

# Genome-Wide Identification, Phylogenetic and Co-Expression Analysis of *OsSET* Gene Family in Rice

Zhanhua Lu<sup>1</sup>, Xiaolong Huang<sup>2</sup>, Yidan Ouyang<sup>3\*</sup>, Jialing Yao<sup>2\*</sup>

**1** College of Horticulture and Forestry Science, Huazhong Agricultural University, Wuhan, PR China, **2** College of Life Science and Technology, Huazhong Agricultural University, Wuhan, PR China, **3** National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, PR China

## Abstract

**Background:** SET domain is responsible for the catalytic activity of histone lysine methyltransferases (HKMTs) during developmental process. Histone lysine methylation plays a crucial and diverse regulatory function in chromatin organization and genome function. Although several *SET* genes have been identified and characterized in plants, the understanding of *OsSET* gene family in rice is still very limited.

**Methodology/Principal Findings:** In this study, a systematic analysis was performed and revealed the presence of at least 43 *SET* genes in rice genome. Phylogenetic and structural analysis grouped SET proteins into five classes, and supposed that the domains out of SET domain were significant for the specific of histone lysine methylation, as well as the recognition of methylated histone lysine. Based on the global microarray, gene expression profile revealed that the transcripts of *OsSET* genes were accumulated differentially during vegetative and reproductive developmental stages and preferentially up or down-regulated in different tissues. *Cis*-elements identification, co-expression analysis and GO analysis of expression correlation of 12 *OsSET* genes suggested that *OsSET* genes might be involved in cell cycle regulation and feedback.

**Conclusions/Significance:** This study will facilitate further studies on *OsSET* family and provide useful clues for functional validation of *OsSETs*.

**Citation:** Lu Z, Huang X, Ouyang Y, Yao J (2013) Genome-Wide Identification, Phylogenetic and Co-Expression Analysis of *OsSET* Gene Family in Rice. PLoS ONE 8(6): e65426. doi:10.1371/journal.pone.0065426

**Editor:** Xiaoyu Zhang, University of Georgia, United States of America

**Received:** January 13, 2013; **Accepted:** April 23, 2013; **Published:** June 7, 2013

**Copyright:** © 2013 Lu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported by grants from the National Natural Science Foundation of China ([http://www.nsf.gov.cn/e\\_nsf/desktop/zn/0101.htm](http://www.nsf.gov.cn/e_nsf/desktop/zn/0101.htm); grant number 30971551), the Fundamental Research Funds for the Central Universities (<http://www.dost.moe.edu.cn/>; grant number 2012MBDX012), and the Specialized Research Fund for the Doctoral Program of Higher Education (<http://www.cutech.edu.cn/cn/index.htm>; grant number 20100146120034). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: yaojlmy@mail.hzau.edu.cn (JLY); diana1983941@mail.hzau.edu.cn (YDO)

## Introduction

SET domain, named after the three *Drosophila* proteins SUPPRESSOR OF VARIATION 3–9 [SU(VAR)3–9], ENHANCER OF ZESTE [E(Z)] and TRITHORAX (TRX) [1], has been known to be involved in the biochemical process of the histone lysine methyltransferases (HKMTs) [2]. It contains an approximately 130-amino acid, presenting as an evolutionarily conserved motif in chromosome proteins from yeast to mammals and higher plants [3]. It consists of two non-contiguous regions formed by N- and C-terminal ends of the primary sequence, known as SET-N and SET-C, respectively, and an insert region (SET-I) [4]. SET domain protein methyltransferases have enormous impacts on the regulation of chromatin structure and function [5,6]. They catalyze the transfer of methyl groups from the cofactor S-adenosylmethionine (AdoMet) to specific lysine residues of protein substrates, such as the N-terminal tails of histone (H3 or H4) and the large subunit of the Rubisco holoenzyme complex [7,8].

Baumbusch [1] first identified 39 SET domain genes in *Arabidopsis thaliana* and divided them into four classes based on the SET domains, cysteine-rich regions and additional conserved

domains. Springer et al. identified 32 SET domain genes in *Arabidopsis* and 22 ones in *Zea mays*, and classified the SET domain proteins into five subfamilies, on the basis of phylogenetic analyses and domain organization [9]. It revealed that the duplication of SET domain proteins in plants was extensive and had occurred via multiple. Ng et al. [10] inferred that there were at least 47 SET genes in *Arabidopsis*, 35 members in maize and 34 ones in rice, based on the annotation in Pfam and ChromDB database, which were classified into seven groups. Pontvianne [11] reported that SET domain genes in *Arabidopsis* can be divided into five classes (I to V), based on their domain architectures and/or differences in enzymatic activity of SET domain-containing proteins.

Presently, a number of *SET* genes have been functionally identified in plants. Mutation of *Arabidopsis* SET domain genes resulted in phenotypic abnormalities due to the improper regulation of important developmental genes [1,9–11]. *Arabidopsis* *CURLY LEAF (CLF)*, a homolog of *Drosophila E(Z)* gene, was involved in the division and elongation of cells during leaf morphogenesis [12–14]. Further evidences revealed that CLF can directly mediate the repression of *AGAMOUS (AG)*, *FLOWERING LOCUS C (FLC)* and *FLOWERING LOCUS T (FT)* via lysine 27 of histone H3 trimethylation (H3K27me3) and thus control floral

organogenesis in a Polycomb repressive complex 2 (PRC2)[15–18]. *MEDEA* (*MEA*), another *E(Z)* homolog in *Arabidopsis*, is a self-controlled imprinting gene and functions in controlling the proliferation of central cell [19–23]. Su(VAR)3–9 homolog (SUVH) proteins KRYPTONITE (KYP, also known as SUVH4), SUVH5 and SUVH6 are shown to function in a locus specific manner to undergo H3K9 methylation and cytosine methyltransferase 3 (CMT3)-mediated non-CG DNA methylation [24–27]. In contrast to the *E(Z)* and SUVH proteins suppressing transcription, ATX1 functions as an activator of homeotic genes by lysine 4 of histone H3 (H3K4) methylation [28–30]. These results uncovered the extensive functions of SET proteins in plant development.

Rice is one of the major staple foods and an ideal model species of monocotyledons for functional genomics analysis. In previous studies, several SET proteins have been characterized in rice. *OsCLF* and *OsiEZ1*, both of which are *E(Z)* homologs, is expressed preferentially in young seedlings and during reproductive development [31–33]. SET Domain Group 714 (SDG714) and SDG728, encoding H3K9me2 histone methyltransferase, display specific functions in chromatin modification and retrotransposon repression [34,35]. The histone methyltransferase SDG724 mediates H3K36me2/3 deposition at *MADS50* and *RFT1* and promotes flowering in rice [36]. Therefore, it is necessary to carry out a comprehensive functional exploration for *OsSET* gene family in rice.

In this study, the members of SET family in rice have been analyzed based on complete genome and protein sequences and annotations. Expression patterns and co-expression analyses were performed to discover the potential functions of *OsSET* gene family. Promoter *cis*-elements identification and the combined analysis of expression correlation suggest that most of *OsSET* genes may be cell cycle modulated and linked to the cell cycle progression by histone modifications. Our results will provide a useful reference for further functional analysis of members of *OsSET* gene family in rice.

## Results and Discussion

### Identification and Classification of *OsSET* Gene Family

In the previous report [10], 34 SET genes in rice genome were inferred, on the basis of annotation in Pfam and ChromDB database. However, based on the update Pfam and ChromDB database and MSU data, 43 SET family genes in rice were identified in our study. These 43 *OsSET* genes were named from *OsSET1* to *OsSET43* according to their positions on chromosomes. Nine genes, *OsSET12*, *OsSET13*, *OsSET18*, *OsSET23*, *OsSET29*, *OsSET31*, *OsSET36*, *OsSET37* and *OsSET39*, were novel *OsSET* genes compared to the ones in previous study. *OsSET* genes vary substantially in the size of their encoded proteins and their physicochemical properties (Table S1). *OsSET33* has 25 exons and 24 introns, whereas *OsSET14*, *OsSET20* and *OsSET30* have no intron. The position of the SET domain also varies within the proteins. The shortest *OsSET* protein is *OsSET9* with 231 amino acids, while the longest one is *OsSET30* with 1292 amino acids. EXPASY analysis showed a large variation in isoelectric point (pI) values (ranging from 4.4119 to 9.2625) and molecular weights (ranging from 110.892 kDa to 138.5423 kDa). Nevertheless, only 3 (*OsSET4*, *OsSET14* and *OsSET29*) of the 43 *OsSET* genes were predicted to be stable proteins. Details on other parameters of protein sequences were shown in Table S1.

### Chromosomal Localization and Gene Duplication

The genomic distribution of *OsSET* genes was determined by their chromosomal positions on rice chromosome. Totally, 43

*OsSET* genes were dispersed on the 12 chromosomes, presenting unevenly in all regions of the chromosomes. A brief diagrammatic representation of chromosomal distribution of *OsSET* genes was described (Figure 1, the exact position on rice chromosome pseudomolecules was given in Table S1). Seven *OsSET* genes are on chromosome 2, six *OsSET* genes are on chromosome 1 and 8, respectively; five *OsSET* genes are on chromosome 9; four *OsSET* genes are on chromosome 3 and chromosome 4; two *OsSET* genes are on chromosome 5, 7, 10, 11, 12, respectively; only one *OsSET* gene is on chromosome 6.

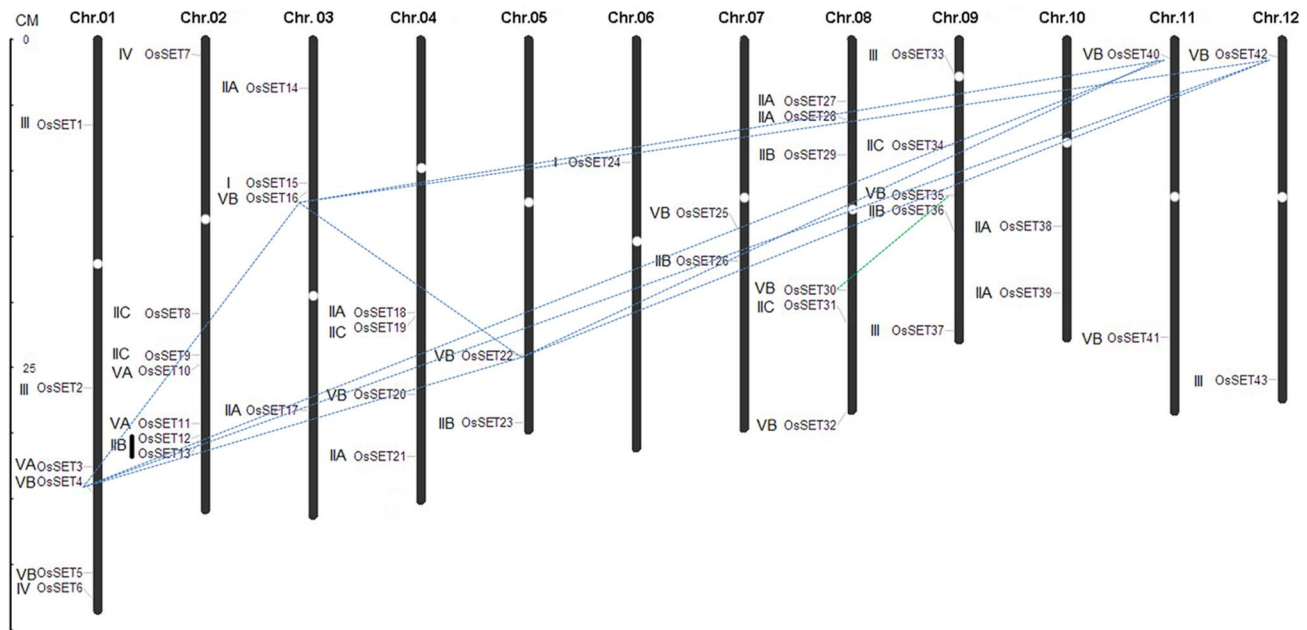
During the evolution of a gene family, segmental duplication and tandem duplication play important roles in generating new members [37]. Therefore, both segmental and tandem duplication events were investigated for elucidating the potential mechanism of evolution of *OsSET* gene family. Analysis of the MSU RGAP rice segmental duplication database revealed that 10 pairs of *OsSET* genes could be assigned to MSU RGAP segmental duplication blocks. The overall similarity of the cDNA sequences of these genes ranged from 25.6% to 77.0% and all of them were found to have their counterparts on duplicated segments (Table S2). Interestingly, these duplicated segments can be clustered in 2 groups. Five *OsSET* genes (*OsSET4*, *OsSET16*, *OsSET22*, *OsSET40* and *OsSET42*), which had high overall identities between each other, belonged to the same group. While the overall identity between *OsSET30* and *OsSET35* was 52.0%, and was included in another group. None of the *OsSET* genes seemed to be generated from tandem duplications in our analysis. These results implicated that much of the diversity of the *OsSET* gene family in rice was mainly due to the segmental duplication events.

### Phylogenetic and Structural Analysis of *OsSET* Gene Family

To determine the evolutionary relationships of SET family genes between rice and *Arabidopsis*, an unrooted phylogenetic tree was constructed from alignments of their full-length protein sequences. The latest data showed that *Arabidopsis* SET family genes can be divided into five classes (I to V), based on their domain architectures and/or differences in enzymatic activity of SET domain-containing proteins [11]. Coincidentally, our phylogenetic analysis and their domain architectures support the classification of rice and *Arabidopsis* SET gene family into five classes (Figure 2, 3; Table S1).

Class I SET proteins include 2 rice *OsSET* proteins and 3 *Arabidopsis* SET proteins. *OsSET15/OsiEZ1* and *OsSET24/OsCLF* are the homologs of *Arabidopsis* SWINGER (SWN)/SDG10 and CLF/SDG1, respectively. No *Arabidopsis* MEDEA (*MEA*)/SDG5 homolog was found in rice. In addition to the C-terminal SET domain, SANT domain (Swi3, Ada2, N-Cor, and TFIIB DNA binding domain) and cysteine rich CXC domain were found in this subfamily. This result is in agreement with previous studies [10]. Recent studies suggested that SANT domains might be a histone-tail-binding module [38,39]. It is reported that *E(Z)*-like proteins are components of PRC2 complexes and function as transcriptional repressors by H3K27me3 in diverse eukaryotes [13,16,21,22]. Therefore, *OsSET15/OsiEZ1* and *OsSET24/OsCLF* may have H3K27me3 activities by these conserved domains.

Class II subfamily can be divided into three clusters of IIA, B and C based on their domains, which is also in accordant with the previous reports [9,10]. Eight members in rice (*OsSET14*, *OsSET17*, *OsSET18*, *OsSET21*, *OsSET27*, *OsSET28*, *OsSET38* and *OsSET39*) and five in *Arabidopsis* (SDG35–39) belonged to Class IIA, which only contain the SET domain. Class IIB proteins (*OsSET12*, *OsSET13*, *OsSET23*, *OsSET26*, *OsSET36*, *SDG40*,



**Figure 1. Chromosomal distribution, and tandem and segmental genome duplications of the *OsSET* gene family.** The scale on the left is in megabases (Mb). The ovals on the chromosomes (vertical bars) indicate the positions of centromeres; the chromosome numbers are shown on the top of each bar. The segmental duplication genes are connected by a dotted line.  
doi:10.1371/journal.pone.0065426.g001

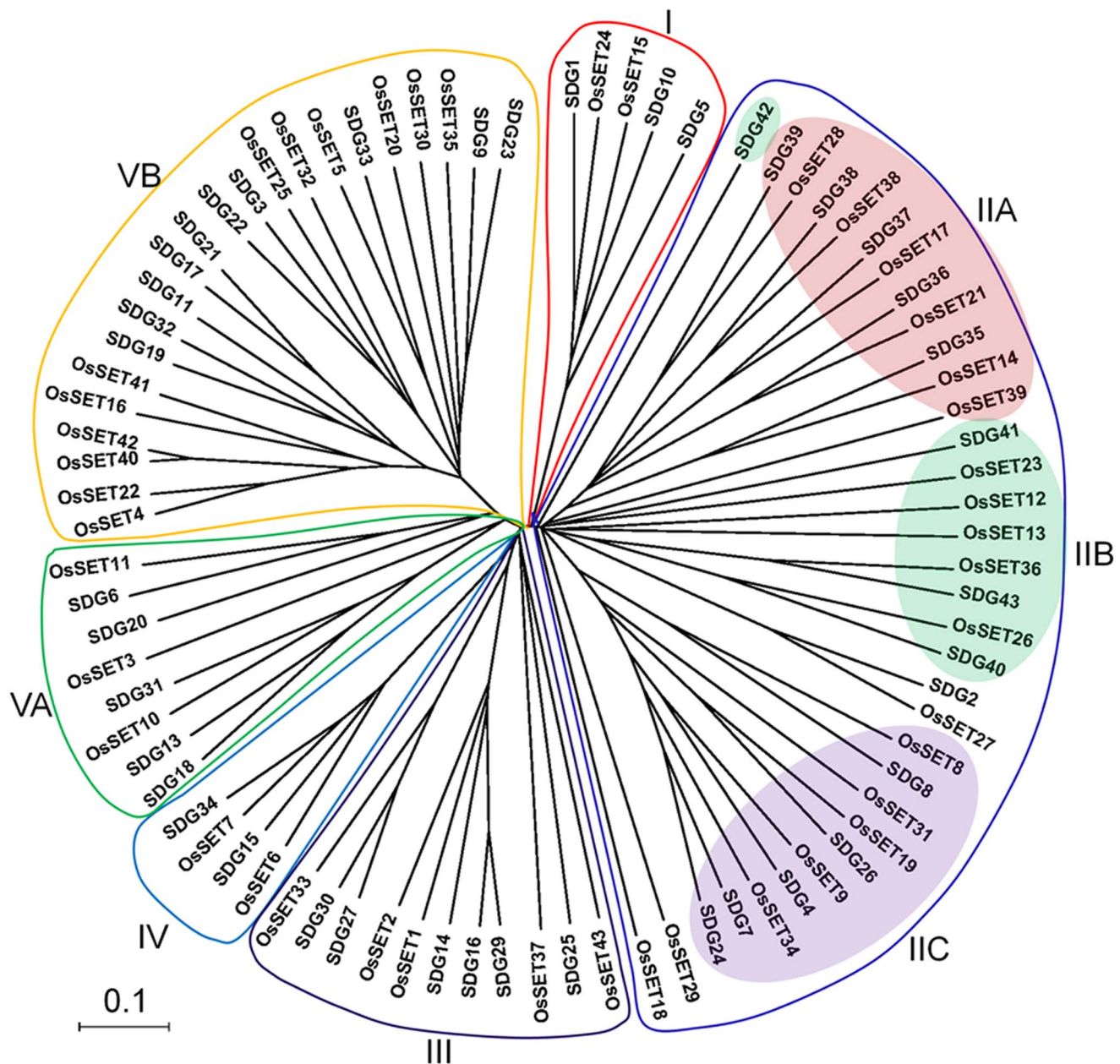
SDG41, SDG42, SDG43) have a Rubisco LSMT substrate-binding domain (RSB domain), which allows the binding of the protein to its substrate, such as the N-terminal tails of histones H3 and H4 and the large subunit of the Rubisco holoenzyme complex [7]. Class IIC has five SET proteins in rice (*OsSET8*, *OsSET9*, *OsSET19*, *OsSET31* and *OsSET34/SDG724*) and five in *Arabidopsis* (*ABSENT*, *SMALL*, *OR HOMEOTIC DISCS 1 HOMOLOG 1 (ASHH1)/SDG26*, *ASHH2/SDG8*, *ASHH3/SDG7*, *ASHH4/SDG24* and *ASH1-related 3 (ASHR3)/SDG4* [40–42]). All proteins except *OsSET31* in class IIC have an AWS domain (**A**ssociated **W**ith **S**ET, a sub unit of pre-SET domain) [10]. Five of the members in class IIC have an additional cysteine-rich post-SET domain. Although some of the class II subfamily genes have been demonstrated to methylate H3K36 at the region of actively transcribed genes [43], the functions of the additional domains are still little known.

Class III HKMTs consist of four rice SET members (*OsSET1*, *OsSET2*, *OsSET33* and *OsSET37*) and seven *Arabidopsis* genes. Five of the *Arabidopsis* genes encode homologs of Trithorax (*trxG*), which named as *Arabidopsis* Trithorax-like protein (*ATX1/SDG27*, *ATX2/SDG30*, *ATX3/SDG14*, *ATX4/SDG16* and *ATX5/SDG29*), while another two genes, *ATXR3/SDG2* and *ATXR7/SDG25*, are *ATX-RELATED (ATXR)* genes [44]. Class III subfamily genes have several additional highly conserved protein domains, including PWWP, FYRN/C and plant homeo-domain (PHD). The PWWP domain was shown to be a DNA or methyl-lysine histone binding domain [45–47]. In *Arabidopsis*, class III proteins are able to methylate H3K4me<sub>2/3</sub>, acting as antagonistic regulators with Polycomb Group (PcG) proteins to maintain transcriptional OFF and ON states of their target genes [48,49]. Interestingly, PHD finger is considered to be specific and highly robust binding modules for H3K4me<sub>2/3</sub> in humans and plants, resulting the recruitment of basal transcriptional active factor(s) [50–53]. *ATX1* has been demonstrated to interact with *ASHH1/SDG26*, suggesting that *trxG* complexes could involve

different sets of histone lysine methyltransferases in *Arabidopsis* [40]. These results provide an efficient insight for functional identification of *trxG* in rice.

*OsSET6*, *OsSET7*, *ATXR5/SDG15* and *ATXR6/SDG34* are included in class IV, which characterized with a PHD domain and a SET domain. *ATXR5* and *ATXR6* are involved in DNA replication [54]. Although class IV proteins are near to class III on their evolutionary relationship, they are distinct from class III for the absence of PWWP domain. The result suggests that the PWWP domain may be crucial for the antagonistic mechanism between PcG and *trxG*.

Class V proteins are characterized by the presence of pre-SET and SET domains. This class can be divided into VA [SU(VAR)3–9 (SUVR)] and VB (SUVH) for the latter having an YDG/SRA domain [1]. Class VA contains three *OsSET* proteins (*OsSET3*, *OsSET10* and *OsSET11*) and 5 SU(VAR)3–9 (SUVR) proteins (*SUVR1/SDG13*, *SUVR2/SDG18*, *SUVR3/SDG20*, *SUVR4/SDG31* and *SUVR5/SDG6*). *OsSET10*, *SDG18* and *SDG31* also have an N-terminal plant-specific domain, WIYLD, which binds ubiquitin and enables conversion of H3K9me<sub>1</sub> to H3K9me<sub>3</sub> [55]. Subfamily VB consists of 12 *OsSET* proteins (*OsSET5/SDG714*, *OsSET14*, *OsSET16*, *OsSET20*, *OsSET22/SDG728*, *OsSET25*, *OsSET30/SDG710*, *OsSET32*, *OsSET35/SDG727*, *OsSET40*, *OsSET41*, *OsSET42*) and 9 SUVH proteins (*SUVH1/SDG32*, *SUVH2/SDG3*, *SUVH3/SDG19*, *SUVH4/KYP/SDG33*, *SUVH5/SDG9*, *SUVH6/SDG23*, *SUVH7/SDG17*, *SUVH8/SDG21*, *SUVH9/SDG22* and *SDG11*). It has reported that the YDG/SRA domain can mediate epigenetic inheritance by recruiting histone deacetylase (HDAC), DNA methyltransferase (DNMT) to methyl-CpG site [56–58]. Similarly, the YDG/SRA domain of KYP, SUVH5 and SUVH6 binds directly to methylated DNA at both CpG and non-CpG site, thereafter, providing a binding site for CMT3 via its chromodomain to CHG methylation [59]. *OsSET5/SDG714*, an H3K9 methyltransferase, is also involved in DNA methylation in rice [35,60], which implies



**Figure 2. Phylogenetic analysis of Arabidopsis and rice SET proteins.** Phylogenetic tree of rice and Arabidopsis SET proteins. An unrooted NJ tree of rice and Arabidopsis SET proteins is shown. The six classes are marked by different colors. Scale bar represents 0.1 amino acid substitution per site.

doi:10.1371/journal.pone.0065426.g002

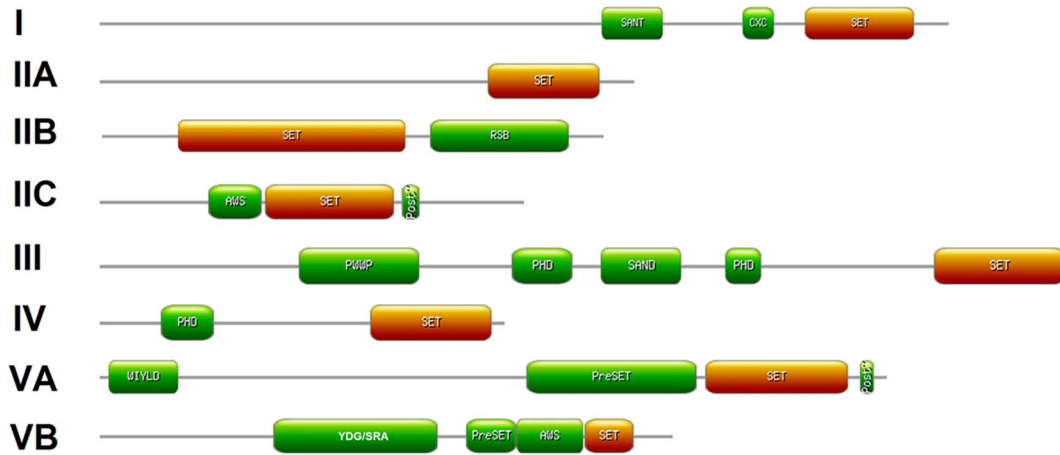
a similar mechanism between class VB OsSET proteins and DNA methylation. SUVR5 establishes the heterochromatic state by H3K9me2 deposition in a DNA methylation-independent manner through zinc fingers [61]. However, no such DNA binding domain was identified in rice SUVR like proteins. Therefore, there must be a distinct mechanism for SUVR in rice.

Because SET domain is essential for the catalytic activity of SET proteins, the MEME motif search tool was employed to identify the conserved motifs of SET domains from 84 SET proteins in rice and Arabidopsis. Three distinct motifs, motifs 2, 3 and 1, were located orderly at SET-N, SET-I and SET-C region of SET domain, respectively (Figure 4). 55 out of 84 (65.5%) SET proteins have motifs 1, 2 and 3. 17 (20.2%) SET proteins only

have motifs 1 and 3. OsSET29 and OsSET39 have motifs 2 and 3. OsSET12, OsSET13 and OsSET23 only have motif 1. OsSET3 and OsSET16 only have motif 2. The other five SET proteins (OsSET26, OsSET36, SDG40, SDG42 and SDG43) have none of the three motifs. Interestingly, 24 out of 29 (82.8%) proteins that have no more than 2 motifs are subfamily IIA or IIB members, which showed diversity in these two subfamilies.

#### Identification of Cis-elements in OsSET Gene Promoters

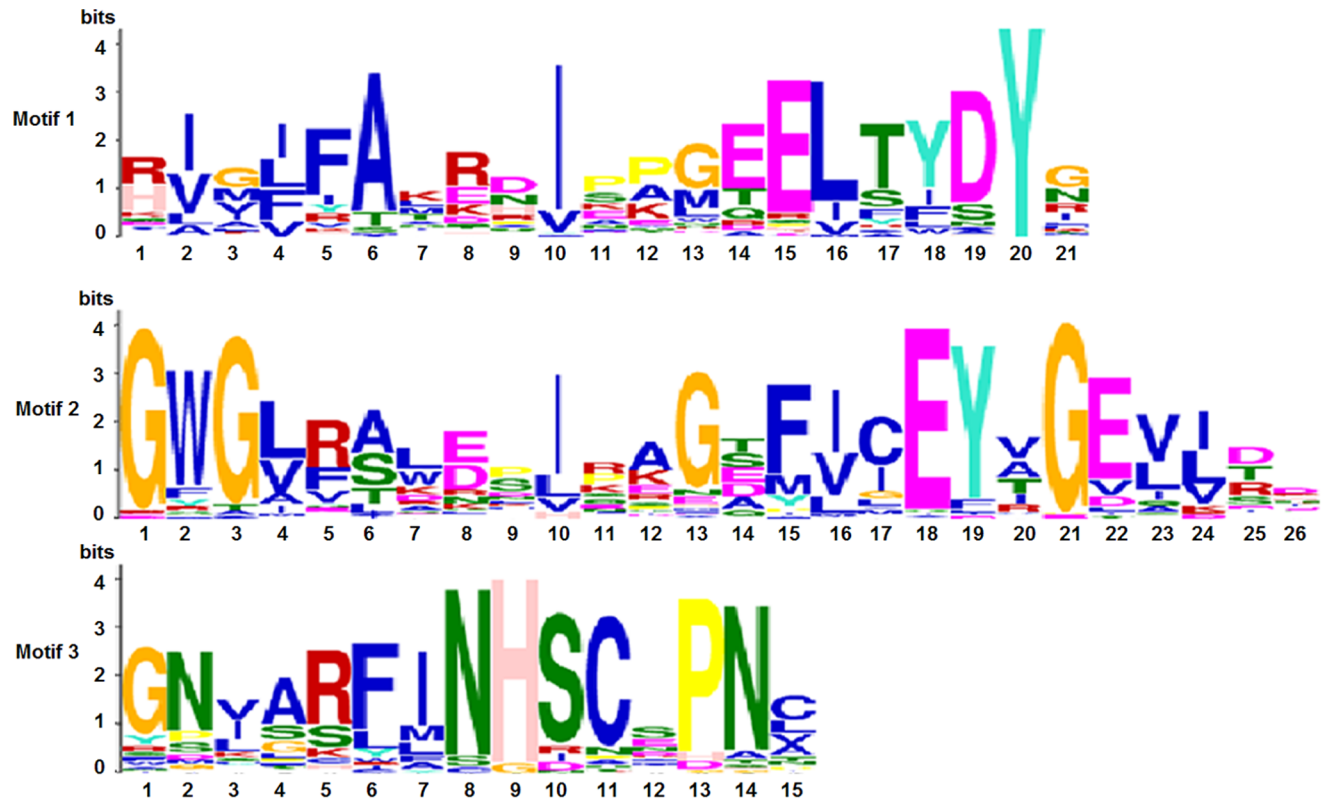
To understand the molecular mechanism of OsSET genes in transcriptional regulation, cis-elements at the promoter regions were identified (Table S3). As a result, 255 cis-elements were obtained. In addition to basic TATA-box and CCAAT box, cis-



**Figure 3. Structure of representative *OsSET* proteins from each subfamily.** The protein structure is based on the presence of *OsSET* and other additional domains as identified by Pfam. Subfamily name of each corresponding protein belonged to are given on the left. The length and order of domains represent actual situation in each protein.  
doi:10.1371/journal.pone.0065426.g003

elements such as MYB recognition, auxin responsive, gibberellin (GA) response, abscisic acid (ABA) responsive and E2F-binding site were found at the promoter regions of *OsSET* genes. It is well known that E2F transcription factors control the cell cycle by regulating transcription of genes required for DNA replication and cell cycle [62]. Many investigations show that the E2F targets have

one or more consensus DNA sequence of E2F binding sites [63–65]. It was reported that *ATX1/SDG27*, *ATXR5/SDG15*, and *ATXR6/SDG34* in *Arabidopsis*, *OsSET6*, and *OsSET7* in rice, and lots of other cell cycle or DNA replication related genes were considered as E2F targets for their E2F binding *cis*-elements [54,63,64]. While, in our analysis, we found 32 *OsSET* genes had



**Figure 4. Sequence LOGOs for each motif of SET domains using the MEME algorithm.** MEME motifs are displayed by stacks of letters at each position. The total height of the stack is the “information content” of that position in the motif in bits. The height of the individual letters in a stack is the probability of the letter at that position multiplied by the total information content of the stack. X- and Y-axis represents the width of motif and the bits of each letters, respectively.  
doi:10.1371/journal.pone.0065426.g004

**Table 1.** Conserved E2F binding *cis*-elements analysis of the *OsSET* gene promoters.

Accession Number <sup>a</sup>	ID <sup>b</sup>	Sequences
S000217	SITEIBOSPCNA	TGGTCCAC
S000224	SITEIOSPCNA	CCAGGTGG
S000366	E2FANTRNR	TTCCCGC
S000396	E2F1OSPCNA	GCGGAAA
S000417	E2FAT	TYTCCGCC; Y = T/C
S000455	PE2FNTRNR1A	ATTGCGC
S000476	E2FCONSENSUS	WTTSSCSS; W = A/T; S = C/G

<sup>a</sup>Accession number of *cis*-element in PLACE database.

<sup>b</sup>Locus identity number of *cis*-element in PLACE database.

doi:10.1371/journal.pone.0065426.t001

E2F binding site (E2FCONSENSUS, SITEIBOSPCNA, SITEIOSPCNA, E2FANTRNR, E2F1OSPCNA, E2FAT, PE2FNTRNR1A) (Table. 1). Therefore, the analyses revealed that most of the *OsSET* genes might be regulated by E2F transcription factors.

### Expression Profiling of *OsSET* Genes in Rice

To investigate the transcript accumulation of *OsSET* genes in the entire life cycle, the expression profiling covering 24 developmental stages (Table S4) in Minghui 63 were analyzed using Affymetrix rice microarray data from CREP database. A hierarchical cluster displaying the logarithm of average signal values for the 40 *OsSET* genes were generated. Distinctly, the expression patterns of *OsSET* genes could be classified into two major groups (**Figure 5**). 20 genes belonged to Group I, most of which showed high transcript accumulations (average expression signal from 777.3 to 4211.1) in the tissues analyzed. *OsSET41* had the highest expression level in the entire life cycle. These genes could be further divided into three subgroups, subgroup A1–3. Subgroup A1 consists of 8 genes, which have high expression level in panicles and/or stamen. Subgroup A2 has 7 *OsSET* genes, all of which show relative high expression level in almost all tissues analyzed. Subgroup A3 has 5 *OsSET* genes, which display higher expression in vegetative tissues than in reproductive tissues. Group B contains 20 genes, exhibiting relative low expression signals in most tissues or preferential expressions in some tissues. *OsSET5*, *OsSET6*, *OsSET10*, *OsSET17*, *OsSET27*, *OsSET28*, *OsSET32* and *OsSET37* showed high expression in panicles; *OsSET6*, *OsSET10* and *OsSET27* expressed relatively higher in stamen. In addition, *OsSET28* in spikelet and seed, *OsSET32* in seed, *OsSET37* in stem showed tissue-specific expressions.

The expression patterns of some *OsSET* genes were further confirmed by real-time PCR analysis. The expression levels of *OsSET3*, *OsSET9* and *OsSET19*, which have no probe set information in CREP database, were detected in different tissues (**Figure 6**). *OsSET3* shows a relative high expression in root, flag leaf, panicles and mature endosperm. Both *OsSET9* and *OsSET19* are included in class II, however, their expression patterns are divergent: *OsSET9* are highly expressed in the later stage of endosperm (21 days after pollination), whereas the expression of *OsSET19* enriched in young leaf and decreased in endosperm. The expression levels of *OsSET11*, *OsSET24/OsCLF* and *OsSET15/OsEz1* are also detected, which are in accordance with the microarray data. *OsSET26* has a low expression pattern in microarray analysis, whereas our quantitative PCR result shows a relative higher expression level in vegetative stage than

reproductive stage. *OsSET39* expresses relatively higher in root, leaf and panicles, and the expression is enriched in the developing endosperm, implying multiple functions in plants development.

On the purpose of revealing more information in *OsSET* gene expression pattern, genes that showed differential expression during various developmental stages in comparison to seed were analyzed. Genes that considered as preferential expression in a given stage showed tremendous differences (Fig 7, Table S5). Up-regulated genes mainly accumulated in panicles and stamen, suggesting that *OsSET* genes may participate in various molecular pathways in flowering development. Surprisingly, although down-regulated genes accumulated in seedlings, they were activated in stamen, either. These microarray and real-time PCR results indicate that *OsSET* genes may play essential roles through the life cycle of rice.

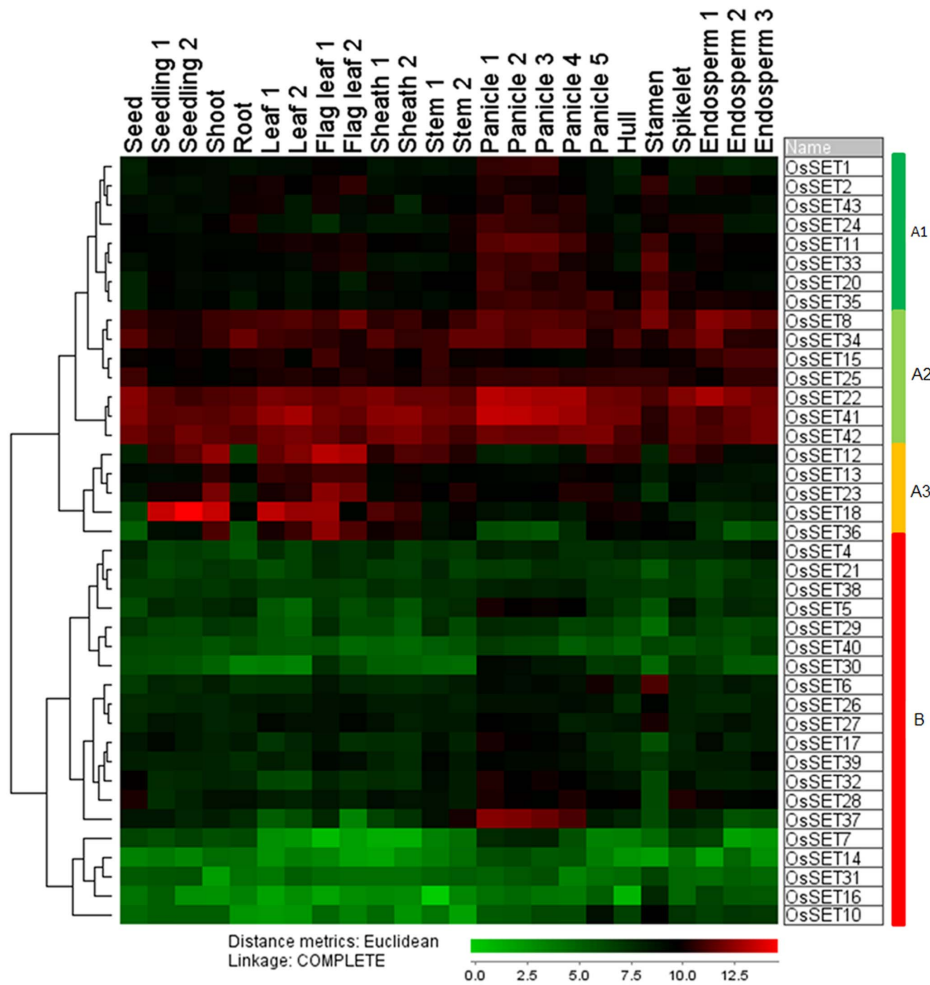
The expression patterns of segmentally duplicated *OsSET* genes were examined by Affymetrix microarray data. Probesets were available for all segmental duplication genes in microarray data. A comparison of expression level revealed that a pair of segmental duplicated genes always showed similar expression pattern, although one of the copy showed low expression level, or was not expressed at significant levels in most of the tissues (**Figure 8**). In the group of *OsSET4*, *OsSET16*, *OsSET22*, *OsSET40* and *OsSET42* (**Figure 8A**), for instance, 2 out of 5 genes had a similarly high expression level. The resemblance also existed in the *OsSET30* and *OsSET35* group (**Figure 8B**). We might therefore infer that immediately after segmental duplication, the two copies of genes might be functionally redundant. However, only one of them is functional retained while the other degenerates into a pseudogene eventually.

### Responses of *OsSET* genes under NAA, KT, and GA3 Treatments

Phytohormones play critical roles in plant growth and development. To investigate the *OsSET* genes in response to phytohormone treatment, differential expression analysis was performed. As a result, 9 *OsSET* genes that were differentially expressed under one or more of the phytohormone NAA, KT, GA3 in seedlings in trefoil stage, compared with the control without treatment. The fold change values with respect to control were given in Table S4. Three *OsSET* genes showed differential expression under all three phytohormone treatments, among which *OsSET23* and *OsSET36* were up-regulated, whereas *OsSET18* was down-regulated. *OsSET2* and *OsSET16* were up-regulated to KT, and *OsSET24* and *OsSET34* were up-regulated to GA3 treatment. In contrast, *OsSET17* and *OsSET37* were down-regulated specifically to KT and GA3 treatment, respectively. The expression profile of the remaining genes in response to NAA, KT, and GA3 was not significant. These results partially in accordance with the *cis*-elements analysis, showing that the above 9 *OsSET* genes have one or more GA responsive elements (Table. S3).

### Identification and Functional Annotation of Genes Co-expressed with *OsSET* Genes

Co-expression analysis has been successfully exploited to identify functional transcription regulators in *Arabidopsis*, rice and other organisms [66,67]. Hence, in order to disinter more information of the *OsSET* genes, 40 *OsSET* genes with matching probeset were selected as “guide genes” to identify the co-expressed genes using expression data from CREP database, with an absolute value of the Pearson correlation coefficient (PCC) greater than 0.75 ( $\alpha = 0.05$ ) [68,69]. As a result, 2390 genes whose



**Figure 5. Expression patterns of *OsSET* genes during the life cycle of the rice plant.** Hierarchical cluster displays the expression profile for 40 *OsSET* genes with matching probesets in the Affymetrix microarray. (Color bar at the base represents log<sub>2</sub> expression values: green, representing low expression; black, medium expression; red, high expression).  
doi:10.1371/journal.pone.0065426.g005

expression pattern tightly correlated with 30 *OsSET* members were extracted (Table S6a).

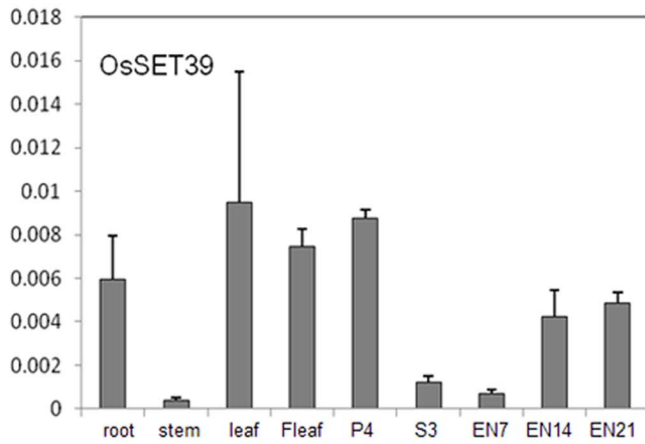
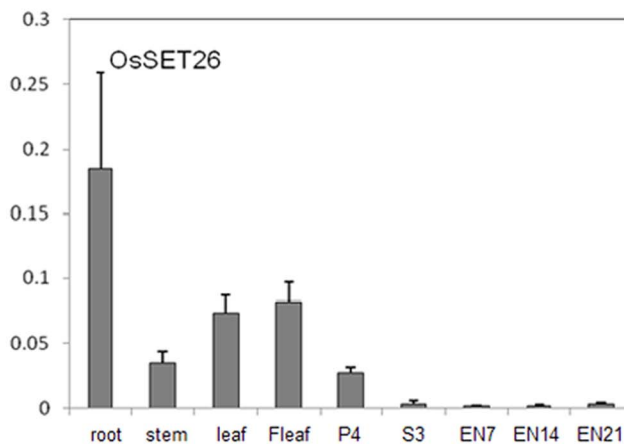
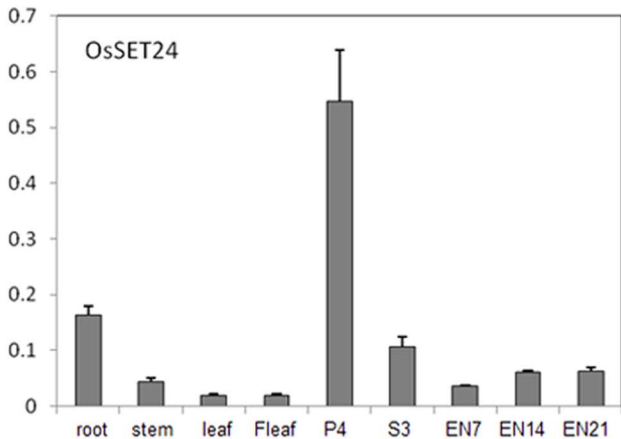
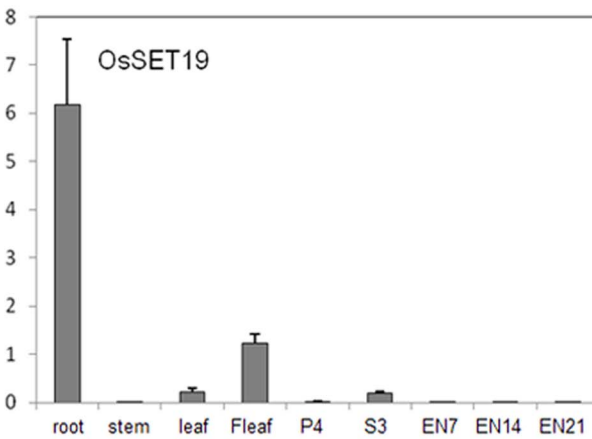
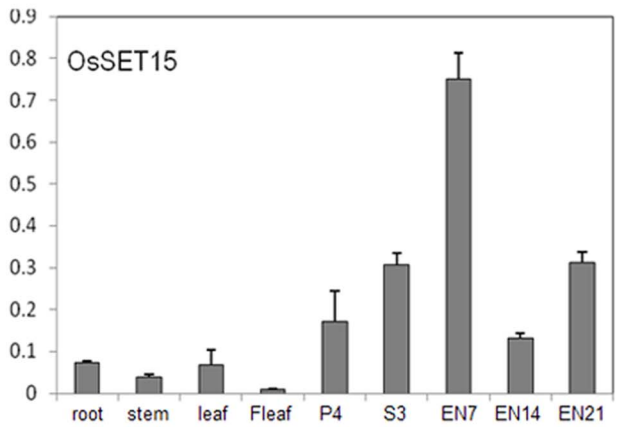
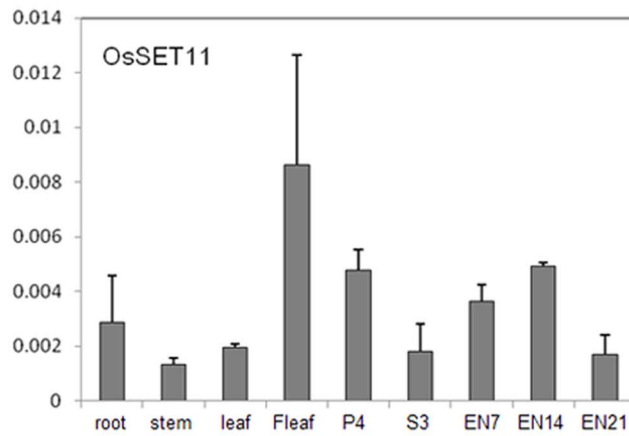
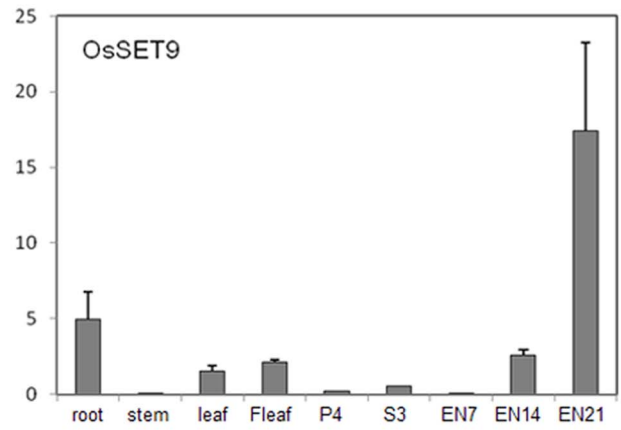
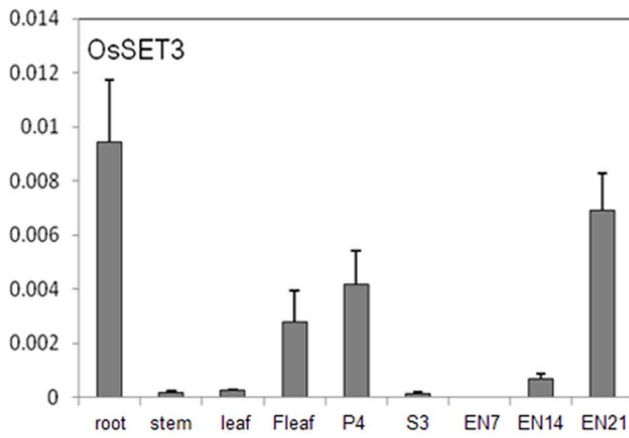
We next analyzed the GO annotations assigned to these genes by agriGO tools. The enriched GO annotations particularly concentrate on cellular process, cellular component biogenesis and organization, biological regulation and metabolic process (Figure 9, Table S6b). They encode proteins as macromolecular complex (protein complex, DNA polymerase, protein-DNA complex) in cell or organelle, substantially. The molecular functions tightly associated with them are catalytic activity, transcription regulator and binding. These results suggest that the functions of *OsSET* genes may be associated with DNA replication and gene transcription.

As is generally known, SET domain proteins have comprehensive impacts on the regulation of chromatin structure and function [5,6]. Therefore, we focused on 503 out of 2390 co-expression genes which may be associated with epigenetic regulation (histone genes, cell cycle related genes, DNA replication, transcription factor, chromatin reassemble related genes and so on. Table S6c). After the recalculation and student-*t* test of PCC, 450 genes were co-expressed with 29 *OsSET* genes. In addition, 13 *OsSET* genes were co-expressed with each other tightly. Except for *OsSET28*, the other 12 *OsSET* genes have close expression relationships with

histone genes, cell cycle control related genes and chromatin assemble factors, anther-specific proline-rich protein (*APG*) genes, DNA replication related genes and so on. The result gives informative clues in functional characterization of these *OsSET* genes.

### *OsSET* Proteins May Be Involved in Cell Cycle Regulating by Histone Modification

The visualized figure in Figure 10A shows the close relationships between 12 *OsSET* genes and 13 histone protein genes (containing H1, H2A, H2B.1, H2B.2, H3 and H4). In this network, *OsSET5* is co-expressed with all of these histone genes, *OsSET24* is co-expressed with 11 histone genes, *OsSET37* has 9 co-expressed histone genes, *OsSET7* has 7 ones, both *OsSET11* and *OsSET44* have 5 ones, while others has 1 or 2 co-expressed histone genes. *OsSET5* and other five genes (*OsSET20*, *OsSET22*, *OsSET30*, *OsSET32* and *OsSET41*) belong to class VB. *OsSET24/OsCLF* is a class I member. *OsSET37* and other two genes (*OsSET1*, *OsSET33*) belong to class III. *OsSET7* is a class IV gene. *OsSET11* is a class VA gene. This co-expression network indicated that *OsSET* protein may not only modify histone lysine, but also be involved in multiple histone site modification directly or indirectly. Because histone proteins are essential for the packaging





**Figure 6. Relative expression of eight *OsSET* genes in Zhonghua11.** root, roots at trefoil stage; stem, stems at heading stage; leaf, leaves at at trefoil stage; Fleaf, flag leaves at heading stage; P4, panicles at meiosis stage of young panicle development; S3, seeds of 3 days after pollination; EN, endosperm, the number followed it means the days after pollination.  
doi:10.1371/journal.pone.0065426.g006

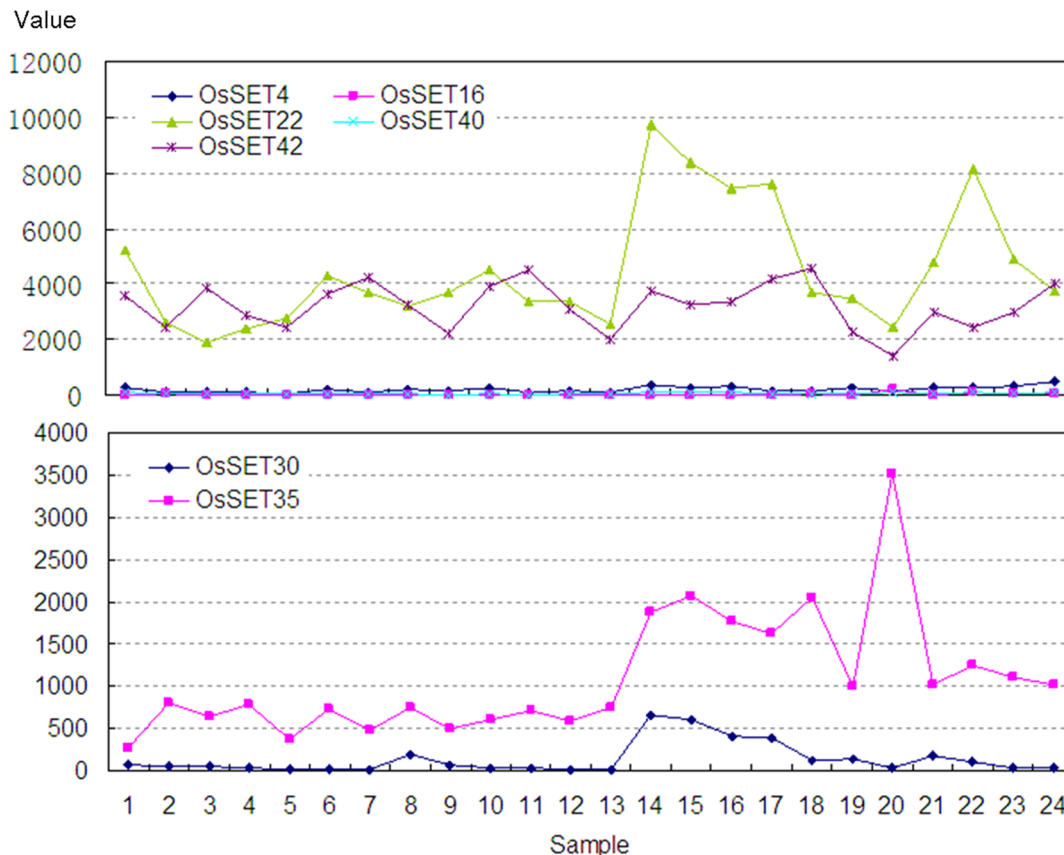
of newly synthesized DNA into chromosomes [70], we speculated that *OsSET* proteins may be relevant to cell cycle regulation.

Subsequently, we found that the 12 *OsSET* genes in **Figure 9A** were co-expressed with 19 cell cycle related genes, simultaneously (**Figure 10B**). These cell cycle related genes include 14 cyclin genes (*Cycs*, A-, B- and D- Type), and 5 cyclin-dependent kinase genes (CDKs, A-, B- and C-Type). It is noted that *OsSET5*, *OsSET24* and *OsSET37*, which are co-expressed with most of histone genes, are also tightly co-expressed with various kinds of *Cycs* and *CDKs*. While *OsSET22* and *OsSET32* are co-expressed with A- and D-type *Cycs*, the *OsSET33* is only co-expressed with D-type *Cycs* (*OsCycD2;2* and *OsCycD3;2*). *OsSET20* and *OsSET33* are co-expressed with *OsCDKB3;2* and *OsCDKC;3*. *OsSET7* is co-expressed with *OsCDKB;1* and *OsCDKB;2*. *OsSET1* and *OsSET32* are co-expressed with *CDKC;3*.

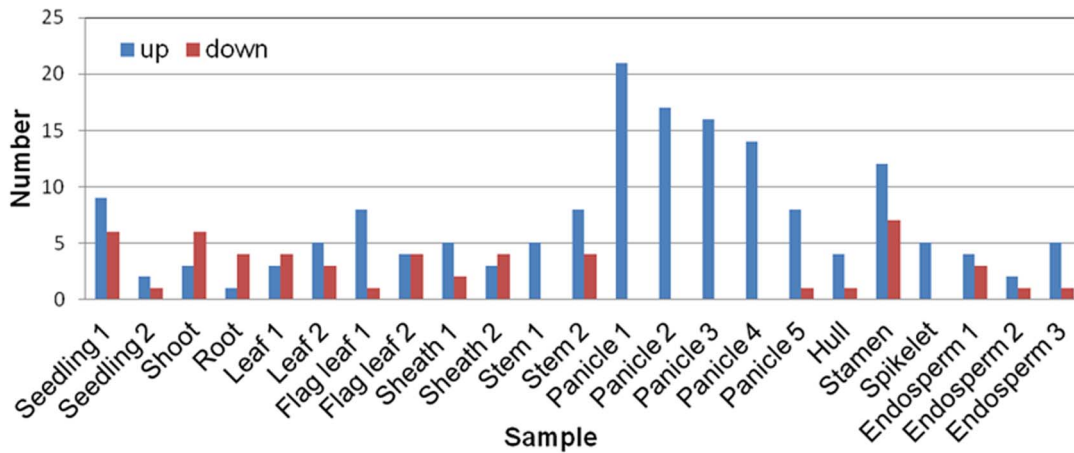
Usually, *Cycs* and *CDKs* act as complex at a precise time and drive the cell cycle progression by phosphorylating downstream target proteins. Cell cycle progression is critical for the maintenance of epigenetic marks and for allowing the daughter products to acquire a distinct epigenetic landscape [71]. Thus, the relevance might exist between SET, histone and cell cycle related protein. In human, it has been confirmed that Retinoblastoma (RB) can target H3K9 methylation to cyclin E promoter by SET-domain protein,

Suv39H1, resulting in heterochromatin protein 1 (HP1) binding and silencing [72]. In higher eukaryotes, cell cycle is mainly controlled by E2F transcription factors, which acts through a conserved RB-E2F/DP pathway (DP, related to the E2F family that can dimerize with E2F members)[64,73–75].

In *Arabidopsis*, CDKAs/*CycDs* complex can activate the expression of E2F/DP targets by phosphorylating RB and further releasing RB from its cooperator E2F/DP complex [76–78]. Genome-wide identification and expression analysis have found conserved *E2F*, *RB*, *Cycs* and *CDKs* in rice [79–81]. In our performances, two homologies of *E2F* (*OsDEL1* and *OsE2F2*) and two *RB* homologies (*OsRB1* and *OsRB2*) were found to be co-expressed with these 12 *OsSET* genes (**Figure 10C**). In consequence, these 12 *OsSET* proteins may be required for the RB-E2F/DP pathway during cell cycle progression. Although there is no direct evidence showing the connection between SET proteins and RB *in vivo* in plant, the CLF of *Arabidopsis* can bind the RB proteins both in maize and human [82], Shen et al proposed that HP1-HKMT-RB-E2F/DP complexes could repress E2F targets in plants [73]. In co-expression network of this study, the rice PcG gene *OsSET24/OsCLF*, a homology of *Arabidopsis CLF*, is co-expresses with *OsE2F2*, *OsDEL1*, *OsRB2* and *OsDP* tightly. Thus, it is possible that a similar RB-PRC2 complex



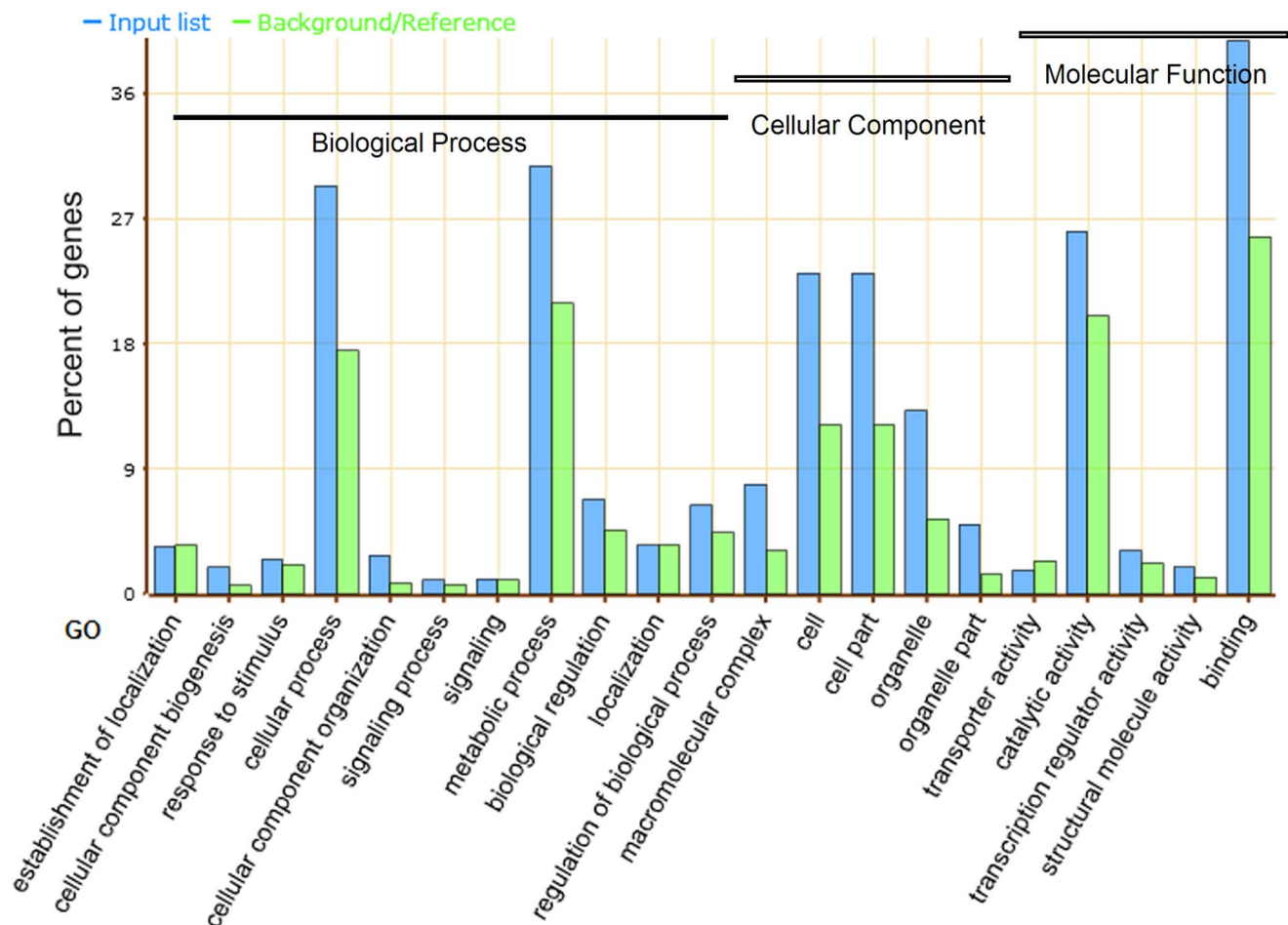
**Figure 7. Expression patterns of *OsSET* genes found in segmentally duplicated regions of the rice genome.** X-axis represents the developmental stages as given in the following table. Y-axis represents the raw expression values obtained from microarray.  
doi:10.1371/journal.pone.0065426.g007



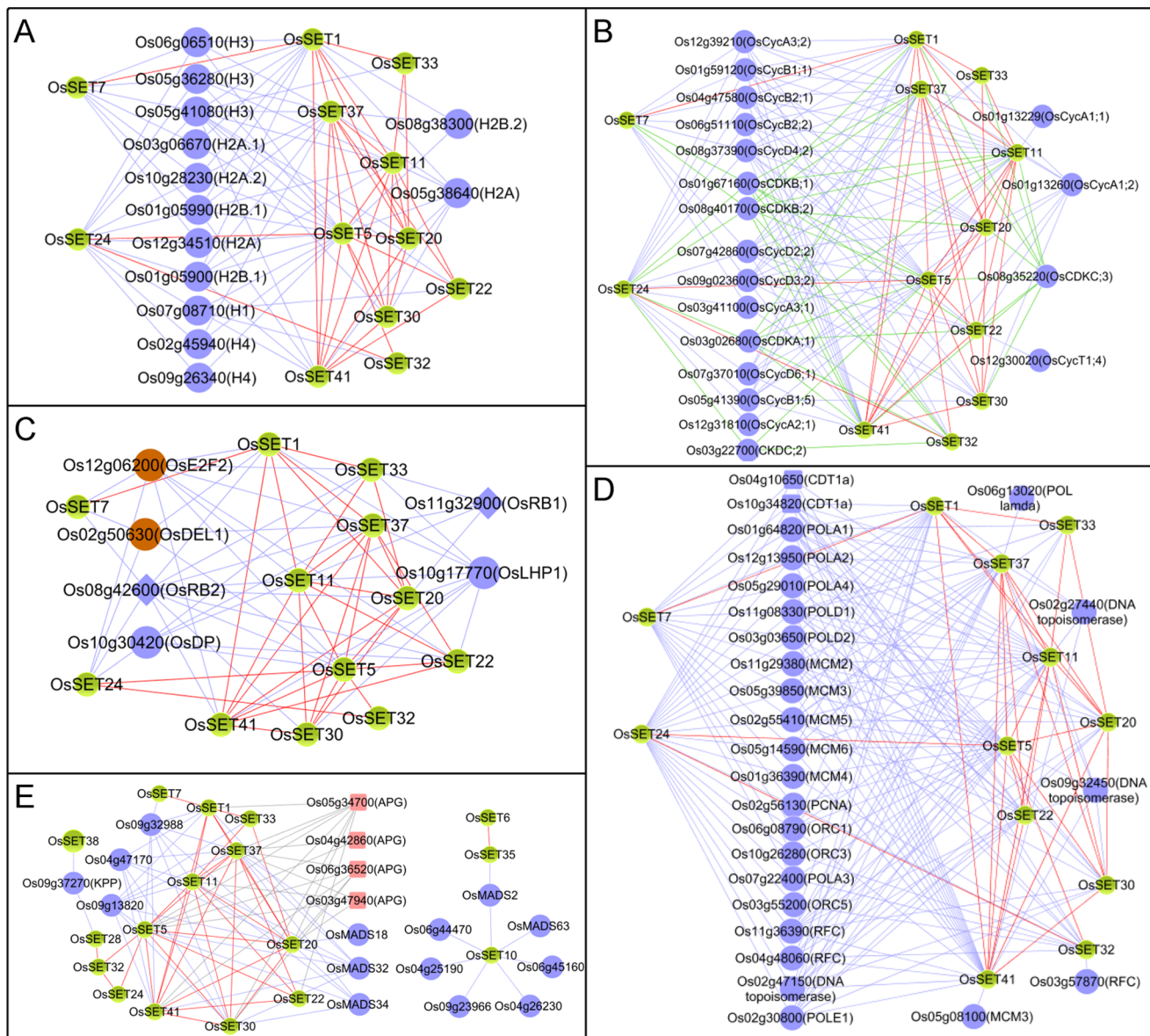
**Figure 8. Differential expressions of *OsSET* genes in different stages in Minghui 63 based on microarray analysis.** Differential expression genes have been taken p value less than 0.05 and fold change  $>2$  or  $<0.5$ . When fold change  $>2$ , regulation is up, and when fold change  $<0.5$ , regulation is down. doi:10.1371/journal.pone.0065426.g008

functions in the cell cycle regulation of rice. *OsSETs* are co-expressed with histone and cell cycle related genes simultaneously,

implying that more HKMTs may be associated with the regulation of cyclins via histone modifications. Therefore, our co-expression



**Figure 9. Enriched GO analysis of genes that co-expressed with *OsSETs*.** X-axis represents the GO annotation. Y-axis represents the percentage of GO annotation. Three categories of GO annotation are biological process, cellular component and molecular function. Input list represents genes analyzed, and the references/background represents all genes in agriGO database. doi:10.1371/journal.pone.0065426.g009



**Figure 10. Co-expression network of *OsSET* genes.** A. The co-expression relationship among *OsSET* genes and histone genes (H2A, H2B.1, H2B.2, H3 and H4). B. Network of *OsSET* genes and cell division related genes. C. Co-expression network of *OsSET* and RB-E2F/DB pathway genes. D. A co-expression network regarding the *OsSET* genes and DNA replication factors. E. A co-expression network concerning *OsSET* genes and genes involved in flower development.

doi:10.1371/journal.pone.0065426.g010

results might provide light in the relationship between these *OsSET* proteins and RB-E2F/DB complex.

### *OsSETs* May Function in Reproductive Development of Rice

Floral organ identity in plants are controlled by combinations of activities mediated by MADS box genes, some of which were identified in our co-expression analysis (Figure 10E). *OsMADS18* is co-expressed with *OsSET20* and *OsSET37*. *OsMADS32* is co-expressed with *OsSET1*, *OsSET5*, *OsSET11*, *OsSET20*, *OsSET30* and *OsSET41*. *OsMADS34* is a member in *SEPALLATA* (*SEP*) subfamily [83,84], which is co-expressed with *OsSET5*, *OsSET11*, *OsSET20*, *OsSET30* and *OsSET37*. Recent research revealed that *OsMADS18*, one of *APETALA1* (*API*)/*FRUITFULL* (*FUL*)-like

genes, was induced in the shoot apical meristem (SAM) during meristem phase transition, which acted co-ordinately in the meristem to specify the identity of the inflorescence meristem (IM) downstream of the florigen signal [85]. The expression domains of *OsMADS32* are mainly restricted to the marginal region of the palea and inner floral organs, showing its contribution on floral organ identity in rice [86]. *OsMADS34* plays a role in the early development of spikelet formation [84]. In the above co-expression network, *OsSET20* is co-expressed with three MADS box genes, while *OsSET1* is co-expressed with one MADS box gene. Besides, another five *SET* genes, *OsSET5*, *OsSET11*, *OsSET20*, *OsSET30* and *OsSET37*, are co-expressed with two MADS box genes. Therefore, it is possible that these *OsSETs* may take part in flowering transition and early floral development in rice.

In the co-expression network (**Figure 10E**), 8 *OsSET* genes are correlated with 4 APG-like protein genes, while *OsSET5* and *OsSET37* are co-expressed with four APG genes. Noticeably, the APG gene *Os05g34700* is co-expressed with 8 *OsSET* genes. The transcripts of these 8 *OsSET* genes accumulate in young panicles. It was reported that APG transcript was confined to anther during microspore development in *Brassica napus* flower buds [87]. It was also suggested that five APG genes in *Silene latifolia* were related to anther fertility, which were required for development of fertile pollen [88]. Hence, we might infer that these *OsSET* genes are involved in rice reproductive development though the regulation of *OsAPGs* during meiosis stage.

Except for the above 13 *OsSET* genes, the other genes were also characterized by co-expression analysis. For example, *OsSET6*, *OsSET10*, *OsSET16* and *OsSET35* have a high expression level in stamen, and their co-expression genes include transcription factor, binding protein, pollen allergen and so on (**Figure 5** and **Figure 10E**). Among them, *OsSET6*, *OsSET10* and *OsSET16* are co-expressed with a cyclin gene. Meanwhile, *OsSET10* is co-expressed with five pollen allergen genes and 2 MADS-box genes (*OsMADS2* and *OsMADS63*). *OsSET35* is co-expressed with *OsMADS2*. The previous report showed that *OsMADS2* transcript was first observed in the region where stamen primordia were formed, and then appeared in the lodicule primordia as well as the stamen primordia [89,90]. *OsMADS63* is the homolog of *Arabidopsis AGL66*, which encodes a MIKC\*-type DNA binding factor as heterodimer affecting pollen viability, germination, and pollen tube growth [91]. Our analysis suggests that these four *OsSET* members might affect on the development of male gametophyte.

## Conclusions

In conclusion, 43 *OsSET* genes can be classified into five classes as supported by phylogeny and conserved domains organization. Phylogenetic and structural analysis indicated that the domains beyond SET domain were significant for their specific functions. The expression analysis revealed that *OsSET* genes might participate in various molecular pathways both in vegetative and reproductive development. GO enrichment analysis showed that the above *OsSET* genes and their co-expressed genes seemed to particularly affect the same or similar GO categories. Promoter *cis*-elements identification and the combined analysis of expression correlation suggested that most of *OsSET* genes might be cell cycle regulated and were associated in the cell cycle progression by histone modifications via E2F. Moreover, we found that some MADS-box and APG proteins may be associated with *OsSET* on the regulation of cell differentiation and reproductive redevelopment in rice.

Although the studies of plant *SET* genes have received much progress, only a minority of *OsSET* genes has been verified in rice. The challenges still exist for the large number of genes in this family. It is a time-consuming process to molecularly characterize the functions and mechanisms of all *OsSET* genes in traditional approach. Thus our studies would provide valuable data for inferring the putative functions and pathways of the *OsSET* genes.

## Materials and Methods

### Identification of *OsSET* Members in Rice

Hidden Markov Model (HMM) profile of SET domain (PF00856) downloaded from Pfam (<http://pfam.sanger.ac.uk/>) were employed to identify the putative *OsSET* genes in rice (*Oryza sativa*) [92]. The BlastP search was carried out using the HMM profile on website of MSU RGAP (<http://rice.plantbiology.msu.edu/>)

and KOME (<http://cdna01.dna.affrc.go.jp/cDNA/>), followed by removal of redundant sequences from the two databases. Meanwhile, the keyword “SET” was also performed in these databases. Additionally, the Pfam and SMART database (<http://smart.embl-heidelberg.de/smart/batch.pl>) were used to confirm and make classification of each predicted SET protein.

### Chromosomal Localization and Gene Duplication

*OsSET* genes were mapped on rice chromosomes according to their positions available in MSU RGAP. The distribution of *OsSET* genes was drawn by MapInspect ([http://www.plantbreeding.wur.nl/UK/software\\_mapinspect.html](http://www.plantbreeding.wur.nl/UK/software_mapinspect.html)). The duplicated genes were elucidated from the segmental genome duplication of rice ([http://rice.plantbiology.msu.edu/segmental\\_dup/500\\_kb/segdup\\_500\\_kb.shtml](http://rice.plantbiology.msu.edu/segmental_dup/500_kb/segdup_500_kb.shtml)), with the maximal length distance permitted between collinear gene pairs of 500 kb [93]. Tandem duplicates were defined as genes separated by five or fewer genes. The distances between these genes on the chromosomes were calculated and the percentage of sequence similarities between the proteins encoded by these genes were determined by MegAlign software 4.0 (MEGA4) [94].

### Phylogenetic Analysis of *OsSET* Family

The protein sequences of *OsSET* family and *Arabidopsis* SET domain group (SDG) were aligned using ClustalX (version 2.0) program. An un-rooted neighbor-joining [93] phylogenetic tree was constructed in ClustalX based on the full sequences of the proteins with default parameters from rice and *Arabidopsis*. Bootstrap analysis was performed using 1,000 replicates. The phylogenetic tree thus obtained was viewed using MEGA 4 software.

### Structural and Sequence Analysis of *OsSET* Genes

Information in gene structures, transcripts, full-length cDNA, BAC accessions for each gene and characteristics of corresponding proteins were procured from MSU RGAP, KOME and GRAMENE. Protein sequences of putative *OsSET* members collected from the MSU RGAP and KOME were analyzed by EXPASY PROTOPARAM tool (<http://www.expasy.org/tools/protparam.html>). Information in the number of amino acids, molecular weight, theoretical isoelectric point (pI), amino acid composition, and instability index (instability index of >40 was considered as unstable) were obtained [95]. The conserved domains of the *OsSET* protein in rice were determined by PFam program.

Protein sequences were analyzed in the MEME program (<http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>) to confirm the conserved motifs. The MEME program was employed using the following parameters: number of repetitions-any, maximum number of motifs-200, optimum motif width set to >2 and <200.

Promoter sequences (−2000 bps) of *OsSET* family genes were obtained from the Rice Annotation Project (RAP) database (<http://rapdb.dna.affrc.go.jp/tools/dump>). The *cis*-elements of promoters were identified using the PLACE Web Signal Scan-PLACE (<http://www.dna.affrc.go.jp/PLACE/signalup.html>) [96,97].

### Genome-wide Expression analysis of *OsSET* Family

Expression profile data of *OsSET* gene family in 24 tissues for Minghui 63 were extracted from CREP database (<http://crep.ncpgr.cn>, Microarray data sets: GSE19024) [98]. Expression values of each gene were logarithm in Microsoft excel 2007 and cluster analyses were performed using J-express 2011 with

euclidean distances and hierarchical cluster method of “complete linkage”. The average signal value of biological replicates for each sample was used for analysis. When more than one probeset was available for one gene, the higher signal value of the probesets was used for analysis. Expression level in each of the tissues was compared against the expression in seed using a student-*t* test. The genes up- or down-regulated by more than two-fold and with *p* values <0.05 were considered to be differentially expressed. The average expression of more than two biological replicates for each sample was used for analysis.

### Identification of Correlated Genes and Network Construction

The co-expression data were downloaded from the CREP database. The standard deviations for the expression level of each *OsSET* gene in 24 tissues were calculated. First, we ranked the genes according to the correlation coefficients and screen ones that were greater than 0.75 positively correlated with *OsSET* gene expression. Then the Pearson correlation coefficient (PCC) and the student-*t* test of candidate genes that we interested in were recalculated with R project (version 2.14.1). As the permutation test done by Ouyang *et al.*, PCC  $\geq 0.7$  were significant ( $\alpha = 0.05$ ), We mapped the correlated genes (at a more strict level, PCC  $\geq 0.75$ , *p* value  $\leq 0.05$ ) to the network with Cytoscape v2.8.1 [69,99]. GO enrichment was performed by Singular Enrichment Analysis (SEA) tool in agriGO (<http://bioinfo.cau.edu.cn/agriGO/index.php>) with default parameters using the rice MSU6.1 genome annotation as background [100]. Statistical significance was determined using Fisher’s exact test and Yekutieli multi-test adjustment.

### Real-time PCR Analysis of Representative Genes in *OsSET* Family

Primers designed for the RT-PCR analysis were listed in Table S6. Samples were ground in liquid nitrogen using a mortar and pestle. Total RNA (4  $\mu$ g) was isolated using a RNAiso (Takara) and treated with RNase-free DNase I (Takara) for 15 min to eliminate possible contaminating DNA. First strand cDNA was then reverse transcribed from total RNA with an oligo(dT)18 primer in a 20  $\mu$ l reaction (diluted to 40  $\mu$ l before use) using an M-MLV Reverse Transcriptase (Promega) according to the manufacturer’s instructions. Real-time quantitative PCR was carried out on ABI StepOne™ Real-time PCR instrument (Applied Biosystems), containing 5  $\mu$ l of 2 $\times$  SYBR Premix EX Taq (Takara), 0.5  $\mu$ l of Rox Reference Dye II (Takara), 0.5  $\mu$ l of the cDNA sample, 2  $\mu$ M of each gene-specific primer, in a final

volume of 10  $\mu$ l. The reactions were carried out according to the following temperature profile: 95°C for 30 seconds, 40 cycles of 95°C for 5 seconds, and 60°C for 34 seconds.

### Plant Materials and Growth Conditions

A *japonica* rice variety of Zhonghua11 was used in this study. Plants were grown at long day under natural light.

### Supporting Information

**Table S1** A list of 43 *OsSET* genes identified in rice and their sequences and protein characteristics. (XLS)

**Table S2** *OsSET* genes that localized on duplicated segments of the rice genome. (XLS)

**Table S3** *Cis*-elements analysis of *OsSET* gene promoters. (XLS)

**Table S4** Average signal values in 24 samples of 40 *OsSET* genes in Minghui 63. (XLS)

**Table S5** Table S5a. Results of differential expression analysis using seed as reference (Minghui63); Table S5b. Results of differential expression analysis in 7 d-old seedlings subjected to three phytohormone (NAA, GA3 and KT) treatments and plumule or radicle with light/dark regulation in Minghui63. Differential expression genes have been taken *p* value less than 0.05 and fold change >2 or <0.5. When fold change >2, regulation is up, and when fold change <0.5, regulation is down. (XLS)

**Table S6** A list of co-expression genes mentioned in network construction with *p* value less than 0.05. (XLS)

**Table S7** Primers used in the research. (XLS)

### Acknowledgments

We would like to thank Prof. Chungun Hu for helpful advice.

### Author Contributions

Conceived and designed the experiments: ZL YO JY. Performed the experiments: ZL XH. Analyzed the data: ZL XH. Contributed reagents/materials/analysis tools: JY. Wrote the paper: ZL.

### References

- Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, et al. (2001) The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res* 29: 4319–4333.
- Gendler K, Paulsen T, Napoli C (2008) ChromDB: the chromatin database. *Nucleic Acids Res* 36: 298–302.
- Jenuwein T, Laible G, Dorn R, Reuter G (1998) SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol Life Sci* 54: 80–93.
- Marmorstein R (2003) Structure of SET domain proteins: a new twist on histone methylation. *Trends Biochem Sci* 28: 59–62.
- Schneider R, Bannister AJ, Kouzarides T (2002) Unsafe SETs: histone lysine methyltransferases and cancer. *Trends Biochem Sci* 27: 396–402.
- Yu Y, Bu ZY, Shen WH, Dong AW (2009) An update on histone lysine methylation in plants. *Prog Nat Sci* 19: 407–413.
- Triebel RC, Flynn EM, Houtz RL, Hurley JH (2003) Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSM1. *Nat Struct Biol* 10: 545–552.
- Hu P, Wang S, Zhang Y (2008) How do SET-domain protein lysine methyltransferases achieve the methylation state specificity? Revisited by ab initio QM/MM molecular dynamics simulations. *J Am Chem Soc* 130: 3806–3813.
- Springer NM, Napoli CA, Selinger DA, Pandey R, Cone KC, et al. (2003) Comparative analysis of SET domain proteins in maize and Arabidopsis reveals multiple duplications preceding the divergence of monocots and dicots. *Plant Physiol* 132: 907–925.
- Ng DW, Wang T, Chandrasekharan MB, Aramayo R, Kertbundit S, et al. (2007) Plant SET domain-containing proteins: structure, function and regulation. *Biochim Biophys Acta* 1769: 316–329.
- Pontvianne F, Blevins T, Pikaard CS (2010) Arabidopsis Histone Lysine Methyltransferases. *Advances in botanical research*. London: Academic Press Ltd-Elsevier Science Ltd. 1–22.
- Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, et al. (1997) A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. *Nature* 386: 44–51.
- Kim GT, Tsukaya H, Uchimiya H (1998) The CURLY LEAF gene controls both division and elongation of cells during the expansion of the leaf blade in Arabidopsis thaliana. *Planta* 206: 175–183.

14. Köhler C, Hennig L (2010) Regulation of cell identity by plant Polycomb and trithorax group proteins. *Curr Opin Genet Dev* 20: 541–547.
15. Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, et al. (2006) Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *The EMBO journal* 25: 4638–4649.
16. Jiang D, Wang Y, He Y (2008) Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components. *PLoS One* 3: e3404.
17. Schatowski N, Stahl Y, Hohenstatt ML, Goodrich J, Schubert D (2010) The CURLY LEAF Interacting Protein BLISTER Controls Expression of Polycomb-Group Target Genes and Cellular Differentiation of Arabidopsis thaliana. *Plant Cell* 22: 2291–2305.
18. Xu L, Shen WH (2008) Polycomb Silencing of KNOX Genes Confines Shoot Stem Cell Niches in Arabidopsis. *Curr Biol* 18: 1966–1971.
19. Grossniklaus U, Jean-Philippe Vielle-Calzada, Marilu A. Hoepfner WBG (1998) Maternal Control of Embryogenesis by MEDEA, a Polycomb Group Gene in Arabidopsis. *Science*: 446–450.
20. Baroux U, Gagliardini V, Page DR, Grossniklaus U (2006) Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis. *Genes Dev* 20: 1081–1086.
21. Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999) Imprinting of the MEDEA Polycomb Gene in the Arabidopsis Endosperm. *Plant Cell* 11: 1945–1952.
22. Köhler C, Page DR, Gagliardini V, Grossniklaus U (2005) The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat Genet* 37: 28–30.
23. Eriova A, Brownfield L, Exner V, Rosa M, Twell D, et al. (2009) Imprinting of the Polycomb Group Gene MEDEA Serves as a Ploidy Sensor in Arabidopsis. *PLoS Genet* 5: 14.
24. Pontvianne F, Blevins T, Chandrasekhara C, Feng W, Stroud H, et al. (2012) Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. *Genes Dev* 26: 945–957.
25. Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, et al. (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. *Chromosoma* 112: 308–315.
26. Ebbs ML, Bartec L, Bender J (2005) H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. *Molecular and Cellular Biology* 25: 10507–10515.
27. Ebbs ML, Bender J (2006) Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *Plant Cell* 18: 1166–1176.
28. Alvarez-Venegas R, Pien S, Sadler M, Witmer X, Grossniklaus U, et al. (2003) ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. *Curr Biol* 13: 627–637.
29. Saleh A, Al-Abdallat A, Ndamukong I, Alvarez-Venegas R, Avramova Z (2007) The Arabidopsis homologs of trithorax (ATX1) and enhancer of zeste (CLF) establish 'bivalent chromatin marks' at the silent AGAMOUS locus. *Nucleic Acids Res* 35: 6290–6296.
30. Saleh A, Alvarez-Venegas R, Yilmaz M, Oahn L, Hou G, et al. (2008) The highly similar Arabidopsis homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions. *Plant Cell* 20: 568–579.
31. Liang YK, Wang Y, Zhang Y, Li SG, Lu XC, et al. (2003) OsSET1, a novel SET-domain-containing gene from rice. *J Exp Bot* 54: 1995–1996.
32. Luo M, Platten D, Chaudhury A, Peacock WJ, Dennis ES (2009) Expression, imprinting, and evolution of rice homologs of the polycomb group genes. *Molecular Plant* 2: 711–723.
33. Thakur JK, Malik MR, Bhatt V, Reddy MK, Sopory SK, et al. (2003) A POLYCOMB group gene of rice (*Oryza sativa* L. subspecies indica), OsE2I, codes for a nuclear-localized protein expressed preferentially in young seedlings and during reproductive development. *Gene* 314: 1–13.
34. Qin F-J, Sun Q-W, Huang L-M, Chen X-S, Zhou D-X (2010) Rice SUVH Histone Methyltransferase Genes Display Specific Functions in Chromatin Modification and Retrotransposon Repression. *Mol Plant* 3: 1674–2052.
35. Ding Y, Wang X, Su L, Zhai JX, Cao SY, et al. (2007) SDG714, a histone H3K9 methyltransferase, is involved in Tos17 DNA methylation and transposition in rice. *Plant Cell* 19: 9–22.
36. Sun C, Fang J, Zhao T, Xu B, Zhang F, et al. (2012) The Histone Methyltransferase SDG724 Mediates H3K36me2/3 Deposition at MADS50 and RFT1 and Promotes Flowering in Rice. *Plant Cell* 24: 3235–3247.
37. Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC plant biology* 4: 10.
38. Horton JR, Elgar SJ, Khan SI, Zhang X, Wade PA, et al. (2007) Structure of the SANT domain from the Xenopus chromatin remodeling factor ISWI. *Proteins* 67: 1198–1202.
39. Boyer LA, Latak RR, Peterson CL (2004) The SANT domain: a unique histone-tail-binding module? *Nature reviews Molecular cell biology* 5: 158–163.
40. Valencia-Morales MdP, Camas-Reyes JA, Cabrera-Ponce JL, Alvarez-Venegas R (2012) The Arabidopsis thaliana SET-domain-containing protein ASHH1/SDG26 interacts with itself and with distinct histone lysine methyltransferases. *J Plant Res* 125: 679–692.
41. Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, et al. (2009) The ASH1 HOMOLOG 2 (ASHH2) Histone H3 Methyltransferase Is Required for Ovule and Anther Development in Arabidopsis. *PLoS One* 4.
42. Thorstensen T, Grini PE, Mercy IS, Alm V, Erdal S, et al. (2008) The Arabidopsis SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS). *Plant Mol Biol* 66: 47–59.
43. Lee JS, Shilatifard A (2007) A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat Res* 618: 130–134.
44. Avramova Z (2009) Evolution and pleiotropy of TRITHORAX function in Arabidopsis. *Int J Dev Biol* 53: 371–381.
45. Stec I, Nagl SB, van Ommen GJ, den Dunnen JT (2000) The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS Lett* 473: 1–5.
46. Wang Y, Reddy B, Thompson J, Wang H, Noma K, et al. (2009) Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Mol Cell* 33: 428–437.
47. Wu H, Zeng H, Lam R, Tempel W, Amaya MF, et al. (2011) Structural and histone binding ability characterizations of human PWWP domains. *PLoS One* 6: e18919.
48. Schuettengruber B, Martinez AM, Iovino N, Cavalli G (2011) Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* 12: 799–814.
49. Papp B, Muller J (2006) Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev* 20: 2041–2054.
50. Aasland R, Gibson TJ, Stewart AF (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* 20: 56–59.
51. Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, et al. (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 442: 100–103.
52. van Ingen H, van Schaik FM, Wienk H, Ballering J, Rehmann H, et al. (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure* 16: 1245–1256.
53. Spiliotopoulos D, Spitaleri A, Musco G (2012) Exploring PHD Fingers and H3K4me0 Interactions with Molecular Dynamics Simulations and Binding Free Energy Calculations: AIRE-PHD1, a Comparative Study. *PLoS One* 7: e46902.
54. Raynaud C, Sozzani R, Glab N, Domenichini S, Perennes C, et al. (2006) Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. *Plant J* 47: 395–407.
55. Veiseth SV, Rahman MA, Yap KL, Fischer A, Egge-Jacobsen W, et al. (2011) The SUVH4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in Arabidopsis. *PLoS Genet* 7: e1001325.
56. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450: 908–912.
57. Unoki M, Nishidate T, Nakamura Y (2004) ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene* 23: 7601–7610.
58. Achour M, Jacq X, Ronde P, Alhosin M, Charlot C, et al. (2008) The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene* 27: 2187–2197.
59. Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, et al. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Current biology* : CB 17: 379–384.
60. Ding B, Zhu Y, Gao J, Yu Y, Cao KM, et al. (2007) Molecular characterization of three rice SET-domain proteins. *Plant Sci* 172: 1072–1078.
61. Caro E, Stroud H, Greenberg MV, Bernatavichute YV, Feng S, et al. (2012) The SET-Domain Protein SUVH5 Mediates H3K9me2 Deposition and Silencing at Stimulus Response Genes in a DNA Methylation-Independent Manner. *PLoS Genet* 8: e1002995.
62. Helin K (1998) Regulation of cell proliferation by the E2F transcription factors. *Curr Opin Genet Dev* 8: 28–35.
63. Vandepoel K, Vlieghe K, Florquin K, Hennig L, Beemster GT, et al. (2005) Genome-wide identification of potential plant E2F target genes. *Plant Physiol* 139: 316–328.
64. Kosugi S, Ohashi Y (2002) E2F sites that can interact with E2F proteins cloned from rice are required for meristematic tissue-specific expression of rice and tobacco proliferating cell nuclear antigen promoters. *The Plant journal : for cell and molecular biology* 29: 45–59.
65. de Jager SM, Menges M, Bauer UM, Murra JA (2001) Arabidopsis E2F1 binds a sequence present in the promoter of S-phase-regulated gene AtCDC6 and is a member of a multigene family with differential activities. *Plant Mol Biol* 47: 555–568.
66. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302: 249–255.
67. Fu F-F, Xue H-W (2010) Coexpression Analysis Identifies Rice Starch Regulator1, a Rice AP2/EREBP Family Transcription Factor, as a Novel Rice Starch Biosynthesis Regulator. *Plant Physiol* 154: 927–938.
68. Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol* 48: 381–390.
69. Ouyang Y, Huang X, Lu Z, Yao J (2012) Genomic survey, expression profile and co-expression network analysis of OsWD40 family in rice. *BMC genomics* 13: 100.

70. Rattray AMJ, Mueller B (2012) The control of histone gene expression. *Biochem Soc Trans* 40: 880–885.
71. Sanchez MdP, Caro E, Desvoyes B, Ramirez-Parra E, Gutierrez C (2008) Chromatin dynamics during the plant cell cycle. *Seminars in cell & developmental biology* 19: 537–546.
72. Nielsen SJ, Schneider R, Bauer UM, Bannister AJ, Morrison A, et al. (2001) Rb targets histone H3 methylation and HP1 to promoters. *Nature* 412: 561–565.
73. Shen WH (2002) The plant E2F-Rb pathway and epigenetic control. *Trends Plant Sci* 7: 505–511.
74. van den Heuvel S, Dyson NJ (2008) Conserved functions of the pRB and E2F families. *Nature reviews Molecular cell biology* 9: 713–724.
75. Dimova DK, Stevaux O, Frolov MV, Dyson NJ (2003) Cell cycle-dependent and cell cycle-independent control of transcription by the *Drosophila* E2F/RB pathway. *Genes Dev* 17: 2308–2320.
76. Qi R, John PC (2007) Expression of genomic AtCYCD2;1 in *Arabidopsis* induces cell division at smaller cell sizes: implications for the control of plant growth. *Plant Physiol* 144: 1587–1597.
77. Zhao XA, Harashima H, Dissmeyer N, Pusch S, Weimer AK, et al. (2012) A General G1/S-Phase Cell-Cycle Control Module in the Flowering Plant *Arabidopsis thaliana*. *PLoS Genet* 8.
78. Nowack MK, Harashima H, Dissmeyer N, Zhao XA, Bouyer D, et al. (2012) Genetic Framework of Cyclin-Dependent Kinase Function in *Arabidopsis*. *Developmental Cell* 22: 1030–1040.
79. Guo J, Song J, Wang F, Zhang XS (2007) Genome-wide identification and expression analysis of rice cell cycle genes. *Plant Mol Biol* 64: 349–360.
80. La H, Li J, Ji Z, Cheng Y, Li X, et al. (2006) Genome-wide analysis of cyclin family in rice (*Oryza Sativa* L.). *Molecular genetics and genomics* : MGG 275: 374–386.
81. Shultz RW, Tatineni VM, Hanley-Bowdoin L, Thompson WF (2007) Genome-wide analysis of the core DNA replication machinery in the higher plants *Arabidopsis* and rice. *Plant Physiol* 144: 1697–1714.
82. Williams L, Grafi G (2000) The retinoblastoma protein - a bridge to heterochromatin. *Trends Plant Sci* 5: 239–240.
83. Gao X, Liang W, Yin C, Ji S, Wang H, et al. (2010) The SEPALLATA-like gene OsMADS34 is required for rice inflorescence and spikelet development. *Plant Physiol* 153: 728–740.
84. Kobayashi K, Mackawa M, Miyao A, Hirochika H, Kyojuka J (2010) PANICLE PHYTOMER2 (PAP2), encoding a SEPALLATA subfamily MADS-box protein, positively controls spikelet meristem identity in rice. *Plant & cell physiology* 51: 47–57.
85. Kobayashi K, Yasuno N, Sato Y, Yoda M, Yamazaki R, et al. (2012) Inflorescence meristem identity in rice is specified by overlapping functions of three AP1/FUL-like MADS box genes and PAP2, a SEPALLATA MADS box gene. *Plant Cell* 24: 1848–1859.
86. Sang X, Li Y, Luo Z, Ren D, Fang L, et al. (2012) CHIMERIC FLORAL ORGANS1, encoding a monocot-specific MADS box protein, regulates floral organ identity in rice. *Plant Physiol* 160: 788–807.
87. Roberts MR, Foster GD, Blundell RP, Robinson SW, Kumar A, et al. (1993) Gametophytic and sporophytic expression of an anther-specific *Arabidopsis thaliana* gene. *Plant J* 3: 111–120.
88. Ageez A, Kazama Y, Sugiyama R, Kawano S (2005) Male-fertility genes expressed in male flower buds of *Silene latifolia* include homologs of anther-specific genes. *Genes Genet Syst* 80: 403–413.
89. Yao SG, Ohmori S, Kimizu M, Yoshida H (2008) Unequal genetic redundancy of rice PISTILLATA orthologs, OsMADS2 and OsMADS4, in lodicule and stamen development. *Plant & cell physiology* 49: 853–857.
90. Kyojuka J, Kobayashi T, Morita M, Shimamoto K (2000) Spatially and temporally regulated expression of rice MADS box genes with similarity to *Arabidopsis* class A, B and C genes. *Plant & cell physiology* 41: 710–718.
91. Adamczyk BJ, Fernandez DE (2009) MIKC\* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. *Plant Physiol* 149: 1713–1723.
92. Ouyang Y, Chen J, Xie W, Wang L, Zhang Q (2009) Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice. *Plant Mol Biol* 70: 341–357.
93. Haas BJ, Delcher AL, Wortman JR, Salzberg SL (2004) DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20: 3643–3646.
94. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
95. Guruprasad K, Reddy BV, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng* 4: 155–161.
96. Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27: 297–300.
97. Prestridge DS (1991) SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. *Computer applications in the biosciences* : CABIOS 7: 203–206.
98. Wang L, Xie WB, Chen Y, Tang WJ, Yang JY, et al. (2010) A dynamic gene expression atlas covering the entire life cycle of rice. *Plant J* 61: 752–766.
99. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504.
100. Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38: W64–70.