



Global transcriptome analysis uncovers the gene co-expression regulation network and key genes involved in grain development of wheat (*Triticum aestivum* L.)

Qing Chi¹ · Lijian Guo¹ · Meng Ma¹ · Lijian Zhang¹ · Hude Mao^{2,3} · Baowei Wu¹ · Xiangli Liu¹ · Ricardo H. Ramirez-Gonzalez⁴ · Cristobal Uauy⁴ · Rudi Appels⁵ · Huixian Zhao^{1,3}

Received: 20 October 2018 / Revised: 24 March 2019 / Accepted: 12 April 2019 / Published online: 21 May 2019
© The Author(s) 2019

Abstract

Wheat grain development is a robust biological process that largely determines grain quality and yield. In this study, we investigated the grain transcriptome of winter wheat cv. Xiaoyan-6 at four developmental stages (5, 10, 15, and 20 days post-anthesis), using high-throughput RNA sequencing (RNA-Seq). We identified 427 grain-specific transcription factors (TFs) and 1653 differentially expressed TFs during grain development as well as a grain co-expression regulation network (GrainNet) of the TFs and their predicted co-expressed genes. Our study identified ten putative key TFs and the predicted regulatory genes of these TFs in wheat grain development of Xiaoyan-6. The analysis was given a firm basis through the study of additional wheat tissues, including root, stem, leaf, flag leaf, grain, spikes (from wheat plants at booting or heading stages) to provide a dataset of 92,478 high-confidence protein-coding genes that were mostly evenly distributed among subgenomes, but unevenly distributed across each of the chromosomes or each of the seven homeologous groups. Within this larger framework of the transcriptomes, we identified 4659 grain-specific genes (SEGs) and 26,500 differentially expressed genes (DEGs) throughout grain development stages tested. The SEGs identified mainly associate with regulation and signaling-related biological processes, while the DEGs mainly involve in cellular component organization or biogenesis and nutrient reservoir activity during grain development of Xiaoyan-6. This study establishes new targets for modifying genes related to grain development and yield, to fine-tune expression in different varieties and environments.

Keywords Wheat · Grain development · Transcriptome · Differentially expressed genes · Transcription factors · Gene co-expression regulation network

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10142-019-00678-z>) contains supplementary material, which is available to authorized users.

✉ Huixian Zhao
hxzhao212@nwfufu.edu.cn

¹ College of Life Sciences, Northwest A & F University, Yangling 712100, Shaanxi, China

² College of Plant Protection, Northwest A & F University, Yangling 712100, Shaanxi, China

³ State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A & F University, Yangling 712100, Shaanxi, China

⁴ John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

⁵ School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia

Introduction

Allohexaploid common wheat (*Triticum aestivum* L., $2n = 6x = 42$, AABBDD) has a complex genome with an overall size of 16 Gb (Appels et al. 2018; Zimin et al. 2017), and is one of the most important cereal crops for human diets worldwide. The increase in global population makes the improvement of wheat yield and quality a consistent and major target of wheat breeding. Wheat grain yield and quality is largely determined by the events occurring during wheat growth and development, and hence, the improvement of the traits of wheat yield and quality requires a better understanding of the biological processes in grain development and their regulation.

Wheat grain development can be broadly divided into three stages: cell division and expansion (0~14 days post-anthesis, DPA), effective grain filling (14~28 DPA), and maturation and desiccation (28 DPA to maturity) (Shewry et al. 2012). Final

grain size and grain weight are largely determined by carpel size (Brinton et al. 2017; Calderini et al. 1999), which is established during the first two developmental stages when the basic structure of grain is generated and reserves (storage compounds like starch and gluten proteins) accumulation is nearly completed (Nadaud et al. 2010; Sabelli and Larkins 2009; Shewry et al. 2012). Grain development relies on gene expression regulated in a very strict chronological order.

Transcriptomics was used in our study to define events during the early grain development and grain filling stages to gain insights into the complex gene regulatory networks that underlies these specific phases. We used winter wheat cv. Xiaoyan-6, an elite Chinese cultivar that is the first cultivar successfully developed by crossing with *Agropyron elongatum* (2n=70) in the early stages of its pedigree, as research materials. Several studies have applied transcriptomics approaches to investigate the gene expression during grain development in wheat (Laudencia-Chinguanco et al. 2007; Shewry et al. 2012; Wan et al. 2008; Yu et al. 2016). However, all these studies were conducted using microarrays, which represent a fraction of the transcriptome and are unable to distinguish between homoeologous genes. High-throughput RNA sequencing (RNA-Seq) has been extensively used to investigate grain development of cereal crops, such as *Oryza sativa* (Gao et al. 2013; Xue et al. 2012), *Zea mays* (Chen et al. 2014; Xiao et al. 2016), and wheat (Pfeifer et al. 2014). However, the accuracy of RNA-seq is dependent on the availability of a high-quality reference sequence and accurate gene models. All the studies mentioned above used either expressed sequence tags (ESTs) or the Chromosome Survey Sequence (CSS) (International Wheat Genome Sequencing Consortium 2014) as references. Recently, the releases of high-quality reference sequences and accurate annotations of hexaploid wheat (Appels et al. 2018; Clavijo et al. 2017; Zimin et al. 2017) have provided novel resources for the analysis of transcriptome from different tissues.

The winter wheat cultivar Xiaoyan-6 has high and stable yield as well as good quality for noodle and bread making, and it has been widely utilized as a donor parent in wheat breeding in China for approximately 30 years. However, the biological process and regulation of grain development of Xiaoyan-6 remain unclear. The main purpose of the present study is to reveal the changes in gene expression and to explore gene co-expression network of transcription factors (TFs) involved in wheat grain development of Xiaoyan-6. We investigated the global transcriptomes of Xiaoyan-6 spikes at two developmental stages and grains at four different developmental stages (5, 10, 15, and 20 DPA) using RNA-Seq, identified grain-specific genes (SEGs), and differentially expressed genes (DEGs) during developmental grains, and conducted GO annotation and KEGG enrichment of these genes. The analysis of vegetative tissues (root, stem, and leaf of five-leaf stage seedling, and flag leaf of wheat plant at heading

stage) provided a useful reference point. We identified differentially expressed transcription factors (TFs) showing large changes in expression during grain development and established a grain co-expression regulatory network (GrainNet) by combining the high connectivity TFs with the genes that were predicted to be regulated by these TFs. The data in this study provides new insights into grain development of an elite donor parent in wheat breeding in China, based on defining functional genes associated with grain yield and quality of wheat.

Materials and methods

Plant materials and sample collection

Wheat cultivar Xiaoyan-6 was planted in the experimental station of Northwest A & F University, Yangling, Shaanxi, China (longitude 108° E, latitude 34° 15' N), during a natural growth season in 2014 to 2015. Cultivation and management of wheat followed local normal production conditions. Ten tissues including root (R), stem (S), and leaf (L) of five-leaf-stage seedling (Zadoks 15) (Zadoks et al. 1974), flag leaf (FL) of wheat plant at booting stage (Zadoks 45), young spike (YS5) from wheat plant at booting stage (Zadoks 45), spike (YS15) from wheat plants at heading stages (Zadoks 53–54), and grains at 5, 10, 15, and 20 DPA (GR5, GR10, GR15, and GR20, respectively) were separately collected. Two independent biological replicates for each sample were collected for transcriptome sequencing, and three were collected for quantitative real-time reverse transcriptase PCR. The collected tissue samples were rapidly frozen in liquid nitrogen and then stored at – 80 °C until use.

Total RNA isolation

Total RNA was extracted using Total RNA Rapid Extraction Kit for Polysaccharides Polyphenol Plant (BioTeke) according to the manufacturer's directions. The quality and quantity of RNA samples were assessed by 1% RNase free agarose gel electrophoresis and NanoDrop 2000 Spectrophotometer (Thermo), respectively. The RNA samples passing the quality control were sent to Gene Denovo (Guangzhou, China) for cDNA library construction and sequencing on the Illumina sequencing platform (Illumina HiSeq™ 2500) after further assessment of accurate quantity of each RNA sample using Agilent 2100 Bio-analyzer (Agilent Technologies, Santa Clara, CA).

cDNA library construction and transcriptome sequencing

cDNA libraries were constructed according to Illumina protocols. Briefly, mRNA was extracted using dynabeads oligo (dT)

and fragmented by fragmentation buffer. Double-stranded cDNAs were synthesized using reverse-transcriptase and random hexamer primers. The cDNA fragments were purified using a QIA quick PCR extraction kit. These purified fragments were washed with EB buffer for end repair of poly (A) addition and then ligated to sequencing adapters. Following agarose gel electrophoresis and extraction of cDNA from gels, the cDNA fragments were purified and enriched by PCR to construct the final cDNA library. The cDNA library was sequenced on the Illumina sequencing platform (Illumina HiSeq™ 2500) using the paired-end technology.

Read alignment and expression quantification

Raw data of RNA sequencing were filtered using a Perl program to get the high-quality clean reads by removing low quality sequences (more than 50% bases with quality lower than 20 in one sequence), reads with more than 10% N bases (bases unknown), and reads containing adaptor sequences. The cleaned reads were aligned to two reference sequences from the same wheat accession (Chinese Spring) that were released at different time: the chromosome-based draft sequence of bread wheat (International Wheat Genome Sequencing Consortium 2014) and the IWGSC RefSeq v1.0 that is the latest released fully annotated reference genome of bread wheat (Appels et al. 2018). The tools used for read alignment and expression quantification included TopHat2 (TopHat 2.1.1, released on Feb. 23, 2016) (Kim et al. 2013) and Cufflinks (version 2.2.1, released on May 5, 2014) (Trapnell et al. 2010). To compare gene expression profiles across different tissues, the transcript levels of individual transcripts in each tissue were normalized as fragments per kilobase of transcript per million mapped reads (FPKM).

Analysis of SEGs and DEGs in developmental grains

We defined a transcript/gene as an SEG if it is expressed only in grain rather than non-grain tissue. Identification of SEGs was conducted by comparing four types of tissue sample group, including RSL (the union of genes expressed in R, S, and L), FL, YS (the union of genes expressed in YS5 and YS15), and GR (the union of genes expressed in GR5, GR10, GR15, and GR20). Statistical tests were applied according to a method previously reported (Wang et al. 2015). DEGs in developmental grains were identified by comparing two consecutive time points (GR5 vs GR10, GR10 vs GR15, and GR15 vs GR20) using EdgeR (Robinson et al. 2010). The DEGs were required to show statistically significant low false discovery rate ($FDR \leq 0.001$) and fold change of absolute value $\log_2 \text{Ratio} \geq 1$ (Wright and Simon 2003).

Expression trend analysis of the DEGs during grain development

Gene expression trend analysis of DEGs during grain development was performed by Short Time-series Expression Miner software (STEM) (version 1.3.11, released on Dec. 26, 2016) (Ernst and Bar-Joseph 2006) on the OmicShare tools platform (www.omicshare.com/tools), a free online platform for data analysis. Before the trend analysis, relative expression level of individual genes across developmental grains was calculated as the logarithm (\log_2) of the fold change, whereas the fold change is the ration of the abundance of a gene in GR10, GR15, or GR20 to the abundance of the same gene in GR5. Then, the trend analysis was conducted. The clustered profiles with P value ≤ 0.05 were considered as significant profiles.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analysis

All the DEGs and the SEGs were subjected to Gene Ontology (GO) (<http://geneontology.org/>) to obtain GO annotation. GO annotation and enrichment analyses were conducted according to a protocol previously described (Zhang et al. 2013). GO enrichment analysis was conducted using the DEGs, the SEGs, and the genes from each expression trends, respectively, against the genes which were expressed across all tissues tested in Xiaoyan-6 (Benjamini and Hochberg 1995). Through the hypothesis test of the P value calculation, the GO terms with P value ≤ 0.05 were defined as significantly enriched GO terms. To visualize GO term enrichment, histograms were generated, where the top ten or 20 significantly enriched GO terms were displayed.

Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.kegg.jp/>) is a major public pathway-related database where functional classification and pathway assignment were provided. The DEGs and the SEGs were separately subjected to KEGG pathway enrichment analysis according to a method previously reported (Zhang et al. 2013). Pathways with P value ≤ 0.05 were defined as significantly enriched pathways.

Weighted gene co-expression network analysis

Co-expression network analysis was performed using all the eight grain samples at four different developmental stages (GR5, GR10, GR15, and GR20), according to the protocol of the weighted gene co-expression network analysis (WGCNA) package in R (version 1.46, released on Mar. 28, 2015) (Langfelder and Horvath 2008), and then sample-clustering tree, modules, and kME (module eigengene-based connectivity) for individual genes were obtained. The parameters used in WGCNA were as follows: $FPKM \geq 1$; cv (Variation of FPKM) ≥ 0.5 ; hierarchal clustering tree: dynamic hybrid tree cut

algorithm; power: 27; minimum module size: 30; minimum height for merging modules: 0.15905. The genes with high degree of connectivity could be potentially major genes in the network. To explore major TFs and their co-expressed genes during wheat grain development, the TFs with high connectivity ($kME \geq 0.98$) and their co-expression genes with top 100 edge weight were used to construct a co-expression network, then the network was visualized by Cytoscape (version 3.4.0, released on May 13, 2016). In the net, circular nodes represent genes and edges represent connection.

Quantitative real-time reverse transcriptase-PCR (qRT-PCR)

Total RNA of individual tissues/organs was extracted using Total RNA Rapid Extraction Kit for Polysaccharides Polyphenol Plant (BioTeke) according to the manufacturer's instructions. GoScript™ Reverse Transcription System (Promega) was used to synthesize cDNA. Three independent biological replicates were included. qRT-PCR was performed on CFX96 real-time PCR (Bio-Rad) with three replicates for each sample using GoTaq® qPCR Master Mix (Promega). Wheat β -actin (GenBank accession number: MF405765.1) was used as a reference gene, and the sequence of all the primers used in this study are listed in Supplementary Table 1. The values of the threshold cycle were analyzed according to the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

Results and discussion

The wheat cv. Xiaoyan-6 transcriptome

The wheat cv. Xiaoyan-6 transcriptome was characterized using RNA libraries from ten tissues sampled throughout the vegetative and reproductive stages, as well as the transition from vegetative to reproductive growth (Supplementary

Table 2). Reads were aligned to the two reference sequences: the IWGSC CSS v2 and the IWGSC RefSeq v1.0, respectively. Consequently, 73% of reads were mapped to the CSS reference, while 85% of reads were mapped to the IWGSC RefSeq v1.0; as we expected, many more reads were mapped to the IWGSC RefSeq v1.0 (Supplementary Table 2). Consequently, 101,202 and 118,264 transcripts were identified when aligned to the CSS and the IWGSC RefSeq v1.0, corresponding to 84,520 and 92,478 annotated genes, respectively (Table 1). Considering the improved and more accurate gene annotations provided by IWGSC Refseq v1.0 rather than the CSS gene models, we continued our data analysis only using the results of the alignment to the IWGSC RefSeq v1.0. Consequently, a total of 92,478 protein-coding genes were detected in the ten tissues tested (in duplicate), covering 83.47% (92,478/110,790) of the annotated high-confidence protein-coding genes in the reference (Table 1).

The heatmap of sample correlations, which based on the expression values (FPKM) of all genes in the 20 samples tested (Supplementary Table 3), showed good reproducibility between the two biological replicates (Supplementary Fig. 1a and Supplementary Fig. 1b). Principal component analysis (PCA) showed that the first principal component could explain 49.6% of total variance and distinguish samples based on tissues identity, discriminating developmental grains (GR5, GR10, GR15, and GR20) from spikes (YS5 and YS15) with vegetative tissues R, S, L, and FL in between; and the second principal component could explain 24.5% of the total variance and separate organs according to developmental stages (Supplementary Fig. 1c). Cluster analysis classified tissue samples into three groups corresponding to their developmental stages, suggesting similar expression patterns in similar tissues (Supplementary Fig. 1c).

To further validate the RNA-Seq data, 20 genes randomly selected were quantified for their expression profiles across all the tested tissues through qRT-PCR with homoeolog-specific primers, including three SEGs (TraesCS2B01G392500,

Table 1 Summary of wheat cv. Xiaoyan-6 RNA Sequencing data from ten tissues analyzed using two different gene models of wheat cv. Chinese Spring as references

Data source	Raw reads data ^{*4}	Clean reads data ^{*4}	Gene number	Transcript number
Wheat cv. Xiaoyan-6 ^{*1}	184.37 Gb	180.59 Gb	92,478 ^{*5} 84,520 ^{*6}	118,264 ^{*5} 101,202 ^{*6}
IWGSC RefSeq v1.0 ^{*2}	–	–	110,790	137,056
IWGSC CSS v2 ^{*3}	–	–	100,934	112,496

^{*1} A Chinese winter wheat cv. Xiaoyan-6 used in the present study

^{*2} IWGSC RefSeq v1.0, the fully annotated reference genome sequence of hexaploid wheat cv. Chinese Spring (Appels et al. 2018)

^{*3} IWGSC CSS v2 presents genome sequence of wheat cv. Chinese Spring: the Chromosome Survey Sequence reference (International Wheat Genome Sequencing Consortium 2014)

^{*4} 1 Gb is equal to ten to the ninth power bp (1 Gb = 10^9 bp)

^{*5} Gene or transcript number using the IWGSC RefSeq annotation v1.0 as a reference

^{*6} Gene or transcript number when using the IWGSC CSS v2 as a reference

TraesCS2A01G013400, and TraesCS4B01G013700) and 17 constitutively (all tissues) expressed genes, *NAC* (TraesCS2A01G338300), *MYB* (TraesCS3A01G336500), and *MIKC_MADS* (TraesCS5B01G286100) also being included (Supplementary Fig. 2 and Supplementary Table 3). It can be shown that the expression patterns of these 20 genes quantified by qRT-PCR were highly consistent with those obtained by RNA-Seq (Supplementary Fig. 2), demonstrating that the RNA-Seq data are reliable.

The overview of different wheat tissue/organ transcriptomes is shown in Fig. 1a and Supplementary Table 4. The average of 67,447 genes was detected in the ten tissues tested, with the least number in FL (63,498) and the most in YS15 (71,224) (Fig. 1a and Supplementary Table 4). About 47.4% (43,871/92,478) of the detected genes were common in all these tissues (Supplementary Table 4), and GO categories of these common genes were over-represented for cellular nitrogen compound metabolic process, cellular biosynthetic process, gene expression, cellular localization, intracellular transport, establishment of localization in cell, protein localization, or organelle organization

(Supplementary Table 5), suggesting that these genes might play general functions essential for the cells of all these tissues. In contrast, only 0.38~2.96% (354/92,478~2734/92,478) of the total genes were tissue-specific or development stage-specific (Fig. 1a and Supplementary Table 4). Moreover, among the vegetative tissue, root had the greatest number of tissue-specific genes (2734), which is approximately 2 to 7 times greater than those of the other tissues (Fig. 1a and Supplementary Table 4). This probably reflects large differences in gene expression profiles between underground and aerial organs. For example, wheat germ agglutinin genes are root-specific, which play an important role in the colonization of plant growth promoting rhizobacteria, enhancing plant growth, preventing disease, and increasing crop yield (Kloepper et al. 1980; Rodriguez-Navarro et al. 2007).

To characterize the subgenome and chromosome distribution of all the expressed genes across wheat tissues, the data of individual gene location and gene expression value in the grain samples (GR5, GR10, GR15, and GR20) were compared across each of the 21 chromosomes (Fig. 1b and Supplementary Table 6), with the 1803 genes of unknown

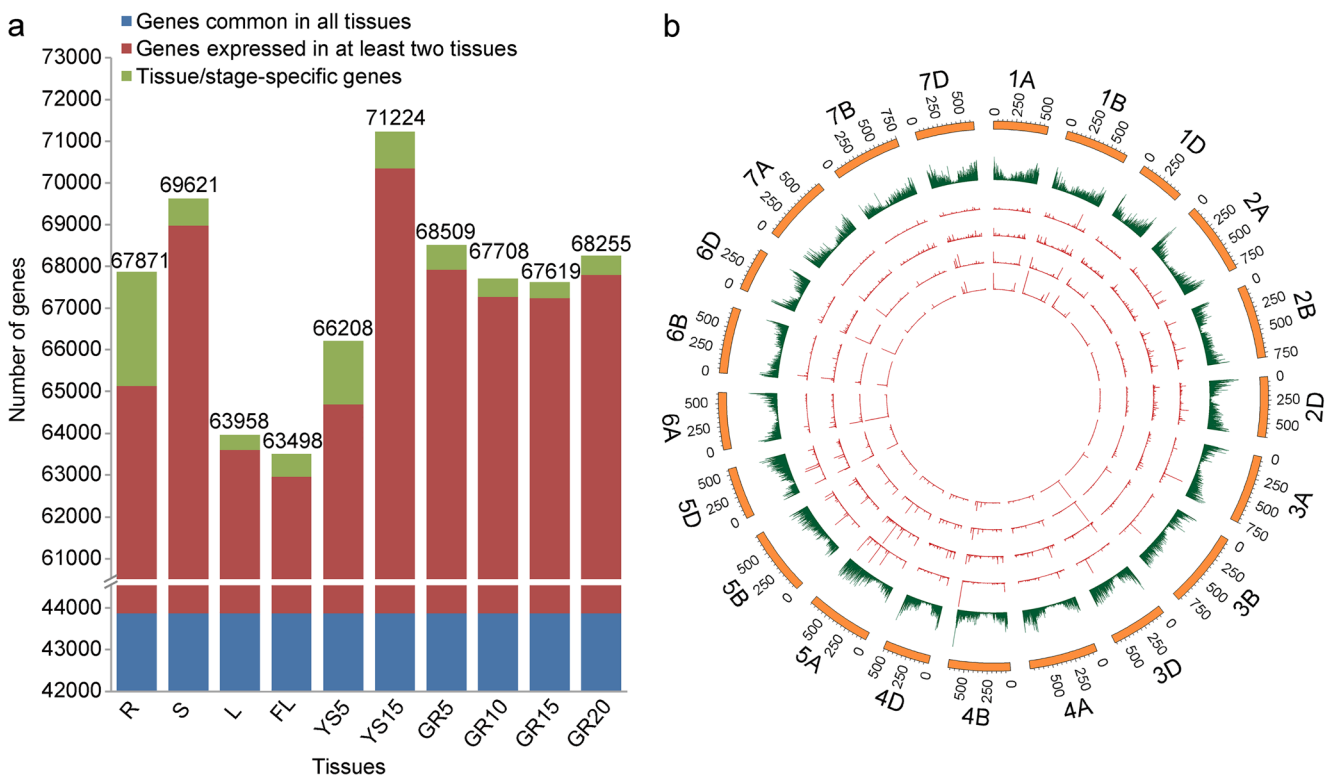


Fig. 1 Overview of the transcriptomes of wheat cv. Xiaoyan-6. **a** Number of genes detected in each of the ten tissues tested. The numbers above individual columns present total number of genes detected in the corresponding tissues. The detail number of the genes common in all tissues, the genes expressed in at least two tissues, and tissue or developmental stage-specific genes were shown in Supplementary Table 4. R, S, and L represent root, stem, and leaf of five-leaf stage seedling, respectively. FL represents flag leaf of wheat at heading stage. YS5 represents young spike of wheat plant at early booting stage. YS15

represents spike of wheat plant at heading stage. GR5, GR10, GR15, and GR20 represent grain at 5, 10, 15, and 20 days post-anthesis, respectively. **b** Circular diagrams of the distribution and the genomic characteristics of all the expressed genes across 21 wheat chromosomes. From outside to inside, each circle represents chromosome name and size (50-Mb tick size), gene density (0 to 47 gene per Mb), and gene expression values (0 to 21,163 FPKM) in grain samples (GR5, GR10, GR15, and GR20, respectively). The length of the line in circles represents the value, FPKM means fragments per kilobase of transcript per million mapped reads

origin (with gene/transcript ID of “TraesCSU” in Supplementary Table 3) being excluded. These expressed genes covered 82.78% (30,050/36,302), 83.19% (30,563/36,738), and 85.84% (30,062/35,021) of the total reference genes in A, B, or D subgenome of the IWGSC RefSeq v1.0, respectively (Supplementary Table 6). Furthermore, the expression levels of genes across individual chromosomes were similar across the grain developmental stages of Xiaoyan-6 (the four most inner circles in Fig. 1b), with exceptions of homoeologous chromosome groups I (1A/1B/1D), II (2A/2B), and VI (6A/6B/6D) that had at least one hotspot (with a median expression level across the grain developmental stages more than 30 FPKM, containing at least five genes) (Supplementary Table 6). It was previously reported that the expression levels of genes across chromosomes of wheat cv. Chinese Spring (CS) were also similar, with the exception of 19 genetic bins that had “hotspots” (with a median expression level > 20 tpm, containing on average 5 genes) across the six tissues, including leaf, root, seedling, stem, spike, and seed (Clavijo et al. 2017). The difference of gene expression distribution across subgenomes and chromosomes between Xiaoyan-6 and CS most likely reflects the differences in the genomes between Xiaoyan-6 and CS (Li et al. 1990). It was demonstrated that Xiaoyan-6 lineage involved an *A. elongatum* cross at the early stages of its pedigree, which resulted in the presence of at least two *Agropyron* chromosome segments within five chromosome arms of Xiaoyan-6, 1AL, 2AS, 5AS, 6AS, and 7BS. In addition, at least two reciprocal interchanges involved in 1A, 2D, 3B, 4D, and 6A present in Xiaoyan-6 (Li et al. 1990).

The high-quality and fully annotated wheat genome sequence IWGSC Refseq v1.0 provided an opportunity to analyze homoeolog-specific gene expression. Transcript levels of all triads were analyzed across the ten tissues tested. It was found that almost one-third of expressed triads exhibited unevenly unbalanced expression among three subgenomes across the tissues of Xiaoyan-6, taking *TaCYP78A* family members (*TaCYP78A3*, *A5*, *A12*, and *A16*) as examples (Supplementary Fig. 3a–d). The homoeologous triads of *TaCYP78A3* (TraesCS7A01G270700, TraesCS7B01G168800, and TraesCS7D01G271100) showed higher expression level in D and B subgenome across tissues (Supplementary Fig. 3a and Supplementary Table 3); the triads of *TaCYP78A5* (TraesCS2A01G175700, TraesCS2B01G201900, and TraesCS2D01G183000) showed unevenly unbalanced expression pattern among three subgenomes across tissues and developmental stages (Supplementary Fig. 3b and Supplementary Table 3); the triads of *TaCYP78A12* (TraesCS5A01G316600, TraesCS5B01G317200, and TraesCS5D01G322900) exhibited higher expression level in A and D subgenome (Supplementary Fig. 3c and Supplementary Table 3), while the expression level of three triads of *TaCYP78A16*

(TraesCS5A01G502000, TraesCS4B01G330500, and TraesCS4D01G327400) were relatively balanced among subgenomes across the tissues (Supplementary Fig. 3d and Supplementary Table 3). Similar unbalanced expressions were also observed in Chinese Spring (Clavijo et al. 2017) and Azhurnaya (Ramirez-Gonzalez et al. 2018).

In general, a total of 92,478 protein-coding genes were expressed across the ten tissues tested (in duplicate) of wheat cv. Xiaoyan-6; almost half of the total expressed genes were common in all the tissues tested, which suggested that these genes play general functions essential for cells, while ~2.96% were tissue-specific or development stage-specific. In addition, the expressed genes across these tissues covered 82.78%, 83.19%, and 85.84% of the total reference genes in A, B, or D subgenome of the IWGSC RefSeq v1.0, respectively, and unevenly distributed within each of the seven homoeologous chromosome groups. Expression levels of genes across individual chromosomes were similar across the grain developmental stages of Xiaoyan-6, with the exception of homoeologous chromosome groups I (1A/1B/1D), II (2A/2B), and VI (6A/6B/6D) that had at least one hotspot (with a median expression level across the grain developmental stages more than 30 FPKM, containing at least five genes). These data complement previous findings on tissue-specific gene expression and gene distribution patterns among subgenomes and chromosomes in hexaploid wheat.

Expression trends and GO enrichment of the genes involved in grain development

Among 92,478 protein-coding genes detected in the ten wheat tissues tested, 79,229 (86%) were identified as being expressed in developmental grains by unifying all the genes expressed in the four grain samples (GR5, GR10, GR15, and GR20) (Supplementary Table 4, Fig. 2a). Further, 4659 SEGs were identified by comparing four types of tissue sample group as described in “Materials and methods” (Fig. 2a), and the results of statistical tests illustrated the significance of the intersections (Supplementary Fig. 4). DEGs were also identified by comparing two consecutive time points of grain development stages, and the results indicated that large numbers of genes (26,500) exhibited differential expression throughout the grain development stages tested and the majority of these variations take place during the early stage of grain development (between 5 and 10 DPA) (Fig. 2b), this is corresponding to the early developmental and metabolic events occurring in the grain of wheat (Shewry et al. 2012) as well as being consistent with the results obtained by using cDNA arrays (Laudencia-Chingcuanco et al. 2007). However, the numbers of the DEGs between two consecutive time points in the present study were much greater than those in previous reports, most likely owing to the improved transcriptome reference and sensitivity of the RNA-Seq

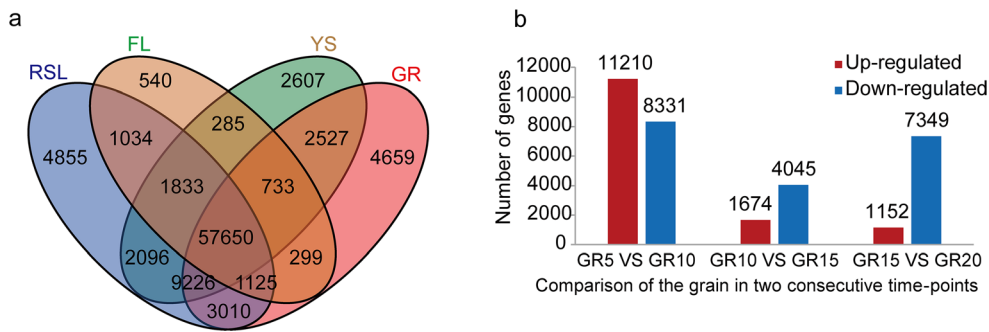


Fig. 2 Distribution of the expressed genes in different types of wheat tissues/organs and differentially expressed genes in developmental grains. **a** Venn diagram showing the distribution of the expressed genes in four types of sample group, including RSL, FL, YS, and GR. RSL, vegetative tissues from seedling including root, stem, and leaf of wheat plant at five-leaf stage; FL, flag leaf of wheat at heading stage; YS, spike

from wheat plant at booting stage and heading stage; GR, grains at 5, 10, 15, and 20 days post-anthesis. **b** Histogram of differentially expressed genes in developmental grains by comparing two consecutive time points. The criteria for differential expression: false discovery rate ($FDR \leq 0.001$) and fold changes (absolute value of $\log_2\text{Ratio} \geq 1$)

methodology. GO and KEGG annotations as well as the pathway mapping of all the SEGs and the DEGs were summarized in Table 2, Supplementary Table 7, Supplementary Table 8, Supplementary Fig. 5, and Supplementary Fig. 6. These data indicated that GO terms were over-represented for the SEGs encoding the components of regulation and signaling-related categories, and GO categories were over-represented for the DEGs encoding the components of cellular component organization or biogenesis and nutrient reservoir activity during grain development of Xiaoyan-6. In contrast, the GO terms

of the genes common in all the tissues tested were over-represented for those genes encoding the proteins involved in general cellular biological process (Supplementary Table 5). A recent study in wheat cv. Chinese Spring indicated that “hotspots” (with a median expression level > 20 tpm, containing on average 5 genes) across the six tissues, including leaf, root, seedling, stem, spike, and seed, tended to be enriched for genes encoding components of cellular component organization or biogenesis in all the tissues as well as nutrient reservoir activity in seed tissues (Clavijo et al. 2017).

Table 2 Summary of GO enrichment of the grain-specific genes (SEGs) and the differentially expression genes (DEGs) as well as their pathways during wheat grain development grouped based on metabolic activities

GO/KEGG analysis	DEGs	SEGs
GO enrichment		
Cellular component	318	10
Molecular function	504	68
Biological process	1285	122
Metabolic activity		
Amino acid metabolism	6	2
Biosynthesis of other secondary metabolites	1	3
Carbohydrate metabolism	3	3
Energy metabolism	2	0
Environmental adaptation	0	2
Folding, sorting and degradation	1	2
Glycan biosynthesis and metabolism	1	1
Lipid metabolism	4	1
Metabolism of cofactors and vitamins	1	0
Signal transduction	0	1
Replication and repair	2	0
Translation	3	0

The DEGs were grouped according to shared expression patterns during grain development (Supplementary Table 7). Expression trend analysis of these DEGs exhibited 18 expression profiles being classified into three groups: up-regulated (group 1), down-regulated (group 2), and modulated with grain development (group 3) (Fig. 3, Supplementary Fig. 7, and Supplementary Table 7), providing the context for understanding the complex metabolic pathways and the molecular control of the quality and nutrition properties of Xiaoyan-6. Among the 18 expression profiles, six were statistically significant ($P \leq 0.05$) expression trends (profiles 15, 0, 2, 8, 14, and 11), and each of these significant expression trends tended to be associated with those biological processes and molecular functions as shown in Supplementary Fig. 8. The profile 15 belongs to the up-regulated group, and the genes in this profile are increasingly expressed with grain development and remain at a stable level after 10 DPA (Fig. 3, Supplementary Fig. 7). The most enriched GO terms in this profile are related to regulation of biological or metabolic process (Supplementary Fig. 8). This is in line with the metabolic processes that are very active at the early developmental stages of wheat grain. Profiles 0, 2, and 8 belong to the down-regulated group, and the most enriched GO terms in these profiles are associated with photosynthesis and secondary metabolic processes (Supplementary Fig. 8). This is consistent with the cellular events taking place during growth and development of wheat grain.

For the list of enriched GO categories and various metabolic pathways that are grouped under various metabolic activities, please refer to Supplementary Table 7 and Supplementary Table 8, respectively

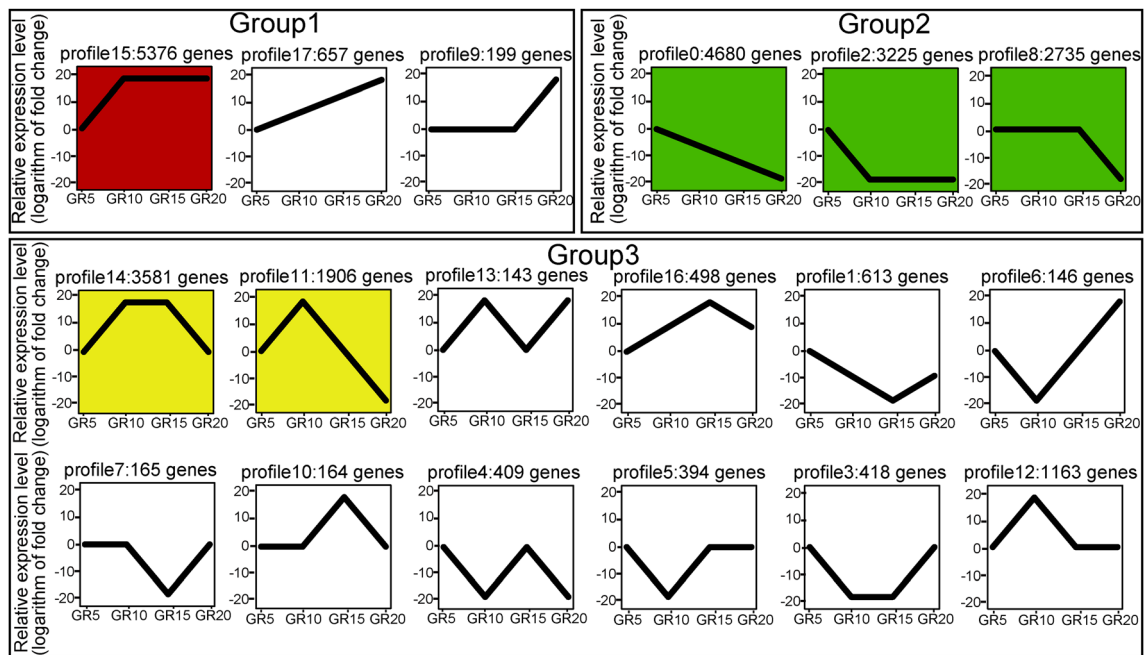


Fig. 3 Expression trends of differentially expressed genes during wheat grain development (DEGs). Each panel represents a model of an expression file. The vertical axes represent relative expression levels (the logarithm (\log_2) of the fold change), whereas the fold change is the ratio of the abundance of a gene in GR10, GR15, or GR20 to the abundance of the same gene in GR5) across the developmental stages of wheat grain. The horizontal axes indicate developmental grains: GR5, GR10, GR15, and GR20 representing grain at 5, 10, 15, and 20 days post-

anthesis, respectively. The DEGs are clustered to 18 expression profiles according to shared expression patterns, which are classified into three groups: groups 1, 2, and 3 corresponding to the trends of up-regulated, down-regulated, and modulated (up-regulation to down-regulation or vice versa) with grain development, respectively. The gene number containing in each expression profile is listed following the profile number. Six statistically significant expression profiles (with $P \leq 0.05$) are marked with colors

Taken together, the above findings in Xiaoyan-6 that the GO terms of the SEGs were over-represented in regulation and signaling-related biological processes and that the individual expression trends of the DEGs during grain development tended to be associated with certain biological processes provide new information to the existing research on wheat transcriptome and enrich our knowledge of functional genes involved in wheat grain development.

TFs involved in wheat grain development

TFs regulate target genes to ensure tightly regulated developmental process. A total of 3606 wheat TFs in 56 families have been collected in Plant Transcription Factor Database (PlantTFDB 4.0, <http://plantfdb.cbi.pku.edu.cn/index.php>) (Jin et al. 2017). In the present study, we detected 4735 TFs expressed in grain samples from all 56 families that were identified using the Transcription Factor Prediction tool in PlantTFDB through Hidden Markov Model-guided method (Jin et al. 2017). We also found 427 grain-specific TFs (called as SE-TFs) and 1635 TFs differentially expressed during grain development (named as DE-TFs) (Table 3 and Supplementary Table 9). Among the 1635 DE-TFs, 425 and 749 showed expression patterns of group 1 and group 2, respectively (Table 3 and Supplementary Table 9),

corresponding to the up-regulated and the down-regulated groups as shown in Fig. 3. By comparison, relatively few of TFs differentially expressed through wheat grain development were identified previously by cDNA microarrays technique (Laudencia-Chingcuanco et al. 2007; Wan et al. 2008), and the previously identified TFs were represented in our set of TFs as summarized in Supplementary Table 10. In rice, 1118 TFs from 55 families were also detected in developmental endosperm using RNA-Seq technique (Gao et al. 2013). A comparison of the TF families previously identified in rice development endosperm with our set of developmental grain TF families (Supplementary Table 10) detected similar TF families (56). One-third of the wheat grain TF genes had close homologs (70% identity, 70% coverage) in rice (Supplementary Table 10).

Gene co-expression network and key TFs involved in grain development

WGCNA is one of the most widely used approaches aimed at the systematic understanding of network instead of individual genes (Langfelder and Horvath 2008). To explore the gene co-expression regulation network and determine the putative key genes in the regulatory pathway, we applied WGCNA across all eight grain samples from the four developmental stages,

Table 3 Statistics of transcription factors (TFs) and their expression patterns

TF family ^{*1}	The number of TFs across tissues	The number of TFs in grain	SE-TFs ^{*2}	DE-TFs ^{*3}	DE-TFs in three expression pattern ^{*4}		
					Group 1	Group 2	Group 3
AP2	71	68	5	19	8	3	8
ARF	66	66	0	30	8	12	10
ARR-B	46	46	23	11	2	1	8
B3	280	237	40	60	24	27	9
BBR-BPC	6	6	0	1	1	0	0
BES1	20	18	1	9	2	6	1
bHLH	476	365	9	108	22	65	21
bZIP	266	243	10	85	42	21	22
C2H2	354	312	34	96	18	31	47
C3H	126	124	5	47	12	6	29
CAMTA	16	16	0	2	0	1	1
CO-like	44	37	1	15	0	9	6
CPP	35	32	2	9	5	0	4
DBB	26	26	0	13	3	5	5
Dof	95	91	3	36	12	19	5
E2F/DP	27	27	0	11	2	3	6
EIL	19	19	1	8	5	3	0
ERF	437	367	46	157	15	92	50
FAR1	137	128	40	19	17	0	2
G2-like	144	125	5	32	11	15	6
GATA	77	73	5	16	2	4	10
GeBP	31	31	0	10	0	2	8
GRAS	156	135	12	44	6	29	9
GRF	29	29	0	13	2	4	7
HB-other	52	50	23	23	19	2	2
HB-PHD	9	9	0	5	4	0	1
HD-ZIP	120	104	6	51	3	39	9
HRT-like	2	2	0	0	0	0	0
HSF	77	73	2	27	7	9	11
LBD	79	62	5	24	1	16	7
LFY	3	2	0	0	0	0	0
LSD	13	13	0	4	0	4	0
MIKC_MADS	106	101	5	34	0	27	7
M-type_MADS	95	76	27	2	0	2	0
MYB	410	352	19	136	37	63	36
MYB_related	216	174	32	70	42	8	20
NAC	429	342	36	125	36	63	26
NF-X1	7	7	0	3	2	0	1
NF-YA	19	19	0	10	9	0	1
NF-YB	53	47	11	25	11	1	13
NF-YC	39	38	2	11	3	6	2
Nin-like	32	30	1	6	3	0	3
RAV	26	23	0	5	0	4	1
S1Fa-like	3	3	0	3	0	0	3
SBP	56	41	0	11	3	7	1
SRS	15	14	0	5	1	3	1
STAT	3	3	0	1	1	0	0

Table 3 (continued)

TF family ^{*1}	The number of TFs across tissues	The number of TFs in grain	SE-TFs ^{*2}	DE-TFs ^{*3}	DE-TFs in three expression pattern ^{*4}		
					Group 1	Group 2	Group 3
TALE	65	62	0	17	0	9	8
TCP	59	48	0	18	0	16	2
Trihelix	91	90	1	40	8	22	10
VOZ	6	6	0	3	2	0	1
Whirly	6	6	0	2	0	0	2
WOX	43	32	9	9	4	1	4
WRKY	287	234	4	107	6	75	26
YABBY	20	19	0	10	0	9	1
ZF-HD	35	32	2	15	4	5	6
Total	5460	4735	427	1653	425	749	479

^{*1} TFs, transcription factors

^{*2} SE-TFs, TFs specially expressed in developmental grains

^{*3} DE-TFs, TFs differentially expressed during grain development

^{*4} Group 1, group 2, and group 3 are corresponding to the groups shown in Fig. 3, which represent up-regulated, down-regulated and modulated with grain development, respectively

and a hierarchical clustering tree with 15 distinct transcription modules was constructed after merging of similar modules (Supplementary Fig. 9). The correlation between modules and different grain developmental stages showed that module 3 and module 4 are positive correlated with GR5 and GR10; module 6 and module 11 are positive correlated with GR15 or G20; while module 5 and module 13 are negative correlated with GR15 or G20 (Supplementary Fig. 10). The connectivity values of individual genes and their expression pattern modules are presented in Supplementary Table 11.

As described above, 4735 of the total 5460 TFs identified across the tissues tested were expressed in developmental grains of wheat cv. Xiaoyan-6 (Table 3). We developed a grain co-expression regulation network (GrainNet) by using the TF genes with high connectivity and their putative regulatory genes with high edge weight in order to identify the putative critical TFs and functional genes involved in wheat grain development (Fig. 4 and Supplementary Table 12). In the GrainNet, ten TF genes (red nodes in Fig. 4) from different families, including *bZIP* (TraesCS2B01G489900), *C2H2* (TraesCS4A01G041400), *MYB* (TraesCS3B01G399300), *WRKY* (TraesCS2A01G489500), *HB-other* (TraesCS7A01G168000), *B3* (TraesCS4A01G055700), *NAC* (TraesCS5D01G148800), *LBD* (TraesCS2A01G271300), *bHLH* (TraesCS1D01G084200), and *GRF* (TraesCS4A01G255000), are considered to be the top ten highly connected TFs (hub TFs) in the GrainNet (Fig. 4 and Supplementary Table 11). The other nodes represent the predicted regulatory genes of these key TFs, the purple nodes indicating the genes predicted to be regulated by five or six of these TFs, the light blue ones indicating the genes predicted

to be regulated by three or four of these TFs, and the dark blue ones representing the genes predicted to be regulated by one or two of the TFs. In general, TFs and their predicted regulatory genes showed a many-to-many relationship, multiple TFs regulating any one predicted gene and individual TFs regulating multiple predicted genes.

To explore the major biological functions of the key TFs involved in grain development, we first analyzed expression profiles of the ten putative key TFs and the seven predicted target genes of these TFs (the purple nodes in Fig. 4) across all wheat tissues tested. The result exhibited that nine of the ten TFs and four of the seven functional genes most highly expressed during early grain development (Supplementary Fig. 11), suggesting that they play important roles in early grain development of Xiaoyan-6. Then, we focused on the molecular functions of the seven predicted target genes of these TFs. TraesCS3B01G524500 encodes expansin B7, a β -expansin that affects leaf elongation, internodal elongation, and grain development by regulating the extension of cell wall in rice (Lee and Kende 2001; Xu et al. 2016), and it is highly expressed during very early grain development of Xiaoyan-6 (Supplementary Fig. 11), these suggested that this gene might be play an important role in cell expansion in early developmental wheat grain. TraesCS4D01G357100 encodes nucleoside triphosphatase (NTPase), an important membrane enzyme participating in substance trans-membrane transport, elongation growth of cells as well as responses to environmental stress (Dietrich et al. 2001; Rober-Kleber et al. 2003; Serrano 1988). TraesCS4D01G357100 is most highly expressed in early grain development (Supplementary Fig. 11). TraesCS2A01G463200 encodes BCL2-associated

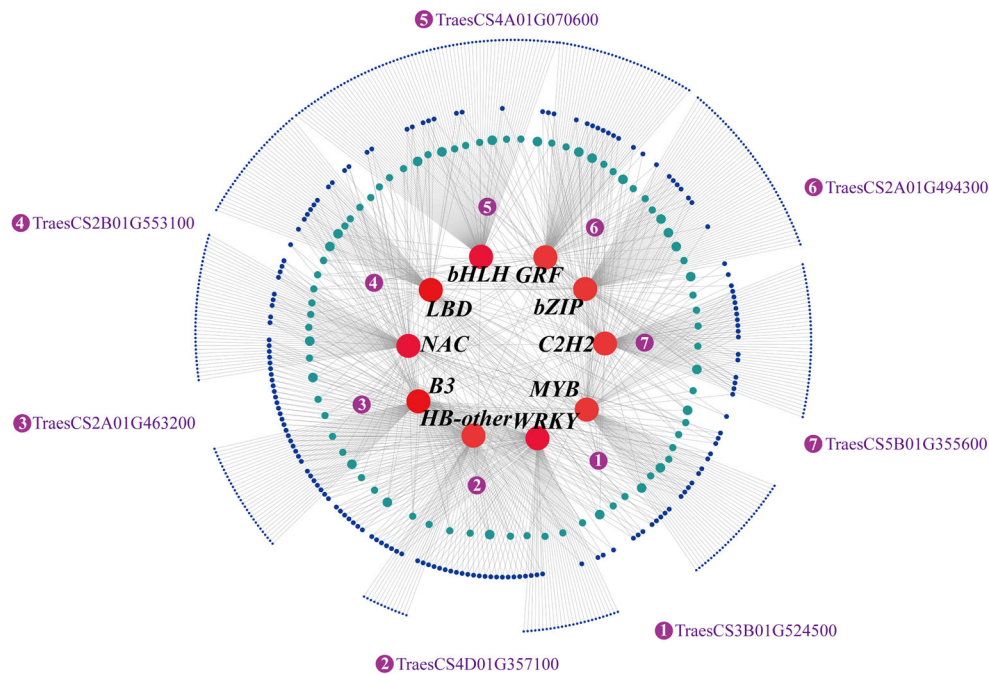


Fig. 4 Wheat grain gene co-expression network (GrainNet). The GrainNet was developed by using the transcription factors (TFs) genes with high module eigengene-based connectivity ($kME \geq 0.98$) and their putative regulatory genes with top 100 edge weight. Circular nodes represent genes in the net, the size of nodes represents the connectivity, and the edges represent interaction between TF genes and their predicted regulatory genes. The nodes with more edges indicate more importance in

the network. The group of the red nodes represent TF genes, the group of the purple nodes represent the genes predicted to be regulated by five or six of these TFs and the gene IDs are shown outside the circles, the group of the light blue ones represent the genes predicted to be regulated by three or four of these TFs, and the group of the dark blue nodes represent the genes predicted to be regulated by one or two of these TFs

athanogene (*BAG*) family molecular chaperone regulator 1-like that can interact with molecular chaperones directly and play a role in a variety of signaling pathways (Brive et al. 2001). *BAG1* in rice is involved in cell elongation and cell cycle regulation (Kabbage and Dickman 2008), indicating that TraesCS2A01G463200 could also be involved in signaling, cell division, expansion, differentiation, and elongation of endosperm cells during the early grain development stage and grain filling of wheat. TraesCS2B01G553100 encoding alpha-L-arabinofuranosidase/beta-D-xylosidase isoenzyme ARA-I was demonstrated to affect nutritional quality and processing quality of wheat grain by controlling the content of non-starch polysaccharide in grain cell wall (Cleemput et al. 1997). TraesCS4A01G070600 encoding UDP-glycosyltransferase 83A1 (*UGT83A1*) can catalyze the glycosylation and play an important role in the modification of secondary metabolites (Zhang et al. 2014). TraesCS2A01G494300 can produce small auxin up RNA 19 (*SAUR19*). *SAUR* is a plant-specific protein family and also the largest family in auxin-responsive factors. *AtSAUR19* in Arabidopsis is capable to positively modulate cell expansion by regulating auxin synthesis and transport (Spartz et al. 2012). While, TraesCS5B01G355600 encoding pentatricopeptide repeat-containing protein (PPR-like) in mitochondrial is expressed at a low level during grain development (Supplementary Fig. 11). The homolog of this PPR-like

gene in Arabidopsis is At5g41170, and its loss-of-function mutant plants result in some abnormal phenotypes, such as development retardation, flowering delay, plant smaller, leaf curl, mature seed volume reduction, seed coat shrinkage, embryo or endosperm deficiency (Liu et al. 2013; Manavski et al. 2012). These infer that TraesCS5B01G355600 might play an important role during wheat plant development. Moreover, we also performed GO enrichment analysis on the genes predicted to be regulated by two or more of the putative key TFs, and the result showed that the GO terms of signal transduction, amino acid transport and metabolic process, cell growth, and hormone transport were over-represented (Supplementary Table 13). Taken together, these predicted functional genes now define network regulated by multiple putative key TFs associated with signaling, cell cycle regulation, cell elongation and expansion, metabolite trans-membrane transport, and the components of the cell wall, this implying their crucial roles in early grain development of wheat cv. Xiaoyan-6.

A previous study in wheat cv. Chinese Spring developed a co-expression network based on homeologous gene expression in different cell types (starchy endosperm, aleurone layer, transfer cells) at different developmental stages of grain (10, 20, 30 DPA) and revealed that the co-expression network consisted of 25 modules displaying distinct co-expression clusters associated with the spatiotemporal progression during endosperm development (Pfeifer et al. 2014). In our study, the

co-expression network consisting of TFs and their predicted regulatory genes was developed based on gene expression at different development stages of wheat grain (5, 10, 15, and 20 DPA), using the latest released fully annotated reference genome of bread wheat (Appels et al. 2018) as a reference. The GrainNet in the present study focused on the TFs and their predicted regulatory genes involved in developmental grains of wheat and allowed identification of the putative critical TFs and functional genes, compared to the previous work. Our GrainNet provides new insights into gene co-expression regulation network during grain development of wheat. Our findings establish new target genes for further study of the functional genes related to grain development and yield and for modifying genes related to grain development and yield, to fine-tune expression in different varieties.

Conclusion

In this study, we investigated the transcriptome of Chinese winter wheat cv. Xiaoyan-6 grains at four developmental stages (5, 10, 15, and 20 DPA) using RNA-Seq, with the IWGSC RefSeq ver1.0 as a reference, to identify 427 SE-TFs and 1653 DE-TFs as well as a GrainNet consisting of the TFs and their predicted regulatory genes that are first being developed in wheat. Previously, in wheat cv. Chinese Spring, a co-expression network consisting of 25 modules that displayed distinct co-expression clusters associated with the spatiotemporal progression during endosperm development was developed based on homeologous gene expressions in different cell types (starchy endosperm, aleurone layer, transfer cells) at three different developmental stages of grain (10, 20, and 30 DPA), using the IWGSC CSS v2 as a reference. In the present study, the GrainNet based on co-expression network focused on the TFs and their predicted regulatory genes involved in developmental grains of wheat and allowed to identify ten putative key TFs and the predicted regulatory genes of these TFs in developmental grain of Xiaoyan-6, compared to the work in Chinese Spring. The analysis was given a firm basis through the study of additional wheat tissues, including root, stem, leaf, flag leaf, and spike at two developmental stages to generate a dataset of 92,478 high-confidence protein-coding genes that were mostly evenly distributed among subgenomes but unevenly distributed across each of the seven homoeologous chromosome groups. Moreover, the expression levels of the genes across chromosomes are similar across developmental grains of Xiaoyan-6, with the exception of homoeologous chromosome group I (1A/1B/1D), II (2A/2B), and VI (6A/6B/6D) that had at least one hotspot (with a median expression level across developmental grains more than 30 FPKM, containing at least five genes). Within this larger framework, the transcriptomes identified 4659 SEGs and 26,500 DEGs throughout grain development stages

tested. The SEGs were mainly associated with regulation and signaling-related biological processes, and the DEGs were mainly involved in cellular component organization or biogenesis and nutrient reservoir activity during grain development of Xiaoyan-6. Gene expression trend analysis of the DEGs revealed six statistically significant expression profiles during grain development of wheat. The study establishes new targets for further study of functional genes related to grain development and yield and for modifying genes related to grain development and yield, to fine-tune expression in different varieties. All the raw data in our study has been deposited in NCBI's Sequence Read Archive under BioProject number PRJNA525250.

Acknowledgments We would like to thank Gene Denovo (Guangzhou, China) for next generation sequencing and reads alignment to the CSS (International Wheat Genome Sequencing Consortium 2014) transcriptome reference. We also would like to give our thanks to all other members in our lab for their help and discussion on the draft of this manuscript.

Authors and contributors HZ and QC conceived and designed the experiments. QC, LG, MM, LZ, XL, and BW conducted the experiments and analyzed the data. RR and CU conducted alignment of RNA-seq data with the IWGSC RefSeq ver1.0 gene models and revised the draft. QC and HZ wrote draft and revised the draft. RA made modification on the manuscript.

Funding This study was financially supported by the Natural Science Foundation of China (31471482), the major project of collaborative innovation in production, learning and research of Yangling Demonstration Zone (2017CX-01), the Natural Science Foundation of Shaanxi province (2018JM3025), and the Scientific Research Foundation of State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A & F University (CSBAAZD2016001).

Compliance with ethical standards

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

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Appels R et al (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. <https://doi.org/10.1126/science.aar7191>
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

- Brinton J, Simmonds J, Minter F, Leverington-Waite M, Snape J, Uauy C (2017) Increased pericarp cell length underlies a major quantitative trait locus for grain weight in hexaploid wheat. *New Phytol* 215:1026–1038. <https://doi.org/10.1111/nph.14624>
- Brive L, Takayama S, Briknarova K, Homma S, Ishida SK, Reed JC, Ely KR (2001) The carboxyl-terminal lobe of Hsc70 ATPase domain is sufficient for binding to BAG1. *Biochem Biophys Res Commun* 289:1099–1105. <https://doi.org/10.1006/bbrc.2001.6087>
- Calderini DF, Abeledo LG, Savin R, Slafer GA (1999) Effect of temperature and carpel size during pre-anthesis on potential grain weight in wheat. *J Agric Sci* 132:453–459. <https://doi.org/10.1017/S0021859699006504>
- Chen J, Zeng B, Zhang M, Xie S, Wang G, Hauck A, Lai J (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol* 166:252–264. <https://doi.org/10.1104/pp.114.240689>
- Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H, Lipscombe J, Barker T, Lu FH, McKenzie N, Raats D, Ramirez-Gonzalez RH, Counce A, Peel N, Percival-Alwyn L, Duncan O, Trösch J, Yu G, Bolser DM, Namaati G, Kerhornou A, Spannagl M, Gundlach H, Haberer G, Davey RP, Fosker C, Palma FD, Phillips AL, Millar AH, Kersey PJ, Uauy C, Krasileva KV, Swarbreck D, Bevan MW, Clark MD (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res* 27:885–896. <https://doi.org/10.1101/gr.217117.116>
- Cleemput G, Hessing M, Van Oort M, Deconynck M, Delcour JA (1997) Purification and characterization of a [beta]-D-xylosidase and an endo-xylanase from wheat flour. *Plant Physiol* 113:377–386. <https://doi.org/10.1104/pp.113.2.377>
- Dietrich P, Sanders D, Hedrich R (2001) The role of ion channels in light-dependent stomatal opening. *J Exp Bot* 52:1959–1967. <https://doi.org/10.1093/jxbbot/52.363.1959>
- Ernst J, Bar-Joseph Z (2006) STEM: a tool for the analysis of short time series gene expression data. *BMC bioinformatics* 7:191. <https://doi.org/10.1186/1471-2105-7-191>
- Gao Y, Xu H, Shen Y, Wang J (2013) Transcriptomic analysis of rice (*Oryza sativa*) endosperm using the RNA-Seq technique. *Plant Mol Biol* 81:363–378. <https://doi.org/10.1007/s11103-013-0009-4>
- International Wheat Genome Sequencing Consortium I (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788. <https://doi.org/10.1126/science.1251788>
- Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* 45:D1040–D1045. <https://doi.org/10.1093/nar/gkw982>
- Kabbage M, Dickman MB (2008) The BAG proteins: a ubiquitous family of chaperone regulators. *Cellular and molecular life sciences* : CMLS 65:1390–1402. <https://doi.org/10.1007/s00018-008-7535-2>
- Kim D, Perteau G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>
- Kloepper JW, Leong J, Teintze M, Schroth MN (1980) Enhanced plant growth by siderophores produced by plant growth-promoting rhizobacteria. *Nature* 286:885–886. <https://doi.org/10.1038/286885a0>
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 9:559. <https://doi.org/10.1186/1471-2105-9-559>
- Laudencia-Chinguanco 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. <https://doi.org/10.1007/s11103-006-9114-y>
- Lee Y, Kende H (2001) Expression of beta-expansins is correlated with internodal elongation in deepwater rice. *Plant Physiol* 127:645–654. <https://doi.org/10.1104/pp.127.2.645>
- Li WL, Li ZS, Mu SM (1990) A cytological study of chromosomal structure changes in a common wheat variety Xiaoyan No. 6. *Acta Genet Sin* 17:430–437
- Liu YJ, Xiu ZH, Meeley R, Tan BC (2013) Empty pericarp5 encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. *Plant Cell* 25:868–883. <https://doi.org/10.1105/tpc.112.106781>
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)). *Method Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>
- Manavski N, Guyon V, Meurer J, Wienand U, Bretschneider R (2012) An essential pentatricopeptide repeat protein facilitates 5' maturation and translation initiation of rps3 mRNA in maize mitochondria. *Plant Cell* 24:3087–3105. <https://doi.org/10.1105/tpc.112.099051>
- Nadaud I, Girousse C, Debiton C, Chambon C, Bouzidi MF, Martre P, Branlard G (2010) Proteomic and morphological analysis of early stages of wheat grain development. *Proteomics* 10:2901–2910. <https://doi.org/10.1002/pmic.200900792>
- Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, International Wheat Genome Sequencing Consortium, Mayer KFX, Olsen OA (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345:1250091. <https://doi.org/10.1126/science.1250091>
- Ramirez-Gonzalez RH et al (2018) The transcriptional landscape of polyploid wheat. *Science* 361:ear6089. <https://doi.org/10.1126/science.ear6089>
- Rober-Kleber N, Albrechtová JT, Fleig S, Huck N, Michalke W, Wagner E, Speth V, Neuhaus G, Fischer-Iglesias C (2003) Plasma membrane H⁺-ATPase is involved in auxin-mediated cell elongation during wheat embryo development. *Plant Physiol* 131:1302–1312. <https://doi.org/10.1104/pp.013466>
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rodriguez-Navarro DN, Dardanelli MS, Ruiz-Sainz JE (2007) Attachment of bacteria to the roots of higher plants. *FEMS Microbiol Lett* 272:127–136. <https://doi.org/10.1111/j.1574-6968.2007.00761.x>
- Sabelli PA, Larkins BA (2009) The development of endosperm in grasses. *Plant Physiol* 149:14–26. <https://doi.org/10.1104/pp.108.129437>
- Serrano R (1988) Structure and function of proton translocating ATPase in plasma membranes of plants and fungi. *Biochim Biophys Acta* 947:1–28. [https://doi.org/10.1016/0304-4157\(88\)90017-2](https://doi.org/10.1016/0304-4157(88)90017-2)
- 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. <https://doi.org/10.1016/j.jcs.2011.11.007>
- Spartz AK, Lee SH, Wenger JP, Gonzalez N, Itoh H, Inzé D, Peer WA, Murphy AS, Overvoorde PJ, Gray WM (2012) The SAUR19 subfamily of SMALL AUXIN UP RNA genes promote cell expansion. *Plant J* 70:978–990. <https://doi.org/10.1111/j.1365-313X.2012.04946.x>
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515. <https://doi.org/10.1038/nbt.1621>

- Wan Y, Poole RL, Huttly AK, Toscano-Underwood C, Feeney K, Welham S, Gooding MJ, Mills C, Edwards KJ, Shewry PR, Mitchell RAC (2008) Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics* 9:121. <https://doi.org/10.1186/1471-2164-9-121>
- Wang M, Zhao Y, Zhang B (2015) Efficient test and visualization of multi-set intersections. *Sci Rep* 5:16923. <https://doi.org/10.1038/srep16923>
- Wright GW, Simon RM (2003) A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* 19:2448–2455. <https://doi.org/10.1093/bioinformatics/btg345>
- Xiao Y, Thatcher S, Wang M, Wang T, Beatty M, Zastrow-Hayes G, Li L, Li J, Li B, Yang X (2016) Transcriptome analysis of near-isogenic lines provides molecular insights into starch biosynthesis in maize kernel. *J Integr Plant Biol* 58:713–723. <https://doi.org/10.1111/jipb.12455>
- Xu Q, Krishnan S, Merewitz E, Xu J, Huang B (2016) Gibberellin-regulation and genetic variations in leaf elongation for tall fescue in association with differential gene expression controlling cell expansion. *Sci Rep* 6:30258. <https://doi.org/10.1038/srep30258>
- Xue LJ, Zhang JJ, Xue HW (2012) Genome-wide analysis of the complex transcriptional networks of rice developing seeds. *PLoS One* 7:e31081. <https://doi.org/10.1371/journal.pone.0031081>
- Yu Y, Zhu D, Ma C, Cao H, Wang Y, Xu Y, Zhang W, Yan Y (2016) Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development. *The Crop Journal* 4:92–106. <https://doi.org/10.1016/j.cj.2016.01.006>
- Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereal. *Weed Res* 14:415–421. <https://doi.org/10.1111/j.1365-3180.1974.tb01084.x>
- Zhang F, Li X, Li Z, Xu X, Peng B, Qin X, Du G (2014) UPLC/Q-TOF MS-based metabolomics and qRT-PCR in enzyme gene screening with key role in triterpenoid saponin biosynthesis of *Polygala tenuifolia*. *PLoS One* 9:e105765. <https://doi.org/10.1371/journal.pone.0105765>
- Zhang J, Wu K, Zeng S, Teixeira da Silva JA, Zhao X, Tian CE, Xia H, Duan J (2013) Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated with floral development. *BMC Genomics* 14:279. <https://doi.org/10.1186/1471-2164-14-279>
- Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL (2017) The first near-complete assembly of the hexaploid bread wheat genome. *Triticum aestivum* *GigaScience* 6:1–7. <https://doi.org/10.1093/gigascience/gix097>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.