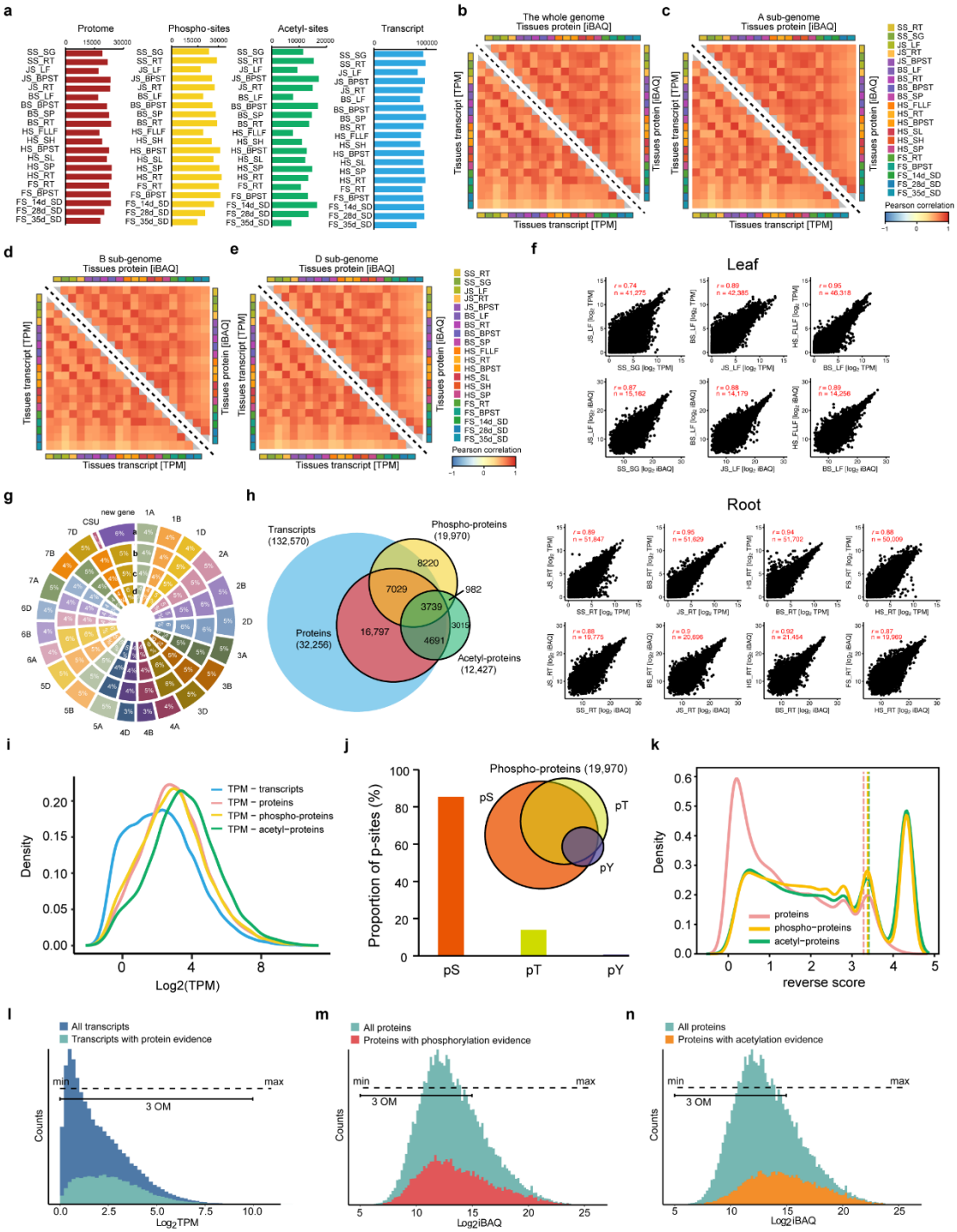
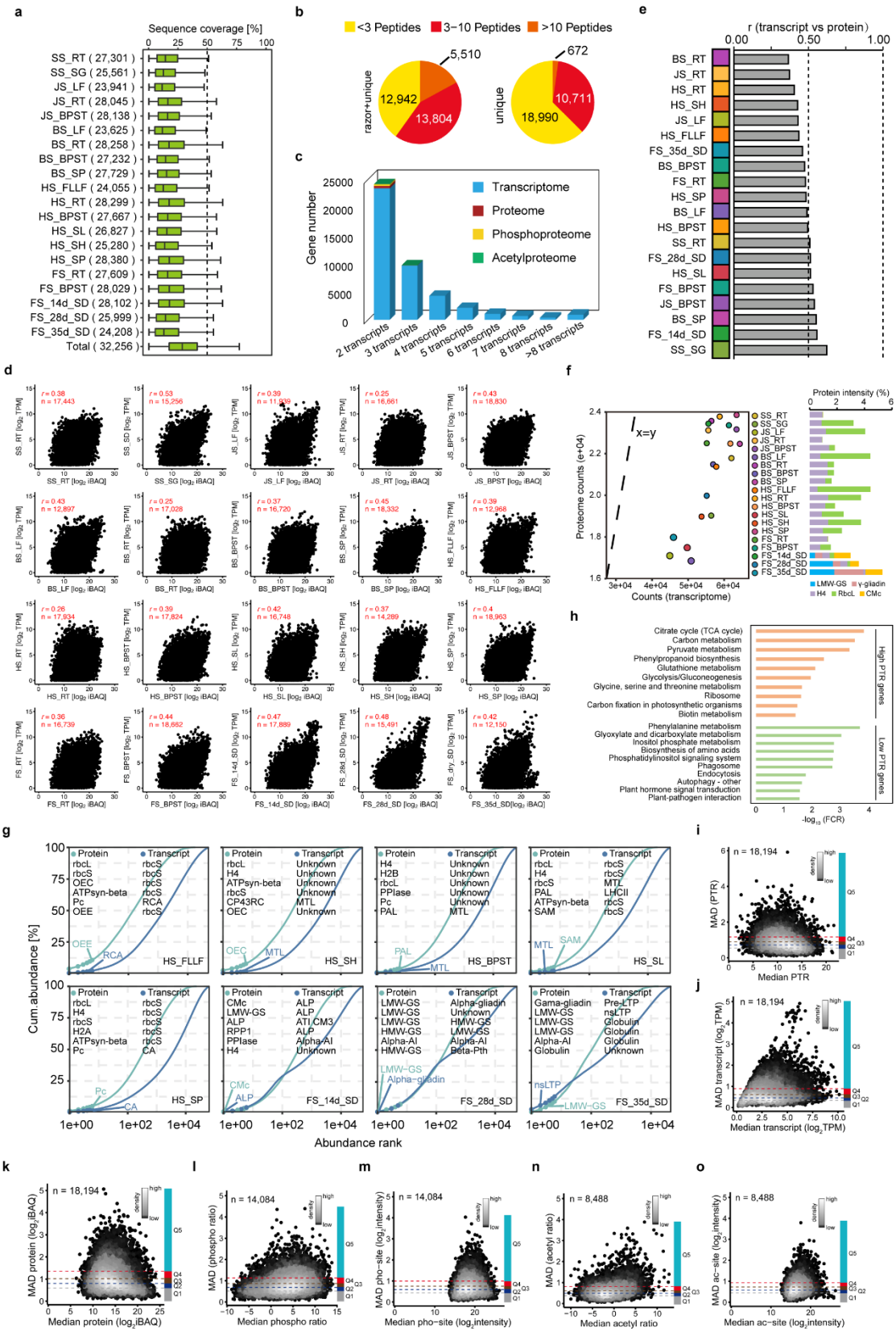
Integration of multi-omics data accelerates molecular analysis of common wheat traits

Zhang *et al.*



Supplementary Fig. 1. Properties of the wheat multi-omics atlas.
a, The number of proteins, phosphorylation sites, acetylation sites, and transcripts from 20 tissues. **b**, Pairwise global Pearson's expression correlation analysis of 20 tissues on transcriptome and proteome levels using genes we identified from whole genome, **c,** A sub-genome, **d**, B sub-genome, **e**, D sub-genome. iBAQ, intensity-based absolute quantification. **f**, Pearson's correlation analysis showed high coefficient of correlation at protein and transcript levels between leaf or root tissues in different stages. *r*, the Pearson's correlation coefficient; n, the number of transcripts or proteins. **g**, Percentage distribution of the number of identified expressed genes at transcript, protein, phosphorylation, and acetylation levels (a-d) in whole chromosomes of wheat. **h**, Total number and overlap of the identified genes in the transcriptome, proteome,

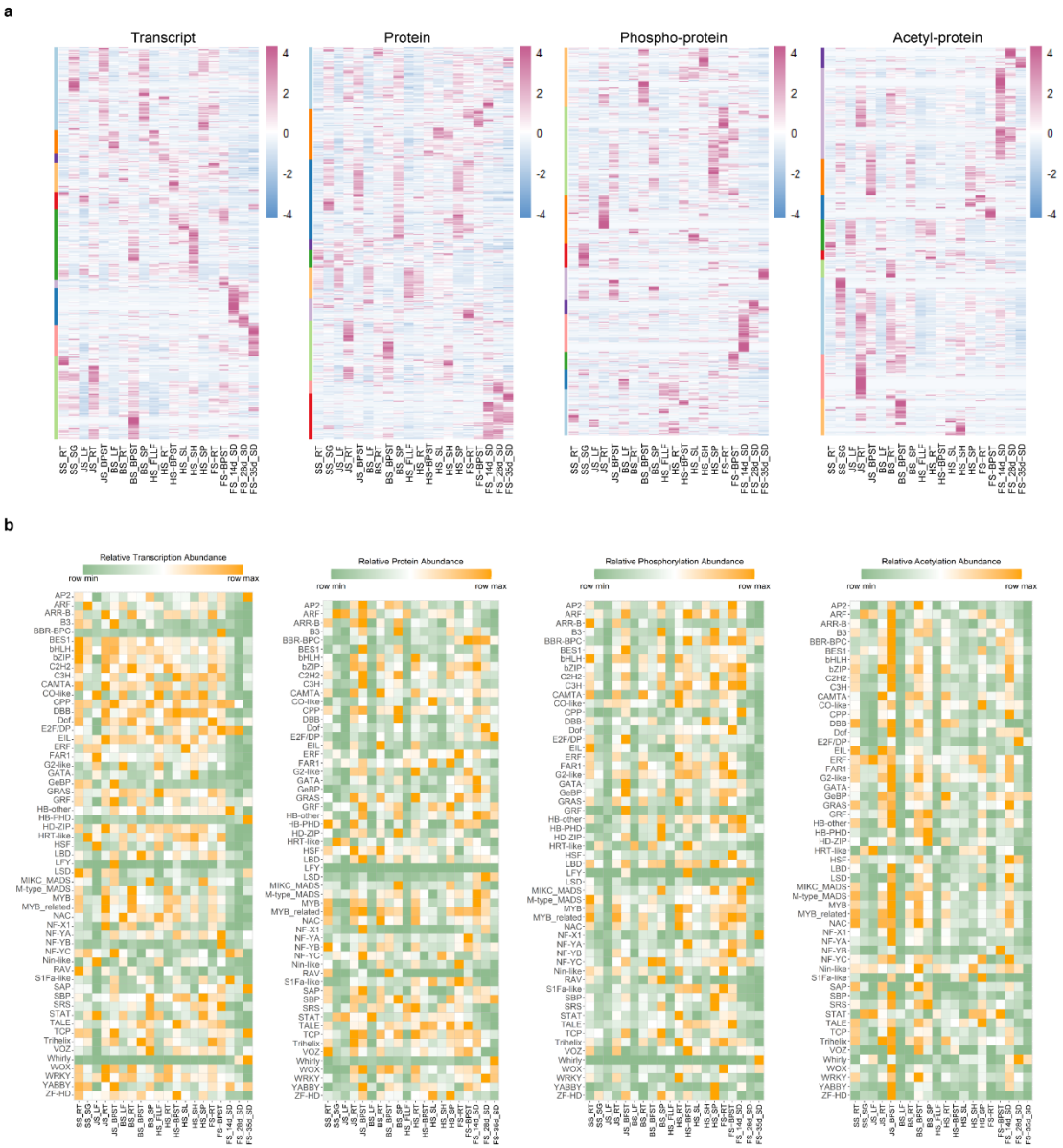
phosphoproteome and acetylproteome datasets. **i**, TPM (transcripts per million) distribution of mRNA abundance at transcript, protein, phosphorylation, and acetylation levels. Data are the average expression from the 20 tissues profiled. **j**, Distribution of proteins with phosphorylation of Ser, Thr or Tyr residues. **k**, Distribution of Shannon entropy scores obtained for all proteins in the atlas. Red, yellow, and green dashed lines indicate the 10%, 50%, 50% quantile cut-off for tissue specificity at protein, phosphorylation and acetylation levels, respectively. **l**, Dynamic range of transcript abundance (blue) and proportion of transcripts that were also identified at the protein level projected into this plot (green). OM, orders of magnitude. **m**, Dynamic range of protein abundance and proportion of proteins with phosphorylation of Ser, Thr or Tyr residues. **n**, Dynamic range of protein abundance and proportion of proteins with acetylation of lysine residue. SS_RT: root of seedling stage; SS_SG: seedling of seedling stage; JS_LF: leaf of jointing stage; JS_RT: root of jointing stage; JS_BPST: basal part of stem of jointing stage; BS_LF: leaf of booting stage; BS_RT: root of booting stage; BS_BPST: basal part of stem of booting stage; BS_SP: spike of booting stage; HS_LF: flag leaf of heading stage; HS _RT: root of heading stage; HS _BPST: basal part of stem of heading stage; HS_ SL: stalk of heading stage; HS_ SH: sheath of heading stage; HS _SP: spike of heading stage; FS _RT: root of filling stage; FS _BPST: basal part of stem of filling stage; FS_14d_seed: seed of 14 day post anthesis (DPA) of filling stage; FS_28d_seed: 28 DPA of filling stage; FS_35d_seed: 35 DPA of filling stage. The same below. Source data are provided as a Source Data file.



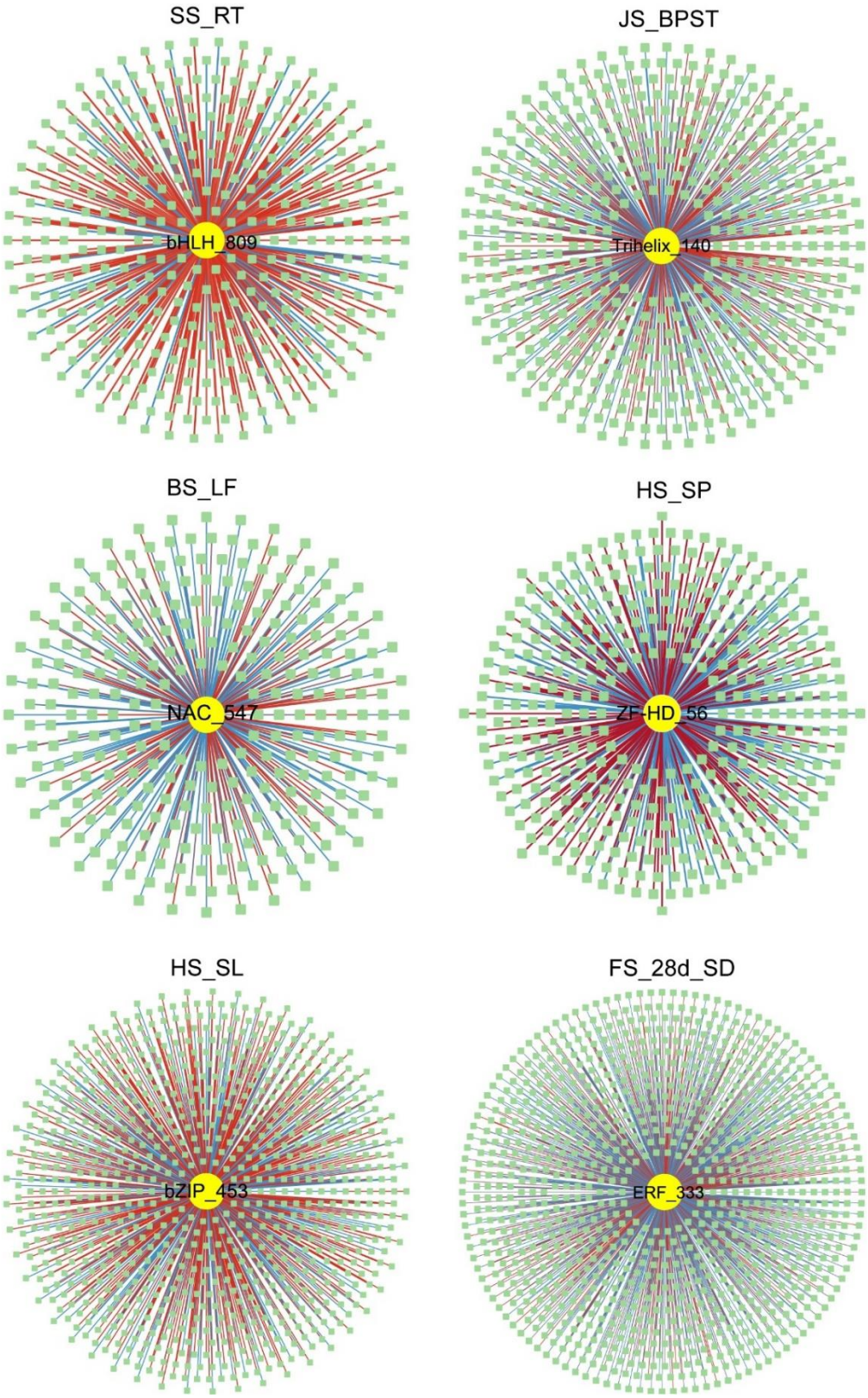
Supplementary Fig. 2. Multi-omics characterization, dynamic range of transcript and protein expressions, and correlations between transcriptome, proteome, phosphoproteome, and acetylproteome.

a, Distribution of peptide-based sequence coverage of proteins in all tissues. **b,** Pie charts shows the percentage of proteins identified with <3, 3–10 or >10 peptides either

razor unique or unique peptides. **c**, Number of genes with multiple transcripts distinguished at transcript, protein, phosphorylation and acetylation levels, respectively. There was substantial evidence that the number of transcripts differed across omics levels. **d-e**, Pearson's correlation coefficient (*r*) between proteins and transcripts of each tissue **(d)** for all data and **(e)** shared_intersection data (*n* = 2,751), respectively. **(f)** (Left) Total number of transcripts and proteins showed that the more transcripts are expressed, the more proteins are detected in each tissue (*r* = 0.79). (Right) Proportion of the top five proteins with the most abundance in at least one tissue. **g,** Cumulative plots of abundance-ranked identifications of transcripts and proteins for eight representative tissues. The top six transcripts and proteins with the most abundance are listed for each tissue. The characteristics of the plots showed the difference between protein and transcript levels for each tissue. **h**, GO (gene ontology) enrichment analysis of high protein-to-mRNA ratio (PTR) and low PTR genes. **i**, Median PTRs across tissues plotted against the inter-tissue variation of these PTRs (expressed as MAD; proteins and transcripts were detected in at least eight matching tissues). Bar plot shows the MAD range segmented into five quantiles, each containing the same number of genes (colored bars and dashed lines). Most genes have stable PTRs across tissues. **j-k**, As in **i** (n = 18,194) but for transcript and protein measurements. 80% of all transcripts and 60% of all proteins show a MAD of <1. **l**, As in **i** but for the ratio of phosphorylation sites (Pho-sites) versus proteins. Phosphorylation sites and proteins had to be detected in at least eight matching tissues (n = 14,084). **m**, As in **k** (n = 14,084) but for Pho-site abundance. 80% of all P-sites show MAD <1. **n**, As in **i** but for the ratio of acetylation sites versus proteins. Acetylation sites (Kac-sites) and proteins had to be detected in at least eight matching tissues (n = 8,488). **o**, As in **k** (n = 8,488) but for Kac-site abundance. 80% of all Kac-sites show MAD <1. MAD: median absolute deviation. Source data are provided as a Source Data file.

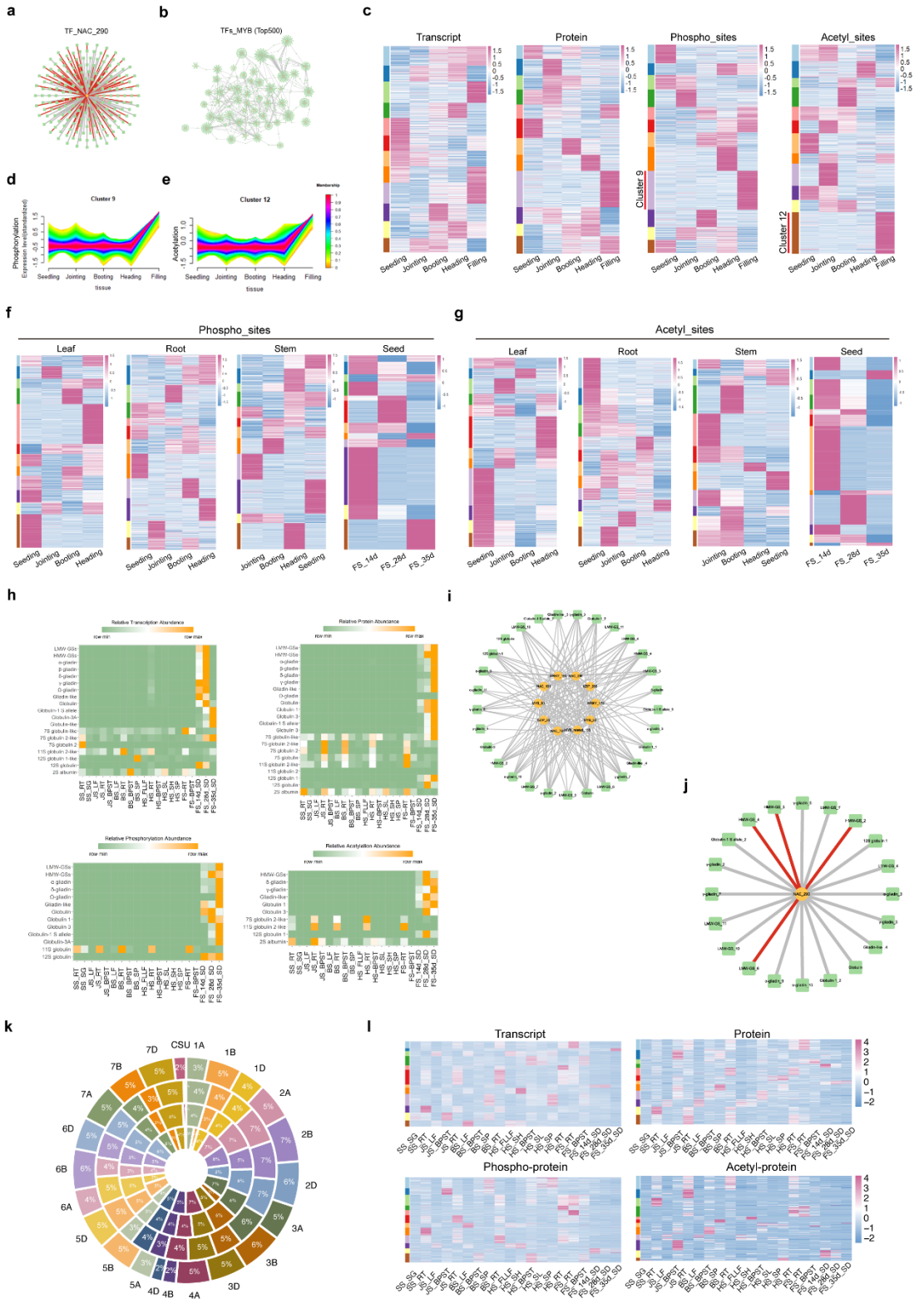


Supplementary Fig. 3. Hierarchical clustering and family-wise expression analysis of transcription factors. a, Hierarchical clustering. **b**, family-wise expression analysis. Source data are provided as a Source Data file.



Supplementary Fig. 4. Gene regulatory networks (GRNs) of six representative transcription factors.

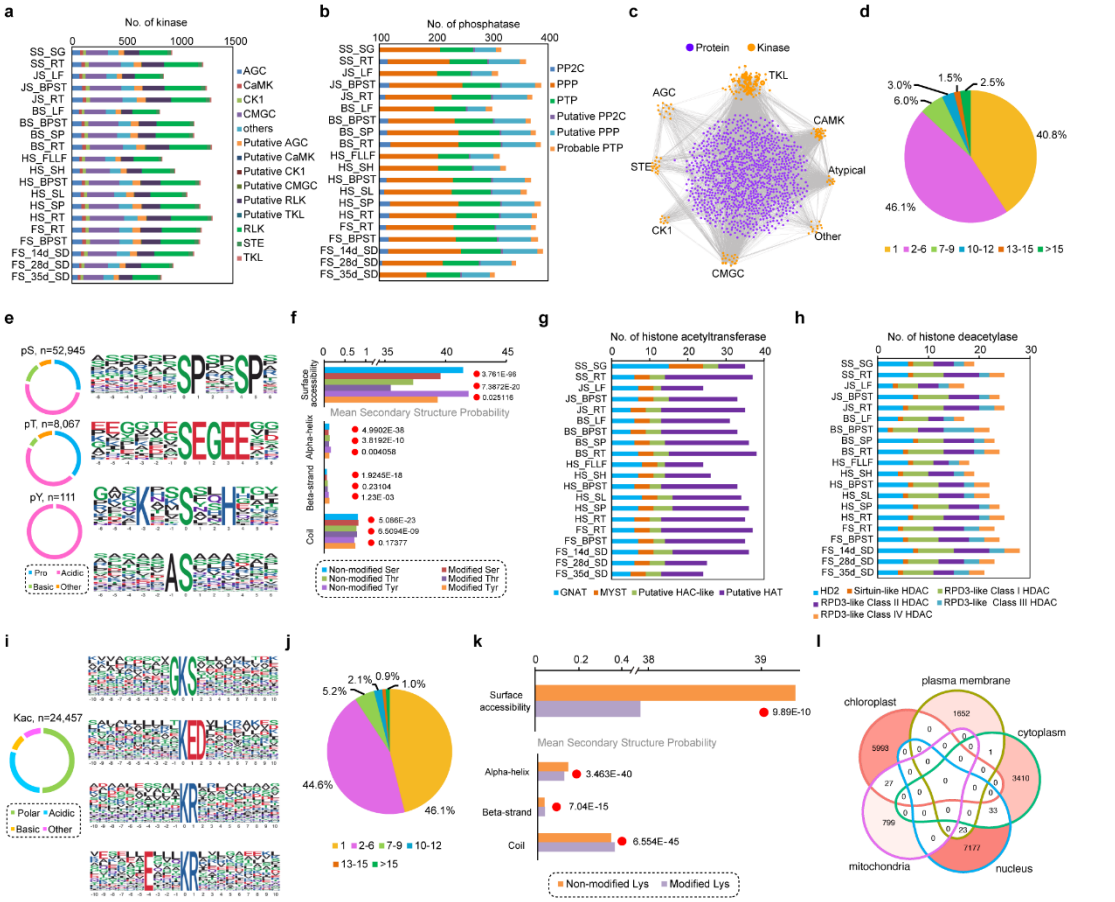
Red and blue lines indicated that the tissue-specific regulatory relationship was enhanced or weakened compared with the total GRN, respectively. ZF-HD: zinc-finger homeodomain; bZIP: basic region/leucine zipper motif. Source data are provided as a Source Data file.



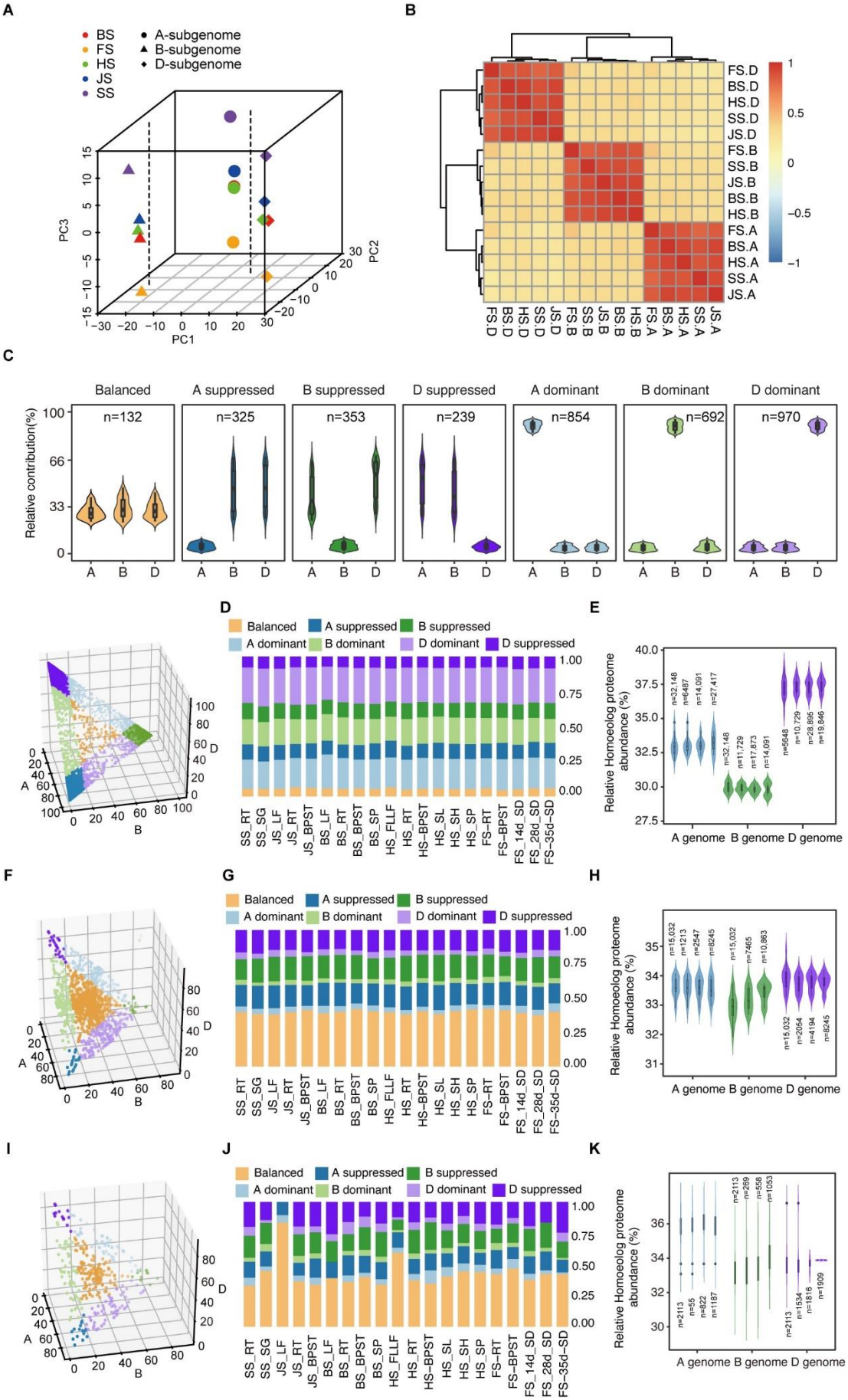
Supplementary Fig. 5. Regulatory network of transcription factors and target genes involved in seed storage proteins and hierarchical clustering of genes at transcript, protein, phosphorylation and acetylation levels.

a, Gene regulatory network (GRN) of transcription factor (TF) NAC_290 and its targets. Red lines indicated the regulatory relationship previously reported. The same below in j). **b**, GRN of 48 MYBs from the top 500 of total TFs and their 1,754 targets (top 2000 edge). **c**, Hierarchical clustering of genes identified from different wheat developmental

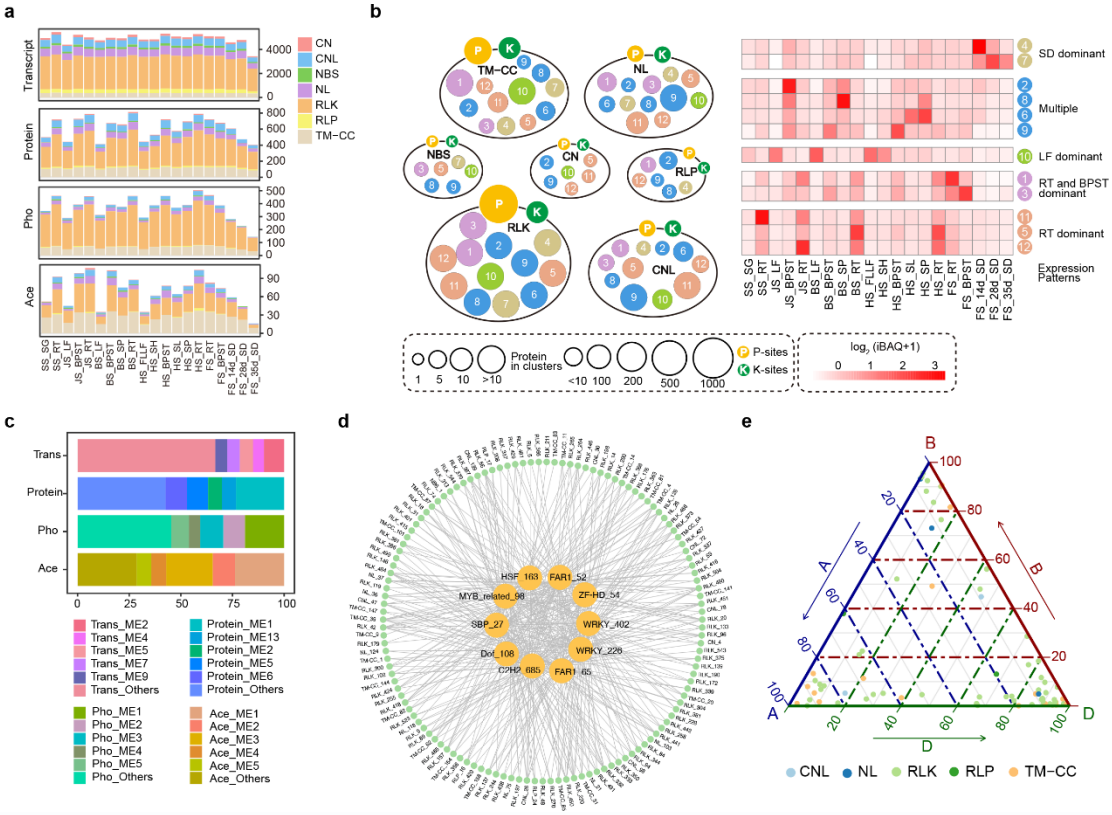
stage at transcript, protein, phosphorylation and acetylation levels, respectively. **d**, The expression pattern of genes from cluster 9 at the phosphorylation level. **e**, The expression pattern of genes from cluster 12 at the acetylation level. **f-g**, Hierarchical clustering of genes from different developmental stages at phosphorylation and acetylation levels, respectively. **h**, Family-wise expression analysis and (**i**) GRN of seed storage proteins (SSPs). **j**, The regulatory relationship of NAC_290 and SSPs. **k**, Chromosome distribution (from outside to inside: transcript, protein, phosphorylation, and acetylation) and (**l)** hierarchical clustering of the disease resistance-related genes (DRRGs). Source data are provided as a Source Data file.



Supplementary Fig. 6. Characteristics of phosphoproteome and acetylproteome. **a-b**, The number of kinase and phosphatase family members annotated in proteome. **c**, Relationship prediction of the substrates and predicted kinases in the same tissue. **d**, The percentage of protein with different numbers of phosphorylation sites. Results show that 59.2% of proteins contained multiple phosphorylation sites. **e**, Numbers and examples of proline-directed, acidic, basic and other motif categories for phosphorylated Ser (pS), Thr (pT) and Tyr (pY) residues. **f**, The predicted surface accessibility and secondary structure probability of phosphorylation residues. Results suggested that modified residues are less surface accessibility compared to unmodified residues, and phosphorylation has a preference of secondary structure, except for beta-strands of pT sites and disordered coils of pY sites. **g-h**, The number of annotated histone acetyltransferases and histone deacetylases family members at the protein level. **i**, Numbers and examples of polar-directed, acidic, basic and other motif categories for acetylated lysine residues. **j**, The percentage of protein with different numbers of acetylation sites. Results show that 53.9% of proteins contained multiple acetylation sites. **k**, The predicted surface accessibility and secondary structure probability of acetylation residues. Results suggested modified residues are less surface accessibility compared to unmodified residues, and acetylation has a preference of secondary structure. **l**, Predicted localization results of acetylproteins. Source data are provided as a Source Data file.

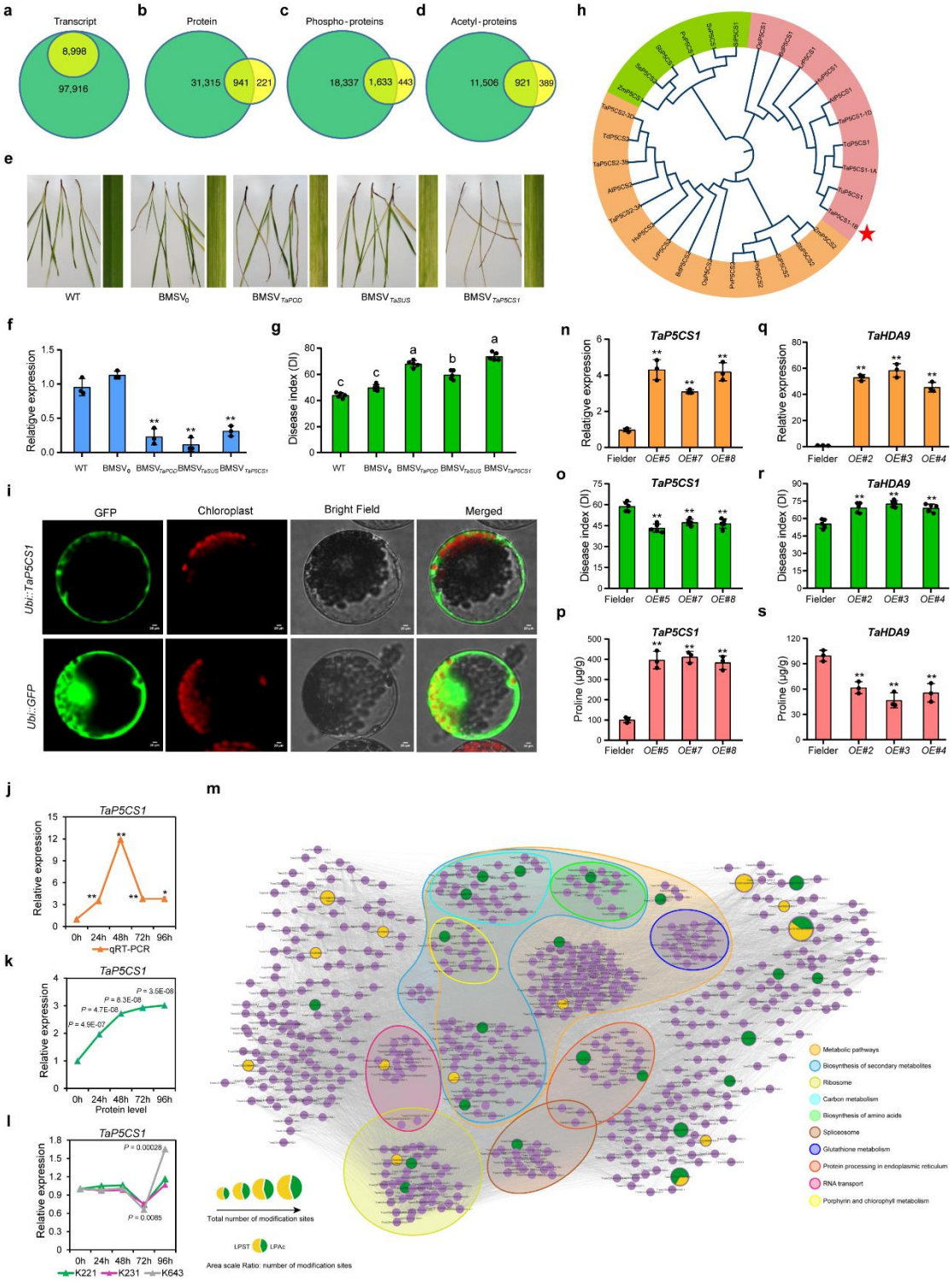

A
BS
FS
HS
JS
SS
A-subgenome
B-subgenome
D-subgenome
PC1
PC2
PC3
B
FS.D
BS.D
HS.D
SS.D
JS.D
FS.B
SS.B
JS.B
BS.B
HS.B
FS.A
BS.A
HS.A
SS.A
JS.A
C
Balanced
A suppressed
B suppressed
D suppressed
A dominant
B dominant
D dominant
Relative contribution(%)
n=132
n=325
n=353
n=239
n=854
n=692
n=970
D
E
Relative Homoeolog proteome abundance (%)
A genome
B genome
D genome
F
G
H
I
J
K
SS_RT
SS_SG
JS_LF
JS_RT
JS_BPST
BS_LF
BS_RT
BS_BPST
BS_SP
HS_FLLF
HS_RT
HS_BPST
HS_SL
HS_SH
HS_SP
FS_RT
FS_BPST
FS_14d_SD
FS_28d_SD
FS_35d_SD

Supplementary Fig. 7. Homoeolog expression bias in syntenic triads at protein, phosphorylation and acetylation levels.
a, Principal component analysis (PCA) of all proteins during SS, JS, BS, HS, and FS stages in our atlas, separating gene expression diversity among A, B and D sub-genomes. b, Clustering analysis of proteins in SS, JS, BS, HS, and FS stages. (**c**, **f**, **i**) Ternary plot (down) showing relative expression abundance of (**c**) 3,565 syntenic protein triads, (**f**) 1,786 syntenic phosphoprotein triads, and (**i**) 425 syntenic acetylprotein triads through the combined analysis of 20 tissues. Each point represents a triad with an A, B, and D coordinate consisting of the relative contribution of each homoeolog to the overall triad expression. Triads in vertices correspond to single sub-genome dominant categories, whereas triads close to edges and between vertices correspond to suppressed categories. Balanced triads are shown in the middle. Violin plots (**c**) indicate the relative contribution of each sub-genome based on protein triad assignment to the seven categories. (**d**, **g**, **j**) Proportion of triads in each category of homoeolog expression bias across the 20 tissues at (**d**) protein, (**g**) phosphorylation and (**j**) acetylation levels. (**e**, **h**, **k**) Genome homoeolog expression bias in wheat. Boxplots of relative expression abundance of A, B, and D genome homoeologs for syntenic triads across 20 tissues at four different minimum riBAQ or rintensity cut off values at (**e**) protein, (**h**) phosphorylation and (**k**) acetylation levels. n is the protein number. In the violin plots, the upper and lower whiskers represent maxima and minima, respectively, the middle line represents the median, and the upper and lower lines represent the third quartile and first quartile, respectively. The bounds of the plot represent data density. riBAQ: Protein abundance was normalized by the ratios of the corresponding protein abundance value (iBAQ) in all protein abundance values ($\sum$iBAQ). Source data are provided as a Source Data file.



Supplementary Fig. 8. Expression and post-translational modifications of disease resistance-related genes.

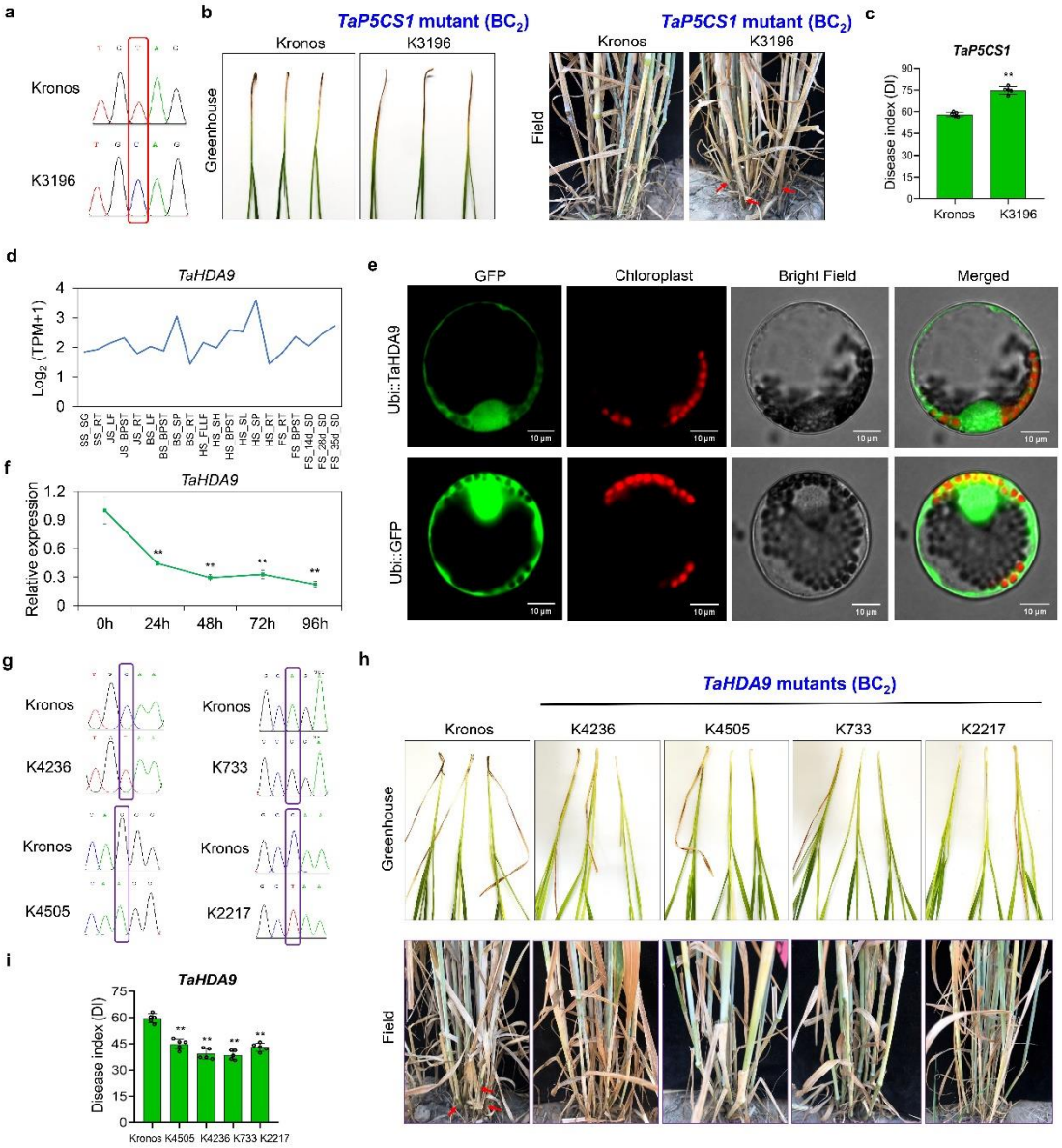
a, The number of disease resistance-related genes (DRRGs) identified across 20 tissues at transcript, protein, phosphorylation (pho) and acetylation (ace) levels, respectively. **b**, DRRG proteins in our atlas were clustered into different modules according to their abundances. Larger circles indicate that more genes from this subfamily fall into this module. Colors indicate different expression patterns: SD dominant (light brown), multiple (blue), LF dominant (green), and RT and BPST dominant (light purple), RT dominant (pink). Roman numbers indicate the different modules. Mean values for the different modules are used for heatmap. **c**, The proportion of the DRRGs number from different modules of weighted gene co-expression network analysis (WGCNA). WGCNA results showed modules with high abundance expression in roots at the five different stages (ME2, ME1, ME1 and ME1). **d**, Gene regulatory network (GRN) for DRRGs and the transcription factors (top 10). **e**, Relative expression abundance of DRRG triads were plotted with each point representing the normalized protein abundance of A, B and D homoeologs. Source data are provided as a Source Data file.



Supplementary Fig. 9. *TaP5CS1* conferred wheat Fusarium crown rot (FCR) resistance.

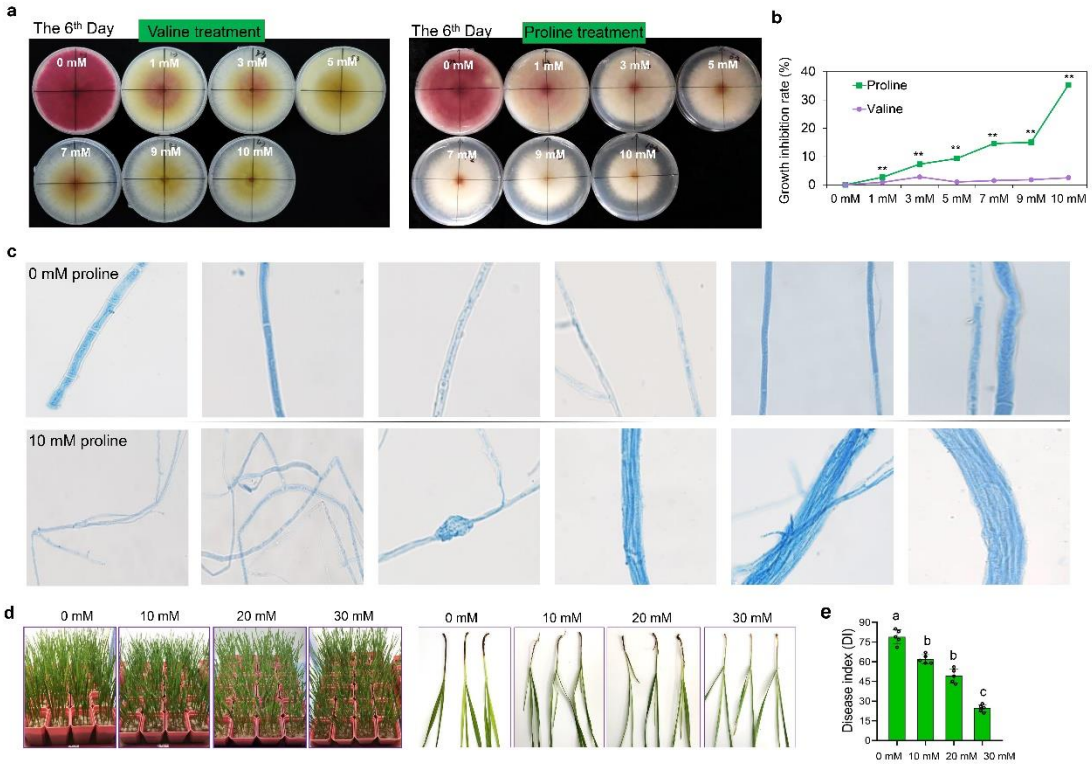
a-d, Venn diagram of the identification number of genes/proteins from 20 tissues and the differentially expressed genes/proteins after *F.pg* inoculation in wheat. **e-g**, VIGS (virus-induced gene silencing) experiment mediated by barley stripe mosaic virus (BSMV) verified functions of *TaPOD*, *TaSUS*, and *TaP5CS1* genes. (**e**) The phenotype (12 plants per replicate), (**f**) relative expression level and (**g**) averaged disease index (DI) (12 plants per replicate) of wild type (WT), $BSMV_0$, $BSMV_{TaPOD}$, $BSMV_{TaSUS}$, $BSMV_{TaP5CS1}$ plants infected by *F.pg*. **h**, Phylogenetic tree for TaP5CS1. **i**, Wheat

protoplasts subcellular localization of TaP5CS1. **J-l**, Expression levels of TaP5CS1 after inoculation with *F.pg* for 0, 24, 48, 72, and 96h at (**j**) transcript, (**k**) protein, and (**l**) acetylation levels in wheat. (**m**) Protein-protein interactions (PPI) network of 542 proteins in response to FCR. **n-s**, Relative expression, FCR DI (12 plants per replicate), and proline content of *TaP5CS*-overexpressed (*OE*) lines (**n-p**) and *TaHDA9-OE* lines (**q-s**) as well as their WTs. In (**f**), (**i**), (**j**), (**n**), (**p**), (**q**), and (**s**), the experiments were conducted with three biological replicates; data are presented as the mean values ± SD (*$P < 0.05$, ** $P < 0.01$; unpaired two-sided *t*-test; $n = 3$ biologically independent experiments). In (**o**) and (**r**), data are presented as the mean values ± SD (*$P < 0.05$, ** $P < 0.01$; unpaired two-sided *t*-test; $n = 5$ biologically independent experiments). In (**g**), different letters indicate significant differences ($P < 0.05$) determined by one-way analysis of variance (ANOVA) with Tukey' multiple comparisons test ($n = 5$ biologically independent experiments). Source data are provided as a Source Data file.



Supplementary Fig. 10. EMS mutants verified both *TaP5CS1* and *TaHDA9* modulating wheat FCR resistance.

a, Mutation sites of *TaP5CS1* in EMS-mutagenized Kronos mutants. (**b**, **c**) Phenotypes and FCR DI of *TaP5CS1* mutants infected by *F. pg* in greenhouse (12 plants per treatment) and field. **d**, The transcript level of *TaHDA9* across 20 tissues. **e**, Protoplast subcellular localization of *TaHDA9* in wheat leaves. **f**, The expression level of *TaHDA9* in response to FCR by qRT-PCR. **g**, Mutation sites of *TaHDA9* in EMS-mutagenized Kronos mutants. (**h**, **i**) Phenotypes and FCR DI of *TaHDA9* mutants infected by *F. pg* in greenhouse (12 plants per treatment) and field. In (**b**), (**c**), (**h**), and (**i**), the experiments were conducted with five biological replicates; data are presented as the mean values ± SD (*$P < 0.05$, ** $P < 0.01$; unpaired two-sided *t*-test; $n = 5$ biologically independent experiments). In (**e**), the experiments were repeated three times with similar results. Source data are provided as a Source Data file.



Supplementary Fig. 11. Exogenous proline enhanced FCR resistance.
(**a**) Hyphal growth of *F. pg* in potato dextrose agar (PDA) medium treated with 0, 1, 3, and 5, 7, 9, and 10 mM valine and proline at sixth day post-inoculation (DPI), respectively. (**b**) The growth inhibition rate of proline and valine. (**c**) Hyphal morphology of *F. pg* treated with 10 mM and 0 mM proline, respectively. (**d-e**) Phenotypes of plants and FCR DI treated with 0, 10, 20, and 30 mM exogenous proline, respectively. In (**a-c**), the experiments were conducted with three biological replicates; data are represented as the mean ± SD (** $P < 0.01$, unpaired two-sided *t*-test; $n = 3$ biologically independent experiments). In (**d-e**), the experiments were conducted with five biological replicates; data are presented as the mean values ± SD, different letters indicate significant differences ($P < 0.05$, one-way ANOVA with Tukey' multiple comparisons test; $n = 5$ biologically independent experiments). Source data are provided as a Source Data file.