

Global involvement of lysine crotonylation in protein modification and transcription regulation in rice

Shuai Liu^{1#}, Chao Xue^{1#}, Yuan Fang², Gang Chen¹, Xiaojun Peng³, Yong Zhou¹, Chen Chen¹,
Guanqing Liu¹, Minghong Gu¹, Kai Wang⁴, Wenli Zhang², Yufeng Wu^{2*}, Zhiyun Gong^{1*}

1: Jiangsu Key Laboratory of Crop Genetics and Physiology/Co-Innovation Center for Modern Production Technology of Grain Crops, Key Laboratory of Plant Functional Genomics of the Ministry of Education, Yangzhou University, Yangzhou 225009, China

2: The State Key Laboratory of Crop Genetics and Germplasm Enhancement, Bioinformatics Center, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing 210095, China

3: Jingjie PTM BioLab (Hangzhou) Co. Ltd., Hangzhou 310018, China

4: Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, China

#: These authors contributed equally to this work.

*Corresponding authors E-mail address: zygong@yzu.edu.cn or yfwu@njau.edu.cn

SUPPORTING INFORMATION

Supplemental_Fig_S1. Quality control validation of MS data. (A) Mass error distribution of all identified peptides. (B) Distribution of crotonylated peptides based on their length. (C) Distribution of crotonylated proteins based on their number of crotonylated peptides.

Supplemental_Fig_S2. Representative MS/MS spectra of validated crotonylated peptides in rice seedling. (A) Crotonylpeptide TIMP-(crotonyl)K-DIQLA with cronylation site at K122 in histone H3. (B) Crotonylpeptide IQGIT-(crotonyl)K-PAIR with cronylation site at K31 in histone H4. (C) Crotonylpeptide LEV-(crotonyl)K-EIAEIM with cronylation site at K817 in chaperone protein C1pC2. (D) Crotonylpeptide LAEEG-(crotonyl)K-VAIR with cronylation site at K205 in ribosome-recycling factor.

Supplemental_Fig_S3. Functional annotation of lysine crotonylome. Representative GO annotations of Kcr sites for (A) biological process, (B) molecular function, and (C) cellular component, and (D) KEGG pathway analysis of lysine-crotonylated proteins.

Supplemental_Fig_S4. WB analysis in green and albino seedlings. (A) The morphology of green seedlings and albino seedlings derived from anther culture of *Oryza sativa* variety “Nipponbare”. (B) WB analysis was performed using anti-Kcr antibody in green seedlings and albino seedlings derived from anther culture.

Supplemental_Fig_S5. Protein-protein interaction network. The balls represent crotonylated proteins. The size of the balls represents the numbers of Kcr modification in each figure.

Supplemental_Fig_S6. IF and WB analysis were performed using anti-H3K14cr antibody. (A) The specificity antibody H3K14cr (green) was detected in two-week-old rice root by immunofluorescence, and nuclei was stained with DAPI (red). Scale bars: 5 μ m. (B) Western blotting analysis in rice seedlings whole protein and histone.

Supplemental_Fig_S7. Distribution of H3K14cr density around differentially expressed

genes. (A) Genome-wide distribution of histone H3K14cr in the rice genome. **(B)** Distribution of H3K14cr density around differentially expressed genes. The H3K14cr modification was calculated by the number of reads per kilobase of the mapped genomic region. The arrow indicates the direction of transcription from transcription start site (TSS). The rice genes were divided into five categories based the expression level from top 20% to bottom 20% (0-20%).

Supplemental_Fig_S8. Expression comparisons of genes associated with different combinations of histone modifications. The non-TE gene expression values (FPKM) of each combination were indicated by box plots. All: all rice genes. Kcr: genes with Kcr modification but not only. Kcr+K9ac: genes with both Kcr and H3K9ac. Kcr only: Genes only with Kcr. The rest may be deduced by analogy. The *** indicated the significant difference between two combinations ($p < 2.2e-16$, Kolmogorov-Smirnov test).

Supplemental_Fig_S9. The distribution of histone Kcr in 12 chromosomes of rice.

Supplemental_Fig_S10. The phylogenic tree of p300/CBP homologous genes in rice, maize, wheat and Arabidopsis genome.

Supplemental_Table_S1. Protein annotation summary

Supplemental_Table_S2. Gene ontology annotation detail.

Supplemental_Table_S3. Protein-protein interaction network.

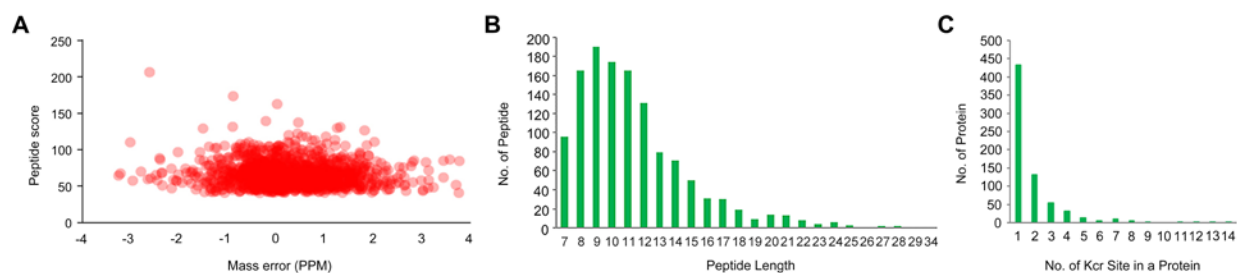
Supplemental_Table_S4. Seven Kcr sites identified in rice histones.

Supplemental_Table_S5. Summary of ChIP-seq data.

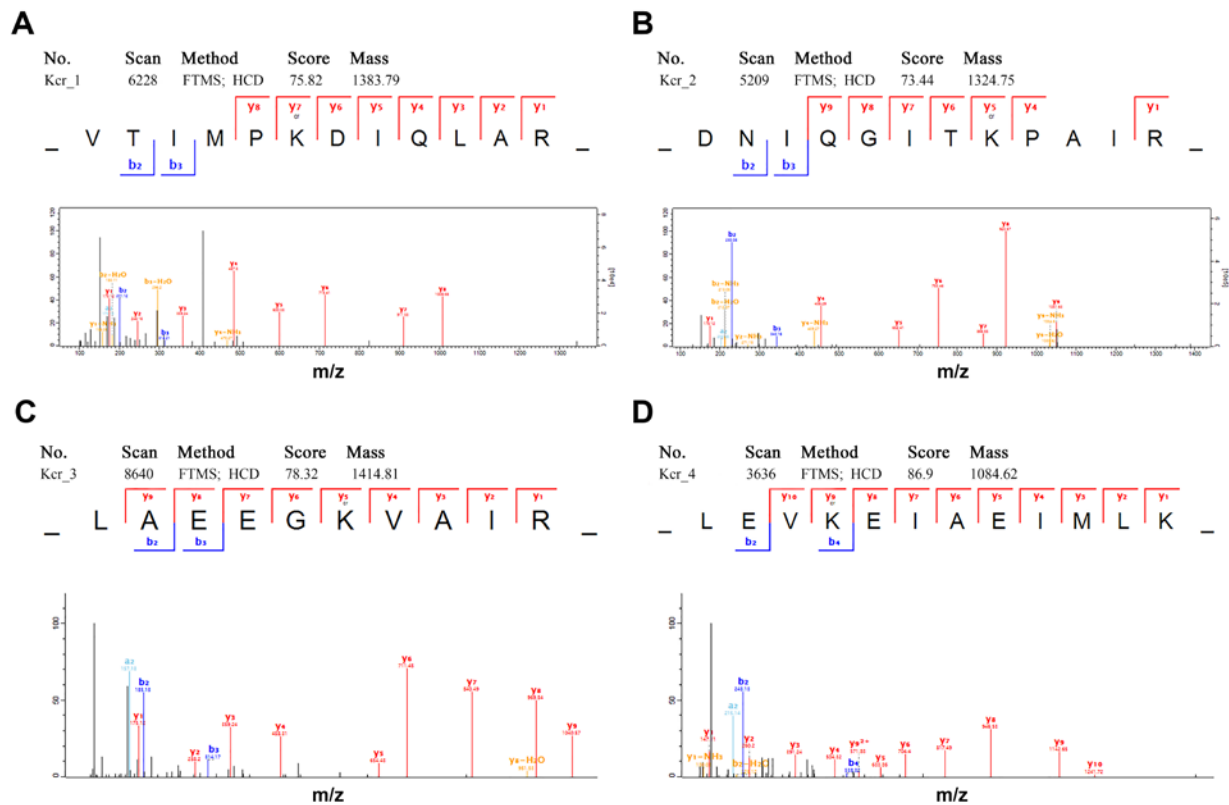
Supplemental_Table_S6. The lists of all enriched regions with Kcr and H3K14cr.

Supplemental_Table_S7. Confirmation of peak sites and non-peak sites by quantitative PCR.

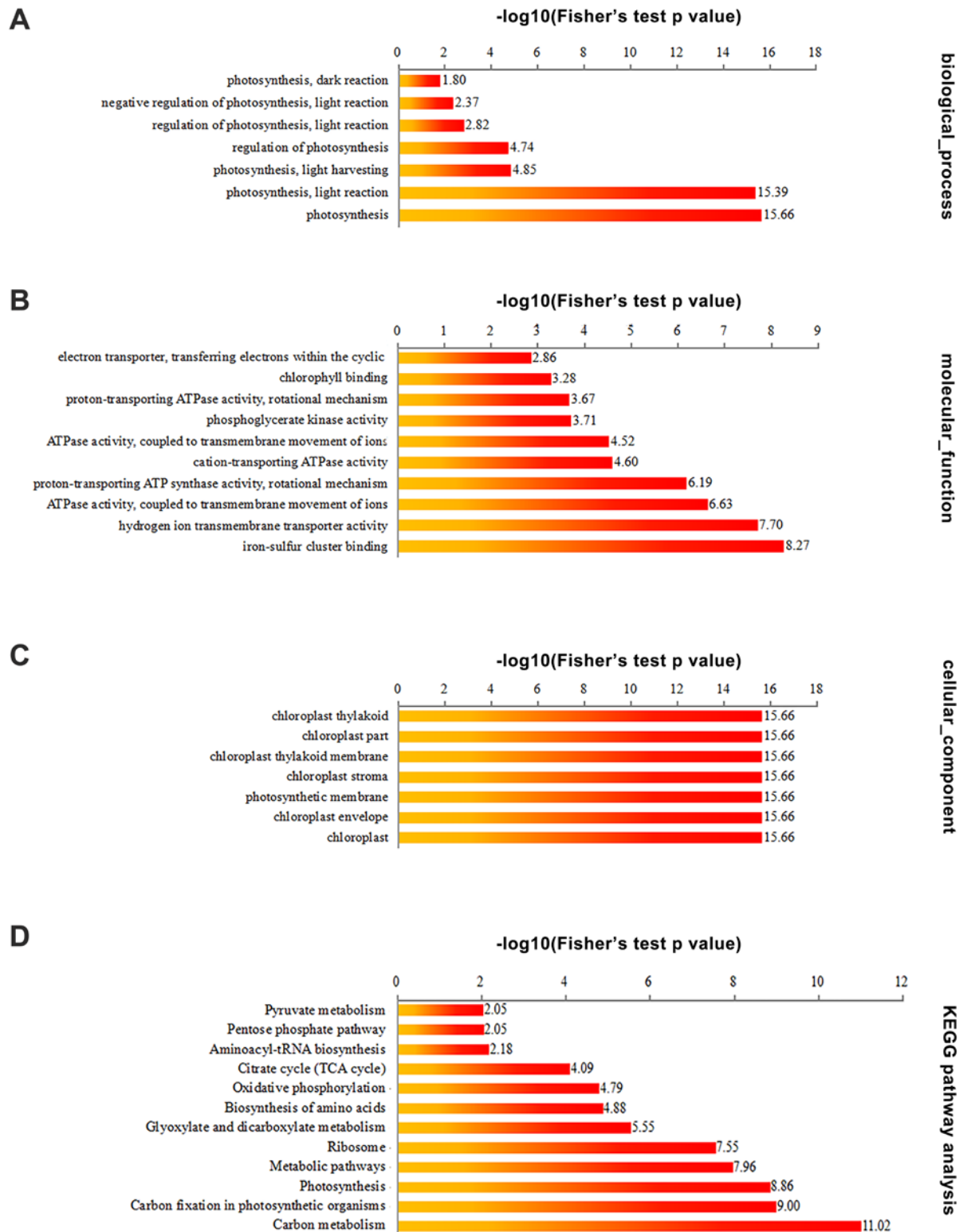
Supplemental_Table_S8. Over-represented functions of genes associated with Kcr modification.



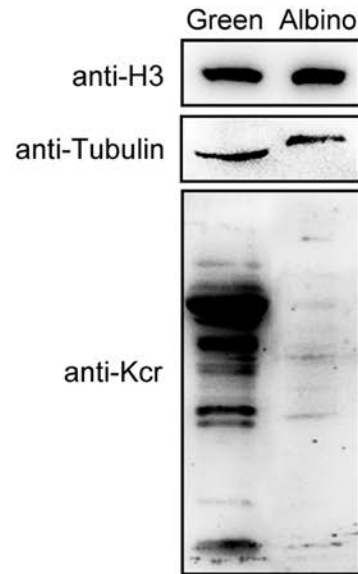
Supplemental_Fig_S1. Quality control validation of MS data. (A) Mass error distribution of all identified peptides. (B) Distribution of crotonylated peptides based on their length. (C) Distribution of crotonylated proteins based on their number of crotonylated peptides.



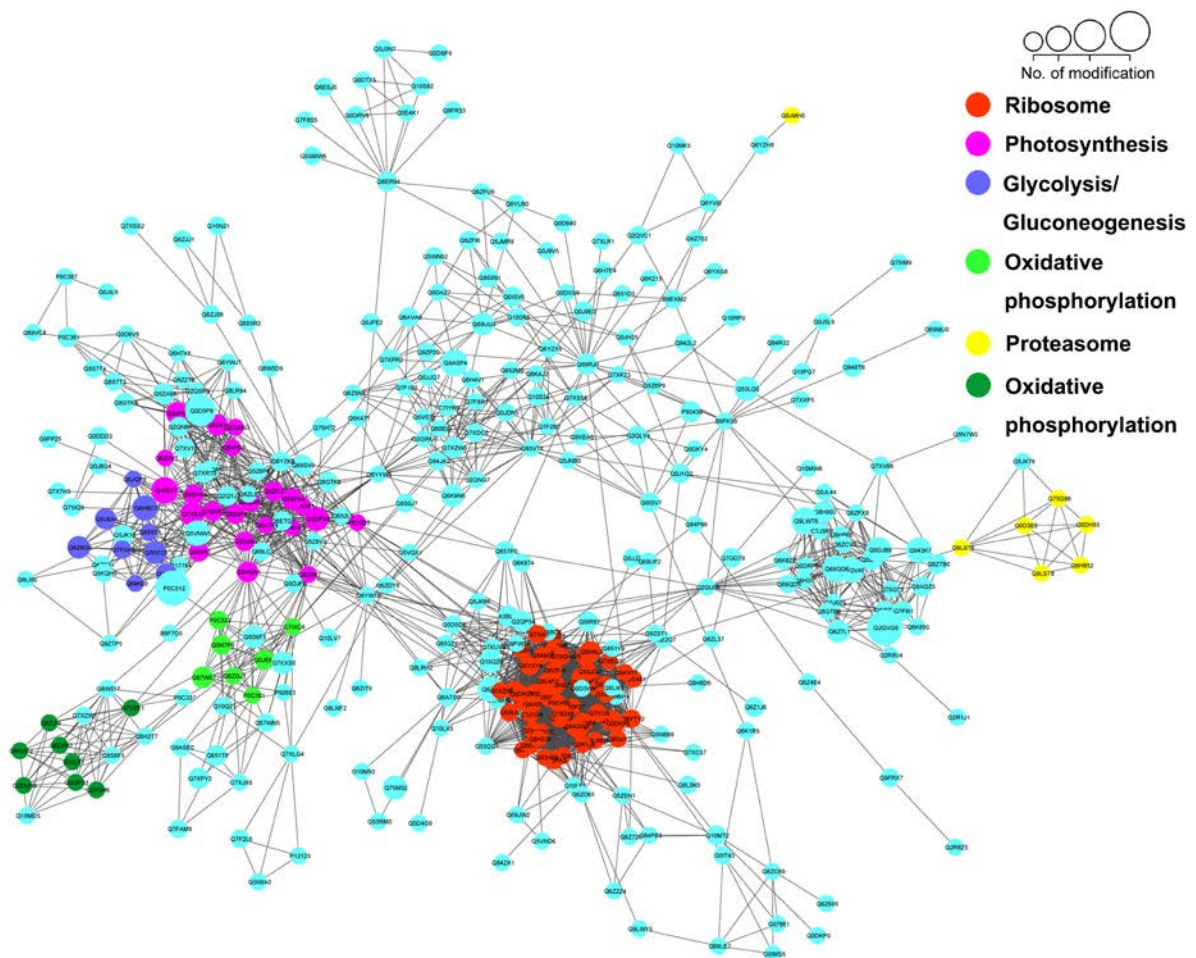
Supplemental_Fig_S2. Representative MS/MS spectra of validated crotonylated peptides in rice seedling. (A) Crotonylpeptide TIMP-(crotonyl)K-DIQLA with crotonylation site at K122 in histone H3. (B) Crotonylpeptide IQGIT-(crotonyl)K-PAIR with crotonylation site at K31 in histone H4. (C) Crotonylpeptide LEV-(crotonyl)K-EIAEIM with crotonylation site at K817 in chaperone protein C1pC2. (D) Crotonylpeptide LAEEG-(crotonyl)K-VAIR with crotonylation site at K205 in ribosome-recycling factor.



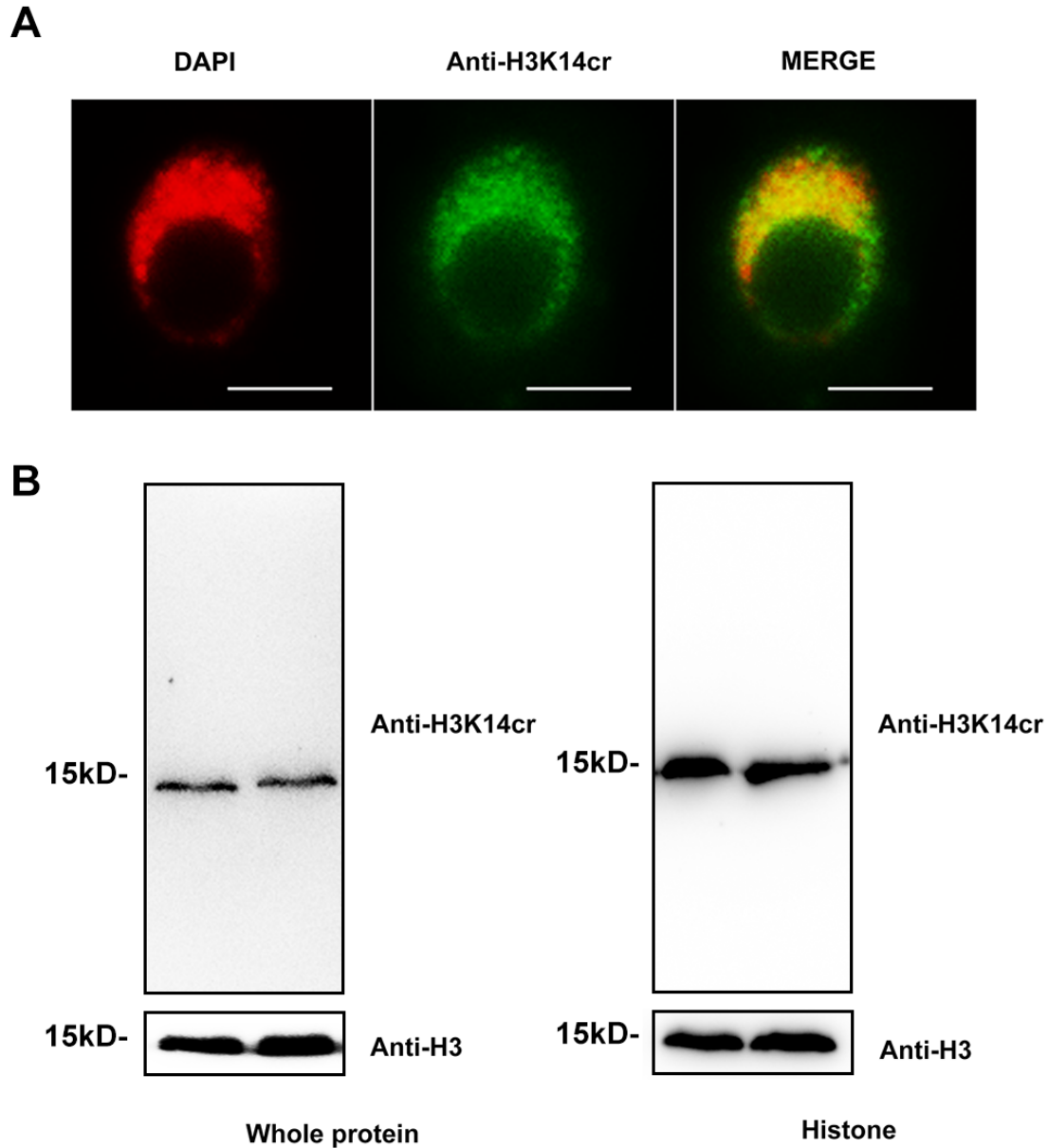
Supplemental_Fig_S3. Functional annotation of lysine crotonylome. Representative GO annotations of Kcr sites for (A) biological process, (B) molecular function, and (C) cellular component, and (D) KEGG pathway analysis of lysine-crotonylated proteins.

A**B**

Supplemental_Fig_S4. WB analysis in green and albino seedlings. (A) The morphology of green seedlings and albino seedlings derived from anther culture of *Oryza sativa* variety “Nipponbare”. (B) WB analysis was performed using anti-Kcr antibody in green seedlings and albino seedlings derived from anther culture.



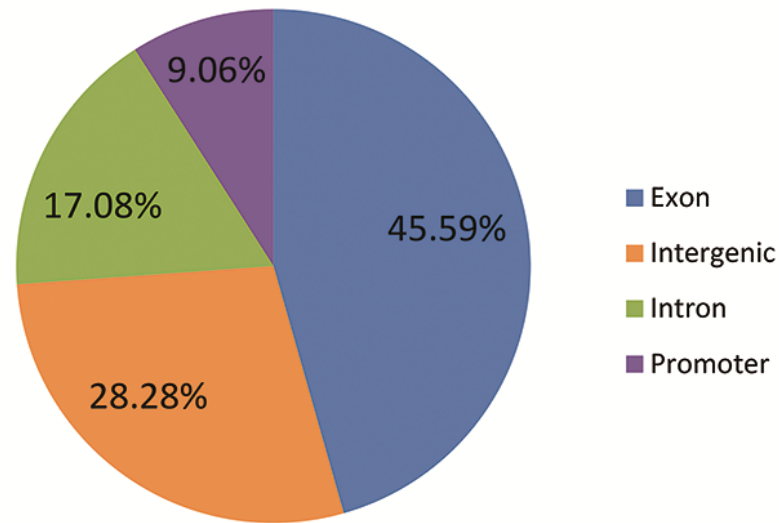
Supplemental_Fig_S5. Protein-protein interaction network. The balls represent crotonylated proteins. The size of the balls represents the numbers of Kcr modification in each figure.



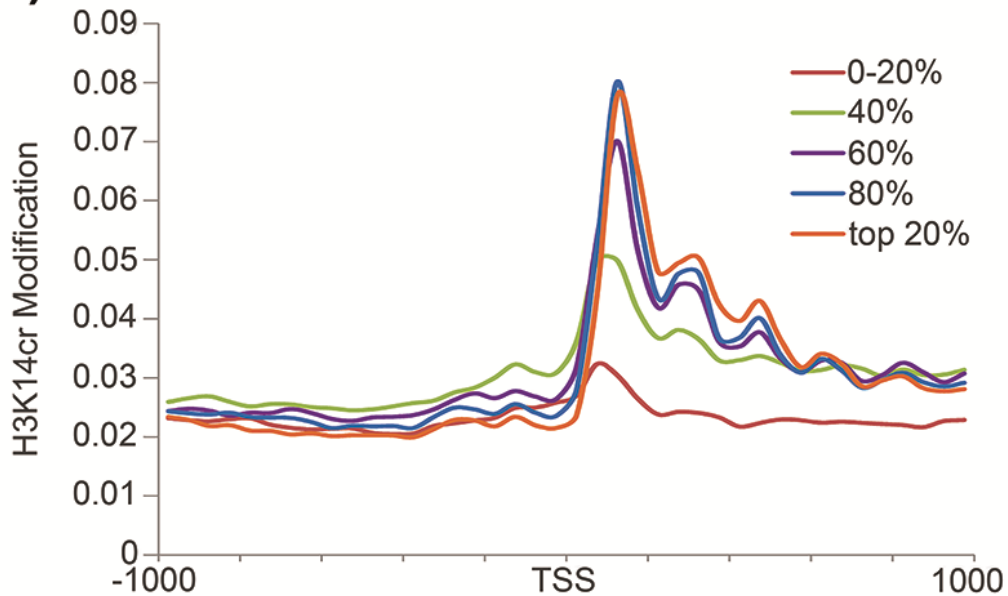
Supplemental_Fig_S6. IF and WB analysis were performed using anti-H3K14cr antibody.

(A) The specificity antibody H3K14cr (green) was detected in two-week-old rice root by immunofluorescence, and nuclei was stained with DAPI (red). Scale bars: 5 μ m. (B) Western blotting analysis in rice seedlings whole protein and histone.

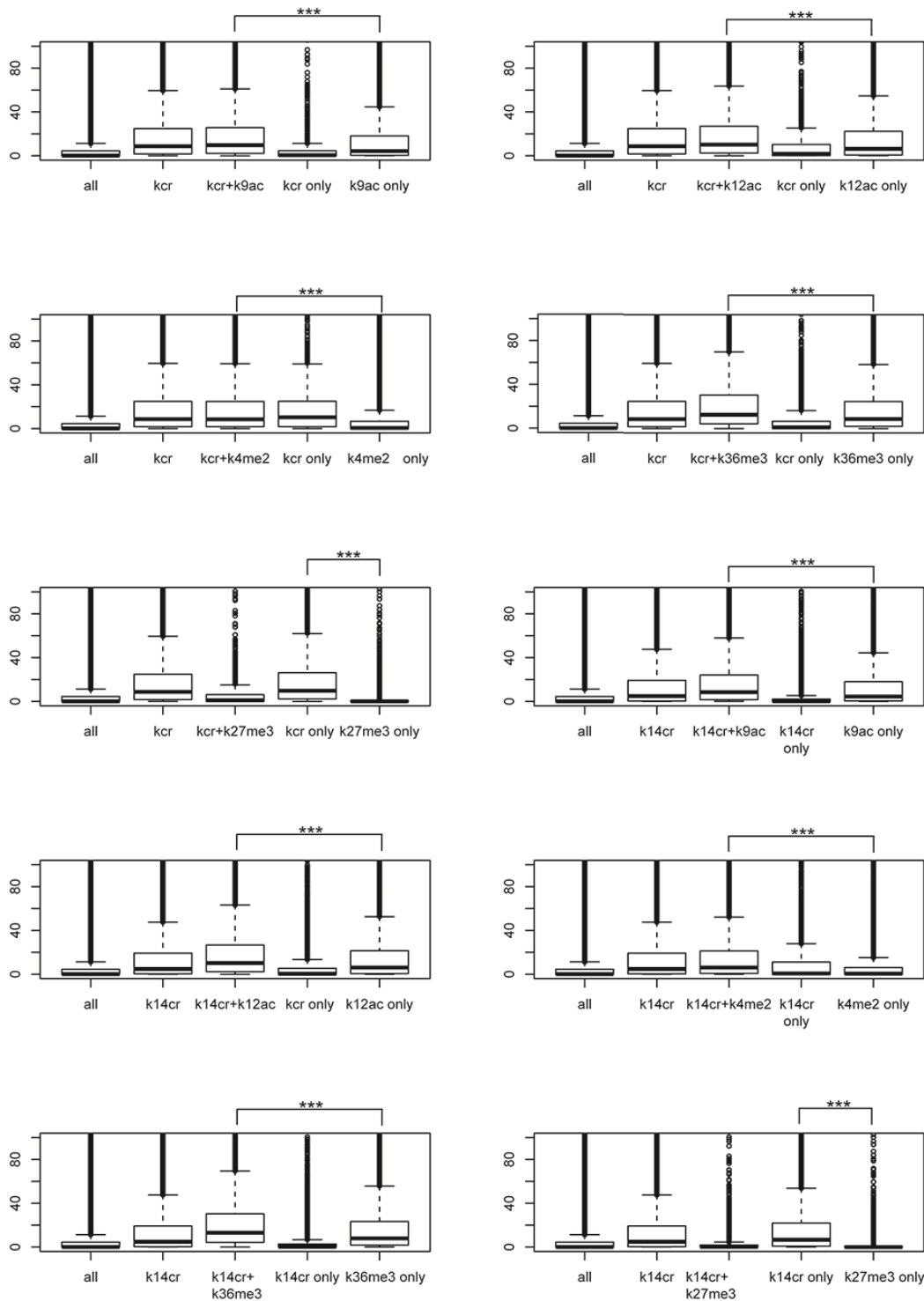
(a)



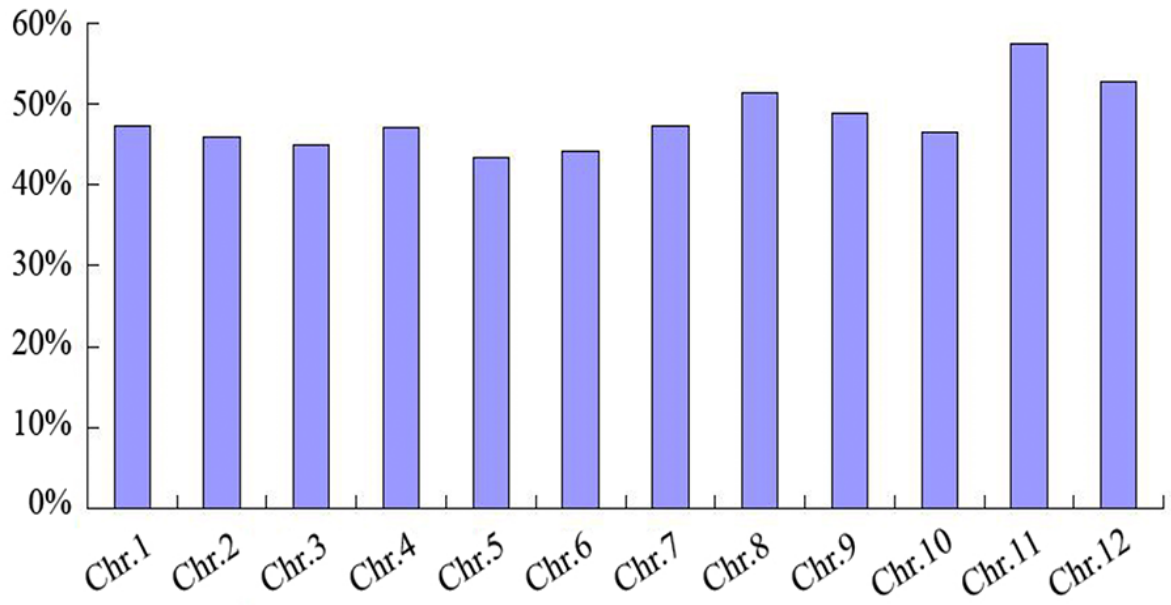
(b)



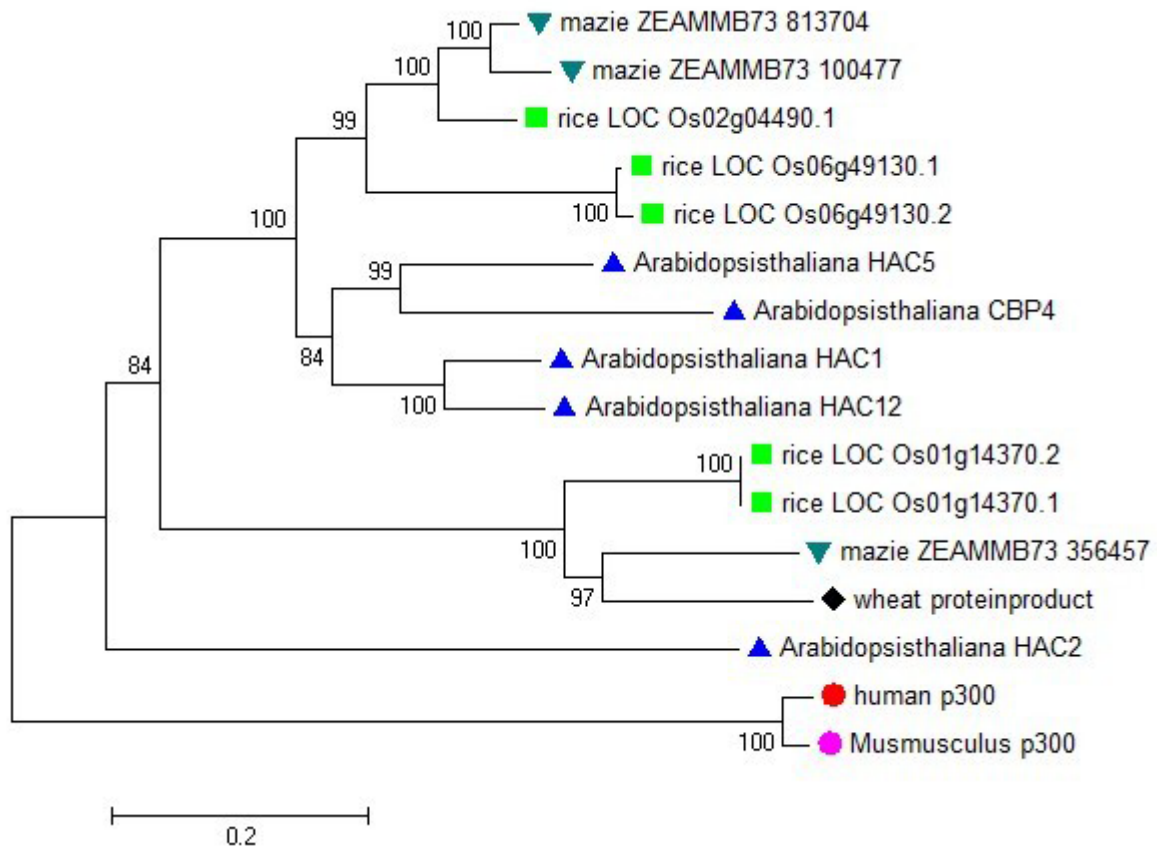
Supplemental_Fig_S7. Distribution of H3K14cr density around differentially expressed genes. (A) Genome-wide distribution of histone H3K14cr in the rice genome. (B) Distribution of H3K14cr density around differentially expressed genes. The H3K14cr modification was calculated by the number of reads per kilobase of the mapped genomic region. The arrow indicates the direction of transcription from transcription start site (TSS). The rice genes were divided into five categories based the expression level from top 20% to bottom 20% (0-20%).



Supplemental_Fig_S8. Expression comparisons of genes associated with different combinations of histone modifications. The non-TE gene expression values (FPKM) of each combination were indicated by box plots. All: all rice genes. Kcr: genes with Kcr modification but not only. Kcr+K9ac: genes with both Kcr and H3K9ac. Kcr only: Genes only with Kcr. The rest may be deduced by analogy. The *** indicated the significant difference between two combinations ($p < 2.2e-16$, Kolmogorov-Smirnov test).



Supplemental_Fig_S9. The distribution of histone Kcr in 12 chromosomes of rice.



Supplemental_Fig_S10. The phylogenetic tree of p300/CBP homologous genes in rice, maize, wheat and Arabidopsis genome.

Supplementary Table 4. Seven Kcr sites identified in rice histones

Protein ID	Modified sequence	Position	Subcellular Location	Domain desc
Q0JQL7	_HAVSEGTK(cr)AVTK_	33	cytosol	Histone-fold; Histone H2B;
Q2RAD9	_EIAQDFK(cr)TDLR_	80	nuclear	Histone core;
Q2RAD9	_STGGK(cr)APR_	15	nuclear	Histone-fold;
Q2RAD9	_VTIMPK(cr)DIQLAR_	123	nuclear	Histone H3;
Q2RAD9	_YQK(cr)STELLIR_	57	nuclear	
Q7XUC9	_DNIQGITK(cr)PAIR_	32	nuclear	Histone H4, conserved site;
Q7XUC9	_K(cr)TVTAM(ox)DVVYALK_	80	nuclear	Histone core;

Supplementary Table 5. Summary of ChIP-seq data

Libraries	Reads number	Mappable reads	Peaks	Common peaks
Kcr (pan-antibody) replicate1	13,308,307	~88%	17514	10923
Kcr (pan-antibody) replicate2	12,219,425	~88%	12343	
H3K14cr replicate 1	77,102,925	~84%	21519	18813
H3K14cr replicate 2	41,081,814	~80%	30856	

Supplementary Table 7. Confirmation of peak sites and non-peak sites by quantitative PCR

Genomic							Peak
Region	qPCR ¹	qPCR ²	Forward Primer	Rverse Primer	-10*log ₁₀ (pvalue)		
Os04g02060	0	5.03	GAACCAGACGAAGACCAT	CGGACAGTTACGAATTGC	138.62	peak sites	
Os09g15850	0	4.23	CAGGAATGGATGAGGAAGAT	GGATTGTGACGAACCTTTG	138.84	peak sites	
Os06g03860	0	3.58	CCTGTGGTAACTGCTAACT	CCTAACTTGTC AACCTATTCC	90.10	peak sites	
Os03g14400	0	3.22	GGTGCCTGCTACTGAATT	GCTGGCTAATGTCGGATT	107.95	peak sites	
Os12g09000	0	3.32	ATTCATGTGGTGCCAGAG	AAGATTCGTGGAACCTGGA	97.77	peak sites	
Os05g03140	0	2.95	AAGAAGGAGAAGAGCAAGAG	GGCGAAGACGATGAAGAA	207.81	peak sites	
Os02g01790	0	1.85	CCTCTTATGAAGCCTGATACT	CGGTTGGTTGAATGATTGG	351.35	peak sites	
Os07g01310	0	4.35	TAACGCTTACTCTGCTCTC	ACCTGTCCTCAATCAATCT	76.20	peak sites	
Os08g01590	0	1.63	CATCAACACAGATACTCATACC	CACAACCTCTAATGTCTACTG	207.85	peak sites	
Os04g02850	0	4.21	GCAGCAGCTAGAAGAAGG	GTACGCCATGATGAGGAC	116.31	peak sites	
Os11g03734	0	1.73	CTGGAGCCAAACATCAA	ACAGACAACCTCTATCCTTAC	157.44	peak sites	
Os02g13430	0	0.57	GTCTGGTGAGTTGCCTGATACG	TGTTGTTGATGGTGTCCGAGTC	165.87	peak sites	
Os01g18630	0	-0.34	TTGCGAGATGGCTGTTGTCT	GCTGGTGCACATAGCAAAG	64.23	peak sites	
Os01g03010	0	0.29	ACTATGGTGGCTGTCTACA	CGATGCTTAACGGTCTTG	-	non-peak sites	
Os02g16260	0	-0.66	ATAATCATCGCTGCTCCATA	CTGTCCACATCAAGTCGTA	-	non-peak sites	
Os03g03150	0	-0.39	GGATGCCACTCGTTGTAA	ATGGTGAAGGATGCTCTTG	-	non-peak sites	
Os04g03080	0	-3.07	CATCCACTATTGCCACTGT	CCTTACTGAGAACCTACCATT	-	non-peak sites	
Os05g03840	0	-0.38	AGAATCATCGTGGCTTGTA	CGTCCTGTTCCATTGTCA	-	non-peak sites	
Os06g02380	0	2.74	AGCCTTATCCTCTCCTCTC	ATCCACCGAATCCTCCAA	-	non-peak sites	
Os07g17160	0	-2.09	GCAATCTGGAAGGAGTTCA	CCGTGATGACTAATGTGTAAC	-	non-peak sites	
Os08g03670	0	-0.15	ACGAAGAATCACCTGACAAT	ACATAGACGCCACGAGAT	-	non-peak sites	
Os09g06770	0	3.61	CCAGCTAATCAATCATCATCG	ATCGTTCGGTTCTATCCAG	-	non-peak sites	
Os10g17630	0	-2.94	AACATCTGCTATGCTAAGGT	TTGCTATTCCGTGTGACTAA	-	non-peak sites	
Os12g19470	0	-1.94	GCACCACATGCTTGATAGG	TAAGTTGCAGTGTGCTGCCT	-	non-peak sites	
Os02g08110	0	-2.10	AATGTCTCGTCCAGTTCTAG	ACTACCTCCATCGTGCTAT	-	non-peak sites	
Os09g03500	0	-2.56	CCTCATACTCCTCCATCC	GGCATAAACCTACTCCAT	-	non-peak sites	
Os11g06490	0	-2.19	GGTAAGGAGGACGGAGTAT	CACAAGAAGTAACGGTAAGAC	-	non-peak sites	

¹Normalized C(t) of input DNA .

²Normalized Δ C(t) of ChIP DNA .

Supplementary Table 8. Over-represented functions of genes associated with Kcr modification.

GO term	Ontology	Description	Number in input list	Number in Ref	FDR
GO:0044249	P	cellular biosynthetic process	256	890	0.0057
GO:0034645	P	cellular macromolecule biosynthetic process	256	890	0.0057
GO:0009059	P	macromolecule biosynthetic process	256	890	0.0057
GO:0006412	P	translation	256	890	0.0057
GO:0010467	P	gene expression	300	1076	0.0073
GO:0019538	P	protein metabolic process	1451	6088	0.023
GO:0044267	P	cellular protein metabolic process	1164	4849	0.035
GO:0043170	P	macromolecule metabolic process	1580	6691	0.035
GO:0005198	F	structural molecule activity	192	672	0.039
GO:0000166	F	nucleotide binding	858	3490	0.039
GO:0005829	C	cytosol	854	3345	0.0017
GO:0005737	C	cytoplasm	2771	11866	0.0021
GO:0005622	C	intracellular	3483	15144	0.0023
GO:0044444	C	cytoplasmic part	2558	10930	0.0023
GO:0031974	C	membrane-enclosed lumen	215	743	0.003
GO:0031981	C	nuclear lumen	215	743	0.003
GO:0043233	C	organelle lumen	215	743	0.003
GO:0044424	C	intracellular part	3333	14514	0.003
GO:0005730	C	nucleolus	152	495	0.003
GO:0005886	C	plasma membrane	1040	4240	0.003
GO:0070013	C	intracellular organelle lumen	215	743	0.003
GO:0044428	C	nuclear part	239	844	0.0031
GO:0044446	C	intracellular organelle part	239	844	0.0031
GO:0044422	C	organelle part	239	844	0.0031
GO:0005794	C	Golgi apparatus	138	460	0.0094
GO:0043232	C	intracellular non-membrane-bounded organelle	326	1237	0.014
GO:0043228	C	non-membrane-bounded organelle	326	1237	0.014
GO:0044464	C	cell part	4386	19532	0.026
GO:0005840	C	ribosome	172	624	0.036
GO:0030529	C	ribonucleoprotein complex	172	624	0.036
GO:0032991	C	macromolecular complex	172	624	0.036