Plant Biotechnology Journal Open Access

Plant Biotechnology Journal (2023), pp. 1–14 doi: 10.1111/pbi.14103

Efficient proteome-wide identification of transcription factors targeting *Glu-1*: A case study for functional validation of TaB3-2A1 in wheat

Lina Xie[1,2,†], Siyang Liu[1,†], Yong Zhang[1,*], Wenfei Tian[1,3], Dengan Xu[1], Jihu Li[1], Xumei Luo[1], Lingli Li[1], Yingjie Bian[1], Faji Li[1], Yuanfeng Hao[1], Zhonghu He[1,3], Xianchun Xia[1], Xiyue Song[2,*] and Shuanghe Cao[1,*]

[1]*Institute of Crop Sciences, National Wheat Improvement Centre, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China*
[2]*College of Agronomy, Northwest A&F University, Yangling, Shaanxi Province, China*
[3]*International Maize and Wheat Improvement Center (CIMMYT) China Office, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China*

Received 19 September 2022;
revised 13 January 2023;
accepted 5 June 2023.
**Correspondence* (Tel +86 10 82108610; Fax +86 10 82108745; email caoshuanghe@caas.cn (S. C.); Tel +86 10 82108745; Fax +86 10 82108745; email zhangyong05@caas.cn (Y. Z.) or Tel +86 029 87082845; Fax +86 29 87082845; email songxiyue@nwafu.edu.cn (X. S.))
[†]Lina Xie and Siyang Liu contributed equally to this work.

Keywords: B3 transcription factor, DNA pull-down plus LC–MS, *Glu-1* promoter, seed storage protein, *Triticum aestivum*.

Summary

High-molecular-weight glutenin subunits (HMW-GS), a major component of seed storage proteins (SSP) in wheat, largely determine processing quality. HMW-GS encoded by *GLU-1* loci are mainly controlled at the transcriptional level by interactions between *cis*-elements and transcription factors (TFs). We previously identified a conserved *cis*-regulatory module CCRM1-1 as the most essential *cis*-element for *Glu-1* endosperm-specific high expression. However, the TFs targeting CCRM1-1 remained unknown. Here, we built the first DNA pull-down plus liquid chromatography-mass spectrometry platform in wheat and identified 31 TFs interacting with CCRM1-1. TaB3-2A1 as proof of concept was confirmed to bind to CCRM1-1 by yeast one hybrid and electrophoretic mobility shift assays. Transactivation experiments demonstrated that TaB3-2A1 repressed CCRM1-1-driven transcription activity. *TaB3-2A1* overexpression significantly reduced HMW-GS and other SSP, but enhanced starch content. Transcriptome analyses confirmed that enhanced expression of *TaB3-2A1* down-regulated SSP genes and up-regulated starch synthesis-related genes, such as *TaAGPL3*, *TaAGPS2*, *TaGBSSI*, *TaSUS1* and *TaSUS5*, suggesting that it is an integrator modulating the balance of carbon and nitrogen metabolism. *TaB3-2A1* also had significant effects on agronomic traits, including heading date, plant height and grain weight. We identified two major haplotypes of *TaB3-2A1* and found that *TaB3-2A1-Hap1* conferred lower seed protein content, but higher starch content, plant height and grain weight than *TaB3-2A1-Hap2* and was subjected to positive selection in a panel of elite wheat cultivars. These findings provide a high-efficiency tool to detect TFs binding to targeted promoters, considerable gene resources for dissecting regulatory mechanisms underlying *Glu-1* expression, and a useful gene for wheat improvement.

Introduction

Wheat (*Triticum aestivum* L.) is a staple food source for approximately 35% of people worldwide. It can be processed into a wide range of food products, such as breads, steamed buns, noodles, cakes, biscuits and cookies (Veraverbeke and Delcour, 2002). The unique processing properties of wheat dough are due to its elasticity and extensibility determined by seed storage proteins (SSP) (Don *et al*., 2006; Shewry *et al*., 1995; Shewry *et al*., 2009; Veraverbeke and Delcour, 2002; Weegels *et al*., 1996; Wieser, 2007). High-molecular-weight glutenin subunits (HMW-GSs), a major component of SSP, are largely responsible for wheat dough elasticity and strength, and quantitative and qualitative differences in HMW-GS explain up to 70% of the variation in processing quality (Eagles *et al*., 2002; He *et al*., 2005; Liu *et al*., 2005; Payne *et al*., 1987).

HMW-GS were encoded by *GLU-1* loci including *GLU-1A*, *-1B* and *-1D* in homoeologous group 1 chromosomes of wheat; each of three loci contains tightly linked *Glu-1-1* and *Glu-1-2* genes encoding x- and y-type HMW-GS, respectively (Lawrence and Shepherd, 1981). As SSP, HMW-GS are primarily controlled at the transcriptional level by interactions between *cis*-acting motifs and transcription factors (TFs) (Kornberg, 2007; Vicente-Carbajosa and Carbonero, 2005). bZIP, DOF, MYB, NAC and B3 family members are major TFs regulating SSPs. For example, bZIP family TFs, including maize O2 (Schmidt *et al*., 1992), BARLEY LEUCINE ZIPPER 1 (BLZ1) and BLZ2 (Oñate *et al*., 1999; Vicente-Carbajosa *et al*., 1998), and rice OsbZIP58/RISBZ1 (Qian *et al*., 2020) activate expression of storage protein genes. SPA, a seed-specific bZIP protein in wheat, can activate transcription driven by the GLM motif of a low-molecular-weight glutenin subunit (LMW-GS) gene promoter (Albani *et al*., 1997), and its heterodimerizing protein (SHP) suppresses glutenin gene expression (Boudet *et al*., 2019). P-box binding factors (PBFs), members of plant-specific Dof TFs, activate expression of genes encoding maize zein (Wu and Messing, 2012), rice seed proteins (Yamamoto

Please cite this article as: Xie, L., Liu, S., Zhang, Y., Tian, W., Xu, D., Li, J., Luo, X., Li, L., Bian, Y., Li, F., Hao, Y., He, Z., Xia, X., Song, X. and Cao, S. (2023) Efficient proteome-wide identification of transcription factors targeting *Glu-1*: A case study for functional validation of TaB3-2A1 in wheat. *Plant Biotechnol. J.*, https://doi.org/10.1111/pbi.14103.

et al., 2006), barley hordein (Mena *et al*., 1998) and wheat α-gliadin and glutenin (Dong *et al*., 2007; Zhu *et al*., 2018). The GA-dependent MYB TF TaGAMyb recruits GCN5-like histone acetyltransferase and activates expression of *Glu-1* alleles by binding to the promoters (Guo *et al*., 2015). Wheat TaNAC100 was recently reported to act as an integrator modulating SSP and starch synthesis (Li *et al*., 2021). Another NAC TF TaNAC019 was also shown to regulate SSP and starch contents (Gao *et al*., 2021). A novel NAC TF, SPR, suppresses SSP synthesis in wheat (Shen *et al*., 2021). B3 TF proteins ABI3, LEC2 and FUS3 play important roles in SSP synthesis in Arabidopsis (Verdier and Thompson, 2008; Xi and Zheng, 2011). Barley HvFUS3 maintains a conserved function in the regulation of SSP genes by recognizing RY motifs (Moreno-Risueno *et al*., 2008). Wheat TaFUSCA3 transactivates *Glu-1Bx7* (Sun *et al*., 2017). Although a few TFs and *cis*-elements modulating the transcription of *Glu-1* genes have been identified, the underlying regulatory mechanism is largely unknown.

We previously defined conserved *cis*-regulatory modules (CCRMs) and functionally confirmed that all CCRMs had significant effects on expression of *Glu-1* (Li *et al*., 2019b). We subsequently performed yeast one hybrid (Y1H) library screening using the CCRMs as 'baits', but detected only two TFs binding to CCRMs (Li *et al*., 2021). Identification of more TFs targeting *Glu-1* is necessary for systematic dissection of the regulatory transcription mechanism conferring highly active endosperm-specific expression. Here we constructed an efficient DNA pull-down plus liquid chromatography-mass spectrometry (LC–MS) platform for wheat and identified many candidate TFs binding to CCRM1-1, an indispensable *cis*-regulatory module determining the spatio-temporal expression of *Glu-1*. The up-regulation of *TaB3-2A1* as a proof of concept reduced accumulation of HMW-GS and other SSP but increased the grain starch content. *TaB3-2A1* overexpression also significantly affected agronomic traits and its major haplotype *TaB3-2A1-Hap1* conferred lower SSP content but higher starch content and grain weight.

Results

High-throughput identification of TFs binding to the *Glu-1* promoter

CCRM1-1 (208 to 101 bp upstream of the start codon) in *Glu-1Dx2* promoter was identified as the most crucial *cis*-active regulatory module for highly active endosperm-specific expression of *Glu-1* in a previous study (Li *et al*., 2019b). The CCRM1-1 was used as a probe to pull down TFs modulating *Glu-1* from nuclear proteins in developing grains at 10 days post-anthesis (dpa) (Figure 1a). A total of 385 unique proteins were identified using LC–MS; these were classified into different groups, such as those annotated as transcription effectors (Figure 1b; Table 1; Datasets S1–S3), kinases (Figure 1b; Dataset S4), metabolic enzymes (Figure 1b; Dataset S5) and others (mainly involved in the protein degradation pathway) (Figure 1b; Dataset S6) according to the UniProt database (https://www.uniprot.org/) and the International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1.1 (https://wheat-urgi.versailles.inrae.fr/Seq-Repository/Assemblies) (IWGSC, 2018). Considering that developing grains contain many abundant metabolism-related enzymes to catalyse seed storage reserves, the detected kinases and metabolic enzymes are probably non-specific binding proteins for CCRM1-1. One hundred and eighteen putative transcription effectors binding to CCRM1-1 were identified, including 31 TFs (Figure 1b; Table 1), 34 TF partners (Figure 1b; Dataset S1), 18 transcription initiation or elongation factors (Figure 1b; Dataset S2) and 35 epigenetic modifiers, suggesting that the presence of multiple layers of transcriptional regulatory machinery underpinning expression of *Glu-1* (Figure 1b; Dataset S3). The identified TFs belonged to different families, including B3, MYB, bHLH, bZIP, MADS, NAC and WRKY (Table 1). We further investigated *cis*-elements in the CCRM1-1 region and identified canonical DNA motifs corresponding to most of the TFs based on the PLACE database (https://www.dna.affrc.go.jp/PLACE/?action=newplace) and previous studies (Figure 1c) (Aerts *et al*., 2018; Boer *et al*., 2014; Cao *et al*., 2007; Li *et al*., 2021; Reidt *et al*., 2000). We also analysed the expression patterns of the TFs in different tissues using RNA-Seq databases in the Wheat Expression Browser (http://wheat-expression.com) and observed that the majority of TFs were highly expressed in spikes or grains (Figure 1d). In addition, some TFs underwent changes in expression pattern during grain development (Figure 1d).

Confirmation of TaB3-2A1 binding to the *Glu-1* promoter

Among all TF subfamily proteins, those containing B3 domains were the most frequent. We chose *TaB3-2A1*, *TaB3-2B1* and *TaB3-2D1*, orthologous genes on chromosomes 2A, 2B and 2D, respectively, for validating the functions of identified TFs in modulating *Glu-1* expression. We designed primers to amplify the full-length coding sequences of the three orthologs according to IWGSC RefSeq v1.1 and observed that *TaB3-2A1* had the highest expression in developing grains. Thus, we used *TaB3-2A1* as a representative for further analysis. To confirm *TaB3-2A1* binding to CCRM1-1, we performed a Y1H assay. The 'bait' and 'prey' vectors containing CCRM1-1 and *TaB3-2A1*, respectively, were co-transformed into the yeast strain EGY48. Compared to negative controls, the co-transformed cells survived and harboured galactosidase activity on SD/−Trp/−Ura/−X-Gal medium, indicating that TaB3-2A1 could bind to CCRM1-1 in yeast (Figure 2a). Transactivation assays in tobacco (*Nicotiana benthamiana*) showed that *TaB3-2A1* significantly decreased the signal strength of the LUC reporter driven by the *Glu-1* promoter containing CCRM1-1 plus a basal region (100 to 1 bp upstream of the start codon) (Figure 2b). Likewise, luminescence activity driven by the *Glu-1* promoter was markedly repressed in the presence of TaB3-2A1 in Arabidopsis (*Arabidopsis thaliana*) protoplasts (Figure 2c). Thus, TaB3-2A1 appeared to behave as a transcriptional repressor to regulate *Glu-1* expression. Electrophoretic mobility shift assays (EMSA) verified the physical interaction between TaB3-2A1 and CCRM1-1 (Figure 2d). These results suggested that TaB3-2A1 directly binds to CCRM1-1 to negatively regulate *Glu-1* expression.

Quantitative PCR (qPCR) analysis showed that *TaB3-2A1* was widely expressed with the highest expression levels in young spikes and developing seeds (Figure 2e). In addition, *TaB3-2A1* transcripts were present at highest abundance at 5 dpa, sharply declined at 10 dpa, increased a little at 15 dpa, and then gradually decreased until 30 dpa, whereas *Glu-1* transcripts had the lowest expression levels at 5 dpa, increasing to a peak at 15 dpa (Figure 2f). Thus, *TaB3-2A1* had an overlapping, but opposing, temporal expression window to *Glu-1* in developing grains.

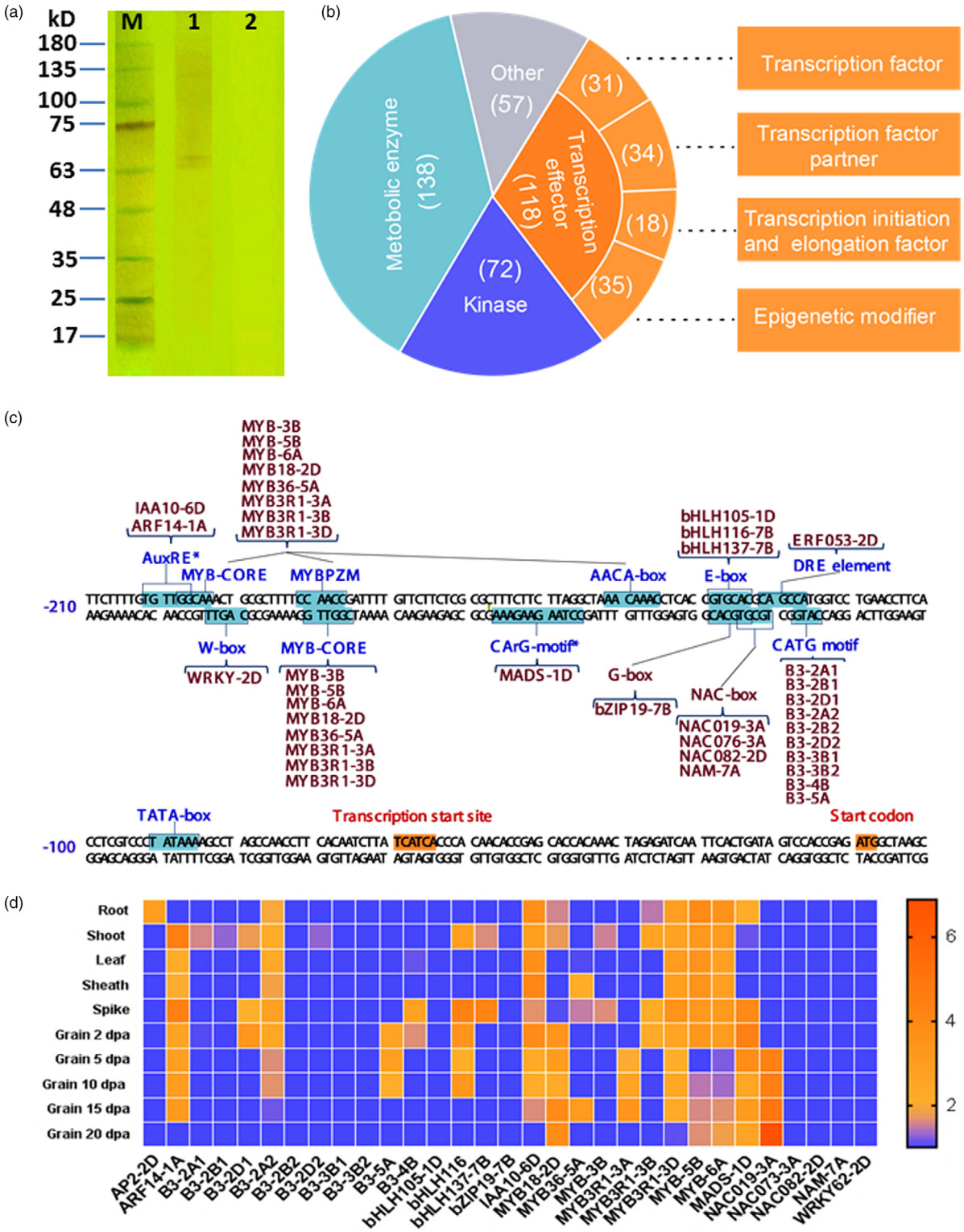

Figure 1 Identification of transcription factors (TFs) binding to the conserved *cis*-regulatory module CCRM1-1 in the *Glu-1* promoter. (a) Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and sliver staining display the proteins captured by DNA pulldown using CCRM1-1 as a probe. M, protein marker (Cat#RM19001, ABclonal); 1: proteins captured by magnetic beads tagged with CCRM1-1 probes; 2: proteins captured by untagged magnetic beads. CCRM1-1 spans 208–100 bp upstream of the start codon of *Glu-1Dx2* (used as a representative of *Glu-1* alleles). (b) Fan chart shows protein components binding to CCRM1-1 identified from DNA pulldown plus liquid chromatography-mass spectrometry. (c) Identified TFs were predicted to bind to *cis*-acting motifs in CCRM1-1 mainly based on PLACE databases (https://www.dna.affrc.go.jp/PLACE/?action=newplace). * motifs with similar canonical sequences; CATG motif is the same as the core sequence of RY motif. (d) Spatio-temporal expression profile of the TF-coding genes based on transcriptome databases in the Wheat Expression Browser (http://wheat-expression.com). dpa, days post-anthesis. The scale at the right was used to quantify relative expression levels of a gene in different tissues and developing grains at different phases.

Table 1 Transcription factors binding to the *Glu-1* promoter detected by DNA pull-down plus LC–MS.

Gene name*	Gene ID†	Protein ID‡	Times**	Gene anotation††
TaARF14-1A	*TraesCS1A02G334900*	A0A3B5Y4M5	3	Auxin response factor 14
TaB3-2A1	*TraesCS2A02G159800*	A0A3B6AVE6	3	B3 domain-containing protein Os06g0194400 && B3 DNA binding domain
TaB3-2B1	*TraesCS2B02G185500*	A0A3B6C268	3	B3 domain-containing protein Os06g0194400 && B3 DNA binding domain
TaB3-2D1	*TraesCS2D02G167000*	A0A7H4LDZ5	3	B3 domain-containing protein Os06g0194400 && B3 DNA binding domain
TaB3-2A2	*TraesCS2A02G144100*	A0A341Q7B8	2	B3 domain-containing protein Os01g0234100 && PF02362:B3 DNA binding domain
TaB3-2B2	*TraesCS2B02G185600*	A0A3B6C4S8	2	B3 domain-containing protein Os06g0194400
TaB3-2D2	*TraesCS2D02G148000*	A0A1D5V095	2	B3 domain-containing protein Os01g0234100 && B3 DNA binding domain
TaB3-3B1	*TraesCS3B02G348400*	A0A3B6FRM7	2	B3 domain-containing protein Os03g0212300 && B3 DNA binding domain
TaB3-3B2	*TraesCS3B02G530300*	A0A3B6FXP4	2	B3 domain-containing protein Os03g0212300 && B3 DNA binding domain
TaB3-5A	*TraesCS5A02G387400*	A0A3B6KMP2	2	B3 domain-containing protein Os11g0197600 && B3 DNA binding domain
TaB3-4B	*TraesCS4B02G249100*	A0A3B6IS61	2	B3 domain-containing protein Os03g0212300 && B3 DNA binding domain
TabHLH105-1D	*TraesCS1D01G280600*	A0A3B5ZYI5	1	Transcription factor ILR3; Full = Basic helix–loop–helix protein 105; Full = Protein IAA-LEUCINE RESISTANT 3
TabHLH116-7B	*TraesCS7B02G211600*	A0A3B6SLU6	1	Transcription factor ICE1; Full = Basic helix–loop–helix protein 116
TabHLH137-7B	*TraesCS7B02G161700*	A0A3B6SG30	1	Transcription factor bHLH137; Full = Basic helix–loop–helix protein 137
TabZIP19-7B	*TraesCS7B02G321300*	A0A3B6SHK7	1	Basic leucine zipper 19; Short = AtbZIP19; Short = bZIP protein 19
TaERF053-2D	*TraesCS2D02G369900*	A0A1D5UG83	1	Ethylene-responsive transcription factor ERF053 && AP2 domain
TaIAA10-6D	*TraesCS6D01G378300*	A0A3B6QP65	3	Auxin-responsive protein IAA10; AltName: Full = Indoleacetic acid-induced protein 10 && AUX/IAA family
TaMADS-1D	*TraesCS1D02G264500*	A0A1D5SWM2	2	WHEAT PISTILLATA-like MADS box protein OS = Triticum aestivum GN=WPI1
TaMYB-3B	*TraesCS3B02G387000*	A0A3B6FWC9	1	Myb/SANT-like DNA-binding domain
TaMYB-5B	*TraesCS5B02G262800*	A0A3B6LNZ0	2	Myb/SANT-like DNA-binding domain
TaMYB-6A	*TraesCS6A02G345400*	A0A3B6NVC3	2	Myb-like DNA-binding domain
TaMYB18-2D	*TraesCS2D02G366800*	A0A1D5URG7	1	Transcription factor LAF1; AltName: Full = Myb-related protein 18
TaMYB36-5A	*TraesCS5A01G217900*	A0A3B6KG82	1	Transcription factor MYB36; AltName: Full = Myb-related protein 36
TaMYB3R1-3A	*TraesCS3A02G359200*	A0A3B6ENC0	2	Myb-related protein 3R-1; AltName: Full = Plant c-MYB-like protein 1; Myb-like DNA-binding domain
TaMYB3R1-3B	*TraesCS3B01G391900*	A0A077S4V4	3	Myb-related protein 3R-1; AltName: Full = Plant c-MYB-like protein 1; Myb-like DNA-binding domain
TaMYB3R1-3D	*TraesCS3D02G353500*	A0A3B6GYJ2	2	Myb-related protein 3R-1; AltName: Full = Plant c-MYB-like protein 1; Myb-like DNA-binding domain
TaNAC073-3A	*TraesCS3A02G377700*	A0A3B6ELE9	1	NAC domain-containing protein 73; Protein SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 2 No apical meristem (NAM) protein
TaNAC082-2D	*TraesCS2D01G568000*	A0A7H4LQS5	1	NAC domain-containing protein 82; AltName: Full = Protein VND-INTERACTING 1 && No apical meristem (NAM) protein
TaNAM-7A	*TraesCS7A02G209300*	A0A3B6RGK3	1	No apical meristem (NAM) protein
TaWRKY62-2D	*TraesCS2D02G497600*	A0A3B6DLR4	3	Probable WRKY transcription factor 62; AltName: Full = WRKY DNA-binding protein 62

*Temporally designated names.

†Gene accession numbers are retrieved from EnsemblPlants (http://plants.ensembl.org/Triticum_aestivum) according to Protein ID.

‡Protein ID are from the non-redundant wheat database in UniProt (https://www.uniprot.org/taxonomy/4565).

**Number of detections of target proteins.

††Gene annotations are from IWGSC RefSeq v1.1 (https://wheat-urgi.versailles.inrae.fr/Seq-Repository/Assemblies). LC–MS, liquid chromatography-mass spectrometry.

TaB3-2A1 suppresses expression of *Glu-1*

We generated lines overexpressing (OE) *TaB3-2A1* in the wheat cultivar Fielder to confirm its function. Nine independent transgenic lines were obtained and positive lines OE2, OE4 and OE5 with the highest expression levels of *TaB3-2A1* were used in subsequent analyses. We measured HMW-GS in mature grains using ultra-performance liquid chromatography (UPLC) and found that *TaB3-2A1*-OE lines had significantly reduced the HMW-GSs Bx14, By15, Dx2 and Dy12 compared with corresponding transgenic null lines (TNL) (Figure 3a–e). The total HMW-GS contents in *TaB3-2A1*-OE lines were significantly reduced by 15.53–23.27% compared with those in the TNL (Figure 3f; Table S1). Transcriptome analysis also validated that all active *Glu-1* alleles in Fielder, including *Glu-1Bx14*, *Glu-1By15*, *Glu-1Dx2* and *Glu-1Dy12*, were down-regulated in developing grains of *TaB3-2A1*-OE lines (Table S2). These results indicate that TaB3-2A1 functions as a transcriptional repressor in *Glu-1* gene expression.

Effects of *TaB3-2A1* on other SSP and starch content

In addition to *Glu-1*, the genes encoding other SSP, such as LMW-GS, gliadin and avenin, were differentially expressed between *TaB3-2A1*-OE lines and TNL according to transcriptome data (Table S2; Dataset S7). Quantitative protein analysis showed that LMW-GS contents in grains of *TaB3-2A1*-OE lines were

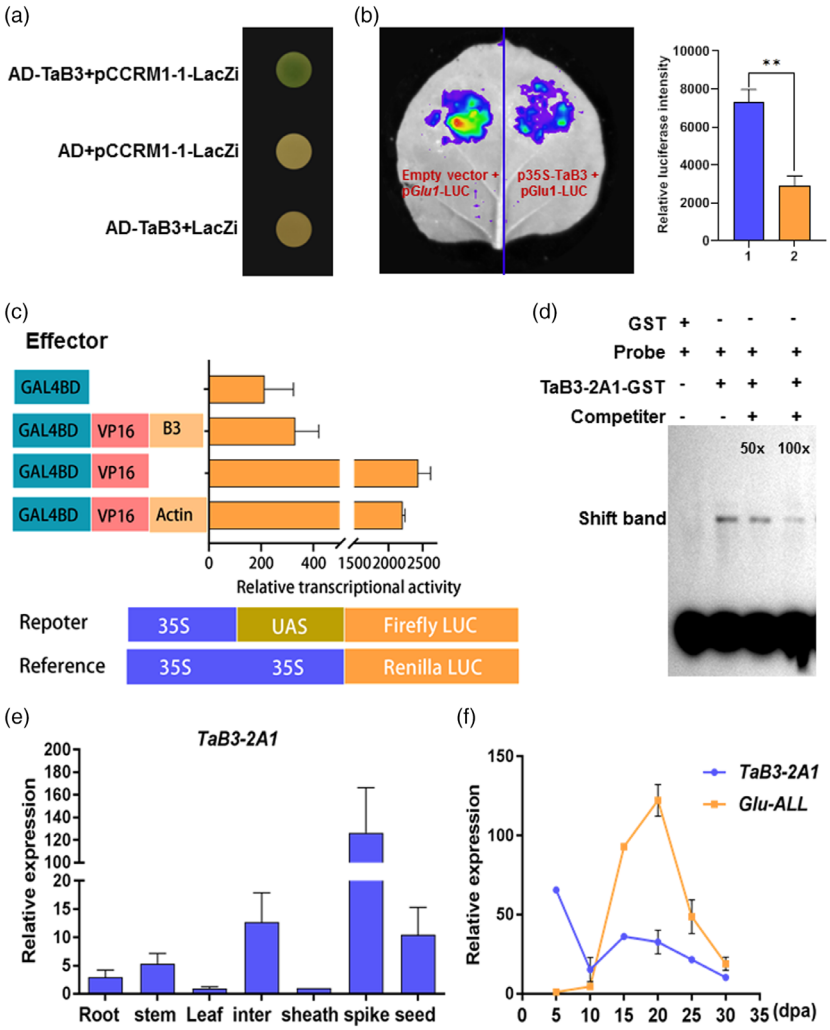

Figure 2 Confirmation of TaB3-2A1 binding to the conserved *cis*-regulatory module CCRM1-1 and its expression patterns. (a) Yeast one-hybrid assay validates TaB3-2A1 binding to CCRM1-1. pCCRM1-1 represents 208 to 100 bp upstream of the start codon of *Glu-1Dx2* (used as a representative of *Glu-1* alleles). AD+pCCRM1-1-LacZi and AD-TaB3 + LacZi were used as negative controls. (b) Transactivation assays for TaB3-2A1 binding to the CCRM1-1 in tobacco. The colour change from red to blue indicates fluorescence signal intensity from high to low. The right panel shows quantitative fluorescence signals produced from two parts of the leaves just as shown in the left panel. $^{**}P < 0.01$. (c) Transactivation assays for TaB3-2A1 binding to CCRM1-1 in Arabidopsis protoplasts. (d) Electrophoretic mobility shift assay (EMSA) confirms physical interaction between TaB3-2A1 and CCRM1-1. (e) Tissue expression analysis of *TaB3-2A1*. (f) Comparisons of dynamic expression of *TaB3-2A1* and *Glu-1* during seed development. *Glu-ALL* represents all active genes at the *GLU-1* loci in wheat cultivar Fielder. dpa, days post-anthesis.

significantly decreased by 11.29–17.79% compared with that of TNL (Figure 3g; Table S3). *TaB3-2A1*-OE lines also had lower seed protein contents (reduced by 9.81–16.31%) than TNL (Figure 3h; Table S3).

Considering the usual trade-off between storage protein and starch accumulations in grains, we compared starch contents between *TaB3-2A1*-OE lines and TNL (Figure 3i); *TaB3-2A1*-OE lines had higher starch contents (increased by 1.94–3.22%) in grains than TNL (Figure 3j; Table S3). Transcriptome analyses also revealed that the highest percentage of differentially expressed genes in developing grains between *TaB3-2A1*-OE lines and TNL was in starch and sucrose metabolism pathways (Table S4). Among them, starch synthesis-related genes *TaAGPL3*, *TaAGPS2*, *TaGBSSI/TaWx1*, *TaSUS1* and *TaSUS5* were up-regulated by *TaB3-2A1* (Table S4; Dataset S7).

TaB3-2A1 has significant effects on agronomic traits

In addition to seeds, spikes and stems had high expression of *TaB3-2A1* (Figure 2e), suggesting that the gene participates in multiple processes of growth and development. We investigated major agronomic traits including plant height (PH), heading date (HD), tiller number, spike number per m^2 (SN), kernel number per spike (KNS) and thousand kernel weight (TKW) in *TaB3-2A1*-OE lines and corresponding TNL (Figure 4a). *TaB3-2A1*-OE lines had higher PH (increased by 3.7–6.3%) and earlier HD (2.6–3.2 days) than TNL (Figure 4b,c; Table S5). *TaB3-2A1* had no significant effect on total tiller or effective tiller (spike) number per plant (Figure 4d,e) and there was no significant difference in SN (Figure 4f) and KNS (Figure 4g) between *TaB3-2A1*-OE lines and TNL. However, *TaB3-2A1*-OE lines had higher TKW than TNL

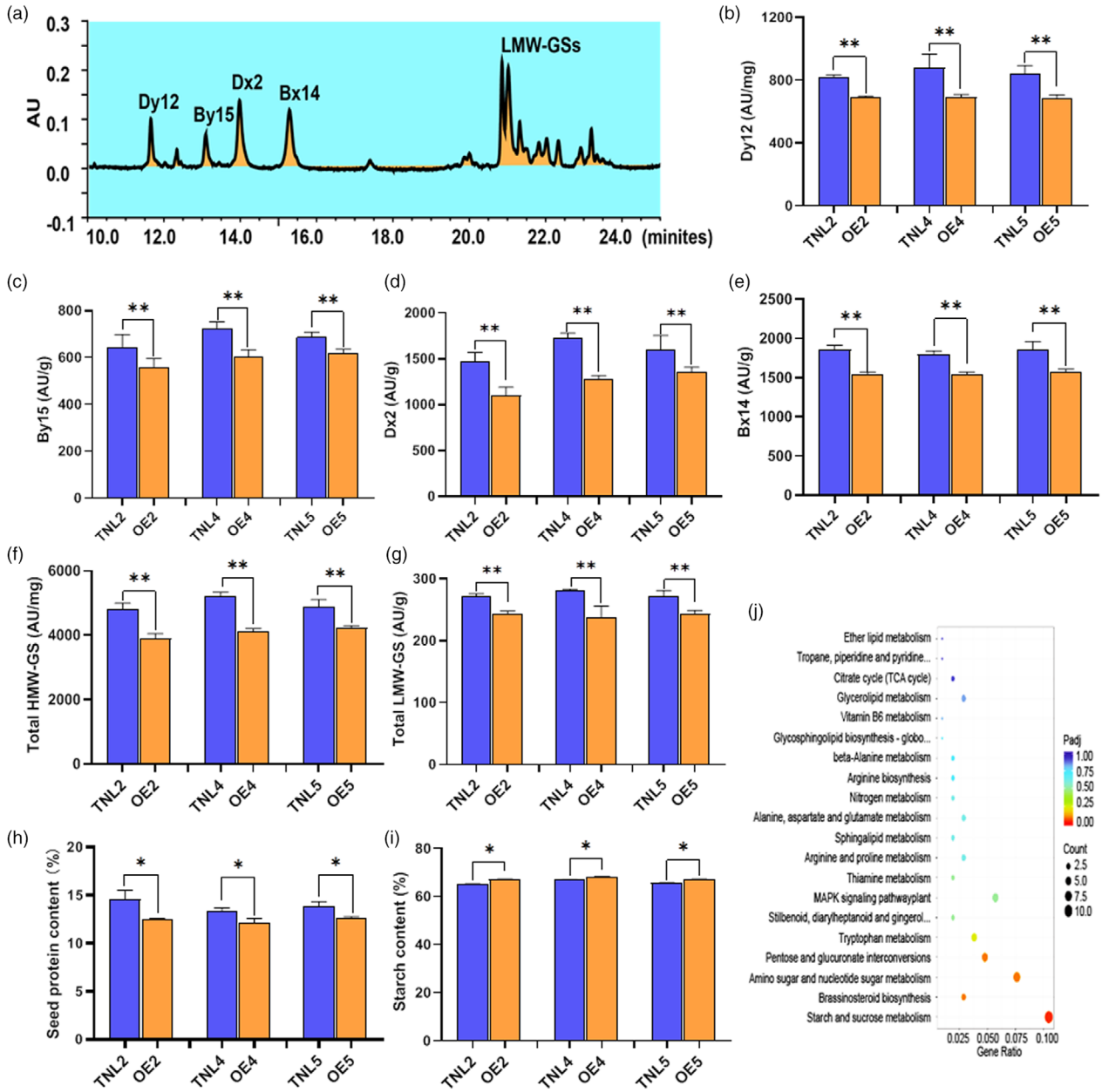

Figure 3 Effects of *TaB3-2A1* overexpression (OE) on seed protein and starch contents. (a) UPLC analysis of high-molecular-weight glutenin subunits (HMW-GSs) and low-molecular-weight glutenin subunits (LMW-GSs). Statistical comparisons of Dx2 (b), Dy12 (c), Bx14 (d), By15 (e), total HMW-GS (f), total HMW-GS (g), seed protein content (h) and starch content (i) in mature grains between *TaB3-2A1*-OE lines and transgenic null lines (TNL). Dx2, Dy12, Bx14 and By15 are the HMW-GSs encoded by *Glu-1Dx2*, *Glu-1Dy12*, *Glu-1Bx14* and *Glu-1By15*, respectively. The HMW-GS and LMW-GS contents are shown as signal strength per mg flour (AU/mg). OE2, OE4, and OE5 are independent lines overexpressing *TaB3-2A1*. $*P < 0.05$, $** P > 0.01$. (j) Transcriptome analysis revealed that the highest percentage of differentially expressed genes in developing grains between *TaB3-2A1*-OE lines and TNL was in the starch and sucrose metabolism pathways.

(Figure 4h) with increased grain length and width, suggesting that *TaB3-2A1* affects grain weight largely by modulating grain size (Figure 4i–l). Collectively, TaB3-2A1 was a positive regulator of HD, PH and TKW.

Haplotype variation and genetic effects of *TaB3-2A1*

To identify sequence variants of *TaB3-2A1*, polymorphic sites in open reading frames (ORF) and flanking regions (i.e., 2-kb upstream of the start codon and 1-kb downstream of the stop codon) were retrieved from the wheat re-sequencing databases in the SnpHub (http://wheat.cau.edu.cn/Wheat_SnpHub_Portal/). Four SNPs in ORF and six SNPs close to the start or stop codons of the *TaB3-2A1* locus formed two major haplotypes, *TaB3-2A1-Hap1* and *TaB3-2A1-Hap2* (Figure 5a). Four SNPs cause amino acid changes (Figure 5a). The amino acid changes caused by the SNP^{A523T} and SNP^{C573T} were in the DNA-binding pseudo-barrel protein fold domain, suggesting they were putatively functional polymorphic sites (Figure 5a). A KASP marker, TaB3-2A1-$KASP^{A523T}$ was developed and used to genotype a panel of 166 elite wheat cultivars from the Huang-Huai River Valley region

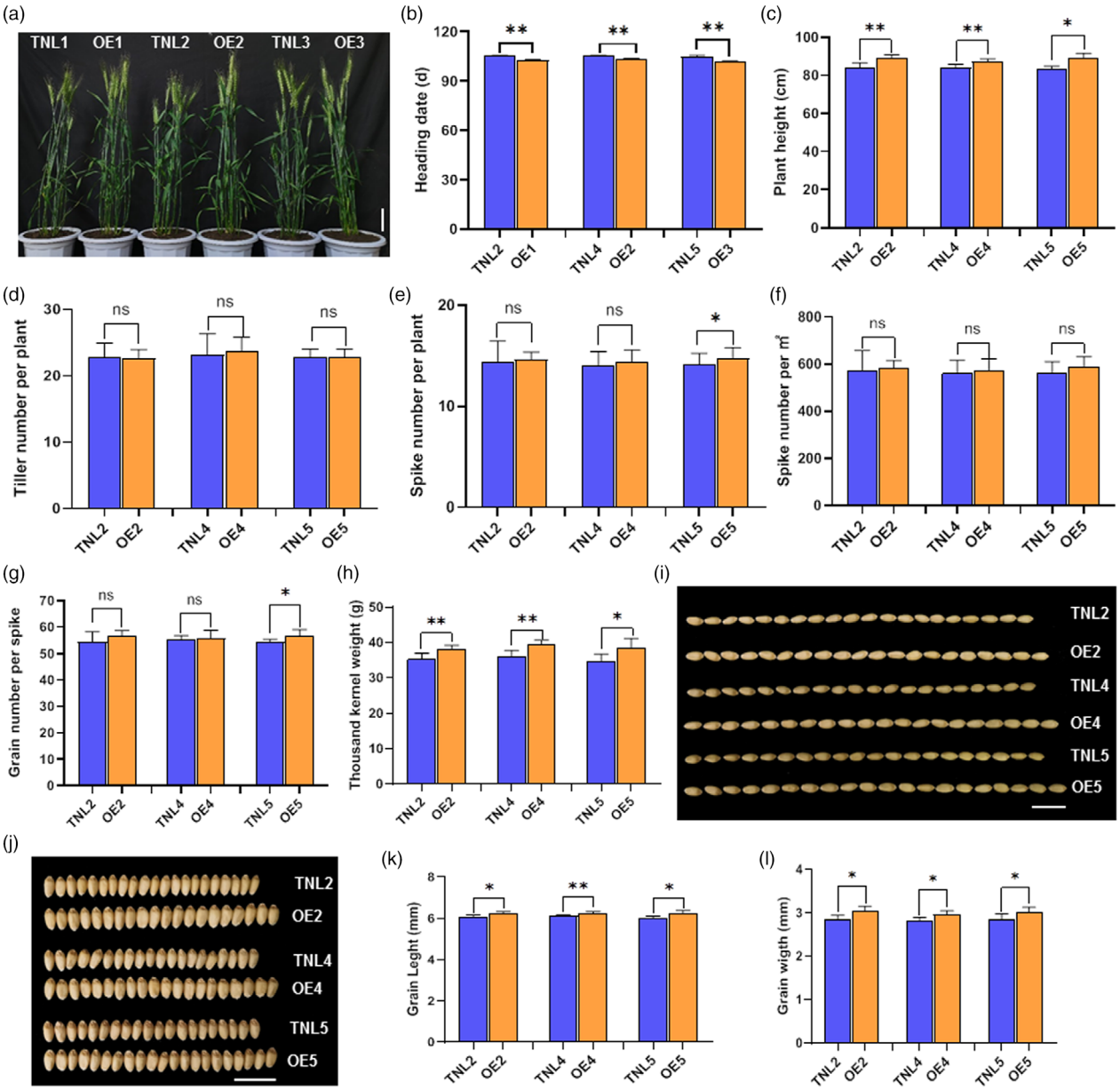

Figure 4 Effects of *TaB3-2A1* overexpression (OE) on agronomic traits. (a) Morphologic visualization of *TaB3-2A1*-OE lines and transgenic null lines (TNL). Bar, 10 cm. Statistical comparisons of heading date (b), plant height (c), tiller number per plant (d), spike number per plant (e), spike number per m^2 (f), grain number per spike (g), thousand kernel weight (h), grain length (k) and grain width (l) between *TaB3-2A1*-OE lines and TNL. OE2, OE4 and OE5 are independent lines overexpressing *TaB3-2A1*. *$P < 0.05$; ** $P > 0.01$; ns, not significantly different. Morphologic comparisons of grain length (i) and grain width (j) between *TaB3-2A1*-OE lines and TNL. Bars, 1 cm.

(Figure 5b; Dataset S8). Association analysis showed that *TaB3-2A1-Hap1* conferred significantly lower seed protein content and earlier heading date than *TaB3-2A1-Hap2* (Figure 5c–e). *TaB3-2A1-Hap1* was also associated with enhanced starch content, PH, TKW and yield, albeit not significantly (Figure 5d, f,h,i; Table S6). We also investigated the effect of *TaB3-2A1* on KNS and did not find a significant difference in this trait of the natural population (Figure 5g). Moreover, *TaB3-2A1-Hap1* occurred at a higher frequency (69.86%) than *TaB3-2A1-Hap2* (30.14%) (Dataset S9). To further display the breeding selection for *TaB3-2A1-Hap1*, we investigated the major haplotypes of *TaB3-2A1* in a wheat genome resequencing database including 308 wheat lines (Dataset S10) and observed that *TaB3-2A1-Hap1* also had higher frequency (65.55%) than *TaB3-2A1-Hap2* (34.45%) (Dataset S11). We calculated the frequencies of *TaB3-2A1* haplotypes in elite wheat cultivars released from different periods in the Huang-Huai Valley of China (Dataset S12). Likewise, *TaB3-2A1-Hap1* had a higher frequency (75.68%) than *TaB3-2A1-Hap2* (24.32%) in the cultivars released after 1985; reversely, the former had a lower frequency (35.00%) than the latter in the cultivars released before 1985 (Dataset S13). These results showed that *TaB3-2A1-Hap1* had been subjected to positive selection in modern breeding programs.

(a)

TraesCS2A02G159800

ATG TGA

SNP position	-1018	-954	-946	-835	48	523	573	1288	1315	2118	2222	No. of accessions
Hap1	G	T	C	T	C/Glu	A/Leu	C/Ala	C/Glu	T	A	A	102
Hap2	T	C	T	C	A/Asp	T/Gln	T/Thr	T/Lys	G	G	G	44

TaB3-2A1-Hap1 74
TaB3-2A1-Hap2 74

TaB3-2A1-Hap1 148
TaB3-2A1-Hap2 148

DNA-binding pseudobarrel domain

TaB3-2A1-Hap1 219
TaB3-2A1-Hap2 219

(b) (c) (d) (e)

(f) (g) (h) (i)

Figure 5 Effects of *TaB3-2A1* haplotypes on seed protein, starch and agronomic traits. (a) Major *TaB3-2A1* haplotypes, Hap1 and Hap2, and the alignment of their deduced amino acid sequences. The polymorphic site shown in red was used to develop a haplotype-specific marker for *TaB3-2A1*. (b) Genotyping display of a panel of wheat cultivars using the haplotype-specific marker for *TaB3-2A1*. KASP, kompetitive allele-specific PCR. Comparisons of seed protein content (c), starch content (d), heading date (e), plant height (f), kernel number per spike (g), thousand kernel weight (h) and grain yield (i) between the cultivars with contrasting haplotypes Hap1 and Hap 2.

Discussion

DNA pull-down plus LC–MS is a highly efficient tool to detect TFs binding to *cis*-regulatory elements in target genes

Glu-1 genes encode HMW-GS with large effects on wheat quality and their expression is mainly regulated at the transcriptional level by interactions between *cis*-acting elements and TFs. We previously identified CCRMs and validated their functions in determining highly active endosperm-specific expression of *Glu-1* (Li *et al*., 2019b). However, the TFs interacting with CCRMs are largely unknown. We tried to detect the TFs binding to CCRMs using yeast one hybrid (Y1H) screening assays, but no TF was captured in a Y1H library prepared from developing grains (Li *et al*., 2021). In the current study, using the DNA pulldown plus LC–MS platform we identified 31 TFs binding to CCRM1-1, the most important *cis*-acting regulatory module for endosperm-specific high expression of *Glu-1*. Among them, TaNAC019 was identified as a transcriptional activator of *Glu-1* (Gao *et al*., 2021). We confirmed that TaB3-2A1 modulated the expression of *Glu-1*. Our DNA pulldown plus LC–MS platform for wheat successfully identified a group of TFs binding to the *Glu-1* promoter, providing not only a useful high-throughput tool to mine TFs modulating target genes, but also a considerable genetic resource for systematic dissection of the regulatory transcriptional mechanism underlying *Glu-1* expression.

TFs targeting *Glu-1* are potential coordinators of quality and yield

Many TFs modulating SSP synthesis have pleiotropic effects on seed development and overall plant metabolism. Storage proteins and starch are synthesized synchronously in the endosperm during grain filling, suggesting coordinated regulation of the two storage products. The promoters of storage protein and starch biosynthesis genes usually share *cis*-active motifs that may allow co-regulation of the storage products by common TFs (Zhang *et al*., 2019). Quite a few TFs function in coordinating the expression of starch biosynthesis and storage protein genes. Maize O2 and PBF1 in maize regulate storage protein accumulation and starch biosynthesis (Zhang *et al*., 2016). O11, an endosperm-specific bHLH TF, directly regulates the expressions of O2 and PBF1 in coordinating the starch biosynthesis and storage protein accumulation (Feng *et al*., 2018). ZmNAC128 and ZmNAC130 activate *16-kD γ-zein* and *AGPS2* (encoding a small subunit of maize AGPase) genes by binding to their common ACGCAA motif. Likewise, OsNAC20 and OsNAC26 modulate *16-kD prolamin* and starch-branching enzyme genes by targeting the same motif (Wang *et al*., 2020). MYB TF HvMCB1 activates expression of SSP genes in barley but inhibits expression of amylase gene *Amy6.4* by recognizing their GATA motifs (Rubio-Somoza *et al*., 2006). TaNAC019 binds directly to a common motif in the promoters of *Glu-1*, *SUS1* (encoding a sucrose synthase) and *SSIIa* (encoding starch synthase IIa) genes to activate their expression (Gao *et al*., 2021). TaNAC100 regulates seed protein and starch synthesis in wheat and also has significant effects on some agronomic traits. Here we confirmed that TaB3-2A1 is an integrator modulating SSP and starch synthesis, and also affecting agronomic traits, such as PH, HD, TGW and grain size. It would be important and challenging to establish the transcriptional regulation network underlying the coordination of starch biosynthesis and storage protein accumulation. The identified TFs targeting *Glu-1* probably function in modulating storage protein and starch, providing substantial clues to orchestrate the biosynthesis of the two storage products for wheat improvement. Moreover, many studies showed that a few TFs, such as FUS3, Q2 and PBF, had conserved function in modulating SSP genes across plant species, especially cereal crops (reviewed in Li *et al*., 2021), so the identified TFs targeting *Glu-1* also provided gene information to ascertain transcriptional controls of SSP genes in other plants.

B3 TFs play important roles in wheat seed development

Plant-specific B3 TFs belong to a large family, that is, involved in multiple biological processes (Swaminathan *et al*., 2008). B3 TFs FUS3, ABI3 and LEC2 modulate seed development in Arabidopsis (Suzuki and McCarty, 2008). HvFUS3 in barley affects expression of SSP genes (Moreno-Risueno *et al*., 2008). TaFUSCA3 in wheat promotes *Glu-1Bx7* expression (Sun *et al*., 2017). Viviparous1 (Vp1), a homologue of ABI3, participates in seed dormancy and preharvest spouting resistance in cereal crops (Giraudat *et al*., 1992; McCarty *et al*., 1989; Yang *et al*., 2007). Here TaB3-2A1 was functionally validated to repress *Glu-1* expression. Enhanced expression of *TaB3-2A1* reduced accumulation of other SSP but enhanced grain starch content. Transcriptome analysis confirmed that TaB3-2A1 down-regulated many genes encoding SSP and up-regulated starch synthesis-related genes, such as *TaAGPL3*, *TaAGPS2*, *TaGBSSI*, *TaSUS1* and *TaSUS5* (Table S6; Dataset S7). These findings indicate that TaB3-2A1 plays a role in the synergistic regulation of SSP and starch synthesis. In addition, *TaB3-2A1* overexpression had significant effects on several agronomic traits, including HD, PH and TKW, suggesting that it acts as a modulator of plant architecture and yield in wheat. As such, we proposed a working module of TaB3-2A1 in modulating SSP, starch and agronomic traits (Figure S1). The more frequent haplotype *TaB3-2A1-Hap1* conferred lower SSP content, but higher starch content, TKW and grain yield than *TaB3-2A1-Hap2*. Clearly, *TaB3-2A1* is a useful gene for wheat improvement. We also developed a breeder-friendly marker to distinguish the two major haplotypes of *TaB3-2A1*, providing an efficient tool for marker-assisted selection in wheat breeding.

Quite a few B3 TFs in addition to TaB3-2A1 were identified to bind to the *Glu-1* promoter. Phylogenetic analysis showed that the B3 TFs belonged to diverse subfamilies, suggesting that such TF family members widely participate in modulating *Glu-1* expression (Figure S2). TaB3-2A1 belongs to the REMD subfamily of B3 TFs, from which no TF was previously reported to regulate grain development. Members of this subfamily were greatly expanded by tandem and segmental duplication events, enriching gene pools for fine-tuning seed development (Figure S3). Since TaB3-2A1 is an important regulator of grain development and plant morphogenesis in wheat, it is necessary to identify the function of other B3 TFs, especially those binding to the *Glu-1* promoter. To further verify the function of *TaB3-2A1* and its orthologs, it is also necessary to generate the loss of function mutants using gene editing or ethyl methanesulfonate (EMS) mutagenesis technology.

Materials and methods

Plant material preparation and agronomic trait investigation

Transgenic recipient Fielder and its transgenic derivatives were grown in the greenhouse and experimental field at Jinan in

Shandong province (N36°39′ E117°06′). Greenhouse conditions were set at 10–22 °C under a 16 h light/8 h darkness cycle with supplementary light provided by high-pressure sodium vapour lamps (Powertone SON-T AGRO 400 W; Philips Electronic UK Ltd, Farnborough, UK). Tobacco (*Nicotiana benthamiana*) and Arabidopsis were grown in a greenhouse at 22 °C and a 16 h light/8 h darkness regime. A natural population including 166 elite wheat cultivars from the Huang-Huai region, a leading wheat production zone in China, were grown at Xinxiang (N 35°18′, E 113°55′) in Henan province and Suixi (N 33°38′, E 116°54′) in Anhui during the 2012–2013 and 2013–2014 cropping seasons.

To investigate agronomic traits, the transgenic lines and corresponding TNL were grown in 2-m-row plots with three replications. Seed protein and starch contents of the natural population above were from Li *et al*. (2021) and the phenotypic data for PH, HD, TKW, KNS, SN and grain size were available in Li *et al*. (2019a).

Nuclear protein extraction and DNA pull-down plus LC–MS assays

Developing grains at 10 dpa were harvested for nuclear protein extraction following the user manual of the Nuclear and Cytoplasmic Protein Extraction kit (Cat#P0028, Beyotime, Beijing). Three grains from each line were ground into powder in liquid nitrogen and mixed with 200 μL of extraction buffer. The protein extraction procedure was performed three times to obtain sufficient nuclear proteins for DNA pull-down assays.

Single-strand probes from *CCRM1-1* for DNA pull-down were synthesized and labelled with biotin tags at Sangon Biotech (Beijing). Double-strand probes were prepared by annealing complementary *CCRM1-1* single-strand probes following the procedure: denatured at 98 °C for 5 min and then incubated at room temperature for 10 min. The probe sequences are listed in Table S7. DNA pull-down assays were performed in four steps (I-IV) as follows: I, preparing beads: (1) pipet 250 μL Dynabeads M-280 Streptavidin (Cat#11205D, Invitrogen, Shanghai, China) into a fresh microfuge tube and vigorously shake Dynabeads to resuspend in the company-supplied preservative; (2) secure the microfuge tube on a PolyATract 1000 magnet (Cat#CS15000, Promega) to pull Dynabeads down; (3) wash the beads with 500 μL of 2 × B/W buffer [10 mM Tris–HCl (pH 7.5), 1 mM EDTA, 2 M NaCl] for three times; II, assembling probe-beans: (1) resuspend the beads in 190 μL of 2 × B/W buffer and add 200 μL biotinylated probe (200–400 ng/μl); (2) incubate the mixture while rolling at room temperature for 20 min; (3) pull beads down by applying to the PolyATract 1000 magnet; (4) repeat the steps (1)–(3) to ensure that the beads are saturated with DNA; (5) wash the probe-beads with 500 μL of BS/THES buffer [22 mM Tris–HCl (pH 7.5), 4.4 mM EDTA, 8.9% sucrose (m/v), 62 mM NaCl, 0.3% protease inhibitor, 0.04% phosphatase inhibitor, 10 mM HEPES, 5 mM $CaCl_2$, 50 mM KCl, 12% glycerol] to ensure that the DNA probe is in the reaction conditions suitable for DNA-protein interactions; III, capturing target proteins binding to the probe: (1) apply 200 μL of BS/THES buffer to probe-bead complex along with 600–750 μL of cleared lysate (supernatant); (2) add 25–100 μg of Poly dI-dC (10 μg/mL) (Cat#20148E, ThermoFisher Scientific) and roll the mixture at 4 °C for 30 min to provide an excess of non-specific competitor DNA; (3) pull bead-probe-protein complex down using magnet, discard supernatant and wash with 500 μL of BS/THES buffer supplemented with 10 μL Poly dI-dC for five times; IV, eluting target: (1) suspend the bead-probe-protein complex in 120 μL of elution buffer (25 mM Tris–HCl, 100 mM NaCl) and roll at room temperature for 3–5 min, (2) pull down beads and save elution protein to store at −20 °C. The captured proteins were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by silver staining. Identification of the proteins binding to the probe was achieved using LC–MS.

LC–MS was performed by Zhongke New Life (Shanghai) Protein digestion was performed using the filter-aided sample preparation method (Wiśniewski *et al*., 2009). The peptides of each sample were purified on Empore™ C18 SPE Cartridges (CDS Analytical, Oxford, Pennsylvania, USA), concentrated by vacuum centrifugation and reconstituted in 40 μL of 0.1% (v/v) formic acid. The peptide content was measured by UV light spectral density (Multiskan SkyHigh, ThermoFisher Scientific) at 280 nm. Analysis of digested peptides was performed on a Q Exactive™ mass spectrometer (Cat#IQLAAEGAAPFALGMAZR, ThermoFisher Scientific, Waltham, MA, USA) coupled to an Easy nLC (Cat#LC120, ThermoFisher Scientific). Mobile phase A was a 0.1% formic acid solution; mobile phase B was a solution containing 84% acetonitrile (ACN) and 0.1% formic acid. The flow rate was 300 nL/min. The mass spectrometer was operated in positive mode. Automatic gain control (AGC) target, maximum inject time and dynamic exclusion duration were set to 1e6, 40 ms and 30 s, respectively. MS data were acquired at a resolution of 70 000 (m/z 200) using a data-dependent method dynamically choosing the most abundant precursor ions from 300–1800 m/z for high-energy collisional dissociation (HCD) fragmentation. The resolution for HCD spectra was set to 17 500 at m/z 200 with 2 m/z isolation width and 50 ms maximum inject time. Normalized collision energy was 27 eV and the underfill ratio was defined as 0.1%. The resulting MS spectra were searched using MASCOT engine (Matrix Science, London, UK; version 2.2) against a non-redundant wheat database (uniprot-Triticum aestivum-142 941–20 190 524.fasta) from UniProt (https://www.uniprot.org/taxonomy/4565).

Expression pattern and evolutionary analyses

Expression patterns of TFs in different tissues and developmental stages were analysed according to the transcriptome database from Wheat Expression Browser (http://www.wheat-expression.com/; Ramirez-Gonzalez *et al*., 2018). Expression patterns were estimated as transcripts per million (TPM) using the log2TPM + 1 value through the heatmap package of R 3.6.3 for Windows.

Roots, stems, leaf sheaths, leaves and spikes were collected at the heading stage to analyse spatiotemporal expression patterns of *TaB3-2A1*. Grains were harvested at 5, 10, 15, 20, 25 and 30 dpa. Total RNA in different tissues were extracted using an RNAprep Pure Plant Kit (TIANGEN). Genomic DNA removal and first-strand cDNA synthesis were performed by a PrimeScript RT Reagent Kit (Takara, Ohtsu, Japan). Quantitative PCR (qPCR) was carried out on a BioRad CFX system (Cat#7500, BioRad, USA) using iTaq Universal SYBR Green Supermix (Cat#172–5272, BioRad). All samples were analysed with three biological replicates. The expression level of *TaB3-2A1* and *Glu-1* was calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). The primers, qGlu-all, were designated to investigate the expression patterns of all active *Glu-1* alleles in Fielder based on their conserved regions (Table S8). Wheat *elongation factor 1 alpha* (*eF1a*) was used as an internal control to normalize the expression level of *TaB3-2A1*.

Gene cloning and transgenic experiments in wheat

The specific primer pair of *TaB3-2A1* (*TraesCS2A02G159800*) was designed and used to amplify the full-length coding sequence (CDS) from developing grains at 10 dpa. Amplified products were cloned into the pEASY-T5 cloning vector (Cat#CT501-01, Transgene, Beijing) and transformed into *Escherichia coli* strain DH5α competent cells for identification of positive clones by sequencing.

To validate the function of *TaB3-2A1* by transgenic assays in wheat, its CDS was inserted into the destination vector pUbiTCK303 (kindly provided by Dr Genying Li, Shandong Academy of Agricultural Sciences). The resultant vector was transformed into wheat cultivar Fielder by the Agrobacterium-mediated method following Li *et al*. (2019b). Primers were listed in Table S8.

Yeast one-hybrid assays

Yeast one-hybrid analysis was performed as described in Lin *et al*. (2007). We inserted the CDS of *TaB3-2A1* into the pB42AD vector (Cat#ZT0295, Clontech) at the *EcoR*I and *Kpn*I restriction sites to produce pB42AD-TaB3-2A1 as 'prey'. The promoter region CCRM1-1 was cloned into the pLacZi reporter vector (Cat#631707, Clontech) at the *Xho*I and *Kpn*I restriction site to generate the 'bait' vector pLacZi-CCRM1-1. The 'bait' and 'prey' vectors containing CCRM1-1 and TaB3-2A1, respectively, were co-transformed into yeast strain EGY48. Transformants were cultured on SD-Trp/-Ura plates at 30 °C for 3 days and then transferred into X-Gal (5-bromo-4-chloro-3-indolyl-b-d-galactopyranoside) plates for analysis of galactosidase activity. Two empty pB42AD and pLacZi vectors combined with pLacZi-CCRM1-1 and pB42AD-TaB3-2A1, respectively, were co-transformed into EGY48 to generate negative controls.

Electrophoretic mobility shift assays

The CDS of *TaB3-2A1* was ligated into the pGEX4T-1 (Cat#ZK132, ZOMANBIO, Beijing) and then transformed into strain *E. coli* (BL21 DE3) to produce the TaB3-2A1-GST fusion protein. TaB3-2A1-GST protein was induced with 0.6 mM isopropyl-b-D thiogalactopyranoside (IPTG) in the LB broth at 23 °C for 8 h. TaB3-2A1-GST and control pGEX4T-1 protein were purified using GST fusion protein magnetic beads (Cat#M2320, Solarbio, Beijing) following the manufacturer's user manual (http://www.novagen.com). The fused protein was eluted from the beads with 50 mM Tris–HCl pH 8.0 containing 10 mM glutathione. The CCRM1-1 double-stranded probe was the same as that in DNA pull-down experiments. The CCRM1-1 probe and the TaB3-2A1-GST fusion protein were added to 20 μL binding buffer [100 mM Tris, 500 mM KCl, 10 mM DTT, 2.5% glycerol, 0.2 mM EDTA, 50 ng/L poly(dI–dC)] at room temperature for 25 min. Samples of 4 μL 5 × protein loading buffer (1 M Tris–HCl, 10% dodecyl sulfate, 25 mg bromophenol blue, 250 μL β-mercaptoethanol) were separated in 6% native polyacrylamide gels and then transferred to nylon membranes. Detection of the biotin-labelled DNA probe was performed using a LightShift Chemiluminescent EMSA Kit (Cat#20148, Thermo Fisher Scientific) according to the manufacturer's instructions.

Transactivation experiments

The promoter of *Glu-1Dx2* as a representative was inserted into the pGreenII 0800-LUC between the *BamH*I and *Kpn*I restriction sites. pGreenII 0800-LUC includes two reporters, firefly luciferase (LUC) and renilla luciferase (REN) driven by the *Glu-1* and *CaMV35S* promoters, respectively. The REN reporter was used as an internal control. Full-length CDS of *TaB3-2A1* was inserted into pGreenII 62-SK to generate an effector. The resultant constructs were transformed into *A. tumefaciens* strain GV3101 and then co-infiltrated into tobacco leaves as described in Song *et al*. (2017). LUC signals were imaged using the Nightshade LB985 apparatus (Berthold, Germany). LUC activity was quantified with a Dual-Luciferase Assay Kit (Promega, Beijing) according to the manufacturer's recommendations. Relative LUC activity was calculated by the ratio of LUC/REN. Each sample was tested in three replications.

For transcription activity assay in *Arabidopsis* protoplasts, the CDS of *TaB3-2A1* was inserted into the GAL4BD vector to generate an effector. Virion protein 16 (VP16) containing an acidic transcriptional activation domain was used as a control to identify the effect of TaB3-2A1 on transcriptional activity driven by the UAS module in the reporter vector 35sLUC. GAL4BD can bind to the UAS in 35sLUC to activate a LUC reporter (Bart *et al*., 2006). The reference vector pRTL containing a 35S promoter-driven REN was used as an internal control. Each effector construct together with the reporter and reference vectors were co-transformed into Arabidopsis protoplasts using the method described in Yoo *et al*. (2007). Fluorescence signals were detected by a GLOMA 20/20 LUMINOMETER detector (Promega). The Dual-Luciferase Reporter Assay System kit (Cat#E1910, Promega) was used to quantify fluorescence signals. Relative LUC activity was calculated by the ratio of LUC/REN.

Measurement of proteins and starch in seeds

Seed protein content was measured by near-infrared reflectance spectroscopy using a Perten DA 7200 analyser (Perten, Springfield, IL, USA). Seed protein components were sequentially extracted and analysed following a modified reported procedure (Zhang *et al*., 2007). Seeds were ground into fine powder by a Cyclotec mill (FOSS, Hilleroed, Denmark) and 0.1 g of whole meal flour from each sample was weighed for glutenin extraction. Add 500 μL 50% (v/v) propanol into each sample, vortex for 1 h and discard the supernatant. Repeat this step for three times. The precipitant was used for glutenin extraction with 500 μL of extraction solution containing 0.2 M Tris–HCl (pH 6.8), 50% (v/v) propanol, and 1% (w/v) DTT. Glutenins were separated and quantified by a Waters ACQUITY UPLC system equipped with a Protein BEH300 C4 (1.7 μm, 2.1 × 100 mm) column (Waters Corporation, Milford, CT, USA). Four μl of glutenin extract was injected into a BEH300 C4 Column for qualitative and quantitative analyses.

Ultrapure water (mobile phase A) and acetonitrile (mobile phase B) were used as eluents and each contained 0.06% trifluoroacetic acid. The flow rate was 0.2 mL /min. Glutenins were separated using a linear gradient, from 15% to 40% mobile phase B over 32 min. Proteins were detected by UV absorbance at 210 nm. Total starch content was measured using the amyloglucosidase-alpha-amylase method (McCleary *et al*., 1997).

RNA-Seq assays

Developing seeds of *TaB3-2A1*-OE lines and TNL at 10 dpa were harvested for transcriptome analysis in Novogene (https://www.novogene.com/). Total RNA was extracted and assessed using an RNA Nano 6000 Assay Kit for a Bioanalyzer 2100 (Agilent Technologies, CA, USA). RNA sequencing libraries were prepared

by a NEBNext Ultra™ RNA Library Prep Kit (NEB, Ipswich, MA, USA) following the manufacturer's guide. The libraries were tested for quality control by Qubit2.0 Fluorometer and sequenced by an Illumina NovaSeq 6000. Effective reads were aligned to IWGSC RefSeq v1.1 (https://wheat-urgi.versailles.inrae.fr/Seq-Repository/Assemblies) using Hisat2 v2.0.5 (IWGSC, 2018; Kim *et al*., 2015). The expected number of fragments per kilobase of transcript sequence per million base pairs sequenced (FPKM) was used for estimating gene expression levels. The resultant *P*-values were adjusted using the Benjamini and Hochberg method for controlling false discovery rates. $P < 0.05$ and |log2(fold change)| >0.8 were set as the thresholds for significantly differential expression. Enrichments of differentially expressed genes in KEGG pathways were tested by clusterProfiler in the R package (Version 3.8.1).

Statistical analysis

Student's *t*-tests were performed to determine significant differences between *TaB3-2A1*-OE lines and TNL using R 3.6.3 for Windows (https://cran.r-project.org/bin/windows/base/old/3.6.3/). GraphPad Prism (version 8.0.1) was used to make charts (https://www.graphpad-prism.cn/).

Author contributions

L.X., S.Y.L., W.F.T., D.A.X., J.H.L. X.M.L., L.L.L., Y.J.B. and F.J.L. performed experiments; S.H.C., Y.Z. and X.Y.S., designed the experiment; S.H.C. and L.X. wrote the draft; X.C.X., Y.F.H., Z.H.H. and X.G.Y. revised the manuscript. All authors read and approved the final version of the manuscript before submission.

Acknowledgements

We are grateful to Prof. Robert McIntosh, Plant Breeding Institute, University of Sydney, for revising the manuscript. The work was supported by National Key Project for Agricultural Science and Technology (NK2022060103), Natural Science Foundation of China (32272182), National Key Research and Development Program of China (2021YFF1000204) and Project for Hebei Scientific and Technological Innovation Team of Modern Wheat Seed Industry (21326318D).

Conflicts of interest

The authors declare no conflicts of interest.

References

Aerts, N., de Bruijn, S., van Mourik, H., Angenent, G.C. and van Dijk, A.D.J. (2018) Comparative analysis of binding patterns of MADS-domain proteins in Arabidopsis thaliana. *BMC Plant Biol.* **18**, 131.

Albani, D., Hammond-Kosack, M.C., Smith, C., Conlan, S., Colot, V., Holdsworth, M. and Bevan, M.W. (1997) The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. *Plant Cell* **9**, 171–184.

Bart, R., Chern, M., Park, C.J., Bartley, L. and Ronald, P.C. (2006) A novel system for gene silencing using siRNAs in rice leaf and stem-derived protoplasts. *Plant Methods* **29**, 2–13.

Boer, D.R., Freire-Rios, A., van den Berg, W.A.M., Saaki, T., Manfield, I.W., Kepinski, S., López-Vidrieo, I. *et al*. (2014) Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell* **156**, 577–589.

Boudet, J., Merlino, M., Plessis, A., Gaudin, J.C., Dardevet, M., Perrochon, S., Alvarez, D. *et al*. (2019) The bZIP transcription factor SPA Heterodimerizing Protein represses glutenin synthesis in *Triticum aestivum*. *Plant J.* **97**, 858–871.

Cao, S., Xu, H., Li, Z., Wang, X., Wang, D., Zhang, A., Jia, X. *et al*. (2007) Identification and characterization of a novel Ag. intermedium HMW-GS gene from T. Aestivum-Ag. intermedium addition lines TAI-I series. *J. Cereal Sci.* **45**, 293–301.

Don, C.G., Mann, G., Bekes, F. and Hamer, R.J. (2006) HMW-GS affect the properties of glutenin particles in GMP and thus flour quality. *J. Cereal Sci.* **44**, 127–136.

Dong, G., Ni, Z., Yao, Y., Nie, X. and Sun, Q. (2007) Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. *Plant Mol. Biol.* **63**, 73–84.

Eagles, H.A., Hollamby, G.J. and Eastwood, R.F. (2002) Genetic and environmental variation for grain quality traits routinely evaluated in southern Australian wheat breeding programs. *Aust. J. Agr. Res.* **53**, 1047–1057.

Feng, F., Qi, W., Lv, Y., Yan, S., Xu, L., Yang, W., Yuan, Y. *et al*. (2018) OPAQUE11 is a central hub of the regulatory network for maize endosperm development and nutrient metabolism. *Plant Cell* **30**, 375–396.

Gao, Y., An, K., Guo, W., Chen, Y., Zhang, R., Zhang, X., Chang, S. *et al*. (2021) The endosperm-specific transcription factor TaNAC019 regulates glutenin and starch accumulation and its elite allele improves wheat grain quality. *Plant Cell* **33**, 603–622.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. and Goodman, H.M. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* **4**, 1251–1261.

Guo, W., Yang, H., Liu, Y., Gao, Y., Ni, Z., Peng, H., Xin, M. *et al*. (2015) The wheat transcription factor TaGAMyb recruits histone acetyltransferase and activates the expression of a high-molecular-weight glutenin subunit gene. *Plant J.* **84**, 347–359.

He, Z.H., Liu, L., Xia, X.C., Liu, J.J. and Peña, R.J. (2005) Composition of HMW and LMW glutenin subunits and their effects on dough properties, pan Bread, and noodle quality of Chinese bread wheats. *Cereal Chem.* **82**, 345–350.

International Wheat Genome Sequencing Consortium (IWGSC) (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191.

Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360.

Kornberg, R.D. (2007) The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12955–12961.

Lawrence, G.J. and Shepherd, K.W. (1981) Chromosomal location of genes controlling seed proteins in species related to wheat. *Theor. Appl. Genet.* **59**, 25–31.

Li, F., Wen, W., Liu, J., Zhang, Y., Cao, S., He, Z., Rasheed, A. *et al*. (2019a) Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* **19**, 168.

Li, J.H., Wang, K., Li, G.Y., Li, Y.L., Zhang, Y., Liu, Z.Y., Ye, X.G. *et al*. (2019b) Dissecting conserved cis-regulatory modules of Glu-1 promoters which confer the highly active endosperm-specific expression via stable wheat transformation. *Crop J.* **7**, 8–18.

Li, J., Xie, L., Tian, X., Liu, S., Xu, D., Jin, H., Song, J. *et al*. (2021) TaNAC100 acts as an integrator of seed protein and starch synthesis exerting pleiotropic effects on agronomic traits in wheat. *Plant J.* **108**, 829–840.

Lin, R., Ding, L., Casola, C., Ripoll, D.R., Feschotte, C. and Wang, H. (2007) Transposase-derived transcription factors regulate light signaling in Arabidopsis. *Science* **318**, 1302–1305.

Liu, L., He, Z.H., Yan, J., Zhang, Y., Xia, X.C. and Peña, R.J. (2005) Allelic variation at the Glu-1 and Glu-3 loci, presence of the 1B.1R translocation, and their effects on mixographic properties in Chinese bread wheats. *Euphytica* **142**, 197–204.

Livak, K.J. and Schmittgen, T. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-DDCt method. *Methods* **25**, 402–408.

McCarty, D.R., Carson, C.B., Stinard, P.S. and Robertson, D.S. (1989) Molecular analysis of viviparous-1: an abscisic acid-insensitive mutant of maize. *Plant Cell* **1**, 523–532.

McCleary, B.V., Gibson, T.S. and Mugford, D.C. (1997) Measurement of total starch in cereal products by amyloglucosidase-alpha-amylase method: collaborative study. *J. AOAC Int.* **80**, 571–579.

Mena, M., Vicente-Carbajosa, J., Schmidt, R.J. and Carbonero, P. (1998) An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. *Plant J.* **16**, 53–62.

Moreno-Risueno, M.A., González, N., Díaz, I., Parcy, F., Carbonero, P. and Vicente-Carbajosa, J. (2008) FUSCA3 from barley unveils a common transcriptional regulation of seed-specific genes between cereals and Arabidopsis. *Plant J.* **53**, 882–894.

Oñate, L., Vicente-Carbajosa, J., Lara, P., Díaz, I. and Carbonero, P. (1999) Barley BLZ2, a seed-specific bZIP protein that interacts with BLZ1 in vivo and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm. *J. Biol. Chem.* **274**, 9175–9182.

Payne, P.I., Nightingale, M.A., Krattiger, A.F. and Holt, L.M. (1987) The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. *J. Sci. Food Agric.* **40**, 51–65.

Qian, D., Qiu, B., Zhou, N., Takaiwa, F., Yong, W. and Qu, L.Q. (2020) Hypotensive activity of transgenic rice seed accumulating multiple antihypertensive peptides. *J. Agric. Food Chem.* **68**, 7162–7168.

Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M. *et al.* (2018) The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089.

Reidt, W., Wohlfarth, T., Ellerström, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L. *et al.* (2000) Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant J.* **21**, 401–408.

Rubio-Somoza, I., Martinez, M., Diaz, I. and Carbonero, P. (2006) HvMCB1, a R1MYB transcription factor from barley with antagonistic regulatory functions during seed development and germination. *Plant J.* **45**, 17–30.

Schmidt, R.J., Ketudat, M., Aukerman, M.J. and Hoschek, G. (1992) Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. *Plant Cell* **4**, 689–700.

Shen, L., Luo, G., Song, Y., Xu, J., Ji, J., Zhang, C., Gregová, E. *et al.* (2021) A novel NAC family transcription factor SPR suppresses seed storage protein synthesis in wheat. *Plant Biotechnol. J.* **19**, 992–1007.

Shewry, P.R., Tatham, A.S., Barro, F., Barcelo, P. and Lazzeri, P. (1995) Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. *Nat. Biotechnol.* **13**, 1185–1190.

Shewry, P.R., Underwood, C., Wan, Y., Lovegrove, A., Bhandari, D., Toole, G., Mills, E.N.C. *et al.* (2009) Storage product synthesis and accumulation in developing grains of wheat. *J. Cereal Sci.* **50**, 106–112.

Song, X., Lu, Z., Yu, H., Shao, G., Xiong, J., Meng, X., Jing, Y. *et al.* (2017) IPA1 functions as a downstream transcription factor repressed by D53 in strigolactone signaling in rice. *Cell Res.* **27**, 1128–1141.

Sun, F., Liu, X., Wei, Q., Liu, J., Yang, T., Jia, L., Wang, Y. *et al.* (2017) Functional characterization of TaFUSCA3, a B3-superfamily transcription factor gene in the wheat. *Front. Plant Sci.* **8**, 1133.

Suzuki, M. and McCarty, D.R. (2008) Functional symmetry of the B3 network controlling seed development. *Curr. Opin. Plant Biol.* **11**, 548–553.

Swaminathan, K., Peterson, K. and Jack, T. (2008) The plant B3 superfamily. *Trends Plant Sci.* **13**, 647–655.

Veraverbeke, W.S. and Delcour, J.A. (2002) Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality. *Crit. Rev. Food Sci. Nutr.* **42**, 179–208.

Verdier, J. and Thompson, R.D. (2008) Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.* **49**, 1263–1271.

Vicente-Carbajosa, J. and Carbonero, P. (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int. J. Dev. Biol.* **49**, 645–651.

Vicente-Carbajosa, J., Oñate, L., Lara, P., Diaz, I. and Carbonero, P. (1998) Barley BLZ1: a bZIP transcriptional activator that interacts with endosperm-specific gene promoters. *Plant J.* **13**, 629–640.

Wang, J., Chen, Z., Zhang, Q., Meng, S. and Wei, C. (2020) The NAC transcription factors OsNAC20 and OsNAC26 regulate starch and storage protein synthesis. *Plant Physiol.* **184**, 1775–1791.

Weegels, P.L., Hamer, R.J. and Schofield, J.D. (1996) Functional properties of wheat glutenin. *J. Cereal Sci.* **23**, 1–17.

Wieser, H. (2007) Chemistry of gluten proteins. *Food Microbiol.* **24**, 115–119.

Wiśniewski, J.R., Zougman, A., Nagaraj, N. and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362.

Wu, Y. and Messing, J. (2012) Rapid divergence of prolamin gene promoters of maize after gene amplification and dispersal. *Genetics* **192**, 507–519.

Xi, D.M. and Zheng, C.C. (2011) Transcriptional regulation of seed storage protein genes in Arabidopsis and cereals. *Seed Sci. Res.* **21**, 247–254.

Yamamoto, M.P., Onodera, Y., Touno, S.M. and Takaiwa, F. (2006) Synergism between RPBF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol.* **141**, 1694–1707.

Yang, Y., Ma, Y.Z., Xu, Z.S., Chen, X.M., He, Z.H., Yu, Z., Wilkinson, M. *et al.* (2007) Isolation and characterization of Viviparous-1 genes in wheat cultivars with distinct ABA sensitivity and pre-harvest sprouting tolerance. *J. Exp. Bot.* **58**, 2863–2871.

Yoo, S.D., Cho, Y.H. and Sheen, J. (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572.

Zhang, P.P., He, Z.H., Chen, D.S., Zhang, Y., Larroque, O.R. and Xia, X.C. (2007) Contribution of common wheat protein fractions to dough properties and quality of northern-style Chinese steamed bread. *J. Cereal Sci.* **46**, 1–10.

Zhang, Z., Zheng, X., Yang, J., Messing, J. and Wu, Y. (2016) Maize endosperm-specific transcription factors O2 and PBF network the regulation of protein and starch synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10842–10847.

Zhang, Z., Dong, J., Ji, C., Wu, Y. and Messing, J. (2019) NAC-type transcription factors regulate accumulation of starch and protein in maize seeds. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11223–11228.

Zhu, J., Fang, L., Yu, J., Zhao, Y., Chen, F. and Xia, G. (2018) 5-Azacytidine treatment and TaPBF-D over-expression increases glutenin accumulation within the wheat grain by hypomethylating the Glu-1 promoters. *Theor. Appl. Genet.* **131**, 735–746.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Dataset S1. *Glu-1* promoter-bound transcription factor partners detected by DNA pull-down plus LC–MS.

Dataset S2. *Glu-1* promoter-bound epigenetic modifiers detected by DNA pull-down plus LC–MS.

Dataset S3. *Glu-1* promoter-bound transcription initiation and elongation factors detected by DNA pull-down plus LC–MS.

Dataset S4. *Glu-1* promoter-bound kinases detected by DNA pull-down plus LC–MS.

Dataset S5. *Glu-1* promoter-bound metabolic enzymes detected by DNA pull-down plus LC–MS.

Dataset S6. *Glu-1* promoter-bound other proteins detected by DNA pull-down plus LC–MS.

Dataset S7. Differentially expressed genes in developing grains between *TaB3-2A1* overexpression lines (OE) and corresponding transgenic null lines (TNL).

Dataset S8. The genotypes and phenotypes of 166 elite wheat lines from Huang-Huai Valley of China.

Dataset S9. The number and frequency of the wheat lines with contrasting *TaB3-2A1* haplotypes from Huang-Huai Valley of China.

Dataset S10. The genotypes of 308 wheat lines from the Wheat SNPHub.

Dataset S11. The number and frequency of the wheat lines carrying contrasting *TaB3-2A1* haplotypes from wheat SNPHub.
Dataset S12. The haplotypes and release years of wheat cultivars with homozygous genotypes of *TaB3-2A1*.
Dataset S13. The number and frequency of *TaB3-2A1* haplotypes in wheat cultivars released before and after 1985.
Figure S1. Phylogenetic analyses of *TaB3-2A1*-upregulated starch synthesis-related genes and their counterparts in rice, Brachypodium and barley.
Figure S2. A model of *TaB3-2A1* regulating seed protein, starch and agronomic traits.
Figure S3. Phylogenetic analysis of B3 transcription factors (TFs) in wheat, barley, maize, rice and Arabidopsis.
Table S1. Differentially expressed genes encoding seed proteins between *TaB3-2A1* overexpression lines (OE) and their transgenic null lines (TNL).
Table S2. Differentially expressed genes involved in starch anabolism and sugar transport pathways between *TaB3-2A1* overexpression lines (OE) and their transgenic null lines (TNL).
Table S3. Transcriptomic data of the identified transcription factors retrieved from the database in Wheat Expression Browser.
Table S4. Analysis of genetic effects of *TaB3-2A1* haplotypic variation on agronomic traits, seed protein and starch contents.
Table S5. Comparisons of high-molecular-weight glutenin subunits (HMW-GSs) between *TaB3-2A1* overexpression lines (OE) and transgenic null lines (TNL).
Table S6. Comparisons of low-molecular-weight glutenin subunits (LMW-GSs), total seed protein and starch between *TaB3-2A1* overexpression lines (OE) and transgenic null lines (TNL).
Table S7. Comparisons of agronomic traits between *TaB3-2A1* overexpression lines (OE) and transgenic null lines (TNL).
Table S8. Primer pairs or probes used in this study.