (Li et al. 2011).

Rice (*Oryza sativa*) is the most important food crop in the world and has also been established as a model species for plant genome research. Rice provides one of the most accurately sequenced genomes from any multicellular eukaryotes (Goff et al. 2002; Matsumoto et al. 2005). Extensive genome-wide DNA methylation and histone modification data sets have recently been generated in rice (Feng et al. 2010; He et al. 2010; Yan et al. 2010; Zemach et al. 2010). Here, we describe high-resolution maps of DH sites in rice from both seedling and callus tissues. We report a number of novel features associated with rice DH sites, including their epigenetic modifications, dynamic response to tissue culture, and association with genes that differentially expressed genes in seedling and callus tissues.

## Results

### Genome-wide identification of DH sites in the rice genome

To generate a high-resolution map of DH sites in the rice genome, we constructed a total of five DNase-seq libraries (see Methods),

<sup>4</sup>These authors contributed equally to this work.

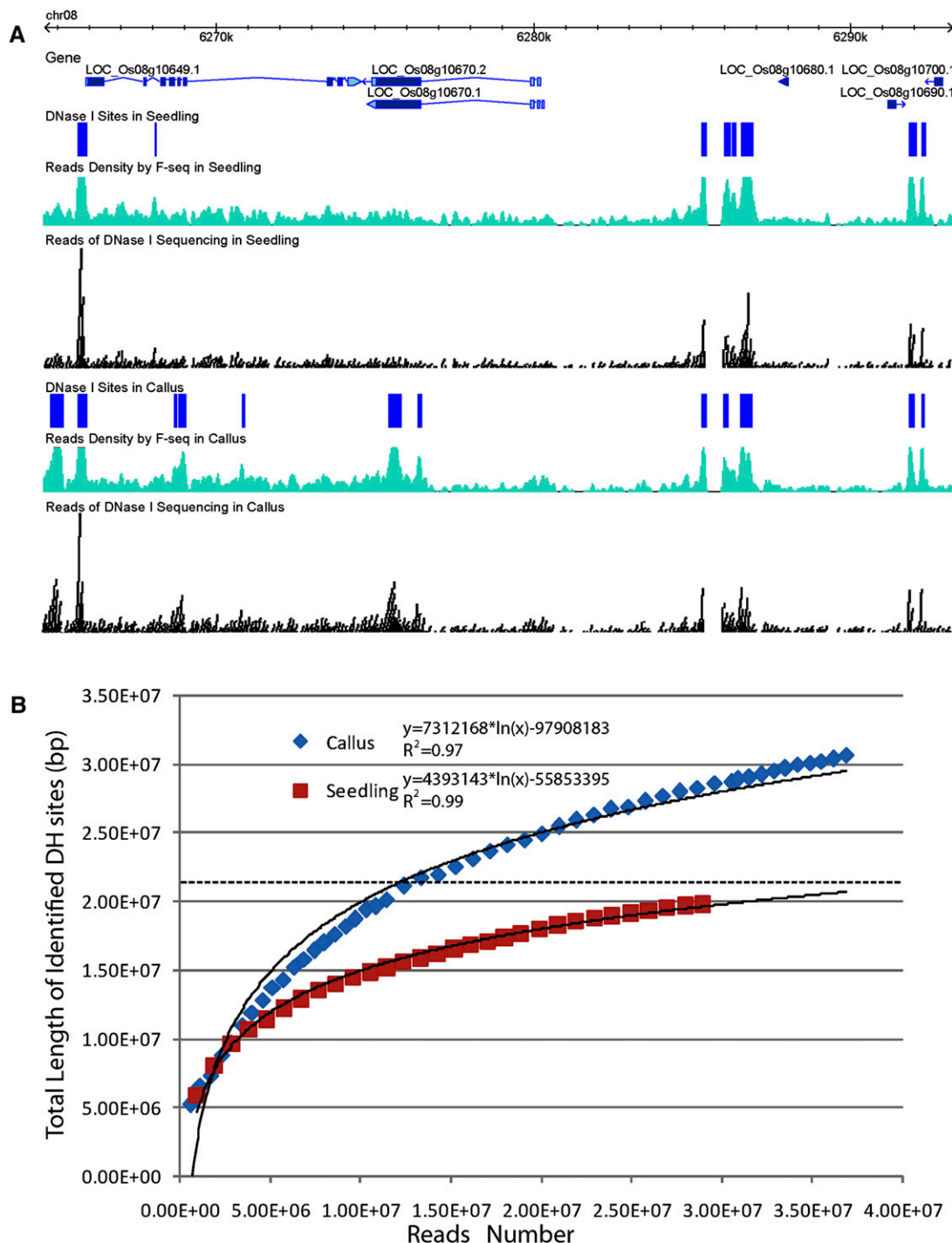
<sup>5</sup>Corresponding author.

E-mail [jjiang1@wisc.edu](mailto:jjiang1@wisc.edu).

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.131342.111>.

including three from seedling tissue (consisting of mostly leaf and a small proportion of stem tissues) and two from callus tissue. These libraries were sequenced using the Illumina Genome Analyzer. We obtained a total of 43 million sequence reads from the seedling libraries and 57 million reads from the callus libraries

(Supplemental Table S1). Approximately 70% of the reads were mapped to unique positions in the rice genome. We used the F-seq software (Boyle et al. 2008a) to generate a DNase I sensitivity value at every base using a kernel density estimation (Fig. 1A) and to identify DH sites (FDR < 0.05).



**Figure 1.** Identification of DH sites in the rice genome. (A) A selected region showing the raw data of DNase-seq (track of reads of DNase I sequencing), density estimation by F-seq (track of reads density by F-seq) and DH sites identified based on these data (track of DH sites). The rice gene models are shown at the top of the tracks. (B) The correlation between the number of DNase-seq reads and total length of DH sites within each subset of data. The dashed line indicates the putative horizontal asymptote of the seedling simulation data.

To examine the reproducibility and reliability of DNase-seq, we compared the data from biological replicates using scatter analysis. The high degree of correlation (Pearson correlation coefficient [PCC] ranged from 0.87–0.93) (Supplemental Fig. S1) indicates that DNase-seq is indeed a reliable method for identifying DH sites. Approximately 70.4%–76.5% of the DH sites were reproducible between biological replicates from seedlings. To test if the <30% unshared DH sites between biological replicates were possibly caused by either sequencing depth or technical variation, we first intentionally sequenced the same callus library twice (Supplemental Table S1) and found a similar frequency (71.9%–79.6%) of overlapping DH sites between the two technical replicates. We also found that most of the unshared DH sites between the biological replicates showed lower DNase I sensitivities than the shared ones (Supplemental Fig. S2). In addition, 92%–96% of the top 70% most-sensitive DH sites identified from one library were detected in the other two biological replicates, respectively. These results further confirm the high reliability of both the DNase-seq method and the quality of the generated data sets. Thus, the variation in DH site identification between biological replicates was most likely caused by lack of depth of sequencing of each library, which would not reveal the DH sites with relatively low sensitivity to DNase I digestion.

We conducted real-time quantitative PCR (qPCR) to verify the DH sites. In the qPCR experiments, we did not detect an unambiguous difference between the nondigested DNA and DNA digested with 0.12–0.16 units of DNase I, which was the same condition for DNase-seq library construction. A similar result was also reported in the qPCR-based DNase I sensitivity analysis of the mouse  $\beta$ -globin locus control region (McArthur et al. 2001). Thus, we used a higher amount of DNase I (4.0 units) in DNA digestion. Under such a condition, the qPCR method is less sensitive than the DNase-seq method, likely resulting in a significant number of false-positive DH sites. We randomly selected 16 DH sites and 13 non-DH sites (seedling tissue data set) to examine the difference of threshold cycles ( $\Delta$ Ct) between DNase I-digested and nondigested DNA. All 16 DH sites had more than two cycles in DNase I-digested sample than nondigested DNA, which corresponds to an approximately fourfold to 103-fold difference in sensitivity to DNase I digestion (Supplemental Table S2). Only four of the 13 non-DH sites had a  $\Delta$ Ct of more than two cycles, two of which were located <1 kb away from a positive DH site. Thus, we conclude that the DH sites identified by DNase-seq are well associated with genomic regions that are highly sensitive to DNase I digestion.

Considering the high correlation between replicates and verification by qPCR, we combined the sequence reads from the same tissue for further analysis. By using a FDR of <0.05, we identified 97,975 and 155,025 DH sites in the seedling and callus, respectively. We simulated the correlation between sequencing depth (the number of reads of a subset of data) and the total length of all DH sites identified (see Methods) (Fig. 1B). The simulation revealed a strong relationship of logarithmic growth between DH sites and the number of reads for DH site identification ( $R^2$ , 0.99 in seedling and 0.97 in callus). We estimated that ~95% of the potential seedling DH sites were identified based on the seedling data simulation.

#### DH sites are significantly associated with conserved noncoding sequences (CNSs) and protein-binding *cis* elements

CNSs have been identified in both mammalian and plant species and are significantly associated with regulatory sequences (Freeling

and Subramaniam 2009; Haeussler and Joly 2011). We were interested in the association between CNS and the DH site since both types of DNA sequences are likely related to regulatory elements. The *liguleless1* gene was extensively studied in grass species, and a total of seven CNSs were identified within this gene (Kaplinsky et al. 2002). A DH site, showing in both seedling and callus tissues, was identified in the 5' UTR of this gene. This DH site is partially overlapped with one of the CNSs ( $P = 0.034$ , binomial test) (Fig. 2A).

We then examined the association of rice DH sites with the 46,355 CNSs conserved between the rice and sorghum genomes (Schnable et al. 2011). These CNSs span a total of 1.6 Mb of rice genome sequences. We found that 25.7% (11,911) and 41.6% (19,281) of the CNSs were associated with DH sites in the seedling and callus, respectively ( $P < 0.001$ , binomial test) (for an example region, see Fig. 2B).

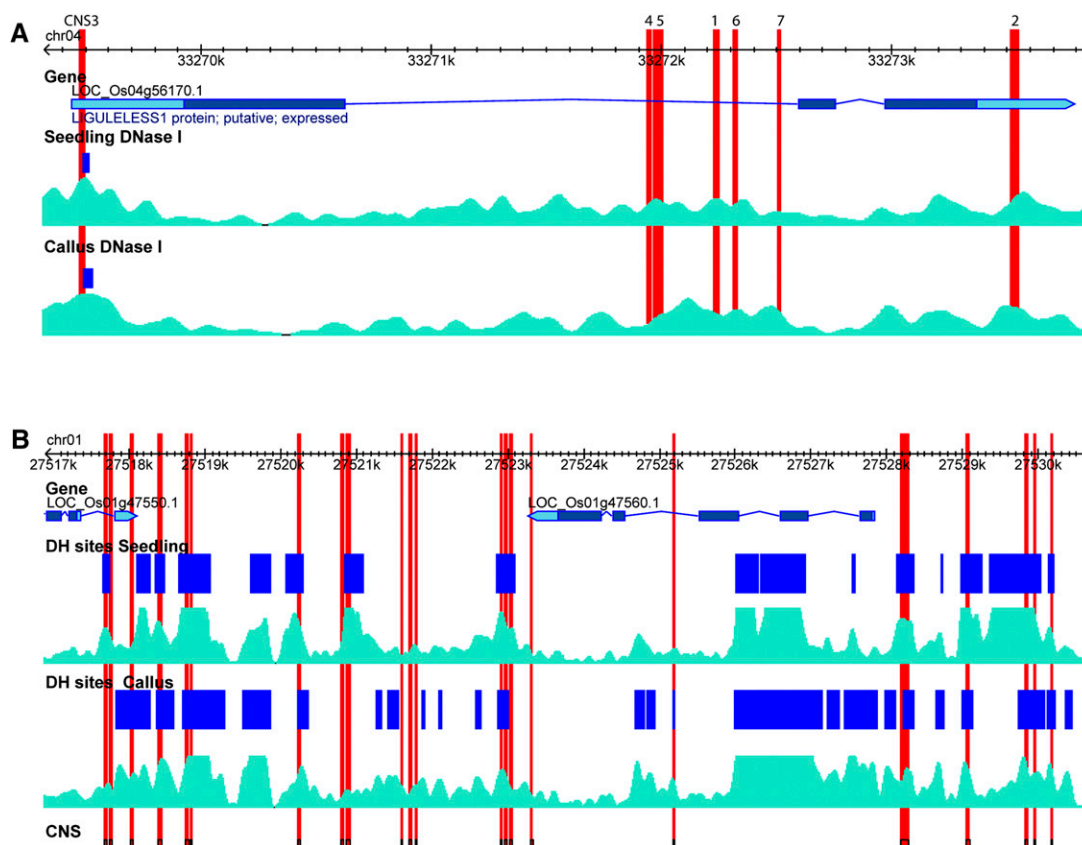
Many regulatory proteins bind to specific DNA motives. The PCF1 and PCF2 proteins in rice, which are essential for regulating meristematic tissue-specific expression of rice genes, bind two *cis* elements (element IIa, AGGTGGGCCCGT, and element IIb, TGTGGGACCATG) (Kosugi and Ohashi 1997). Introducing two mutated bases in element IIa (AGGTGGGCGAGT) resulted in loss of binding affinity to both proteins (Kosugi and Ohashi 1997). We identified a total of 110 regions that have 100% match to the two *cis* elements in the rice genome. Most of these regions were located outside of genes (82 of 110) or in introns (16 of 110). DH sites were associated with 43 (39%) of these regions (33 of 57 element IIa and 10 of 53 element IIb). Almost all of these 43 DH sites were located in outside of genes (37) or in introns (3). In contrast, we identified 22 regions with 100% match to the AGGTGGGCGAGT sequence. DH sites were associated with only two (9%) of these regions ( $P < 0.001$ , binomial test).

#### Distribution of DH sites in the rice genome

We plotted all DH sites along the pseudo-molecules of 12 rice chromosomes. The distribution patterns of the DH sites derived from leaf and callus tissues were highly similar (Fig. 3B; Supplemental Fig. S3). A trend of increased densities of DH sites was clearly visible toward the distal parts of most rice chromosomes. The lowest densities of the DH sites were observed surrounding the centromeres of several rice chromosomes (Fig. 3B; Supplemental Fig. S3). This distribution of the DH sites is parallel to the gene distribution but is opposite of the repeat distribution within rice chromosomes (Fig. 3A).

The distribution of heterochromatin and euchromatin along rice chromosomes 4 can be well visualized by 4',6-diamidino-2-phenylindole (DAPI) staining of the meiotic pachytene chromosome (Cheng et al. 2001). Almost the entire short arm and the proximal portion of the long arm of chromosome 4 can be defined as heterochromatin cytologically (Fig. 3C). The distribution of the DH sites on chromosome 4 is highly similar to the euchromatin distribution pattern (Fig. 3B,C). We found that the percentage of DH site-associated sequences among uniquely mappable sequences (without a duplicated copy in the rice genome) in euchromatin was significantly higher than in heterochromatin for chromosome 4 (Kolmogorov-Smirnov test,  $P < 2.2 \times 10^{-16}$  for both the seedling and callus data sets). Thus, the DH sites were significantly enriched for euchromatic regions in the rice genome.

We analyzed the locations of the DH sites relative to rice genes and found that 24%–27% of the DH sites mapped within 1 kb upstream of the transcription start sites (TSSs) at putative promoter



**Figure 2.** Association of DH sites with CNSs. (A) A DH site associated with a CNS in the *liguleless1* gene. Red bars indicated seven CNSs identified in the *liguleless1* gene. DH sites are indicated by blue blocks. (B) Association of DH sites with CNSs in a genomic region in rice chromosome 1.

regions for both tissues (Fig. 4). Approximately 30% of DH sites were associated with different gene partitions (exons, introns, 5' UTR, and 3' UTR), while the remaining 42%–45% of the DH sites located in the intergenic regions. Even though there were many more DH sites detected in the callus, the DH sites identified in the seedling and callus tissues showed an overall similar distribution pattern among different genomic regions (Fig. 4).

#### Highly expressed rice genes are frequently associated with DH sites

To exploit the relationship between the gene expression levels and DH sites, we conducted massively parallel pyrosequencing of mRNAs (RNA-seq) using the same seedling and callus tissues that were used for DNase-seq. Two biological replicates of each tissue were applied in RNA-seq. We obtained a total of 40.1 million reads from seedling and 29.6 million reads from callus. The RNA-seq data derived from the replicates showed high correlations (Spearman's rank correlation coefficient [SC]: 0.96 for seedling, 0.88 for callus). Thus, the RNA-seq data from the same tissue were combined for the calculation of gene expression value (see Methods). Our RNA-seq data derived from the seedling also showed a high correlation with the recently published rice RNA-seq data that were also from 2-wk-old seedling tissue (SC = 0.87) (Lu et al. 2010).

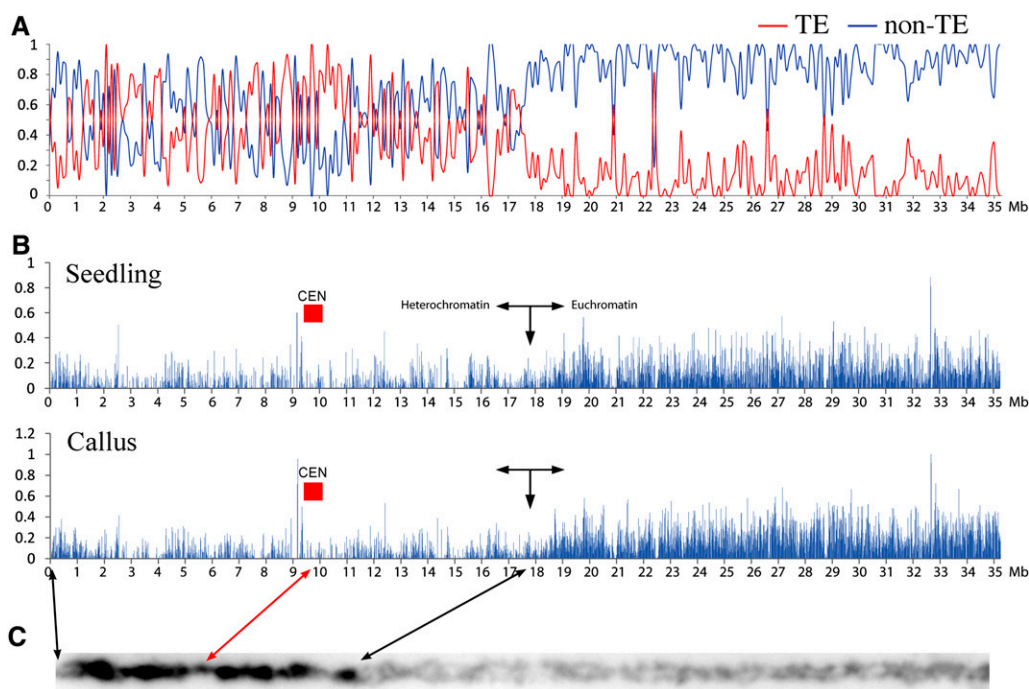
More than 93% of the highly expressed genes in the seedling (top 20%) (Fig. 5A) were associated with DH sites within the gene body as well as 200-bp upstream regions from the TSS. In contrast, only 23.0% and 16.7% of the nonexpressed genes (value = 0 FPKM

[fragments per kilobase of exon model per million mapped fragments]) showed a DH site within gene body and the 200-bp upstream region from TSS, respectively (Fig. 5A). Regions 500–1000 bp upstream of the TSS were much less frequently associated with DH sites than were regions 200 bp upstream of the TSS, and this trend was similar for all highly and low expressed genes (Fig. 5A). We observed same pattern in callus (Supplemental Fig. S4). These results suggest that the regulatory DNA elements associated with most rice genes are located within 200 bp from the TSS.

The expression levels of the rice genes were also related to the DNase I sensitivity levels (i.e., intensity of DNase-seq signal) of the DH sites in the 200-bp regions upstream of the TSS in both tissues (Fig. 5B; Supplemental Fig. S4). Genes with higher levels of expression displayed higher levels of DNase I sensitivity detected within this region. A similar trend was also observed in the 200-bp regions downstream from the genes (Fig. 5B; Supplemental Fig. S4), although the overall DNase I sensitivity level in this region is much lower.

#### Different numbers of DH sites associated with seedling and callus tissues

Comparison of the DH sites associated with the seedling and callus tissues revealed a significant difference in both the number and the total lengths of the DH site-related DNA sequences identified in the two tissues. We identified 155,025 DH sites in the callus tissue, which is 58% more than the 97,975 DH sites found in the seedling tissue. In addition, only 76,057 (49.1%) of the callus DH sites



**Figure 3.** Distribution of DH sites on rice chromosome 4. (A) Distribution of transposable element (TE)-related genes and non-TE genes along rice chromosome 4 (35 Mb). *y*-axis indicates the ratio of TE genes (red line) and non-TE genes (blue line) to total genes in 10 kb windows. (B) Distribution of DH sites on rice chromosome 4. *y*-axis indicates the coverage of the DH sites, which was calculated as the percentage of mappable regions covered by DH sites in 10-kb non-overlapped window bins. Red boxes (CEN) indicate the positions of the centromeres (Yan et al. 2008). (C) A digitally straightened pachytene chromosome 4. The chromosome was stained by DAPI and was converted into a black–white image to enhance the differentiation between heterochromatin (dark color) and euchromatin (light color). The positions of the centromere and the euchromatin-heterochromatin boundary on the long arm are marked by arrows.

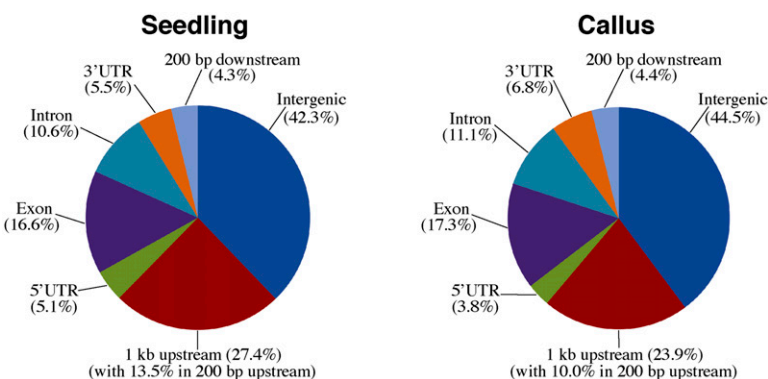
overlapped with the DH sites in seedling. The callus DH sites covered 31.0 Mb (7%) of the rice genome (420 Mb), while the seedling DH sites covered only 20.5 Mb (5%). Based on sequence read saturation simulation (Fig. 1B), we expect that the substantial increase of DH sites observed in the callus was not caused by the different levels of sequence read saturation. We found that the vast majority of tissue-specific DH sites were not a result of threshold artifacts (Supplemental Fig. S5), which is supported by both cumulative plot analysis (Supplemental Fig. S6) and visual inspection (Fig. 1A).

We examined the levels of DNase I sensitivity of a total of 63,838 DH sites that were detected in both tissues. Interestingly, 19,628 (30.7%) of these DH sites showed a significant difference of the sensitivity level between the seedling DH site and the callus DH site ( $FDR < 0.01$ ), suggesting that many DH sites that are present in both tissues show significant variation in sensitivity to DNase I (Supplemental Fig. S7).

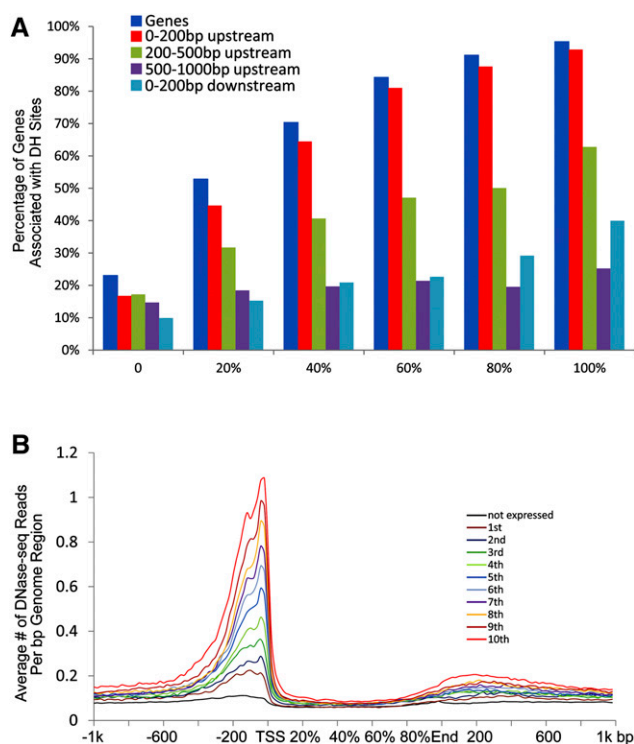
#### High frequency of DH sites associated with genes that are differentially expressed in seedling and callus

To exploit a potential relationship between change of gene expression and change of DNase I sensitivity, we first identified genes differentially expressed in the seedling and callus tissues from our

RNA-seq data set. Differentially expressed genes were also identified based on the Agilent microarray data (82,000 probes) with dye-swap (Satoh et al. 2007). Fold change of gene expression (seedling/callus) showed a moderate correlation between the two data sets ( $R = 0.64$ , PCC). A gene was considered differentially expressed only if it showed a significant expression difference in the two tissues in both experiments. The same criterion was used for identification of nondifferentially expressed genes (see Methods). We identified a total of 16,544 nondifferentially expressed genes and 7000 differentially expressed genes, of which 2813 and 4187 genes were significantly up-regulated in the seedling and callus, respectively.



**Figure 4.** Distribution of DH sites in the rice genome. (1 kb upstream) 1-kb region upstream of TSS. (200 bp downstream) 200-bp region downstream from the end of gene transcription.



**Figure 5.** The relationship between DNase I sensitivity and the levels of gene expression in seedling tissue. (A) The percentage of genes with different expression levels (based on RNA-seq data, FPKM) associated with DH sites in different locations (from 1-kb upstream regions of genes to 200-bp downstream regions of genes). Genes were divided into six bins based on their expression levels, including no expression (FPKM = 0), and from bottom 20% to top 20% (indicated as 100% in figure). (B) The profile of DNase I sensitivity (indicated by the number of DNase-seq reads) among the genes with different expression levels. The expressed genes were divided into 10 bins from low expression (first) to high expression (10th) based on the expression levels.

We sorted the genes based on fold change of expression (seedling/callus) and analyzed the DNase I sensitivity levels within the 1000-bp regions both upstream of and downstream from the TSS. Genes up-regulated in the seedling and callus were associated with high DNase I sensitivity within ~200 bp upstream of and 100 bp downstream from the TSS (Fig. 6A). Interestingly, the genes down-regulated in either of the two tissues were also associated with high DNase I sensitivity in the same regions, indicating that many of these down-regulated genes are repressed by regulatory factors in the respective tissue.

We found that 79.6% of the genes up-regulated in the seedling were associated with a DH site that maps within 200 bp surrounding the TSS in the seedling. Likewise, 77.7% of the same sets of genes were also associated with a DH site surrounding the TSS in callus. Similarly, 89.1% and 71.7% of the genes up-regulated in callus were associated with DH sites that map within 200 bp surrounding the TSS in the callus and seedling, respectively. These numbers were significantly higher than those associated with nondifferentially expressed genes ( $P < 0.001$ , binomial test): Only 44.1% and 52.9% of the nondifferentially expressed genes were associated with DH sites that map within 200 bp surrounding the TSS in the seedling and callus, respectively.

It is possible that the nondifferentially expressed gene set was relatively enriched with weakly expressed genes and/or if the dif-

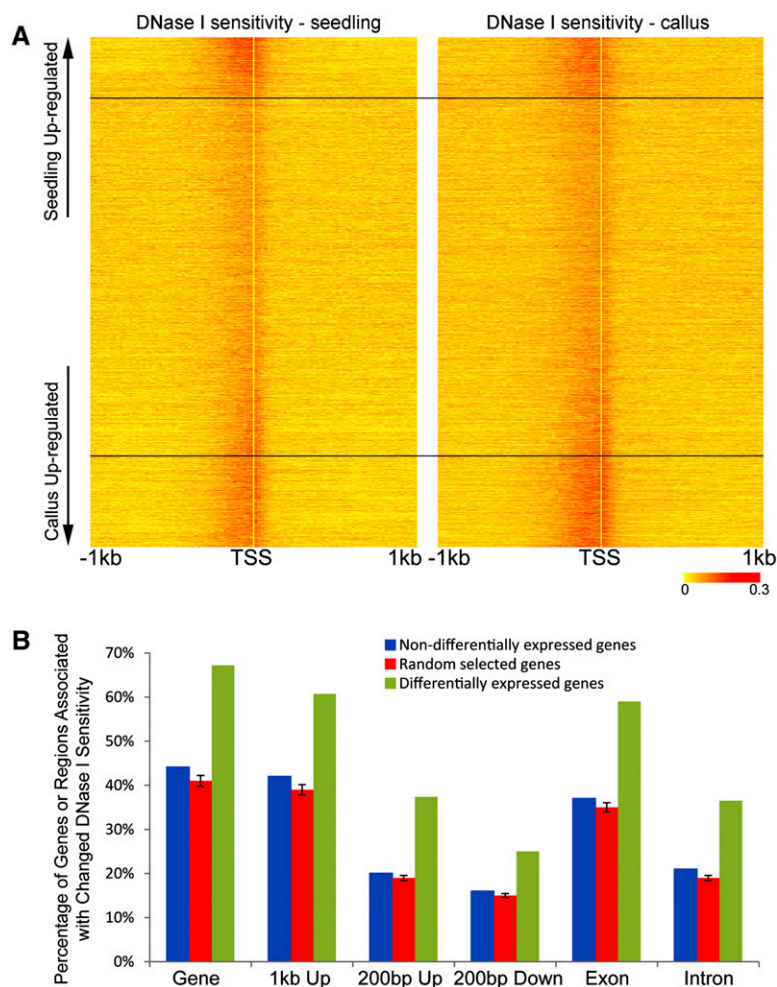
ferentially expressed gene set was relatively enriched with highly expressed genes. We removed weakly expressed genes or both weakly and highly expressed genes in the data sets. Similar conclusions were obtained from analyses of the new data sets after filtration (Supplemental Table S3). We divided the genes into 14 bins based on their expression value and then examined the percentage of genes associated with a DH site in the 200 bp surrounding the TSS. Differentially expressed genes overall showed a higher chance to be associated with a DH site than nondifferentially expressed genes across all bins (Supplemental Fig. S8). Particularly, significantly higher percentages of differentially expressed genes contained DH sites than did nondifferentially expressed genes within bins containing low and intermediate levels of expression (Supplemental Fig. S8).

We also examined the changes of DNase I sensitivities in the two different tissues. We found that 67.2% and 60.7% of the differentially expressed genes were associated with changes of DNase I sensitivity in gene body and 1-kb upstream region, respectively, which were significantly more than the genes with nondifferentially expression levels in the two tissues or randomly selected genes ( $P < 0.001$ , binomial test) (Fig. 6B). The similar pattern of association was also found in gene exons, introns, 200-bp upstream regions, and 200-bp downstream regions (Fig. 6B).

#### Histone modifications associated with rice DH sites

The DH sites in the human genome were associated with specific patterns of histone modifications (The ENCODE Project Consortium 2007; Boyle et al. 2008b). To examine histone modifications associated with the rice DH sites, we performed chromatin immunoprecipitation followed by Illumina sequencing (ChIP-seq) to identify the positions of three histone modifications (H3K4me2, H3K36me3, and H4K12ac) in the rice genome and integrated our data with the recently published H3K4me3, H3K9ac, and H3K27me3 ChIP-seq data from the 4-wk-old rice seedling (He et al. 2010). We divided the rice seedling DH sites into four categories based on their locations relative to genes: (1) DH sites associated with promoters (1–1000 bp upstream of the TSS), (2) DH sites downstream from genes (200 bp downstream from transcription end), (3) DH sites within genes, and (4) DH sites in intergenic regions (all remaining DH sites).

The six histone modifications, except H3K27me3, are associated with active gene expression (Kouzarides 2007). The seedling DH sites associated with promoters, genes, and 200 bp downstream from the genes showed an overall decrease in histone modifications compared with the immediate adjacent regions, consistent with DH sites representing nucleosome depleted/dynamic regions (Fig. 7A–C). Thus, the overall histone modification patterns associated with the regions around these DH sites are consistent with the patterns of the same modifications associated with gene-surrounding regions in both plant and human genomes (Barski et al. 2007; Wang et al. 2009). The DH sites within the intergenic regions showed similar decreases of H3K4me3 and histone acetylation (both H4K12ac and H3K9ac), but a distinct increase of H3K27me3 (Fig. 7D). We obtained the same results from analysis of a set of 48,935 DH sites that are shared within all three biological replicates of seedling tissue (Supplemental Fig. S9). Around all DH sites, we detect significant phasing of nucleosomes, similar to what was previously shown in humans around TSSs (Schones et al. 2008) and CTCF binding sites (Fu et al. 2008).



**Figure 6.** High DNase I sensitivity of the DH sites associated rice genes that differentially expressed in the seedling and callus. (A) Heatmap of DNase I sensitivity surrounding the TSSs of genes. Genes were sorted out based on fold change of gene expression of seedling relative to callus. The genes *above/under* the black lines were significantly up-regulated in the seedling and callus, respectively. DNase I sensitivity was indicated by DNase-seq read number per bp genome per million reads. (B) Percentage of differentially expressed genes associated with DNase I sensitivity changes between the seedling and callus in their genic region, 1 000-bp and 200-bp upstream region of TSS, 200 bp downstream from the gene, exon, and intron. We used similar expressed genes and genes randomly selected (10 times) as a control, with the same sample size of differentially expressed genes.

We then examined the histone modifications (data all from seedling tissue) that are associated with the tissue-specific DH sites. The seedling-specific DH sites showed similar histone modification patterns to those of all seedling DH sites (Fig. 7E–H). Surprisingly, the callus-specific DH sites within the intergenic regions showed a clearly increased level of histone modification compared with the flanking regions, including H3K27me3 (Fig. 7I). We generated a control data set by randomly selecting 78,968 rice genomic regions that have similar lengths and sequence uniqueness as the 78,968 intergenic callus-specific DH sites. The histone modifications associated with these randomly selected regions were not different from the flanking region (Supplemental Fig. S10). These results suggest that these regions are well occupied by nucleosomes in the seedling tissues. Genome-wide histone modification mapping using callus tissue will reveal if these genomic regions are associated with a decreased level of nucleosome occupation after becoming DH sites.

## DNA methylation associated with rice DH sites

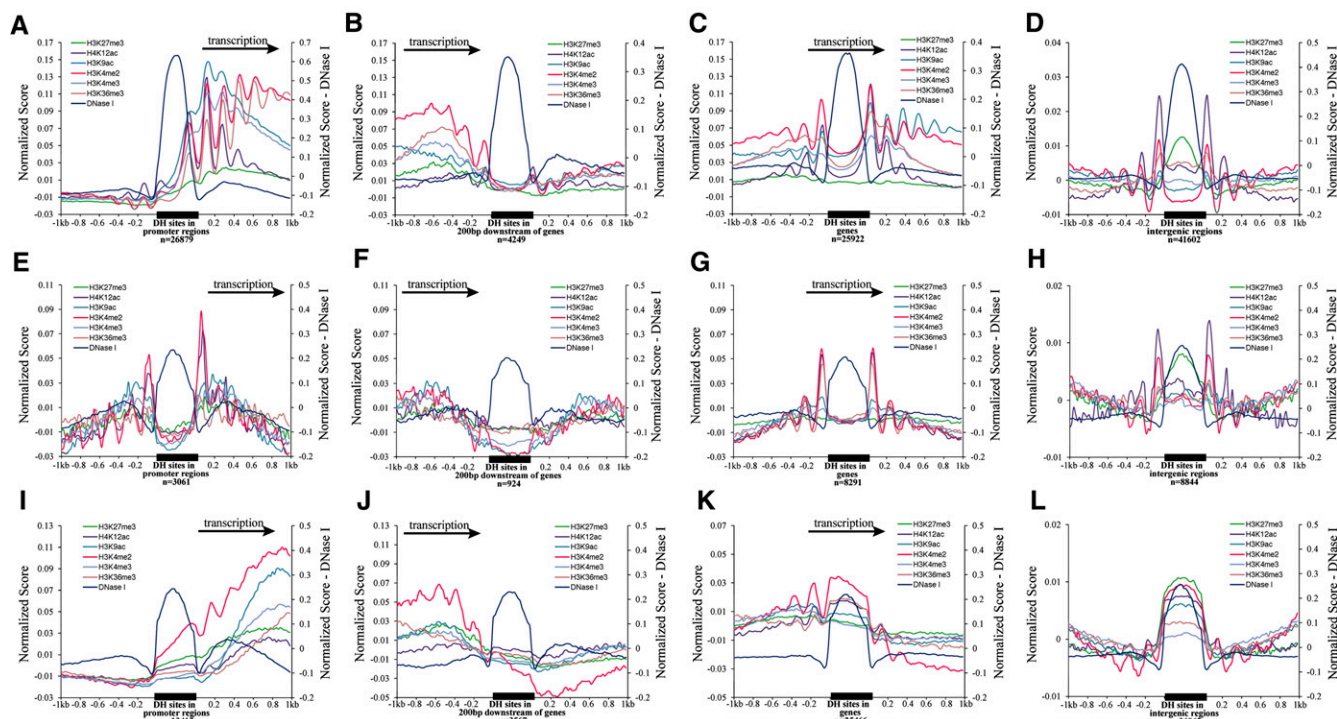
Genome-wide cytosine methylation data in single base pair resolution has recently been generated from rice seedling tissue by sequencing of bisulfite-converted genomic DNA (Zemach et al. 2010). We integrated this DNA methylation data with our seedling DH site data set. Strikingly, the DNA methylation levels of all DH sites were drastically lower than the flanking DNA sequences (Fig. 8A). A similar result was obtained from analysis of the DH sites shared within all three biological replicates of seedling tissue (Supplemental Fig. S9). In comparison, randomly selected rice genomic regions showed a similar methylation level as the flanking sequences (Supplemental Fig. S10). This result is correlated with the fact that the promoter regions of most rice genes are hypomethylated (Yan et al. 2010; Zemach et al. 2010) and with the recent discovery that DNA methyltransferases preferentially target nucleosome-bound DNA (Chodavarapu et al. 2010).

We further analyzed the methylation levels of the tissue-specific DH sites. Both the seedling-specific and callus-specific DH sites located within the genes and the 200 bp downstream regions of the genes showed a similar hypomethylation pattern as all the seedling DH sites (data not shown). Surprisingly, the tissue-specific DH sites located in the promoters showed significantly higher levels of methylation than the average methylation level of all the seedling DH sites located within promoters (Kolmogorov-Smirnov test,  $P < 2.2 \times 10^{-16}$  for both the seedling-specific DH sites and callus-specific DH sites compared with all the seedling DH sites within promoters) (Fig. 8B).

## Discussion

### Differences of distribution of DH sites in human and rice genomes

We describe the first genome-wide maps of the DH sites of a plant species from two different tissues. We found 97,975 DH sites in the rice seedling, which is strikingly similar to the 94,925 DH sites found in human CD4<sup>+</sup> T cells (Boyle et al. 2008a). There are a number of other similar findings between the rice and human DH sites. For example, highly expressed genes in both the human and rice genomes are more frequently associated with DH sites than weakly expressed genes (Boyle et al. 2008a). Similarly, ~39% and 42% of the DH sites are located in the intergenic regions in the human and rice genomes, respectively. However, the DH sites associated with the genes show different distribution patterns in the two genomes. Approximately twice the number of rice DH sites



**Figure 7.** Histone modifications associated with all seedling DH sites (A–D), seedling-specific DH sites (E–H), and callus-specific DH sites (I–L). Six histone modifications associated with DH sites located in promoters (A, E, I), in 200 bp downstream of the genes (B, F, J), within genes (C, G, K), and in intergenic regions (D, H, L). All start (close to 5' of gene) and end sites (close to 3' of gene) of genic DH sites (A–C, E–G) were aligned together. Note that the direction of intergenic DH sites was from short arm toward long arm of the chromosomes. The direction of genic DH sites is the same as the direction of gene transcription. Normalized score indicated the modification level and was calculated by reads number per bp genomic region per million reads with Z-score transformation.

were associated exons, and only one-quarter of the rice DH sites are located in introns compared with human DH sites. These differences are likely caused by the different structures of the rice and human genes. The average intron size of human genes is 6155 bp, which is 15-fold longer than the average intron size of rice genes (398 bp). Intron sequences account for an average of 38% of the rice genes compared with an average of 76% of the human genes (Supplemental Table S4). Thus, a significantly higher proportion of regulatory elements reside in introns in humans than in rice.

Most noticeably, ~13% of the human DH sites are located within 2 kb upstream of the genes. In contrast, 27% of rice DH sites are within 1 kb upstream of the rice genes in seedling tissue. This indicates that a higher percentage of promoters may be active or poised in the rice seedling compared with a single human cell type. A possible explanation is that rice seedling is a heterogeneous tissue or is possibly due to differences in TSS mapping accuracy within the human and rice genomes. In the future, similar DH maps from more homogenous plant cell types, as well as other plant species, will be useful to draw additional comparisons between humans and plants.

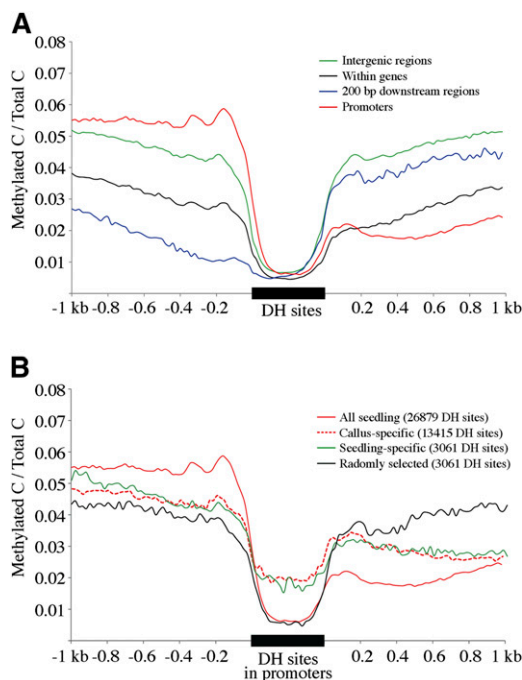
### Tissue-specific changes of DNase I sensitivity

The higher number of DH sites detected in the callus compared with seedling tissue was unexpected but is not incompatible with the literature. McClintock (1984) listed tissue culture as one of few major classes of stress that can induce significant responses of plant genomes to the challenge (McClintock 1984). It has been extensively documented that tissue culture can induce “somaclonal

variation” in plants (Lee and Phillips 1988). Some of the variations are likely caused by tissue culture–induced change of DNA methylation (Kaeppler and Phillips 1993) and transposon activation (Hirochika 1993), which may also impact the changes of DH sites.

In addition to tissue culture increasing the number of DH sites compared with the seedling, it also alters the overall sensitivity to DNase I of a significant proportion of the DH sites. These results demonstrate that the numbers, locations, and levels of DNase I sensitivity of DH sites can vary significantly among different tissues or organs and may be induced by biotic and abiotic stresses. Thus, development of regulatory maps associated with different tissue/organs or under different stresses in model organisms will be an essential future effort for the comprehensive understanding of transcription and regulation of gene expression.

Rice genes differentially expressed in seedling and callus tissues are more likely associated with a DH site than those equally expressed in both tissues (Fig. 6A). In addition, the DH sites associated with such differentially expressed genes also show higher sensitivity than those associated with genes equally expressed in both tissues (Fig. 6A) and also tend to change the DNase I sensitivity in different tissues (Fig. 6B). These results are relevant to the fact that genes differentially expressed in different tissues may be associated with different regulators depending on whether the genes are under an active or repressive status. The enrichment of the DH sites in regions surrounding the TSS of a repressed gene may be related to poised status of transcription, such genes are generally related to development control and have been well characterized in *D. melanogaster* (Muse et al. 2007; Zeitlinger et al.



**Figure 8.** DNA methylation associated with rice DH sites. We plotted the ratio of methylated C to total C sites in DH sites and in 1000-bp regions flanking the DH sites. The ratio of methylated C was adjusted by the length. (A) DNA methylation associated with all rice DH sites associated with seedling tissue. (B) Seedling-specific and callus-specific DH sites within promoters showed higher levels of methylation than the average of all seedling DH sites within promoters. Randomly selected 3061 DH sites within promoters were used as a control.

2007). Binding of regulatory proteins to repressed genes has also been demonstrated in plants (De Lucia et al. 2008).

While preliminary, our finding that genes differentially expressed in seedling and callus were frequently associated with DH sites in both tissues is consistent with the hypothesis that genes under more complex patterns of regulation will require larger more regulatory elements—and therefore larger regions of open chromatin—than “housekeeping” genes that maintain a constant level of expression in all tissues under all conditions. Consistent with this hypothesis, we found that the distribution of DH sites is correlated with that of CNSs identified between rice and sorghum. In *Arabidopsis thaliana*, genes associated with greater numbers of noncoding sequences are enriched in genes expressed in response to environmental stimuli (Freeling et al. 2007). Genes that exhibit complex patterns of expression also tend to be those with larger total promoters (Sun et al. 2010).

### Epigenetic features associated with DH sites

A compelling question is what causes differential DNase I sensitivity of the same genomic loci in two different tissues. The epigenetic modification patterns of these genomic regions provide some clues for this question. It has been recently demonstrated in *A. thaliana* that nucleosomal positioning and deposition of histone H2A.Z can influence DNA methylation pattern (Zilberman et al. 2008; Chodavarapu et al. 2010). DH sites represent genomic regions that are nucleosome free or under dynamic nucleosomal modifications or displacements. Thus, DH sites can be potentially modified for their DNA methylation and histone modification/

replacement. We observed that the DNA associated with all the seedling DH sites are generally hypomethylated (Fig. 8A), consistent with a recent report in humans that chromosomal regions bound to specific DNA-binding proteins are generally hypomethylated (Lister et al. 2009). However, tissue-specific DH sites located in the promoters showed higher levels of DNA methylation than the average methylation level associated with all seedling DH sites located in promoters (Fig. 8B). Interestingly, genes methylated in their promoters in *A. thaliana* tended to be expressed in a tissue-specific manner (Zhang et al. 2006). Thus, some tissue-specific DH sites may represent partially poised DH sites and may undergo a demethylation process in different tissue/organs to become fully activated. This hypothesis is supported by a recent study showing that rice DNA isolated from suspension-cultured cells is less methylated than DNA isolated from rice seedling (Li et al. 2008b). A link between tissue-specific chromatin accessibility, DNA methylation, and transcription factor binding has recently been demonstrated in mammalian species (Wiench et al. 2011).

The intergenic DH sites are distinctly associated with elevated levels of H3K27me3 (Fig. 7D,H). H3K27me3 is controlled by the Polycomb repressive complex and plays a key role in the development of both animals (Tolhuis et al. 2006) and plants (Zhang et al. 2007; Bouyer et al. 2011). H3K27me3 is involved in regulation (silencing) of a large number of genes in *A. thaliana* (Zhang et al. 2007). Importantly, most H3K27me3-targeted genes are expressed in a tissue-specific manner (Zhang et al. 2007; Lafos et al. 2011). Since Polycomb-mediated repression can be easily reversed, H3K27me3 may facilitate the repression of these genes in appropriate tissues (Zhang et al. 2007). We propose that H3K27me3 may play a similar role in repression/activation of DH sites in different tissues. Interestingly, H3K27me3 was also enriched within the intergenic DH sites in humans (Boyle et al. 2008b).

## Methods

### Development of DNase-seq and ChIP-seq libraries

Mixed leaf and stem tissues of 2-wk-old rice cultivar “Nipponbare” seedlings grown in a greenhouse were collected and ground into a fine powder in liquid nitrogen. The resulting powder was suspended in nuclear isolation buffer (NIB; 20 mM Tris-HCl, 50 mM EDTA, 5 mM Spermidine, 0.15 mM Spermine, 0.1% mercaptoethanol, 40% Glycerol at pH 7.5) and followed the standard protocol for nuclei isolation. Rice callus tissue was induced from sterilized “Nipponbare” seeds in rice calli induction medium (NB basal medium plus vitamin, glutamine, proline, casein hydrolysate, sucrose, and phytoigel as well as 3 mg/L 2,4-D at pH 5.8) under 28°C–29°C with dark conditions. Three-week-old calli were collected for nuclei isolation using the same method as for leaf tissue. The prepared nuclei pellet was suspended in RSB buffer (10 mM Tris at pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>) for DNase I (Roche) digestion with increasing concentrations (0–4 units) for 10 min at 37°C.

DNase-seq library construction followed published protocols with only minor modifications (Boyle et al. 2008b). Specifically, the degree of DNase I digestion was assessed by pulsed-field gel electrophoresis (PFGE; 20–60 switch time, 18 h, 6 V/cm; Bio-Rad). High-molecular-weight (HMW) DNA after DNase I digestion was isolated and blunt ended with T4 DNA polymerase. Biotinylated adaptor I (5′ Bio ACAGGTTTCAGAGTCTACAGTCCGAC and 5′ P- GTCG GACTGTAGAAGCTGAAC) was ligated to the DNA molecules. Dynal M-280 beads (Invitrogen) were used for enriching DNase I-digested DNA ends after MmeI digestion.

Adaptor II (5' P-TCGTATGCCGCTTCTGCTTG and 5'-CAAGCA GAAGACGGCATAACGANN) was then ligated to the MmeI-treated ends. The DNA sample was amplified by PCR using linker-specific primers (5'-CAAGCAGAAGACGG CATACGA and 5'-AATGATA CGGCGACCACCGACAGGTTCTACAGTTCTACAGTCCGA) and purified by PAGE for isolation of DNA fragments with ~90 bp in size. The final Illumina sequencing was performed using a primer specific to linker I (5'-CCACCGACAGGTTCTACAGTTCTACAGTCCGAC).

ChIP was performed followed published protocols (Nagaki et al. 2003). Commercial antibodies against H3K4me2 (07-030, Upstate Biotechnology), H3K36me3 (ab9050, Abcam), and H4K12ac (ab1762, Abcam) were used in ChIP experiments. Construction of ChIP-seq libraries was according to the protocol of "preparing samples for ChIP sequencing of DNA" provided by Illumina. Briefly, ChIPed DNA was end repaired using an End-It DNA End Repair Kit (Epicentre, ER0720). The "dA" base was then added to the 3' ends of the blunt-ended DNA fragments using Klenow fragment (3'-5' exo-; NEB, M0212S). Linker for single read was ligated to the DNA molecules with dA overhang using LigaFast kit (Promega, M8221). After 2% agarose gel running with TAE buffer, the gel region containing ~150–300 bp DNA was excised for adapter-ligated DNA size selection. The adapter-modified DNA fragments were enriched by PCR and purified by running 2% agarose gel in TAE buffer for isolation of DNA fragments in the range of 200–300 bp. The purified DNA sample was ready for Illumina sequencing after quality validation. We downloaded the ChIP-seq data of H3K4me3, H3K9ac, and H3K27me3 (He et al. 2010), which was derived from ChIP using mixed leaf and stem tissues from 4-wk-old rice seedling.

### RNA-seq

Rice seedlings and callus at the same growth stage and under the same growing condition as those for DNase-seq experiments were used for RNA-seq analysis. Two biological replicates from both the seedling and callus were prepared for RNA-seq. Total RNA was extracted using a RNeasy Plant Kit (Qiagen catalog no. 74904). High-quality (RNA integrity number >8) was used for downstream DNase I treatment to remove genomic DNA. Approximately 10 µg total RNA was converted to cDNA using the mRNA-seq kit from Illumina and barcoded cDNA library was sequenced on an Illumina Genome Analyzer 2 platform. We used TopHat software (Trapnell et al. 2009) to align the sequence reads to reference genome and used Cufflink (Roberts et al. 2011) to call the expression values (FPKM) of annotated rice genes and to detect differentially expressed genes. We obtained 40.1 M of RNA-seq reads from seedlings, and 89.2% of them were mapped to the rice genome (Tigr release 5). Similarly, we obtained 29.6 million reads from calli, and 89.1% of them were mapped to rice genome. We also downloaded recently published rice RNA-Seq data that were also derived from 2-wk-old rice seedlings (Lu et al. 2010) and RNA-seq data from 4-wk seedling (He et al. 2010) for a comparative analysis with our RNA-seq data sets using same method. All rice genome and annotation data came from TIGR database (release 5). To accurately identify the genes differentially expressed in callus and seedling, we used the limma package of bioconductor (<http://www.bioconductor.org/>) to analyze the published microarray data in NCBI (Satoh et al. 2007). We controlled the FDR of differentially expressed genes to under 0.05 for both microarray and RNA-seq data. Only genes that show a significant expression difference in the seedling and callus tissues in both experiments were considered as differentially expressed genes. Similarly, only genes that do not show significant expression difference in the seedling and callus tissues in both experiments were considered as nondifferentially expressed gene. The programs of data process and statistical test were written in Perl or R (<http://www.r-project.org/>).

### Data analysis

The reads from DNase-seq or ChIP-seq were aligned to rice genome (TIGR release 5) with a 1-bp mismatch allowed using MAQ program (Li et al. 2008a). Only the reads mapped to a unique position of rice genome were used for further analysis. We used F-seq (Boyle et al. 2008a) to identify DH sites with 200-bp bandwidth. To detect the FDR of identified DH sites, we randomly generated 10 data sets, the amount of reads of which is the same as the data sets from DNase-seq. The FDR was calculated as ratio of DH sites identified from random data sets to DH sites from DNase-seq. We set a cutoff in F-seq to control the FDR <0.05. To investigate the saturation of DNase-seq data, we simulated the correlation between sequencing depth (the number of reads) and the total length of all DH sites identified. We randomly sampled the reads from our data set. For the callus, we sampled ~2%, ~4%, ~6%...98%, 100% reads from callus DNase-seq data set, consisting of 50 data points in Figure 1B. For seedling, we sampled ~3%, ~6%, ~9%...96%, ~99% reads from seedling DNase-seq data set, consisting of 30 data points in Figure 1B. We identified DH sites based on these randomly sampled data and examined the correlation between the read number and the total length of DH sites for each data point. For each data point, we sampled 10 times and averaged the total length of DH sites. We fitted the data into logarithmic growth model for the regression analysis. We used DESeq (Anders and Huber 2010) to identify the change of DNase I sensitivity in DH sites shared between callus and seedling (FDR = 0.01). The variance used in DESeq was estimated based on three replicates of DNase-seq in seedling and two lanes of data from replicate 1 of the callus (Supplemental Table S1).

To detect the density of DH sites on an entire chromosome, we first identified all the uniquely mappable regions by generating 20-bp reads starting on every base and examining uniqueness of each read. We then divided each chromosome into 10-kb non-overlapping windows and calculated the ratio of uniquely mappable regions covered by DH sites in each window. To compare the density of DH sites between euchromatin and heterochromatin, we used Kolmogorov-Smirnov test to check the difference of the ratio of two data sets: the ratios of unique region covered by DH sites in euchromatin and the ratios in heterochromatin. To draw the histone modification distribution from ChIP-seq data, an offset (half of the estimated length of ChIP fragment) toward the downstream from reads was set in each data set as the modification signal (75 for H3K4me2, H3K4me3, 97 for H3K9ac, 67 for H3K27me3, 65 for H3K36me3, 77 for H4K12ac).

All rice genome and annotation data came from TIGR database (release 5). The expression data of rice seedling, which were based on Affymetrix microarrays, were downloaded from <http://www.ricearray.org/>. We remapped the probe sets to the genes defined by TIGR release 5. We only kept the probe sets of which all probes were only mapped into single transcribed region. To identify the genes differentially expressed in callus and seedling, we used limma package of bioconductor (<http://www.bioconductor.org/>) to analyze the published microarray data in NCBI (Satoh et al. 2007). We controlled FDR of differentially expressed genes under 0.01. The programs of data process and statistical test were written in Perl or R (<http://www.r-project.org/>).

### Data access

The rice DNase I hypersensitive site and ChIP-seq data sets have been submitted to the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under accession nos. GSE26610 and GSE26734. The data can also be viewed by a Generic Genome Browser: [https://mywebspace.wisc.edu/groups/jiang/Web/Jiming\\_Jiang\\_Lab/Resource.html](https://mywebspace.wisc.edu/groups/jiang/Web/Jiming_Jiang_Lab/Resource.html).

## Acknowledgments

We thank the Dale Bumpers National Rice Research Center for providing the Nipponbare seeds, Dr. Zhukuan Cheng for the rice pachytene chromosome 4 image in Figure 2, and Dr. Robin Buell for assistance with the rice RNA-seq experiments. This research was supported by grants DBI-0603927 and DBI-0923640 from the National Science Foundation.

## References

- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106. doi: 10.1186/gb-2010-11-10-r106.
- Barski A, Cuddapah S, Cui KR, Roh TY, Schones DE, Wang ZB, Wei G, Chepelev I, Zhao KJ. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–837.
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. 2010. The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* **28**: 1045–1048.
- Bouyer D, Roudier F, Heese M, Andersen ED, Gey D, Nowack MK, Goodrich J, Renou JP, Grini PE, Colot V, et al. 2011. Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet* **7**: e1002014. doi: 10.1371/journal.pgen.1002014.
- Boyle AP, Guinney J, Crawford GE, Furey TS. 2008a. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics* **24**: 2537–2538.
- Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. 2008b. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**: 311–322.
- Boyle AP, Song LY, Lee B-K, London D, Keefe D, Birney E, Iyer VR, Crawford GE, Furey TS. 2011. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cell. *Genome Res* **21**: 456–464.
- Cheng ZK, Buell CR, Wing RA, Gu M, Jiang JM. 2001. Toward a cytological characterization of the rice genome. *Genome Res* **11**: 2133–2141.
- Chodavarapu RK, Feng SH, Bernatavichute YV, Chen PY, Stroud H, Yu YC, Hetzel JA, Kuo F, Kim J, Cokus SJ, et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**: 388–392.
- Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, Green R, Meltzer PS, Wolfsberg TG, Collins FS. 2006a. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods* **3**: 503–509.
- Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen YD, Bernat JA, Ginsburg D, et al. 2006b. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res* **16**: 123–131.
- De Lucia F, Crevillen P, Jones AME, Greb T, Dean C. 2008. A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of *FLC* during vernalization. *Proc Natl Acad Sci* **105**: 16831–16836.
- The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799–816.
- Feng SH, Cokus SJ, Zhang XY, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, et al. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci* **107**: 8689–8694.
- Freeling M, Subramaniam S. 2009. Conserved noncoding sequences (CNSs) in higher plants. *Curr Opin Plant Biol* **12**: 126–132.
- Freeling M, Rapaka L, Lyons E, Pedersen B, Thomas BC. 2007. G-boxes, bigfoot genes, and environmental response: characterization of intragenomic conserved noncoding sequences in *Arabidopsis*. *Plant Cell* **19**: 1441–1457.
- Fu YT, Sinha M, Peterson CL, Weng ZP. 2008. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet* **4**: 13. doi: 10.1371/journal.pgen.1000138.
- Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**: 1775–1787.
- Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92–100.
- Gross DS, Garrard WT. 1988. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* **57**: 159–197.
- Haeussler M, Joly JS. 2011. When needles look like hay: How to find tissue-specific enhancers in model organism genomes. *Dev Biol* **350**: 239–254.
- He GM, Zhu XP, Elling AA, Chen LB, Wang XF, Guo L, Liang MZ, He H, Zhang HY, Chen FF, et al. 2010. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* **22**: 17–33.
- Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. 2009. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* **19**: 460–469.
- Hesselberth JR, Chen XY, Zhang ZH, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, et al. 2009. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* **6**: 283–289.
- Hirochika H. 1993. Activation of tobacco retrotransposons during tissue culture. *EMBO J* **12**: 2521–2528.
- Jin CY, Zang CZ, Wei G, Cui KR, Peng WQ, Zhao KJ, Felsenfeld G. 2009. H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. *Nat Genet* **41**: 941–945.
- John S, Sabo PJ, Thurman RE, Sung M-H, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**: 264–268.
- Kaeppler SM, Phillips RL. 1993. Tissue culture-induced DNA methylation variation in maize. *Proc Natl Acad Sci* **90**: 8773–8776.
- Kaplinsky NJ, Braun DM, Penterman J, Goff SA, Freeling M. 2002. Utility and distribution of conserved noncoding sequences in the grasses. *Proc Natl Acad Sci* **99**: 6147–6151.
- Keene MA, Corces V, Lowenhaupt K, Elgin SCR. 1981. DNase I hypersensitive sites in *Drosophila* chromatin occur at the 5' ends of regions of transcription. *Proc Natl Acad Sci* **78**: 143–146.
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu TT, et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**: 480–485.
- Kodama Y, Nagaya S, Shinmyo A, Kato K. 2007. Mapping and characterization of DNase I hypersensitive sites in *Arabidopsis* chromatin. *Plant Cell Physiol* **48**: 459–470.
- Kosugi S, Ohashi Y. 1997. PCF1 and PCF2 specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* **9**: 1607–1619.
- Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**: 693–705.
- Lafos M, Kroll P, Hohenstatt ML, Thorpe FL, Clarenz O, Schubert D. 2011. Dynamic regulation of H3K27 trimethylation during *Arabidopsis* differentiation. *PLoS Genet* **7**: e1002040. doi: 10.1371/journal.pgen.1002040.
- Lee M, Phillips RL. 1988. The chromosomal basis of somaclonal variation. *Annu Rev Plant Physiol Plant Mol Biol* **39**: 413–437.
- Li H, Ruan J, Durbin R. 2008a. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**: 1851–1858.
- Li XY, Wang XF, He K, Ma YQ, Su N, He H, Stolc V, Tongprasit W, Jin WW, Jiang JM, et al. 2008b. 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.
- Li XY, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, Biggin MD. 2011. The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. *Genome Biol* **12**: 17. doi: 10.1186/gb-2011-12-4-r34.
- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**: 315–322.
- Lu TT, Lu GJ, Fan DL, Zhu CR, Li W, Zhao QA, Feng Q, Zhao Y, Guo YL, Li WJ, et al. 2010. Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res* **20**: 1238–1249.
- Matsumoto T, Wu JZ, Kanamori H, Katayose Y, Fujisawa M, Namiki N, Mizuno H, Yamamoto K, Antonio BA, Baba T, et al. 2005. The map-based sequence of the rice genome. *Nature* **436**: 793–800.
- McArthur M, Gerum S, Stamatoyannopoulos G. 2001. Quantification of DNase-sensitivity by real-time PCR: quantitative analysis of DNase-hypersensitivity of the mouse  $\beta$ -globin LCR. *J Mol Biol* **313**: 27–34.
- McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792–801.
- McGhee JD, Wood WI, Dolan M, Engel JD, Felsenfeld G. 1981. A 200 base pair region at the 5' end of the chicken adult  $\beta$ -globin gene is accessible to nuclease digestion. *Cell* **27**: 45–55.
- Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet* **39**: 1507–1511.
- Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang JM. 2003. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics* **163**: 1221–1225.

- Nedospasov SA, Georgiev GP. 1980. Non-random cleavage of SV40 DNA in the compact minichromosome and free in solution by micrococcal nuclease. *Biochem Biophys Res Commun* **92**: 532–539.
- Roberts A, Pimentel H, Trapnell C, Pachter L. 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**: 2325–2329.
- Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma LJ, Lin MF, et al. 2010. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**: 1787–1797.
- Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Hua C, Man Y, Rosenzweig E, Goldy J, Haydock A, et al. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* **3**: 511–518.
- Satoh K, Doi K, Nagata T, Kishimoto N, Suzuki K, Otomo Y, Kawai J, Nakamura M, Hirozane-Kishikawa T, Kanagawa S, et al. 2007. Gene organization in rice revealed by full-length cDNA mapping and gene expression analysis through microarray. *PLoS ONE* **2**: e1235. doi: 10.1371/journal.pone.0001235.
- Schnable JC, Pedersen BS, Subramaniam S, Freeling M. 2011. Dose-sensitivity, conserved noncoding sequences and duplicate gene retention through multiple tetraploidies in the grasses. *Front Plant Sci* **2**: 2. doi: 10.3389/fpls.2011.00002.
- Schones DE, Cui KR, Cuddapah S, Roh TY, Barski A, Wang ZB, Wei G, Zhao KJ. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**: 887–898.
- Sun XL, Zou Y, Nikiforova V, Kurths J, Walther D. 2010. The complexity of gene expression dynamics revealed by permutation entropy. *BMC Informatics* **11**: 607. doi: 10.1186/1471-2105-11-607.
- Tolhuis B, Muijters I, de Wit E, Teunissen H, Talhout W, van Steensel B, van Lohuizen M. 2006. Genome-wide profiling of PRC1 and PRC2 polycomb chromatin binding in *Drosophila melanogaster*. *Nat Genet* **38**: 694–699.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.
- Wang XF, Elling AA, Li XY, Li N, Peng ZY, He GM, Sun H, Qi YJ, Liu XS, Deng XW. 2009. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell* **21**: 1053–1069.
- Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ, et al. 2011. DNA methylation status predicts cell type-specific enhancer activity. *EMBO J* **30**: 3028–3039.
- Wu C. 1980. The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**: 854–860.
- Yan HH, Talbert PB, Lee HR, Jett J, Henikoff S, Chen F, Jiang JM. 2008. Intergenic locations of rice centromeric chromatin. *PLoS Biol* **6**: 2563–2575.
- Yan HH, Kikuchi S, Neumann P, Zhang WL, Wu YF, Chen F, Jiang JM. 2010. Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. *Plant J* **63**: 353–365.
- Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA. 2007. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* **39**: 1512–1516.
- Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D. 2010. Local DNA hypomethylation activates genes in rice endosperm. *Proc Natl Acad Sci* **107**: 18729–18734.
- Zhang XY, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen HM, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**: 1189–1201.
- Zhang XY, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J, Jacobsen SE. 2007. Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol* **5**: e129. doi: 10.1371/journal.pbio.0050129.
- Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S. 2008. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**: 125–129.

Received August 31, 2011; accepted in revised form September 20, 2011.