ORIGINAL ARTICLE

Probing early wheat grain development via transcriptomic and proteomic approaches

Mingming Yang[1,2] · Yang Liu[1] · Jian Dong[1,2] · Wanchun Zhao[1,2] · Samneet Kashyap[3] · Xiang Gao[1,2] · Sachin Rustgi[3] · Shanshan Wen[1]

Abstract

To understand the molecular changes taking place during the early grain development in common wheat, we profiled transcriptome and proteome of two cultivars, "P271" and "Chinese Spring" (CS) with large and small grains, respectively. More than 85,000 genes and 7500 proteins were identified to express during early grain development in two wheat cultivars. We observed enrichment in the number of genes falling in the functional categories—carbohydrate metabolism, amino acid metabolism, lipid metabolism, and cofactor as well as vitamin metabolism with progression in grain development, which indicates towards the importance of these metabolic pathways during grain maturation. Many genes showed inconsistency between transcription and translation, which suggested a role of post-transcriptional events that determine the fate of nascent transcript/protein, in the early grain development. In silico localization of differentially expressed genes/proteins between CS and P271 to wheat chromosomes, exhibited a biased genomic distribution with chromosomes 1A, 4B, and 5B contributing primarily to it. These results corroborated the earlier findings, where chromosomes 1A, 4B, and 5B were reported to harbor genes/QTLs for yield contributing traits such as grain length and thickness. Collectively, this study reveals the molecular changes taking place during early grain development, through light on the regulation of these processes, and allows identification of the gene candidates contributing to the contrasting grain characteristics of CS and P721. This information has implications in the future wheat breeding for the enhanced grain yield.

Keywords *Triticum aestivum* · Grain development · Transcriptome · Proteome

✉ Xiang Gao
gx@nwsuaf.edu.cn

✉ Sachin Rustgi
srustgi@clemson.edu

✉ Shanshan Wen
sswen@nwsuaf.edu.cn

[1] State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University, Yangling 712100, People's Republic of China

[2] Wheat Engineering Research Center of Shaanxi Province, Yangling 712100, People's Republic of China

[3] Department of Plant and Environmental Sciences, Pee Dee Research and Education Center, Clemson University, Florence, SC 29506, USA

Introduction

Common wheat (*Triticum aestivum* L.) is an important crop, as over 2000 million tons of it is used for food or feed. Seeds are the edible component of the wheat plant and the major commodity; hence, a large body of research is dedicated to better understand the wheat grain development (Wan et al. 2008; Pfeifer et al. 2014; Yang et al. 2016, 2017; Rangan et al. 2017).

Wheat grain development can be arbitrarily partitioned into three distinct phases: pre-grain filling, grain filling, and desiccation (Shewry et al. 2012; Yang et al. 2016). Earlier studies have defined the first 2 weeks after anthesis, as the early grain developmental or pre-grain filling phase. However, its precise duration varies among genotypes and is mostly dependent on the ambient environmental conditions (Wan et al. 2008; Olsen 2004). The anatomy and chemical composition of the wheat caryopsis has been studied in great detail, albeit a handful of studies are dedicated to the understanding of the significance

of transcriptional differences among cultivars during early wheat grain development, specifically, in the determination of their respective final grain weights (Rangan et al. 2017). In this direction, transcript profiling of developing wheat grains using microarray or RNA sequencing provided valuable information, and shed light on the changes in gene expression taking place during early grain development (Laudencia-Chingcuanco et al. 2006, 2007; Wan et al. 2008; Pfeifer et al. 2014; Rangan et al. 2017; Ramírez-González et al. 2018).

Proteomics approaches that provide insight at the translational level have also been used widely in wheat grain development studies (Skylas et al. 2005; Mak et al. 2006; Dupont et al. 2011; Juhász et al. 2012; Bancel et al. 2015; Yang et al. 2017). In fact, effects of different nutrient regimes (Altenbach et al. 2011), and stressors such as salinity, drought, heat, and hydrogen peroxide, as well as pathogens on grain development were also studied using proteomics approaches (Peng et al. 2009; Guo et al. 2012; Ge et al. 2013; Gu et al. 2015; Ma et al. 2014; Yang et al. 2016; Kumar et al. 2019). Additionally, changes taking place in the aleurone layer of wheat grains at fifteen developmental time points were studied using two proteomics approaches, 2-dimensional electrophoresis (2-DE) and two-dimensional differential gel electrophoresis (2D-DIGE) (Gao et al. 2009; Nadaud et al. 2015). In these studies, a nice correlation between differential expressions of albumins and globulins and the end-use quality of wheat grains was reported. Subsequently, the iTRAQ (isobaric tag for relative and absolute quantitation) approach of protein quantification was also used to study the wheat grain development (Ma et al. 2014; Yang et al. 2017). These studies unrevealed many more proteins which expressed during grain development and provided a more quantitative overview of the whole process.

Collectively, the transcriptomics and proteomics studies provided a broader overview of the transcriptional and translational changes taking place during wheat grain development, but these studies were confined to the analysis of single wheat genotypes, and none of them have focused on the differences in grain development of wheat varieties with contrasting grain characteristics, such as large and small grains. As a result, despite using a combination of genetic and genomic approaches, only a handful of genes that control wheat grain development have been so far characterized (Zhou et al. 2013), and their mutual interactions and/or cumulative contribution towards the grain development remained elusive.

In the present manuscript, we studied expression level differences in the two wheat cultivars; it is a continuation of our earlier research on the identification of proteins that contribute to differences in early grain development of wheat cultivars with contrasting grain characteristics (Yang et al. 2017). Specifically, in this study, we provide an integrated analysis of transcriptomic and proteomic data collected over three time

points during the early grain development of common wheat and identified several differentially expressed genes (DEGs) and differentially expressed proteins (DEPs). The study unravels the central metabolic changes taking place during the early grain development and sheds light on the developmental differences between the two wheat cultivars with contrasting grain characteristics.

Materials and methods

Wheat materials and sample preparation

Two wheat cultivars P271 (winter type) and Chinese Spring (spring type) respectively with large and small grains were grown at the Yangling Experimental Station (34.26° N and 108.14° E) in the Shaanxi Province. Plants were fertilized with urea (60 kg ha^{-1}) and watered periodically. The main culm spikes were tagged upon anthesis, and the seeds from the labeled spikes were sampled at 4-, 8-, and 12-day post-anthesis (DPA). Seed samples were collected from the four central spikelets. Samples from each stage consisted of at least 200 seeds from 30 spikes. Samples were snap-frozen in liquid nitrogen and stored at – 80 °C for later use (Yang et al. 2016).

RNA extraction, sequencing, and de novo assembly

Total RNA which was extracted separately from the immature wheat grains collected 4, 8, and 12 DPA using Trizol reagent following manufacturer's recommendations. The yield and purity of each RNA sample were determined via spectrophotometry, and the integrity of RNA samples was monitored on 1% agarose gels.

cDNA library was constructed for each of the six RNA samples (3 developmental stages × 2 genotypes) and sequenced on the Illumina HISeq 2000 platform (Illumina Inc. CA, USA). Before assembly, adapter sequences were removed from the raw sequence reads. Low-quality sequence, i.e., reads with > 50% bases with quality scores of 20 or below were removed from each data set. Subsequently, the high-quality reads from all samples were pooled together and assembled using Trinity package to construct a pool of unique consensus sequences to serve as a reference sequence set. DEGs were identified at the false discovery rate (FDR) of < 0.05 and log2FC (fold change) of > 1 and the transcriptomics and proteomics data were compared at an FPKM (Fragments Per Kilobase Million) value of < 5.

Trend analysis

Data on expression analysis was used to cluster genes with similar expression patterns. To examine the expression pattern of DEGs, the expression data of each sample (in the order of

treatment) were normalized to 0, log2(v1/v0), log2(v2/v0), and then clustered by Short Time-series Expression Miner software (STEM) (Ernst and Bar-Joseph 2006). The parameters were set as follows: maximum unit change in model profiles between time points is 1; maximum output profiles number is 20 (similar profiles will be merged); the minimum ratio of the fold change of DEGs is no less than 2.0. The clustered profiles with p value < 0.05 were considered as significant profiles. After identification, the DEGs were subjected to gene ontology (GO) and KEGG pathway enrichment analysis (Benjamini and Hochberg 1995). The GO terms or pathways with Q value < 0.05 were defined as significantly enriched.

Protein preparation

Frozen wheat grains were pulverized in liquid nitrogen, and the proteins were extracted as described by Yang et al. (2016). Briefly, the pulverized grain samples were suspended in 10 ml of ice-cold phenol extraction buffer (0.7 M sucrose; 0.1 M KCl; 50 mM EDTA, 0.5 M Tris-HCl, 1% (w/v) DTT, 0.1 mM PMSF, pH 7.5), and incubated at 4 °C for 30 min. After incubation, the mixture was centrifuged at 5000×g for 30 min under refrigeration. Following centrifugation, the phenolic phase was collected and precipitated overnight with five volumes of 100 mM ammonium acetate prepared in methanol at − 20 °C. Subsequently, the mixture was centrifuged at 5000×g for 30 min at 4 °C, and the supernatant was discarded. The resultant pellet was rinsed with ice-cold acetone with 0.2% DTT (w/v) for two times. Following this step, the pellet was air-dried and resuspended in 200 μl of RIPA lysis buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1% SDS, and 0.1% Triton 100). Subsequently, the protein concentration was determined by standard BCA (bicinchoninic acid) assay using bovine serum albumin as the standard (Beyotime Company, Shanghai).

iTRAQ labeling and LC-ESI-MS/MS analysis

Proteins (100 μg) from each sample were digested overnight with trypsin (Promega) at 37 °C. Following digestion, the protein samples were labeled with 8-plex iTRAQ kit (Applied Biosystems). Peptides from different treatments were labeled with different iTRAQ tags, respectively, by incubation at room temperature for 2 h. The labeled peptides were then pooled and dried by vacuum centrifugation. The pooled mixture of iTRAQ-labeled peptides was fractionated by high-pH reversed-phase chromatography. iTRAQ labeled samples were separated on a Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) equipped with a Gemini® 3 μm NX-C18, 110 Å, 75 × 2 mm LC Column (Phenomenex).

Subsequently, the MS/MS analysis was performed on a Q Exactive system (Thermo Fisher Scientific) in the Information Dependent Mode, as previously described (Yang et al. 2017). iTRAQ proteomics analysis was performed twice for all samples.

Proteome data analysis

For protein identification, the MS/MS spectra were processed by ProteinPilot™ Software 4.5 (AB Sciex). The criteria used for protein identification were as in Yang et al. (2017) (for details, see SI Methods).

Annotation

The function of all DEGs or DEPs were predicted by BLASTP searches (E-value cutoff $1e^{-5}$) against different databases, including NCBI non-redundant protein database (Nr), Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa et al. 2012), and Swiss-Prot. For the gene matching to multiple protein sequences, the protein with the highest similarity score was considered as the best annotation.

Correlation analysis of transcriptome and proteome data

To find concordance between grain transcriptome and proteome, we calculated Pearson's correlation for these data. And ggplot2 package of R (version 3.2.1) software was used to create scatterplots with the expression ratios.

miRNA target gene prediction

The existing miRNA sequences were used to predict the target genes among the DEGs identified in the present study. The predictions were made using the following parameters on the patmatch (v1.2) software: (1) no more than four mismatches between miRNA and/or target (G-U bases count as 0.5 mismatches); (2) no more than two adjacent mismatches in the miRNA-target duplex; (3) no adjacent mismatches at positions 2 to 12 of the miRNA-target duplex (5′ of miRNA); (4) no mismatches at positions 10 and 11 of the miRNA-target duplex; (5) no more than 2.5 mismatches in positions 1–12 of the miRNA-target duplex (5′ of miRNA); and (6) minimum free energy (MFE) of the miRNA-target duplex should be ≥ 60% of the MFE of the miRNA bound to its perfect complement.

Chromosome assignment of DEGs/DEPs and their locations in relation to the known QTLs/genes for grain development traits

Genomic distribution of DEGs/DEPs was studied by assigning them to wheat sub-genomes, homoeologous chromosome groups, and specific chromosomes. For this purpose, sequences of DEGs/DEPs were blasted against the wheat

genomic DNA sequences available in the public domain. To eliminate the bias of genome size or gene number (predicted per chromosome), expected value of DEGs/DEPs was calculated for each chromosome following Yang et al. (2017). Expected and observed numbers of DEGs/DEPs per chromosome were plotted in the form of bar diagrams. The sub-chromosomal locations of the DEGs/DEGs was determined using the wheat chromosome zippers, and the location of genes and QTLs were identified based on the former publications (Gupta et al. 2006; Gegas et al. 2010; Rustgi et al. 2013; Wu et al. 2015; Nadolska-Orczyk et al. 2017; Zhang et al. 2018; Zhai et al. 2018; Cao et al. 2019; Yang et al. 2019; Sakuma et al. 2019).

Results

Summary of RNA-seq data

Approximately 58.42–115.45 million 150 bp paired-end reads were generated from the wheat grain samples through RNA sequencing (Table 1). After removal of the adapter and low-quality sequences, a range of 85,056–87,701 genes was identified from the remaining high-quality reads of the three-grain developmental stages in two wheat cultivars (Table 1). Of 85,056–87,701 genes, 65,265–68,234 were known genes, and 18,353–19,467 were treated as the novel genes (Table 1).

Gene expression pattern analysis, clustering, and functional enrichment of differentially expressed genes (DEGs)

Based on the gene expression patterns at three developmental stages, using STEM software, DEGs from the two wheat cultivars P271 and CS were clustered into eight groups (expression profiles). The expression profiles displayed considerable differences in gene expression patterns of two wheat cultivars over time during grain development.

In CS, the DEGs were significantly over-represented in the profiles at 8 DPA (profiles 0, 1, 3, $P < 0.05$) (Fig. 1b). In both cultivars, we observed that the genes follow two main expression patterns: (i) consistent downregulation and (ii) consistent

Table 1 Genes detected in different groups

Group name	No. of known genes	No. of new genes	All gene num
C12	66,992 (66.76%)	18,920	85,912
C4	66,703 (66.47%)	18,353	85,056
C8	68,126 (67.89%)	19,014	87,140
P12	65,265 (65.04%)	18,690	83,955
P4	66,757 (66.53%)	18,668	85,425
P8	68,234 (68.00%)	19,467	87,701

expression followed by downregulation, but these two expression patterns were recorded to appear at different time points in two cultivars. A third, but less common expression pattern observed specifically in CS was downregulation followed by a consistent expression (see profile 1 in Fig. 1). In both Chinese Spring and P271, we found more downregulated genes than the upregulated genes, except for 4 DPA vs 8 DPA comparison in P271 (Fig. 1).

Functional annotation and KEGG pathway analysis of DEGs

A gene ontology (GO) category enrichment analysis was conducted to study the overall trend of enrichment for the specific functional categories in wheat grain transcriptome between two wheat cultivars. Distribution of genes among different GO categories biological process, cellular component, and molecular function are shown in Fig. 2.

In the biological process, DEGs mainly belonged to sub-categories cellular process and metabolic process at all developmental time points (4, 8, and 12 DPA) in both cultivars P271 and Chinese Spring. While in P271 compared to CS, more genes related to cellular and metabolic processes were found downregulated at 4 and 8 DPA and upregulated at 12 DPA. In cellular component category, DEGs mainly belonged to sub-categories cell, cell part, and organelle at all developmental time points (4, 8, and 12 DPA). More genes were found to show downregulation at 4 DPA in P271 compared to CS at 4 DPA, whereas more genes showed upregulation in P271 at 8 and 12 DPA compared to CS at 8 and 12 DPA. In the molecular function category, DEGs mostly belonged to sub-categories binding and catalytic activity at 4, 8, and 12 DPA. In both sub-categories, more downregulated genes were found at 4, and 8 DPA and more upregulated genes were found at 12 DPA in both P271 and CS.

According to KEGG pathway enrichment analysis (Fig. 3), in CS vs P291 comparison at 4DPA, more genes were found to be involved in starch and sucrose metabolism, biosynthesis of amino acids, and plant-pathogen interaction. Whereas, in CS vs P271 comparison at 8 DPA, more genes were reported to be involved in phenylpropanoid biosynthesis, plant hormone signal transduction in addition to the starch and sucrose metabolism, and plant-pathogen interaction. On the other hand, in CS vs P271 comparison at 12 DPA, more genes were found to be involved in ribosome biogenesis, pyrimidine metabolism, purine metabolism, and spliceosome in addition to starch and sucrose metabolism and plant hormone signal transduction.
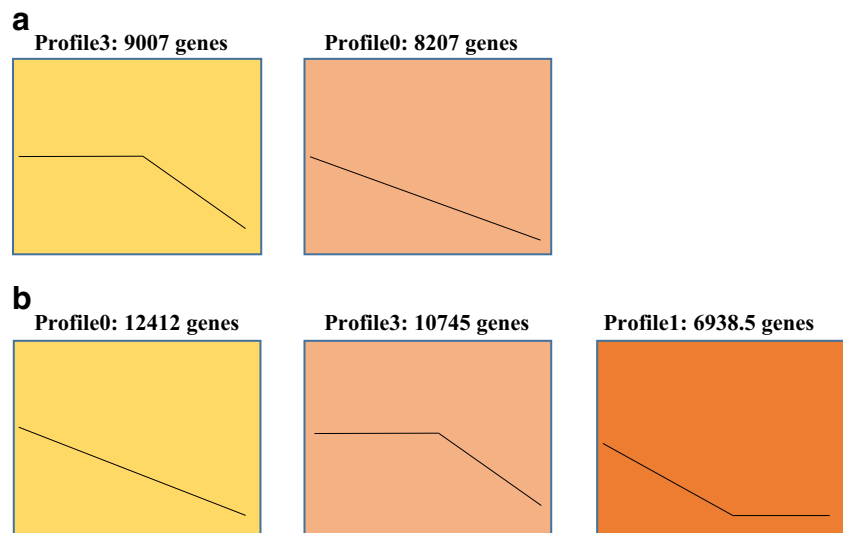
Differences in the protein profiles of P271 and Chinese Spring (CS) during early grain development

In this study, an iTRAQ-based quantitative proteome characterization approach was used to investigate the early grain

Fig. 1 Trend analysis of differentially expressed genes

developmental differences in two wheat cultivars P271 and CS with contrasting grain characteristics. For this purpose, protein profiles at three grain-developmental stages 4, 8, and 12 DPA were obtained. A total of 7548 proteins were identified in CS, and 7525 proteins were identified in P271.

Integrative analysis of the proteome and transcriptome during early grain development

Both transcriptomic and proteomic data are important in determining the molecular changes taking place during early grain development. Same samples were used for both transcriptomic and proteomic analysis; therefore, by comparing two data sets, it is possible to study the relationship between transcription and protein accumulation during early grain development in wheat. The analysis was performed at three levels to study concordance between transcriptional and translational changes: (i) the number of identified proteins or genes, (ii) the quantitation of proteins and transcripts, and (iii) their differential accumulation. In a direct comparison of transcript and protein abundance, a total of 5468, 5526, and 4964 sequences showed correspondence at three grain developmental stages. The distribution of the corresponding mRNA:protein ratios ($\log_2$-transformed ratios) is shown by a scatterplot analysis. Concordance tests revealed an insignificant relationship between mRNA and protein ratios (Pearson's correlation, $R = -0.1573$–0.1713; Fig. 4). However, we observed some concordant dots, representing correspondence between protein abundance and transcript accumulation (red dots in Fig. 4). On the other hand, green and blue dots in Fig. 4 respectively represent expression only at transcriptional or protein level. From the results, we can see that the RNA:protein ratios across the three developmental stages were found to fall in quadrants b, d, h, and f, where the

mRNA:protein level showed a poor correspondence between the transcripts and proteins. In summary, based on these results, a substantial degree of post-transcriptional regulatory activity could be anticipated during early wheat grain development, which was not reported in earlier studies that focused either on transcript or protein analysis.

Carbon metabolism, ribosome, starch, and sucrose metabolism play an important role during the early grain development. It is noteworthy that, at 4 DPA, compared to CS, genes involved in starch and sucrose metabolism are highly expressed in P271, while the abundance of corresponding proteins is quite low. Interestingly an opposite effect was observed at 12 DPA, where proteins involved in starch and sucrose metabolism exhibited high abundance, but a low expression at the transcript level (Figs. S1, S2, and S3).

We compared the DEGs (2-fold difference) and DEPs (1.5-fold difference) identified from P271-CS comparison. The analysis unraveled 61, 81, and 103 DEGs/DEPs respectively at 4, 8, and 12 DPA (see Tables S1, S2, and S3). It is noteworthy that more DEGs/DEPs were identified with the advancement in the developmental stage, and more of these genes/proteins belonged to carbohydrate metabolism, amino acid metabolism, lipid metabolism, cofactor and vitamin metabolism, and translation (Tables S1, S2, and S3).

miRNA target site analysis

Comparison with the wheat miRNA sequences allowed identification of a total of 4346 genes with miRNA binding site(s) in the genes expressing during endosperm development in two wheat cultivars, CS and P271. Out of a total of 4346 genes, 355, 56, and 311 genes respectively expressed differentially at 4, 8, and 12 DPA between CS and P271. When the number of DEGs and DEPs was compared in context of the genes with
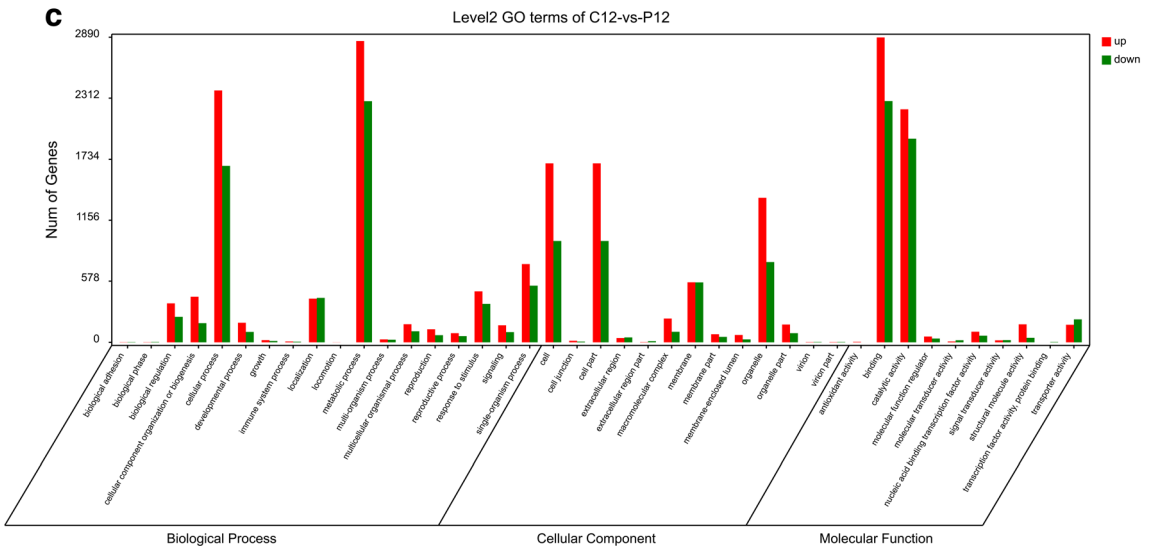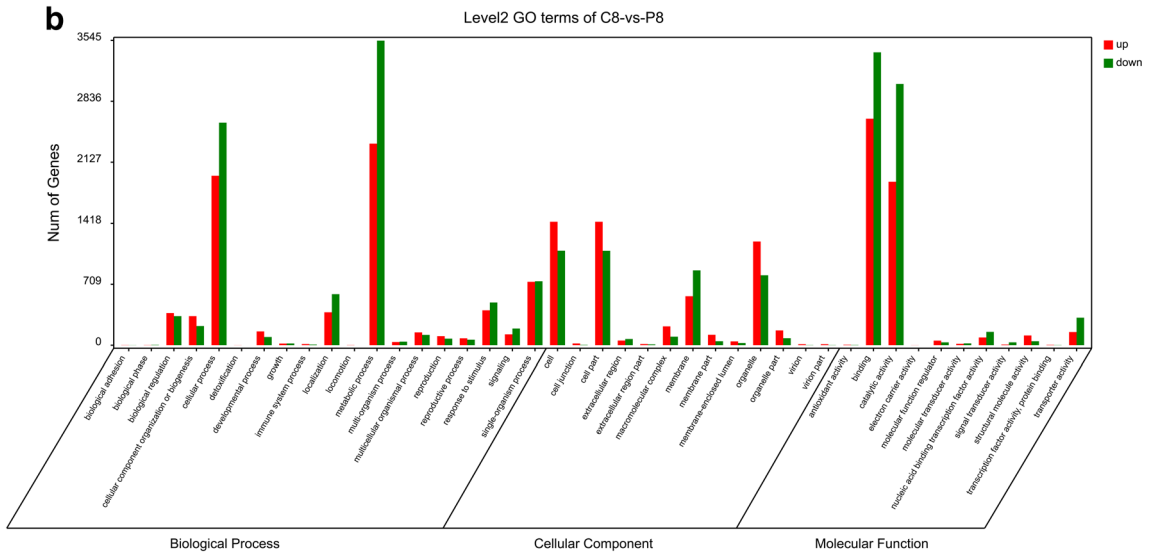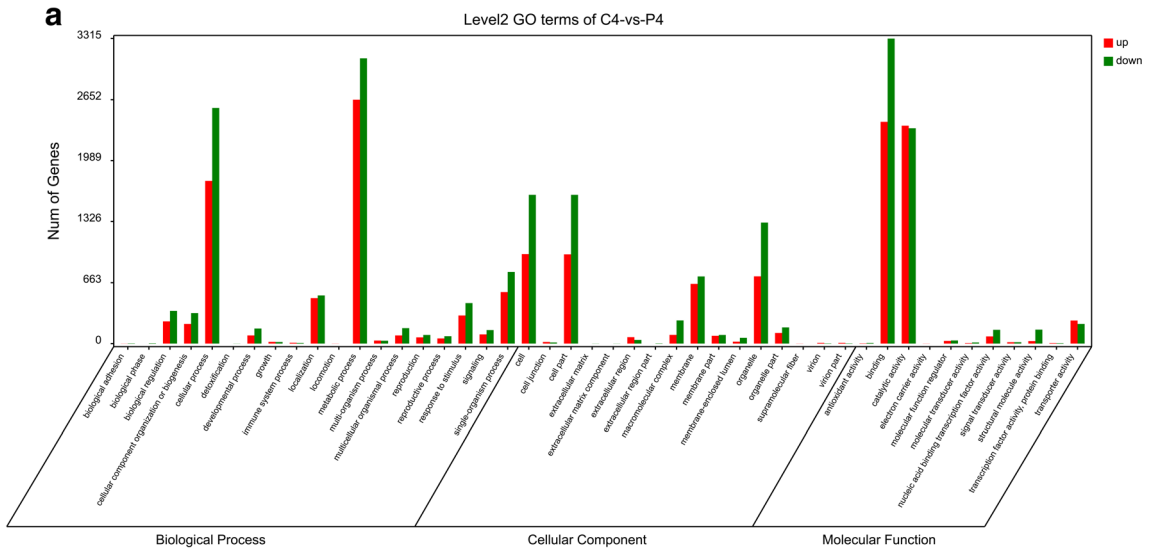
Fig. 2 Bar diagram showing distribution of differentially expressed genes at 4 (**a**)-, 8 (**b**)-, and 12 (**c**)-day post-anthesis (DPA) in wheat cultivars Chinese Spring and P271 comparisons. The proteins were classified based on their predicted functions into biological process, cellular component, and molecular function

known miRNA target-site, this result indicated that the post-transcriptional regulations of gene expression might have played an important role during early grain development in common wheat.

Genomic distribution of DEGs/DEPs

Gene showing differential expression at both transcriptional and translational levels between CS and P271 were assigned respectively to the common wheat chromosomes and sub-chromosomal locations, and their genomic distribution in relation to the known genes/QTLs (quantitative trait loci) for grain traits was studied. As evident from Fig. 5a, 75.4% of DEGs/DEPs at 4 DPA localized to sub-genomes A (34.4%) and B (41%) of common wheat. This trend continued at 8 DPA, where a maximum number of DEGs/DEPs localized to the A sub-genome, but less than the expected number of genes mapped to the B sub-genome. In contrast, more DEGs/DEPs mapped to the B sub-genome at 12 DPA (Fig. 5a). A more detailed analysis performed at the level of wheat homoeologous chromosome groups unveiled that more DEGs/DEPs map to group 3 and 6 chromosomes at 4 DPA, whereas more DEGs/DEPs map to group 1, 3 and 4 chromosomes at 8 DPA, and group 1 and 5 chromosomes at 12 DPA (Fig. 5b). At the chromosome level, more DEGs/DEPs were mapped to chromosomes 1A, 4B, and 5B at all studied time points during grain development, to chromosomes 3A, 3B, 4A, and 6A at 4 and 8 DPA, as well as to chromosomes 1B and 1D at 8 and 12 DPA (Fig. 5c). On the other hand, DEGs/DEPs mapping to chromosome 6D were exclusively identified at 4 DPA, to chromosomes 2A and 7D at 8 DPA, and

chromosome 2B at 12 DPA (Fig. 5c). The relative contribution of each chromosome to the total number of DEG/DEPs identified per developmental stage was also plotted and the pattern showed that at 4 DPA more DEGs/DEPs were mapped to chromosomes 3B and 5B, whereas at 8 DPA to chromosomes 3B and 4A, as well as at 12 DPA to chromosomes 1B, 2B, and 5B (Fig. S4).

The DEGs/DEPs were assigned to the wheat chromosome zippers, and their respective genetic locations on the map (in centiMorgan) were based on the high-density consensus map of SNP makers, which was used to align genes in the wheat chromosome zippers. 88.5% (54 out of 61) of DEGs/DEPs at 4 DPA, 91.4% (74 out of 81) of DEGs/DEPs at 8 DPA, and 90.3% (93 out of 103) of DEGs/DEPs at 12 DPA could be assigned to the wheat chromosome zippers. The distribution patterns of DEGs/DEPs identified at different grain developmental stages (4, 8, and 12 DPA) were also studied along the chromosome axis. A closer look on the distribution patterns suggested that at 4 DPA, most DEGs/DEPs map to centromeric and pericentromeric regions of the chromosomes, whereas the distribution of DEGs/DEPs at 8 and 12 DPA was much more spread across the chromosomes (Fig. 6; Fig. S5). However, the peaks in the chromosome plots, which represent the number of DEGs/DEPs mapping to arbitrary chromosome bins of 50 cM, appear at different locations in the plots at 8 and 12 DPA (Fig. S5). Analysis of the wheat grain transcriptome by Pfeifer et al. (2014) also revealed cell type and stage-dependent genome dominance and asymmetric expression for some groups of genes. Occasionally, the peaks in the plots of different grain developmental stages overlapped, which suggested that the same gene identified as DEG/DEP at two or three developmental stages (Table S4; Fig S5). The homoeologous relationships of DEGs/DEPs were also studied, in 14 cases either 2 or 3 homoeologous genes were identified as DEGs/DEPs at similar or different developmental stages (Table S5). Identification of homoeologous genes as
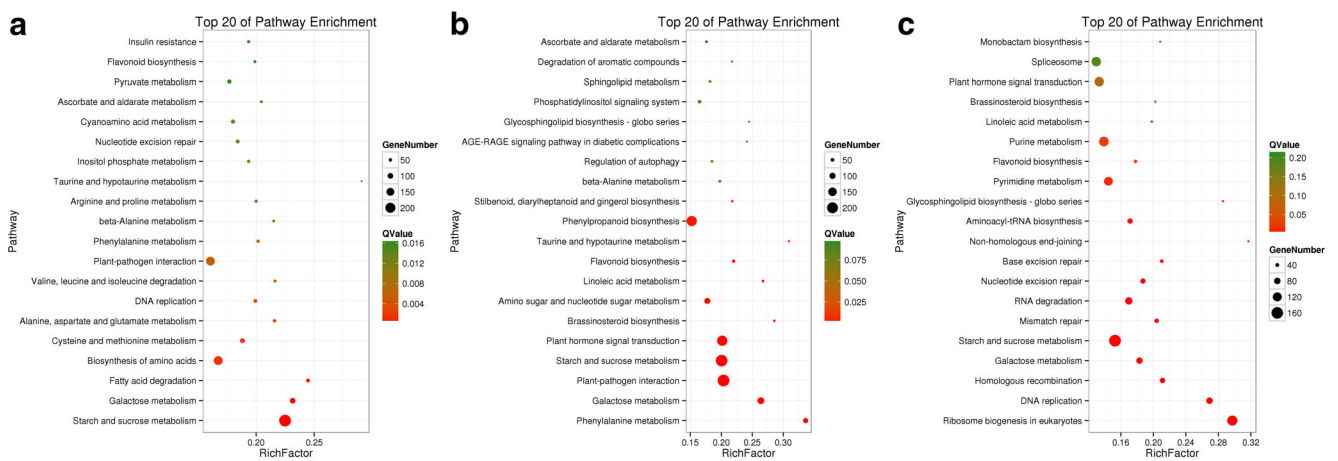

Fig. 3 Bubble diagram showing distribution of differentially expressed genes at 4 (**a**)-, 8 (**b**)-, and 12 (**c**)-day post-anthesis (DPA) in P271 and Chinese Spring comparisons. The DEGs were classified based on the KEGG pathway analysis. The top 20 metabolic pathways are shown here
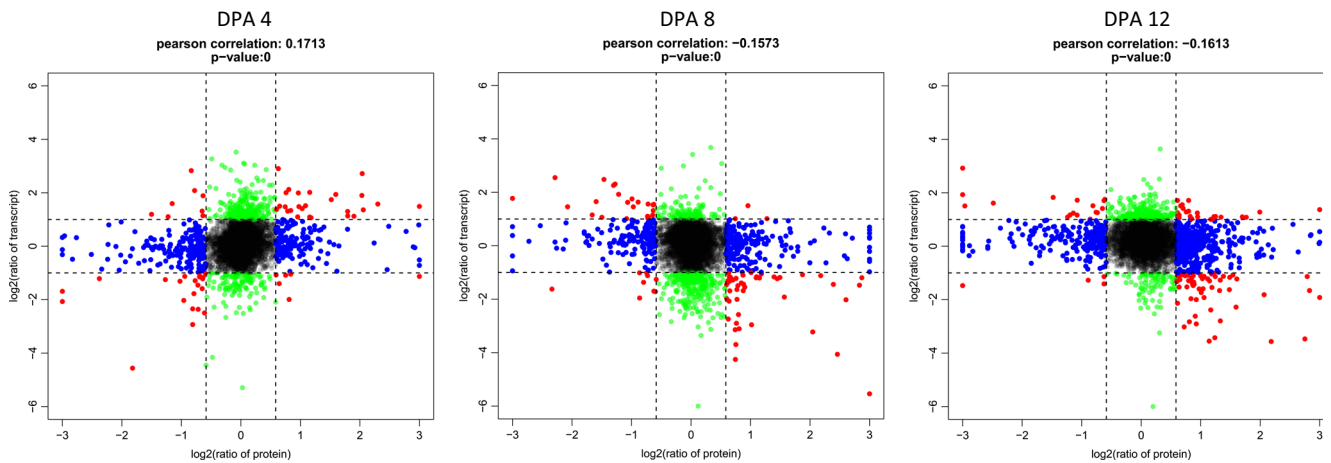
Fig. 4 Comparison of expression ratios of the transcriptomics (*y*-axis) and the proteomics (*x*-axis) data. Log2 expression ratios were calculated from Chinese Spring and P271 comparisons made on data collected at 4 (**a**), 8 (**b**), and 12 (**c**) DPA. Significant changes in expression are color-coded: blue, proteins only; green, transcripts only; red, both

DEGs/DEPs at a specific developmental stage signifies their cumulative expression towards a function whereas their expression at different developmental stages signifies subfunctionalization. In 2 out of 14 cases, the three homoeologous DEPs express at a developmental stage, whereas, in the remaining 12 cases only two homoeologous express at similar or different developmental stages (Table S5).

Genomic locations of 24 out of the 221 chromosomally localized DEGs/DEPs overlapped with known quantitative trait loci (QTLs) for the grain yield traits such as thousand grain weight (TGW), grain length (GL), grain width (GW), and grain thickness (GT) (Table S6; Fig. S6). Out of 23 QTLs that coincided with DEGs/DEPs, 5 QTLs have PVE (phenotypic variation explained) values larger than 10%. Out of these 5 QTLs, 3 QTLs (*QGl.cau-1B.1*, *QGl.cau-2A.2*, and *QGl.cau-3B.1*) that
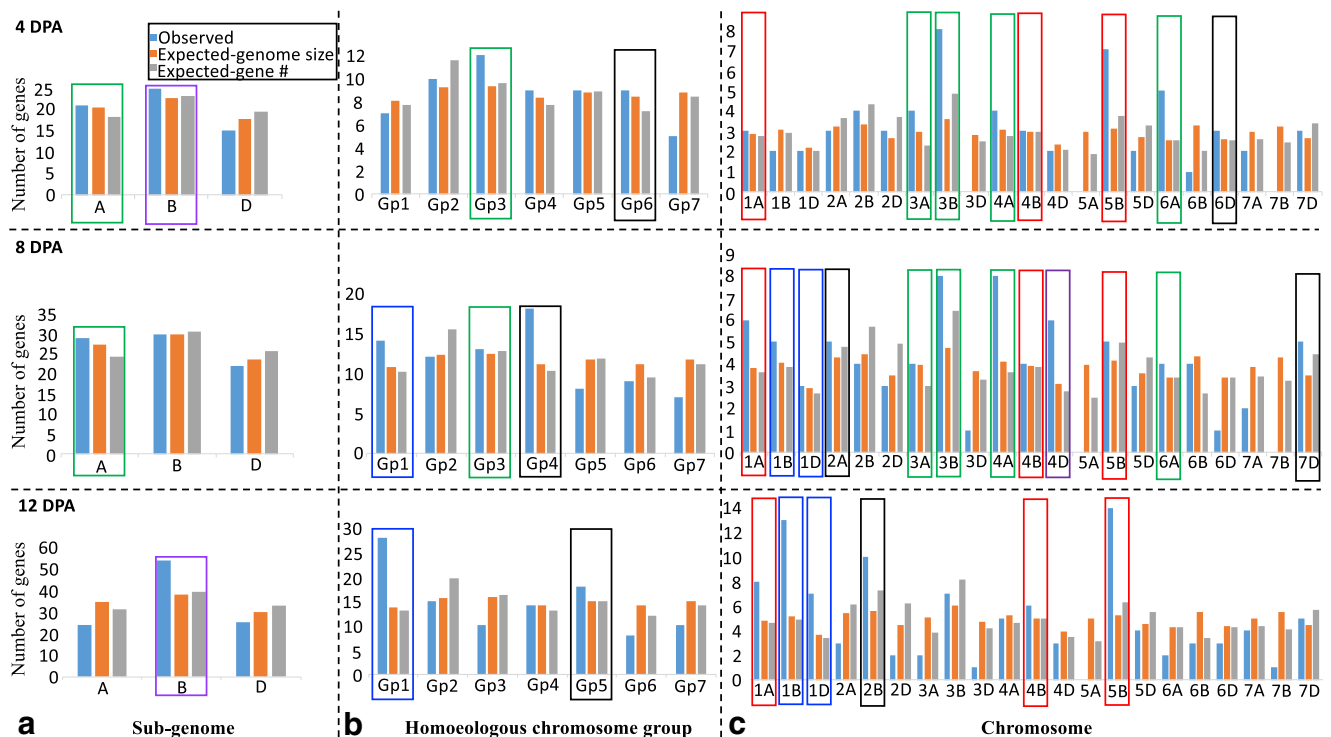

Fig. 5 Bar diagrams showing genomic distribution of genes (total 245) expressed differentially at both transcriptional and translational level between common wheat cultivars Chinese Spring and P271: **a** subgenomes, **b** homoeologous chromosome groups, and **c** specific chromosomes. Common genomic distribution patterns across different grain developmental stages were marked by different color boxes such as red box indicates DEGs identified in A, B, and D sub-genomes; green, A and B sub-genomes; blue, B and D sub-genomes; purple, A and D sub-genomes, and black, either A, B or D sub-genome
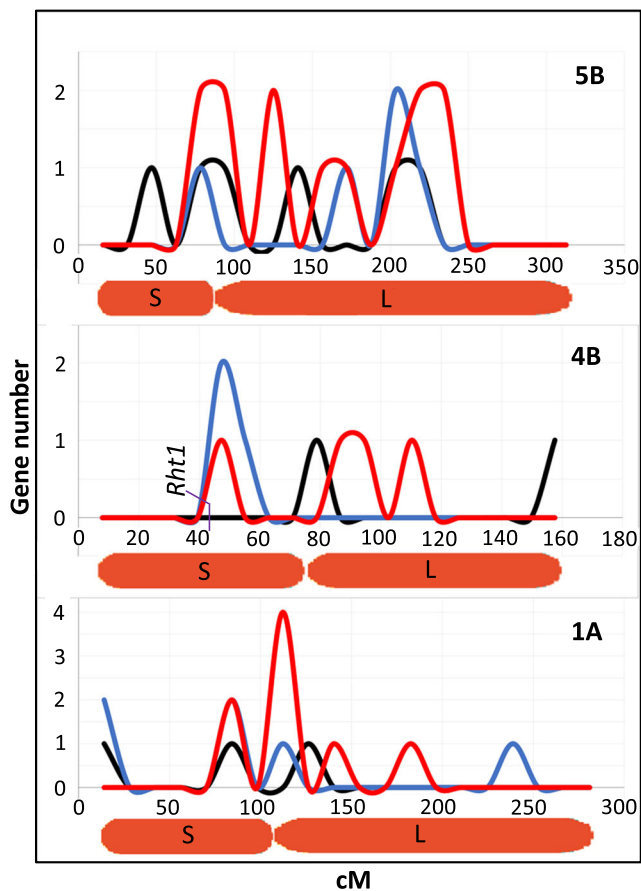
Fig. 6 Chromosomal distribution of differentially expressed genes/proteins. cM location of genes and the centromere on each chromosome were determined on the basis of the wheat chromosome zippers. Peaks in the chromosome plots represent the number of DEGs/DEPs mapping to arbitrary chromosome bins of 50 cM each. Location of the wheat *Rht1* gene on chromosome 4B was determined on the basis of Cabral et al. (2018)

cumulatively explained 59.7% phenotypic variation for GL coincided respectively with DEGs Traes_1BL_301F6C93C, Traes_2AS_61CA61AD3, and TRAES3BF067800180CFD_g, and two QTLs (*QGt.cau-3B.3* and *QGt.cau-5B.4*) that cumulatively explained 36.57% phenotypic variation for GT coincided respectively with DEGs TRAES3BF053500050CFD_g and Traes_5BL_C1F897A4A (Table S6; Fig. S6).

Discussion

Wheat grain development is a highly coordinated and genetically programmed process that involves a series of physiological, biochemical, and molecular changes. In the present study, RNA-seq and iTRAQ technologies were used to investigate the differences in the transcriptome and proteome between the large- and small-grained wheat cultivars. Thousands of genes that differentially expressed during the three grain developmental stages were identified by transcript profiling.

A large number of genes showed differential expression/accumulation at both transcriptional and translational levels (Tables S1, S2, and S3). On the other hand, genes that exhibited inconsistency between transcription and translation suggested that post-transcriptional regulations might play an important role in the regulation of grain development (Wu et al. 2014). Therefore, it is important to obtain both the transcriptomic and proteomic data to decipher the molecular processes involved in grain development. The integrated transcriptomic and proteomic data increase the possibility of identifying genes/proteins that are possibly involved in grain development.

In the present study, however, the correspondence between mRNA levels and protein abundance was low (Pearson's correlation, $R = -0.1613$–0.1713, see Fig. 4), which endorsed the previous findings (Vogel and Marcotte 2012; Hu et al. 2013; Walley et al. 2013). From transcription to translation, many factors influence the correspondence between mRNA level and protein abundance, such as post-transcriptional regulation and differential protein and mRNA degradation rates (Pechanova et al. 2013; Ponnala et al. 2014). The spatiotemporal distribution of the protein might reflect towards the gene function. Another possible explanation for this result is the translational and/or post-translational modifications or processing of the proteins that determine their final quantities, which do not necessarily correspond with RNA expression. Comparative analysis of DEGs and DEPs also revealed only a small amount of overlap at transcriptomic and proteomic levels, which further confirmed the previous findings.

Notably, more DEGs/DEPs were predicted to be involved in carbohydrate metabolism, amino acid metabolism, lipid metabolism, cofactor and vitamin metabolism, and translation during the early grain developmental stages. These results are in correspondence with the previous studies (Zhang et al. 2009; Jia et al. 2013; Wu et al. 2014), where the metabolic pathways of plant hormone signal transduction, starch and sucrose metabolism, and plant-pathogen interaction were shown to be differentially regulated during grain development and fruit ripening. Recently, RNA sequencing and comparative transcriptome analysis were performed during the early grain development in two barley landraces with contracting starch biosynthesis traits. In line with previous studies, the differentially expressed genes in this study belonged to α-amylases, lipid-transport proteins, transcription factors of HD-Zip and MYB families, translational machinery (Nuclear Factor-Y), and starch biosynthesis enzymes (SuSy, AGPase, and SBE2b) (Tang et al. 2017). More recently, a global transcriptome analysis performed in a winter wheat cultivar Xiaoyan-6 at early grain developmental stages also identified a grain-specific co-expression regulation network of transcription factors (TFs) which play a role in the signaling-related biological processes, cellular component organization, and nutrient deposition during grain filling (Chi et al. 2019).

Grain size of wheat is generally described in terms of grain length and grain width. Almost all wheat chromosomes have been identified to carry QTLs for grain size and/or grain shape (for reviews, see Nadolska-Orczyk et al. 2017; Gupta et al. 2006), albeit the majority of the QTL for grain width (W), length (L), grain area, the ratio L/W, and the FFD ((grain weight/(grain length × grain width)) were identified to localize on five wheat chromosomes, 1A, 3A, 4B, 5A, and 6A (Gegas et al. 2010; Rustgi et al. 2013; Gupta et al. 2006 and references cited therein). Similar chromosomes except 5A were also identified to carry DEGs/DEPs in the present study. However, only a few of the genes underlying QTLs for grain size or shape have been so far cloned. For instance, the grain size locus *TaGS-D1* on chromosome 7DS (an ortholog of the rice *OsGs3* gene located on chromosome 3) associated with grain length and grain weight (Zhang et al. 2014). The grain width and grain weight locus *TaGw2* on chromosome 6A (an ortholog of the rice *GW2* locus on chromosome 2) (Su et al. 2011), and the two popular "green revolution" genes *Rht1* (Rht-B1) and *Rht2* (Rht-D1) (Peng et al. 1999) on chromosomes 4B and 4D. Interestingly, the wheat chromosomes 4B, 4D, 6A, and 7D that house the abovementioned grain size/shape genes were also identified to carry DEGs/DEPs identified in the present study. In fact, the peaks representing the distribution of DEG/DEP on wheat chromosomes overlap with the respective locations of these genes on the chromosome (Fig. 6).

Collectively, the results of the genome analysis correspond with the earlier studies that identified wheat chromosomes 1A, 3A, 4B, 5B, 6A, and 7D to carry major genes/QTLs controlling grain length/width in common wheat. Additionally, the genome localization of DEGs/DEPs gave an idea about the genomic locations of the genes contributing to the differences between CS and P271, and this information could be used to develop genetic maps of specific chromosomes and map the expression QTLs contributing to the differences in the grain characteristics of the two common wheat cultivars. The coincident localization of DEGs/DEPs and the QTLs for grain yield traits identified in a recombinant inbred line population of two winter wheat genotypes Yanda1817 and Beinong6 also supported our conclusion about the involvement of the identified DEGs/DEPs in grain development traits (Wu et al. 2015). The functions of these genes could be further validated by induced mutagenesis, allele mining, and/or production of transformants, which is out of the scope of the present study. Briefly, this study has opened up the possibility of identifying previously uncharacterized genes contributing to grain development traits, which will add to the repertoire of formerly characterized grain yield genes in common wheat.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Altenbach SB, Vensel WH, Dupont FM (2011) The spectrum of low molecular weight alpha-amylase/protease inhibitor genes expressed in the US bread wheat Butte 86. BMC Res Notes 4:242

Bancel E, Bonnot T, Davanture M, Branlard G, Zivy M, Martre P (2015) Proteomic approach to identify nuclear proteins in wheat grain. J Proteome Res 14:4432–4439

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Cabral AL, Jordan MC, Larson G, Somers DJ, Humphreys DG, McCartney CA (2018) Relationship between QTL for grain shape, grain weight, test weight, milling yield, and plant height in the spring wheat cross RL4452/'AC Domain'. PLoS ONE 13:e0190681

Cao R, Guo L, Ma M, Zhang W, Liu X, Zhao H (2019) Identification and functional characterization of squamosa promoter binding protein-like gene *TaSPL16* in wheat (*Triticum aestivum* L.). Front Plant Sci 10:212

Chi Q, Guo L, Ma M, Zhang LJ, Mao HD, Wu BW, Liu XL, Ramirez-Gonzalez RH, Uauy C, Appels R, Zhao HX (2019) Global transcriptome analysis uncovers the gene co-expression regulation network and key genes involved in grain development of wheat (*Triticum aestivum* L.). Funct Integr Genomic. https://doi.org/10.1007/s10142-019-00678-z

Dupont FM, Vensel WH, Tanaka CK, Hurkman WJ, Altenbach SB (2011) Deciphering the complexities of the wheat flour proteome using quantitative two-dimensional electrophoresis, three proteases and tandem mass spectrometry. Proteome Sci 9:10

Ernst J, Bar-Joseph Z (2006) STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics 7(1):191

Gao LY, Wang AL, Li XH, Dong K, Wang K, Appels R, Ma WJ, Yan YM (2009) Wheat quality related differential expressions of albumins and globulins revealedd by two-dimensinoal difference gel electrophoresis (2-D DIGE). J Proteomics 73: 279-296

Ge P, Hao P, Cao M, Guo G, Lv D, Subburaj S, Li X, Yan X, Xiao J, Ma W, Yan Y (2013) iTRAQ-based quantitative proteomic analysis

reveals new metabolic pathways of wheat seedling growth under hydrogen peroxide stress. Proteomics 13:3046–3058

Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Sayers L, Doonan JH, Snape JW (2010) A genetic framework for grain size and shape variation in wheat. Plant Cell 22:1046–1056

Gu A, Hao P, Lv D, Zhen S, Bian Y, Ma C, Xu Y, Zhang W, Yan Y (2015) Integrated proteome analysis of the wheat embryo and endosperm reveals central metabolic changes involved in the water deficit response during grain development. J Agric Food Chem 63:8478–8487

Guo G, Ge P, Ma C, Li X, Lv D, Wang S, Ma W, Yan Y (2012) Comparative proteomic analysis of salt response proteins in seedling roots of two wheat varieties. J Proteome 75:1867–1885

Gupta PK, Rustgi S, Kumar N (2006) Genetic and molecular basis of grain size and grain number and its relevance to grain productivity in higher plants. Genome 49:565–571

Hu GJ, Koh J, Yoo MJ, Grupp K, Chen S, Wendel JF (2013) Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. New Phytol 200:570–582

Jia H, Wang Y, Sun M, Li B, Han Y, Zhao Y, Li X, Ding N, Li C, Ji W, Jia W (2013) Sucrose functions as a signal involved in the regulation of strawberry fruit development and ripening. New Phytol 198:453–465

Juhász A, Moolhuijzen P, Bellgard M, Appels R, Békés F (2012) Wheat grain proteomics for the food industry. In: Toldrá F, Nollet LML (eds) Proteomics in Foods. Springer, pp 341–377

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:109–114

Kumar RR, Singh K, Ahuja S, Tasleem M, Singh I, Kumar S, Grover M, Mishra D, Rai GK, Goswami S, Singh GP, Chinnusamy V, Rai A, Praveen S (2019) Quantitative proteomic analysis reveals novel stress-associated active proteins (SAAPs) and pathways involved in modulating tolerance of wheat under terminal heat. Funct Integr Genomics 19:329–348

Laudencia-Chingcuanco DL, Stamova BS, Lazo GR, Cui X, Anderson OD (2006) Analysis of the wheat endosperm transcriptome. J Appl Genet 47:287–302

Laudencia-Chingcuanco DL, Stamova BS, You FM, Lazo GR, Beckles DM, Anderson OD (2007) Transcriptional profiling of wheat caryopsis development using cDNA microarrays. Plant Mol Biol 63:651–668

Ma C, Zhou J, Chen G, Bian Y, Lv D, Li X, Wang Z, Yan Y (2014) iTRAQ-based quantitative proteome and phosphoprotein characterization reveals the central metabolism changes involved in wheat grain development. BMC Genomics 15:1029

Mak Y, Skylas DJ, Willows R, Connolly A, Cordwell SJ, Wrigley CW, Sharp PJ, Copeland L (2006) A proteomic approach to the identification and characterisation of protein composition in wheat germ. Funct Integr Genomics 6:322–337

Nadaud I, Tasleemtahir A, Chateigner-Boutin AL, Chambon C, Viala D, Branlard G (2015) Proteome evolution of wheat (*Triticum aestivum* L.) aleurone layer at fifteen stages of grain development. J Proteome 123:29–41

Nadolska-Orczyk A, Rajchel IK, Orczyk W, Gasparis S (2017) Major genes determining yield-related traits in wheat and barley. Theor Appl Genet 130:1081–1098

Olsen OA (2004) Nuclear endosperm development in cereals and *Arabidopsis Thaliana*. Plant Cell 16:S214–S227

Pechanova O, Takac T, Samaj J, Pechan T (2013) Maize proteomics: an insight into the biology of an important cereal crop. Proteomics 13:637–662

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256–261

Peng Z, Wan M, Li F, Lv H, Li C, Xia G (2009) A proteomic study of the response to salinity and drought stress in an introgression strain of bread wheat. Mol Cell Proteomics 8:2676–2686

Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, International Wheat Genome Sequencing Consortium, Mayer KF, Olsen OA (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345:1250091

Ponnala L, Wang Y, Sun Q, van Wijk KJ (2014) Correlation of mRNA and protein abundance in the developing maize leaf. Plant J 78:424–440

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, Khedikar Y, Robinson SJ, Cory AT, Florio T, Concia L, Juery C, Schoonbeek H, Steuernagel B, Xiang D, Ridout CJ, Chalhoub B, KFX M, Benhamed M, Latrasse D, Bendahmane A, International wheat genome sequencing consortium, Wulff BBH, Appels R, Tiwari V, Datla R, Choulet F, Pozniak CJ, Provart NJ, Sharpe AG, Paux E, Spannagl M, Bräutigam A, Uauy C (2018) The transcriptional landscape of polyploid wheat. Science 361:eaar6089

Rangan P, Furtado A, Henry RJ (2017) The transcriptome of the developing grain: a resource for understanding seed development and the molecular control of the functional and nutritional properties of wheat. BMC Genomics 18:766

Rustgi S, Shafqat MN, Kumar N, Baenziger PS, Ali ML, Dweikat I, Campbell BT, Gill KS (2013) Genetic dissection of yield and its component traits using high-density composite map of wheat chromosome 3A: bridging gaps between QTLs and underlying genes. PLoS ONE 8:e70526

Sakuma S, Golan G, Guo Z, Ogawa T, Tagiri A, Sugimoto K, Bernhardt N, Brassac J, Mascher M, Hensel G, Ohnishi S, Jinno H, Yamashita Y, Ayalon I, Peleg Z, Schnurbusch T, Komatsuda T (2019) Unleashing floret fertility in wheat through the mutation of a homeobox gene. Proc Natl Acad Sci USA 116:5182–5187. https://doi.org/10.1073/pnas.1815465116

Shewry PR, Mitchell RAC, Tosi P, Wan Y, Underwood C, Lovegrove A, Freeman J, Toole GA, Mills ENC, Ward JL (2012) An integrated study of grain development of wheat (cv. Hereward). J Cereal Sci 56:21–30

Skylas DJ, Van Dykc D, Wrigley CW (2005) Proteomics of wheat grain. J Cereal Sci 41:165–179

Su Z, Hao C, Wang L, Dong Y, Zhang X (2011) Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 122:211–223

Tang YW, Zeng XQ, Wang YL, Bai LJ, Xu QJ, Wei ZX, Yuan HJ, Nyima T (2017) Transcriptomics analysis of hulless barley during grain development with a focus on starch biosynthesis. Funct Integr Genomic 17:107–117

Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13:227–232

Walley JW, Shen Z, Sartor R, Wu KJ, Smith LG, Briggs SP (2013) Reconstruction of protein networks from an atlas of maize seed proteotypes. Proc Natl Acad Sci U S A 110:E4808–E4817

Wan Y, Poole RL, Huttly AK, Toscano-Underwood C, Feeney K, Welham S, Gooding MJ, Mills C, Edwards KJ, Shewry PR (2008) Transcriptome analysis of grain development in hexaploid wheat. BMC Genomics 9:121

Wu J, Xu Z, Zhang Y, Chai L, Yi H, Deng X (2014) An integrative analysis of the transcriptome and proteome of the pulp of a spontaneous late-ripening sweet orange mutant and its wild type improves our understanding of fruit ripening in citrus. J Exp Bot 65:1651–1671

Wu QH, Chen YX, Zhou SH, Fu L, Chen JJ, Xiao Y, Zhang D, Ouyang SH, Zhao XJ, Cui Y, Zhang DY, Liang Y, Wang ZZ, Xie JZ, Qin JX, Wang GX, Li DL, Huang YL, Yu MH, Lu P, Wang LL, Wang L, Wang H, Dang C, Li J, Zhang Y, Peng HR, Yuan CG, You MS, Sun QX, Wang JR, Wang LX, Luo MC, Han J, Liu ZY (2015) High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817 x Beinong6. PLoS ONE 10:e0118144

Yang M, Dong J, Zhao W, Gao X (2016) Characterization of proteins involved in early stage of wheat grain development by iTRAQ. J Proteome 136:157–166

Yang M, Gao X, Dong J, Gandhi N, Cai H, von Wettstein DH, Rustgi S, Wen S (2017) Pattern of protein expression in developing wheat grains identified through proteomic analysis. Front Plant Sci 8:1–13

Yang J, Zhou Y, Wu Q, Chen Y, Zhang P, Zhang Y, Hu W, Wang X, Zhao H, Dong L, Han J, Liu Z, Cao T (2019) Molecular characterization of a novel *TaGL3-5A* allele and its association with grain length in wheat (*Triticum aestivum* L.). Theor Appl Genet 132:1799–1814. https://doi.org/10.1007/s00122-019-03316-1

Zhai H, Feng Z, Du X, Song Y, Liu X, Qi Z, Song L, Li J, Li L, Peng H, Hu Z, Yao Y, Xin M, Xiao S, Sun Q, Ni Z (2018) A novel allele of *TaGW2-A1* is located in a finely mapped QTL that increases grain weight but decreases grain number in wheat (*Triticum aestivum* L.). Theor Appl Genet 131:539–553

Zhang M, Yuan B, Leng P (2009) The role of ABA in triggering ethylene biosynthesis and ripening of tomato fruit. J Exp Bot 60:1579–1588

Zhang Y, Liu J, Xia X, He Z (2014) *TaGS-D1*, an ortholog of rice *OsGS3*, is associated with grain weight and grain length in common wheat. Mol Breed 34:1097–1107

Zhang Y, Li D, Zhang D, Zhao X, Cao X, Dong L, Liu J, Chen K, Zhang H, Gao C, Wang D (2018) Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. Plant J 94:857–866

Zhou SR, Yin LL, Xue HW (2013) Functional genomics based understanding of rice endosperm development. Curr Opin Plant Biol 16: 236–246