



OPEN

## RNA-Seq transcriptome profiling of immature grain wheat is a technique for understanding comparative modeling of baking quality

Hossein Ahmadi-Ochtapeh<sup>1</sup>, Hassan Soltanloo<sup>2✉</sup>, Seyyede Sanaz Ramezanzpour<sup>2</sup>, Ahad Yamchi<sup>2</sup> & Vahid Shariati<sup>3</sup>

Improving the baking quality is a primary challenge in the wheat flour production value chain, as baking quality represents a crucial factor in determining its overall value. In the present study, we conducted a comparative RNA-Seq analysis on the high baking quality mutant "O-64.1.10" genotype and its low baking quality wild type "Omid" cultivar to recognize potential genes associated with bread quality. The cDNA libraries were constructed from immature grains that were 15 days post-anthesis, with an average of 16.24 and 18.97 million paired-end short-read sequences in the mutant and wild-type, respectively. A total number of 733 transcripts with differential expression were identified, 585 genes up-regulated and 188 genes down-regulated in the "O-64.1.10" genotype compared to the "Omid". In addition, the families of HSF, bZIP, C2C2-Dof, B3-ARF, BES1, C3H, GRF, HB-HD-ZIP, PLATZ, MADS-MIKC, GARP-G2-like, NAC, OFP and TUB were appeared as the key transcription factors with specific expression in the "O-64.1.10" genotype. At the same time, pathways related to baking quality were identified through Kyoto Encyclopedia of Genes and Genomes. Collectively, we found that the endoplasmic network, metabolic pathways, secondary metabolite biosynthesis, hormone signaling pathway, B group vitamins, protein pathways, pathways associated with carbohydrate and fat metabolism, as well as the biosynthesis and metabolism of various amino acids, have a great deal of potential to play a significant role in the baking quality. Ultimately, the RNA-seq results were confirmed using quantitative Reverse Transcription PCR for some hub genes such as alpha-gliadin, low molecular weight glutenin subunit and terpene synthase (gibberellin) and as a resource for future study, 127 EST-SSR primers were generated using RNA-seq data.

**Keywords** RNA-Seq, qRT-PCR, Transcriptome, Gene ontology, Glutenin subunit

Wheat (*Triticum aestivum* L.,  $2n = 6x = 42$ ) is an important food crop in the world and possesses unique flour quality that can be used to make various food products<sup>1</sup>. However, current wheat varieties need improvement in processing quality to meet the increasing demand for better quality food products<sup>2</sup>. Processing quality of wheat flour is determined by grain protein concentration and its composition which confer wheat dough with unique rheological properties, making it possible to produce a series of quality foods for human consumption<sup>3-5</sup>. Baking quality is one of the most important parameters throughout the value chain of wheat flour<sup>6</sup>. Seed is the primary storage organ in plants for storing nutrients such as starch, lipids, and proteins<sup>7</sup> and the baking quality is largely influenced by these nutrients. The content of gliadins and glutenins<sup>6</sup>, optimal water absorption, rheological parameters<sup>8</sup> and grain hardness<sup>9</sup> are correlated to high baking quality. Determination of flour water absorption, as well as the results of the farinographic analysis of the flour play a key role in the assessment of wheat flour baking quality<sup>10</sup>. However, improving baking quality is a challenge for wheat breeders due to the

<sup>1</sup>Crop and Horticultural Science Research Department, Agricultural Research, Education and Extension Organization (AREEO), Golestan Agricultural and Natural Resources Research and Education Center, Gorgan, Iran. <sup>2</sup>Plant Breeding and Biotechnology Department, Gorgan University of Agricultural Sciences and Natural Resources (GUASNR), Gorgan, Iran. <sup>3</sup>Department of Plant Molecular Biotechnology, Assistant Professor in National Institute of Genetic Engineering and Biotechnology, Karaj, Iran. ✉email: soltanlooh@gau.ac.ir

time-consuming and costly testing forcing breeders to postpone sophisticated quality tests to the very last phases of variety development<sup>11</sup>. The genome of a plant is the most critical factor to control baking quality trait in wheat<sup>12</sup>. For example,<sup>2</sup> processing quality related key genes such as glutenin and gliadins, puroindolines, grain softness protein, alpha and beta amylases, proteases, were identified, and many other candidate genes related to cellular and molecular functions. Therefore, an understanding of the molecular basis of this trait would be a major advantage in wheat breeding<sup>13</sup>. Advances in molecular techniques offer new opportunities to identify the genetic and molecular basis of bread quality in wheat<sup>1,2,14</sup>.

RNA-Seq analysis of various wheat genotypes has opened up new avenues for research on grain development and its composition, which ultimately determines the nutritional and functional quality of wheat<sup>15,16</sup>. The gene expression during wheat grain development plays a crucial role in determining the yield and nutritional properties of the crop. Identifying genes expressed during the grain filling stage can be useful for improving yield and grain quality<sup>17,18</sup>. Several studies have applied transcriptomics approaches to investigate the gene expression during grain development stages in wheat<sup>2,15–17,19–21</sup>. The grain development process of wheat can be categorized into three stages, including cell division and expansion (0 ~ 14 DPA (days post-anthesis)), effective grain filling (14 ~ 28 DPA), and maturation and desiccation (28 DPA to maturity)<sup>22</sup>. Understanding these stages and the associated gene expression can lead to the development of new and improved wheat varieties with superior nutritional and functional properties<sup>17,18</sup>. Early grain development, specifically the first 14 days after pollination, is critical for the final yield and quality of wheat. During this time, rapid grain expansion occurs through continuous cell division and endosperm filling with starch and gluten<sup>16,23</sup>. At 30 DPA, a diverse range of genes are expressed at low levels with a predominance of genes associated with seed defense and stress tolerance<sup>13</sup>.

Wheat flour quality is highly influenced by protein content and composition, with gluteins being the most abundant storage proteins, comprising about 80% of total grain proteins. Glutenins consist of high and low-molecular-weight glutenin subunits (HMW-GS and LMW-GS). Gene expression from these protein families occurs early in grain development<sup>16,23</sup>. Zamani et al.<sup>24</sup> and Izadi-Darbandi et al.<sup>25</sup> reported that mutant, “O-64.1.10” and its parent, “Omid” wheat cultivar have two HMW-GS genes, Dx2 + Dy12 and Bx7 + By8 but the improvement of baking quality in the mutant, “O-64.1.10” was unknown. A negative effect of Dx2 + Dy12 was published previously by several studies<sup>26–31</sup>. The allelic pair Bx7 + By8\* (Glu-B1a1) strongly associated with dough strength<sup>32</sup>. The complex processing quality of wheat is controlled by many genes, which have not been completely explored<sup>2</sup>. For example, phytohormones are the predominant biochemical basis for grain morphogenesis<sup>33</sup> and reported to be important signals in controlling seed development, maturation, and nutrients accumulation<sup>34</sup>. Gene ontology (GO) enrichment and pathway enrichment analysis showed that many transcription products and transcription factors (TFs) associated with carbohydrate and protein metabolism were abundantly expressed in the grain<sup>35</sup>. TFs regulate target genes to ensure tightly regulated developmental processes<sup>15</sup>. bZIP TFs are regulators of important plant processes such as seed storage protein gene regulation<sup>36</sup>, energy metabolism<sup>37</sup>, unfolded protein response<sup>36</sup>, and hormone and sugar signalling<sup>38</sup>. NAC TF play a significant role in regulating the accumulation of glutenin and starch in wheat endosperm, which are crucial for the grains quality<sup>39</sup>.

Rahemi et al.<sup>40</sup> reported that water absorption percentage, valorimeter value, farinograph quality number, zeleny number, hardness, wet gluten, protein content in high baking quality mutant “O-64.1.10” exhibited greater than “Omid”. We compared RNA-Seq data from immature grains of a “O-64.1.10” wheat genotype and its wild type cultivar, “Omid,” which have significantly different baking qualities. The purpose of our study was to identify genes that are primarily expressed in wheat grain and are related to baking quality. Through this study, we gained new insights into the molecular mechanisms that underlie seed quality and identified potential candidate genes related to wheat grain quality. Our findings provide a solid foundation for future research aimed at improving bread-making quality.

## Material and methods

At all stages, the research complied with relevant institutional, national, and international guidelines and legislation.

### Plant materials and sampling

To investigate the transcriptome of immature seeds of *Triticum aestivum* L, mutant “O-64.1.10” genotype and its low baking quality wild type “Omid” cultivar were obtained from the GenBank of Nuclear Science and Technology Research Institute in Karaj, Iran. We used two genotypes of Iranian bread wheat, namely “O-64.1.10” (with high baking quality) and its parent cultivar “Omid” (with poor baking quality), to conduct a comparative analysis of their baking quality (Table 1). Mutant genotype “O-64.1.10” was produced by the gamma irradiation approach and evaluated for baking quality through rheological and proteomics trials<sup>25,40</sup>. The “Omid” and its mutant “O-64-1-10”, were cultivated at the research field of Gorgan University of Agricultural and Natural Resources Gorgan, Iran. The main stem heads were marked with the anthesis date. The grains from the middle

Genotype	Protein content	Wet gluten	Hardness	Zeleny number	Valorimeter value	Water absorption percentage	Farinograph quality number	Bread volume
Omid	11.5 <sup>a</sup>	20.5 <sup>b</sup>	51 <sup>b</sup>	31 <sup>b</sup>	41 <sup>b</sup>	57.65 <sup>b</sup>	41.25 <sup>a</sup>	568. 5 <sup>a</sup>
O-64.1.10	13 <sup>a</sup>	36 <sup>a</sup>	54.5 <sup>a</sup>	36.5 <sup>a</sup>	54 <sup>a</sup>	66.75 <sup>a</sup>	75 <sup>a</sup>	505. 5 <sup>b</sup>

**Table 1.** Comparison the baking quality traits in mutant genotype “O-64.1.10” and its wild type “Omid” cultivar<sup>12</sup>. <sup>a, b</sup>Mean values marked with different letters are significantly different ( $p \leq 0.05$ ) by LSD test.

ear were collected at 5, 10, 15, 20, and 30 DPA between 9:00 and 10:00 am. For dynamic comparison of developing grains three biological replications of samples were used for the five stages. Also for RNA-seq samples, the seeds of “Omid” and “O-64.1.10” were used at 15 DPA. For any genotypes the seeds of 9 spikes (10 grains per spike) harvested from 9 different plants (three plants per three replications) and pooled. The samples were promptly frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  for subsequent total RNA extraction.

### RNA isolation and cDNA library construction

We utilized 100 mg seed samples at 15 DPA and p-Biozol Buffer (BioFlux, Japan) to extract total RNA, following the manufacturer’s recommendations. Additionally, 1% (w/v) RNase-free agarose gel electrophoresis and a Nanophotometer (Implen, Germany) were used to evaluate the amount and quality of RNA samples. The RNA samples of 15 DPA, passing the quality and quantity control, were sent to Beijing Genomics Institute (BGI), Hong Kong, China for cDNA library construction and sequencing on the Illumina sequencing platform (Illumina HiSeq™ 2500). A comprehensive quantitative evaluation of each RNA sample was performed using a Nanodrop 8000 Spectrophotometer (ThermoScientific, USA) and an Agilent 2100 Bioanalyzer System (Agilent Technologies, USA) to produce information on RNA concentration. The high-quality RNA (OD 260/280 = 2.06–2.08; OD 260/230 = 1.9–2.09; RIN value 7.5) was further processed for cDNA library creation using the Illumina TrueSeq RNA Sample Prep kit. To produce 150 bp paired-end reads, NGS sequencing of cDNA from seed tissue was performed.

### Data processing and analysis

The quality of paired-end reads from raw sequencing data was assessed both before and after trimming. The sequencing quality of the raw reads from each sequenced sample was assessed using FASTQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) accessed on January 8th, 2019) and CLC Genomic Workbench 7.5.1 (CLC Bio-Qiagen, Denmark). Trimmomatic software (version 0.36) was used to remove adapter sequences, low-quality nucleotides/sequences, and reads shorter than 150 bp. The remaining high-quality reads with a Phred score of  $\geq 30$  was used for downstream analyses. These reads were aligned against the wheat reference genome (version IWGSC RefSeq v1.0, [http://plants.ensembl.org/Triticum\\_aestivum/](http://plants.ensembl.org/Triticum_aestivum/)) using Hisat2 software (version 2.2.1.0)<sup>41</sup>. The mapped reads from each sample were assembled using Cufflinks v2.0.2 and htseq. These Cufflink assemblies were merged using Cuffmerge. The outputs were then used for differential expression analysis by Cuffdiff and EdgeR packages<sup>42</sup>. The normalization of gene expression values of two samples was estimated as fragments per kilobase of transcript per million fragments mapped (FKPM) and fold change (FC). Data analysis with an adjusted p-value threshold of 0.001 and Log2 Fold Change (log2FC) of  $\geq 2$  were assigned as differentially expressed genes (DEGs). GO enrichment analysis for DEGs was performed using goseq and AgriGO version 2.0. The Benjamini and Hochberg approach for controlling the false discovery rate (FDR) was used to adjust P-values. The threshold of 0.01 was set for the FDR value. The GO analysis was carried out to functionally categorize the DEGs from 15 DPA into three major aspects: "cellular component (CC)," "molecular function (MF)," and "biological process (BP)". Pathway enrichment analysis of DEGs was performed using the KEGG (Kyoto Encyclopedia of genes and genomes) database (<http://www.genome.ad.jp/kegg/>). To identify TFs, transcriptional regulators (TRs), and protein kinases (PKs) encoding genes, The DEGs were screened using the Plant Transcription Factor & Protein Kinase Identifier and Classifier database (iTAK v1.6). Simple Sequence Repeats (SSRs) were identified from RNA-seq data using the Perl script of MISA (MICroSatellite identification tool)<sup>43</sup> with default parameters<sup>44</sup>.

### Validation of RNA-seq data by real-time PCR

The RNA-Seq results were validated through the use of quantitative Reverse Transcription PCR (qRT-PCR). The qRT-PCR was performed on three genes that were chosen from the DEG analysis. The primers were designed based on the 3-UTR region of the sequence (Table 2) using the Primer3 online software ([http://www.embnet.sk/cgi-bin/primer3\\_www.cgi](http://www.embnet.sk/cgi-bin/primer3_www.cgi))<sup>45</sup>.

Before first-strand cDNA synthesis, possible genomic DNA contamination was removed by RNase-free DNaseI (Thermo Scientific, USA) at a ratio of 1 U DNaseI to 2  $\mu\text{g}$  RNA, 1X DNaseI buffer, 10 U Ribolock RNase inhibitor (Thermo Scientific, USA), and DEPC water up to 9  $\mu\text{l}$ , followed by incubation at  $37^{\circ}\text{C}$  for 30 min. To finalize DNaseI activity, 25 mM EDTA buffer was added to the reaction and the RNA was then heat-denatured

Primer name	Gene ID	Primer sequence	Tm	GC%	Product size (bp)
LMW-GS	TraesCS1B01G011600.1	Forward 5'-ACAACAGGTTTCAGGGTTCCA-3'	59.08	50	158
		Reverse 5'-CTATCTGGTGTGGCTGCAAA-3'	58.17	50	
Alpha-gliadin	TraesCSU01G160200.1	Forward 5'-ATGTTGTCAGCAGTTGTGGC-3'	59.33	50	154
		Reverse 5'-TTACTGAGGCTGCTGGTAGG-3'	58.80	55	
Gibberellin	TraesCSU01G099900.1	Forward 5'-TGGACGAGAGAATTGAGGCA-3'	58.73	50	152
		Reverse 5'-ATCGAACACATGGGGAGGG-3'	59.08	57.9	
GAPDH	AK359500.1	Forward 5'-GTTGGCAAGGTGCTCCCAGA-3'	62.70	60	121
		Reverse 5'-GCTCATAGGTGCTGGCTTG-3'	61.10	60	

**Table 2.** List of primers used in the experiment for quantitative Reverse Transcription PCR amplification.

at 65 °C for 10 min. First-strand cDNA synthesis was carried out using 1 µg of total RNA in the reaction mixture containing 0.5 µg Oligo (dT) primer and DEPC water (nuclease-free) up to 11 µl. To eliminate secondary structures, the liquid was gently mixed, quickly centrifuged, then incubated at 70 °C for 5 min. Then, 4 µl 5X cDNA reaction buffer, 10 mM dNTP mix, and 20 U Ribolock RNase inhibitor were added to the reaction mixture. The final volume was adjusted to 19 µl with DEPC-treated water, and the mixture was incubated for 5 min at 37 °C. 200 U Revert Aid enzyme (Thermo Scientific, USA) was added to the reaction mixture and incubated at 42 °C for 1 h. A qRT-PCR was performed using an iCycler thermal cycler (Bio-Rad, iQ5, USA) with a reaction volume containing 3 µl of diluted cDNA, 10 µl of 2X SYBR Bio Pars PCR Master Mix (Gorgan University of Agricultural Sciences and Natural Resources, Iran), and 1 µl of each gene-specific primer (10 pmol) in a final volume 20 µl with double distilled water. The time course of qRT-PCR were done at the following conditions: 3 min at 95 °C for 1 cycle; 10 s at 95 °C, 10 s at 62 °C and 10 s at 72 °C for 35 cycles, and 2 min at 72 °C for 1 cycle. Upon that, PCR was done on each sample in three technical and biological replications. The fluorescence signal was detected at 72 °C. The REST software was used for the gene expression analysis<sup>46</sup>, and the relative expression was computed utilizing the comparative Ct ( $2^{-\Delta\Delta C_t}$ ) method<sup>47</sup> and compared with the expression levels of RNA-Seq. For qRT-PCR normalization, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was utilized as the reference gene.

### Gene ontology analysis

The major biological activities of DEGs were determined using a GO enrichment analysis. DEGs from the 15 DPA stage were annotated and grouped into three major groups based on their functional characteristics: CC, MF, and BP categories.

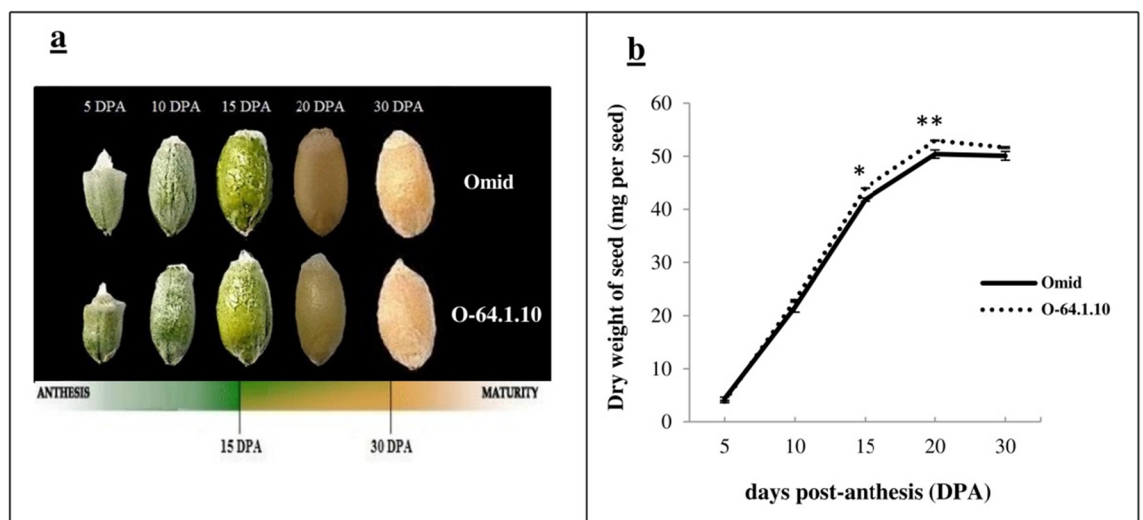
## Result

### Grain morphology changes during seed development

The dynamic comparison of developing grains was examined at 5, 10, 15, 20, and 30 DPA stages between the “O-64.1.10” genotype and its “Omid” counterpart (Fig. 1a). The dry weight of the seeds showed a continuous increase from 5 to 20 DPA and reached its highest point at 20 DPA in both the mutant genotype “O-64.1.10” and its parent genotype “Omid” (Fig. 1b). Our results showed that mutant “O-64.1.10” exhibited greater grain weight growth than “Omid” at 15, 20 and 30 DPA. Also highest rate of dry weight accumulation per day occurred between 10 and 15 DPA. The 15 DPA stage are beginning of effective grain filling stage in this stage the grain dry weight increases by about two-fold. The 15 DPA stage for the RNA-seq analysis. This finding is in agreement with earlier study conducted by Shewry et al.<sup>12</sup>.

### Identification of differentially expressed genes

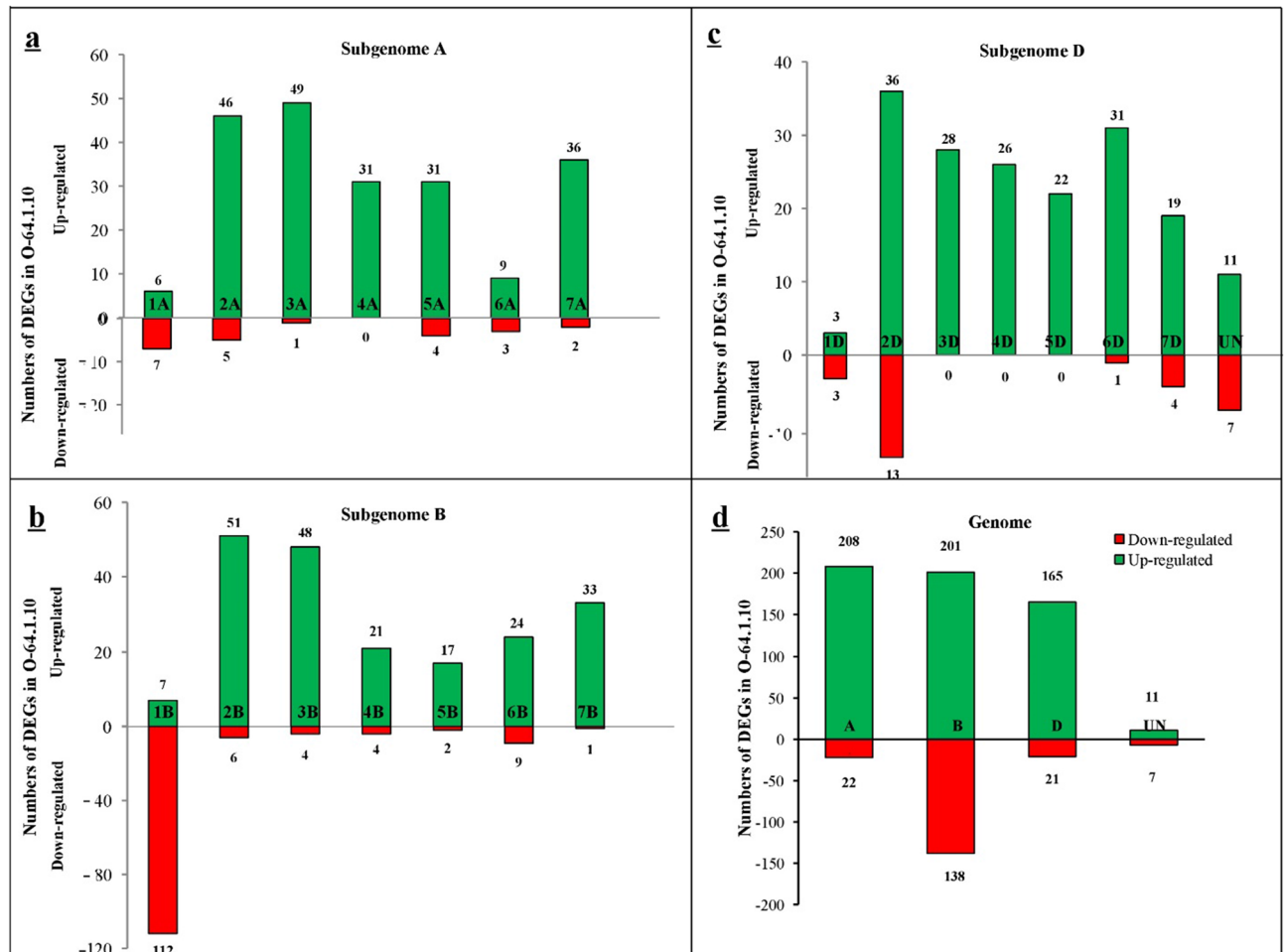
The quality of RNA and quality of high-throughput sequencing data showed in supplementary data (S1 Figure). A total of 18.97 and 16.24 million paired-end short-read sequences with high-quality (Q30 > 99%) obtained by Illumina sequencing of the “Omid” and “O-64.1.10” cDNA libraries, respectively. The GC (guanine-cytosine) content for the clean data was 51% and 52% for the “Omid” and “O-64.1.10”, respectively (Table 3). DEGs were found in 773 genes between “O-64.1.10” and “Omid” (supplemental data 1). 585 genes up-regulated and 188 genes down-regulated in the “O-64.1.10” genotype compared to the “Omid”. Our study showed that most of the DEGs (339 genes, 43.85%) were anchored on B-genome (Fig. 2d, supplemental data 2). In addition, we



**Figure 1.** Dynamic comparison of grains at different developmental stages between in high baking quality mutant “O-64.1.10” and its wild type genotypes “Omid”. (a) Changes in grain morphology at 5, 10, 15, 20 and 30 DPA of “O-64.1.10” and “Omid”. (b) Changes in dry grain weight during the development process. Three biological replications and t-test were done to obtain the curve. \*—significant at  $P < 0.05$  and \*\*—significant  $P < 0.01$ . I: standard error.

Sample	Sequence length (base pairs)	Number of sequences (total reads)	Total nucleotides in data set (base)	% sequence duplication levels	% GC-content	<sup>a</sup> Q30 (%)
Omid	150	18,976,039	2,846,405,850	0.5335	51	99.78
O-64.1.10	150	16,246,625	2,436,993,750	0.672	52	99.73

**Table 3.** Summary of high-throughput sequencing data in “Omid” and mutant at 15 DPA. <sup>a</sup>paired-end short-read sequences with high-quality (Q30 > 99%).

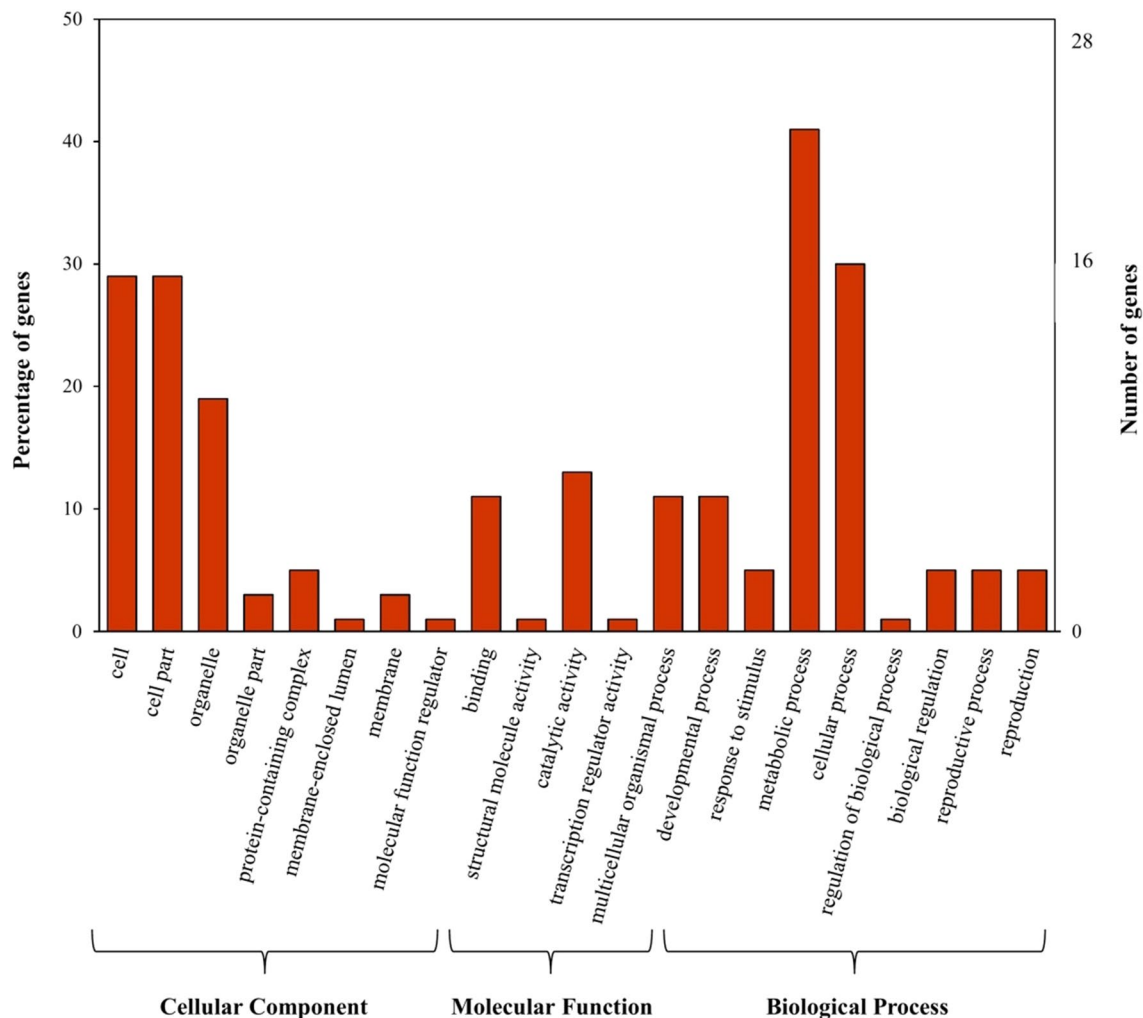


**Figure 2.** The numbers of DEGs on the A, B and D genomes in the “O-64.1.10” genotypes relative to the “Omid” cultivar. UN: unassigned chromosome.

found that up-regulated genes in “O-64.1.10” were enriched on chromosomes 2A (5.95%) and 3A (6.33%) and chromosomes 2B (6.59%) and 3B (6.2%), whereas down-regulated genes in “O-64.1.10” were overrepresented on chromosome 1B (14.5%) (Fig. 2).

### Functional classification and gene ontology analysis

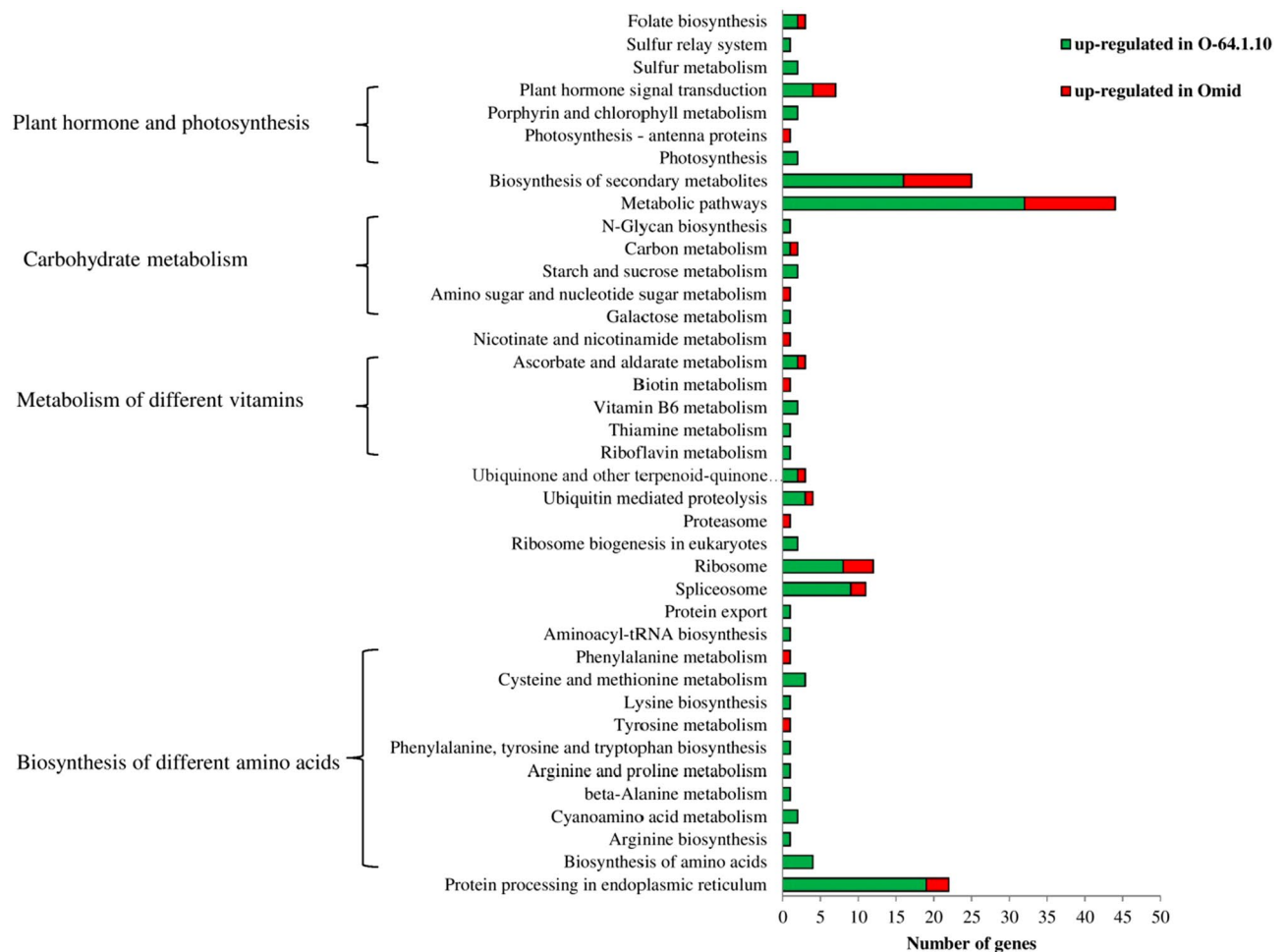
GO enrichment analysis was carried out to characterize the main biological functions of DEGs in seeds 15 DPA. The DEGs from 15 DPA stage were functionally annotated and classified ontologically into three principal categories: CC, MF, and BP. Furthermore, 62 GO terms were enriched in the DEG from O-64.1.10 and Omid, including 19 CC, 15 MF, and 28 BP. A total of 733 DEGs were assigned to three main GO functional categories and then divided into 21 sub-categories. Metabolic process, cellular process, developmental process, and multicellular organismal process were found to be dominant in the BP category. The MF category was primarily dominated by catalytic activity and binding. Cell, cell part, and organelle were foremost in the cellular CC category (Fig. 3). The “O-64.1.10” genotype had the highest number of genes involved in specific BP, such as cellular protein catabolic process, diterpenoid biosynthetic process, gibberellin metabolic process, cellular nitrogen compound biosynthetic process, cellular amide metabolic process, lipid biosynthetic process and response to abiotic stimulus. Overall, these results suggest that the genes associated with wheat early grain development play crucial roles in encoding diverse regulators and proteins.



**Figure 3.** Histogram of Gene Ontology (GO) classifications of DEGs between “Omid” and “O-64.1.10” genotypes in 15 DAP. The results are summarized in three principal GO functional categories: biological processes, cellular component, and molecular function.

At the same time pathway enrichment analysis using KEGG database was performed to reveal the active biological pathways involved in baking quality of immature wheat grains. In this study, 68 and 34 pathways were identified in the “O-64.1.10” and the wild-type, respectively. Among all the pathways, the largest pathways were metabolic pathway, protein processing in endoplasmic reticulum and biosynthesis of secondary. An analysis using the KEGG database on biological pathways in the “O-64.1.10” genotype showed that a total of 187 annotated DEGs were assigned to 68 pathways. Among the 68 pathways, the largest pathway was metabolic pathways, which contained 32 genes, followed by protein processing in endoplasmic reticulum (19 genes), biosynthesis of secondary metabolites (16 genes), spliceosome (9 genes), ribosome (8 genes), plant hormone signal transduction (4 genes), and biosynthesis of amino acids (4 genes). The KEGG analysis in the “O-64.1.10” genotype indicated that 39 DEGs were involved in protein processing in endoplasmic reticulum (19 genes), biosynthesis of different amino acids pathways (14 genes), and different vitamins metabolism pathways (6 genes), which might participate in regulating baking quality (Fig. 6). Furthermore, the carbohydrate metabolism pathways (5 genes) in the wild-type genotype might participate in regulating baking quality. Pathway enrichment analysis of “Omid” genotype displayed the total of 68 annotated DEGs were assigned to 34 pathways. The top three enriched pathways include metabolic pathways (12 genes), biosynthesis of secondary metabolites (9 genes), and ribosome (4 genes) were largest pathways (Fig. 4, Supplemental data 4).

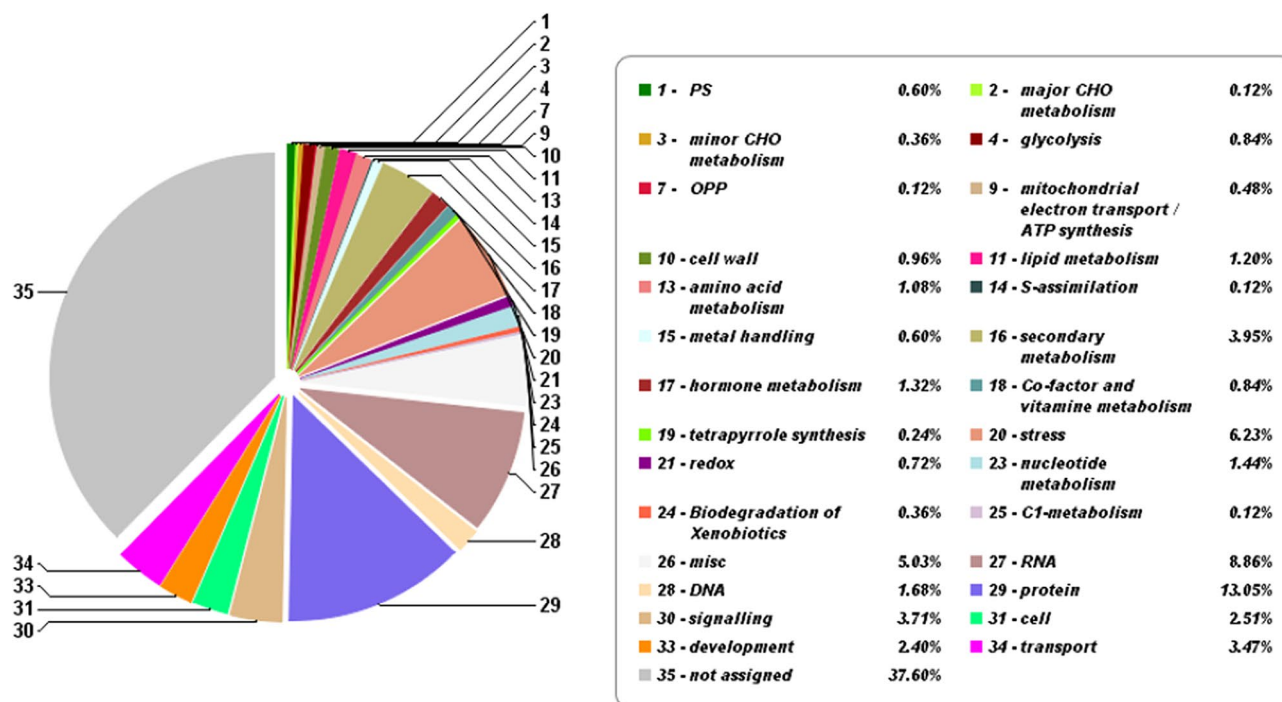
Functional annotations of these DEGs were achieved by using Mercator tool (Mercator V3.6). Results of differential expression showed that out of 773 differentially expressed genes, 109 (13.05%) BINs assigned to protein process. The number of genes were assigned to the functional category “protein” was highest (103), followed by “RNA” (74), “stress” (52) and “misc” (42) (Fig. 5). Lastly, In the “O-64.1.10”, 82 out of 103 DEGs involved in protein processing (Fig. 5), were up-regulated compared to the wild-type (Table 4). Elongation factor 1-alpha (EF1a) play a crucial role in the protein synthesis machinery of cells, and its function extends to various physiological processes in plants, including wheat<sup>48</sup>. Our results indicated that this gene were highly expressed in “O-64.1.10” genotype.



**Figure 4.** Bar diagrams displayed the KEGG significantly metabolic pathways in the “O-64.1.10” and “Omid” at 15 DPA.

### Transcription factor, protein kinases and transcriptional regulators analysis

From DEGs, it was found that out of 22 TF families that the highest expression occurred for heat shock factor (HSF), AP2/ERF-ERF and basic leucine zipper (bZIP). Additionally, the HSF, bZIP, C2C2-Dof, B3-ARF, BES1, C3H, GRF, HB-HD-ZIP, PLATZ, MADS-MIKC, GARP-G2-like, NAC, OFP, and TUB TF families were exclusively expressed in the “O-64.1.10” genotype at the 15 DPA stage (Fig. 6a, supplemental data 3). Our findings revealed that, the bZIP TFs was related to storage protein gene regulation. TFs of the bZIP family are regulators of many central developmental and physiological processes including storage protein gene regulation<sup>49</sup>, energy metabolism<sup>37</sup>, light responses and oxidative stress signaling<sup>50</sup>. The regulations of the grain-filling process may be regulated by several TF families, such as bZIP and HSF TFs. The HSF TFs are regulators in response to stress<sup>51</sup>, early grain filling<sup>16</sup>, seed maturation<sup>52</sup> and pollen development<sup>53</sup>. The HSF family is a transcriptional activator of heat shock protein (HSP) genes<sup>54</sup>. The HSP family is an important factor ensuring correct protein folding that plays a significant role in degradation pathways, such as endoplasmic reticulum-associated degradation<sup>55</sup>. Our results showed that LMW-GS genes and ERF TF had a higher expression in “Omid” and “O-64.1.10”, respectively. Hasrak et al.<sup>56</sup> reported there is a negative correlation between ERF TFs and lower expression of LMW-GS. Our data indicated that the NAC TF (TraesCS2B01G359200.1) were found from highly expressed in “O-64.1.10” genotype. The NAC019 TF regulates glutenin and starch accumulation and its elite allele improves wheat grain quality<sup>39</sup>. The TKL, STE, CAMK, and Others of PKs were found from highly expressed in the



**Figure 5.** The amount of hub genes expression in 35 segregate biological groups (35 main BINs) using the Mercator web server.

Code BIN	Class	No. BIN	No. Unigene	Up-regulated in O-64.1.10	down-regulated in O-64.1.10
29.1	Protein.aa activation	3	3	3	0
29.2	Protein.synthesis	26	23	18	5
29.3	Protein.targeting	9	9	7	2
29.4	Protein.postranslational modification	25	23	17	6
29.5	Protein.degradation	41	40	33	7
29.6	Protein.folding	4	4	3	1
29.7	Protein.glycosylation	1	1	1	0
Total	-	109	103	82	21

**Table 4.** Categorization of differentially expressed hub genes (up/down-regulated) in “O-64.1.10” at 15 DPA into protein process; BIN code 29.

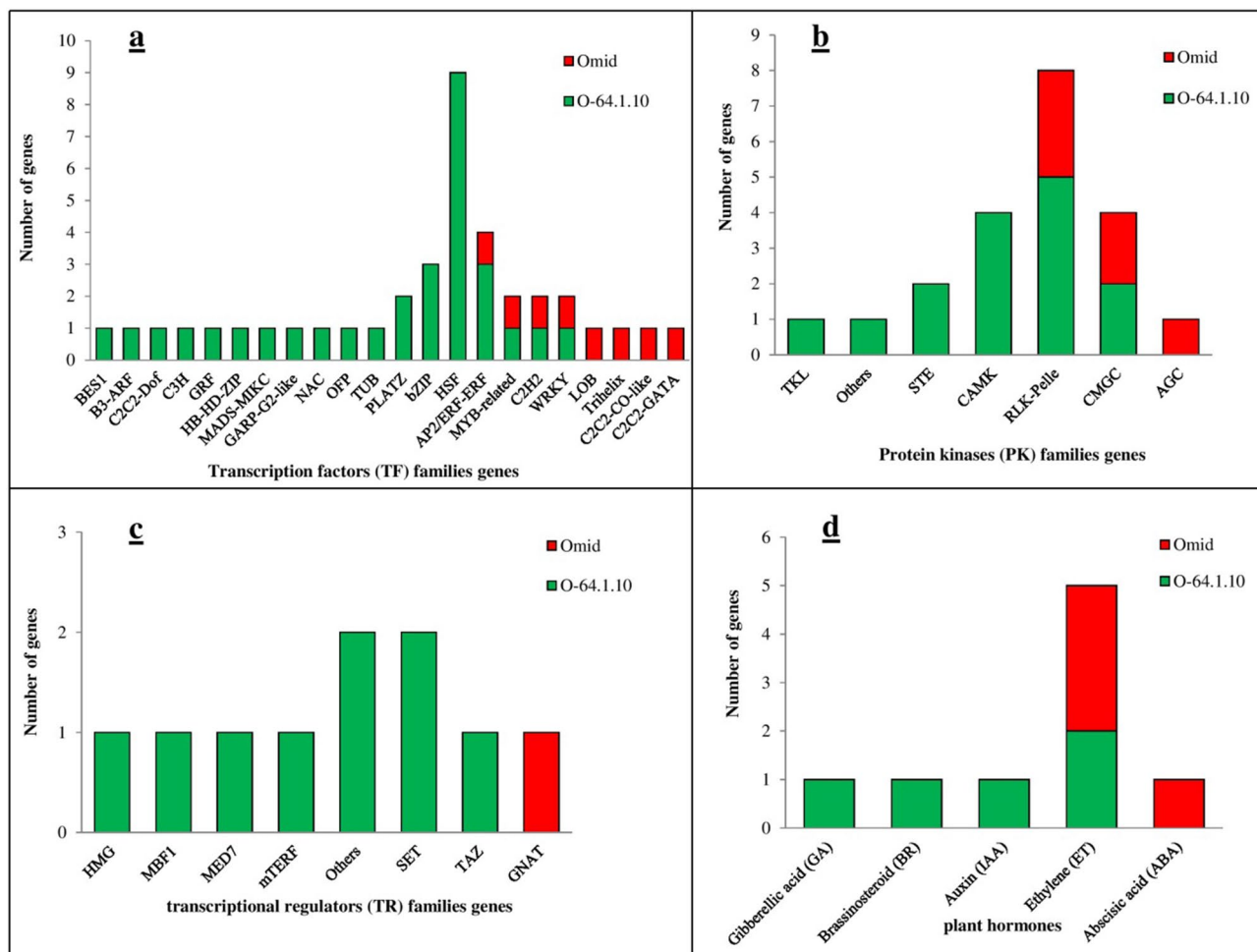
“O-64.1.10” genotypes, whereas the AGC of PK families was expressed only in the “Omid” genotypes (Fig. 6b, supplemental data 3). The PKs may cooperate with the transcriptional networks to refine the regulation of genes during seed development<sup>34</sup>. In this study, among all genes with differential expression, seven important families of TRs such as HMG, MBF1, MED, mTERF, SET, TAZ, and others genes were found from highly expressed in the “O-64.1.10” genotype. In contrast, one family of TRs (GNAT) only was found from highly expressed in the wild-type genotype (Fig. 6c, supplemental data 3).

### Plant hormones and transposable elements (TEs) analysis

Among of the five hormone-related genes that were up-regulated in each of genotypes, three genes associated with gibberellic acid, brassinosteroid, and auxin were expressed only in “O-64.1.10” genotypes (Fig. 6d, supplemental data 3). The plant hormones regulate multiple biological processes in early grain development of wheat and the comprehensive expression profiling provides useful information for hormone regulatory mechanisms<sup>16</sup>. The hormone response genes affect the wheat grain quality through indirect pathways<sup>57</sup>.

Our research displayed eleven TEs that were expressed only in “O-64.1.10” genotypes. These genes related to transposon may be responsible for causing the high baking quality in the mutant genotype “O-64.1.10”. So, further research is needed to find mutations governing genes related baking quality. TEs are mobile genetic elements in the eukaryotic genome which alter the expression of neighboring genes via insertion into promoter regions, or disrupt the function of protein-coding genes when inserted into the genes, or even change gene structure by altering its splicing and polyadenylation patterns<sup>58–60</sup>.





**Figure 6.** Distribution of transcription factor (TF) genes of 22 TF families (a), protein kinases (PK) genes of seven PK families (b), transcriptional regulators (TR) genes of eight families (c), and plant hormones (d) in comparison between Omid and O-64-1-10 genotypes with different baking quality at 15 DPA.

	Omid	O-64.1.10	Total
Total number of sequences examined	189	584	773
Total number of identified SSRs	27	100	127
Number of SSR containing sequences	27	82	109
Number of sequences containing more than 1 SSR	0	17	17
Number of SSRs present in compound formation	0	8	8
Mono	0	0	0
Di	0	3	3
Tri	26	96	122
Tetra	0	0	0
Penta	0	0	0
Hexa	1	1	2

**Table 5.** Details of SSR markers obtained from DEGs of “Omid” and “O-64.1.10”.

### Identification of SSRs

In the context of differential gene expression between the “O-64.1.10” and wild-type genotypes, 127 SSRs were identified. The repeat pattern of the identified SSRs indicated a prevalence of trinucleotide repeats (122) were the most abundant, followed by di- (3) and hexa- (2) nucleotide repeats (Table 5, supplemental data 3). Within the trinucleotide repeats, CCG/CGG (54) was the most frequent motif, followed by AGG/CCT (26) and AGG/CCT (22). The majority of motifs (96.06%) consisted of 5–7 repeats, while motifs with 8–10 repeats were rare (3.94%). For instance, gamma-gliadin contained AAC motifs with 10 repeats. A total of 109 sequences containing SSRs, 17 sequences containing more than one SSR, and 8 sequences containing SSRs in compound formation were identified. For instance, alpha-gliadin contained (ACA)<sub>9</sub> and (CAA)<sub>8</sub> motifs. These genic SSRs/RNAseq-SSR markers were developed from transcriptome sequences and can be used for marker-assisted selection (MAS).

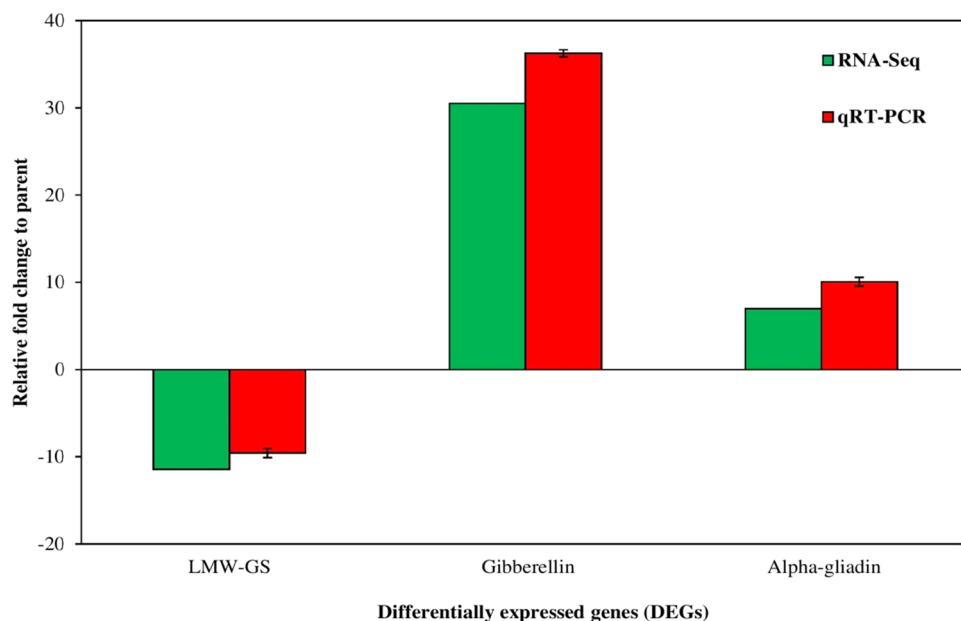
### Quantitative Reverse Transcription PCR Validation

qRT-PCR was used to validate the RNA-seq results. Therefore, out of DEGs, three important genes involved in baking quality were selected from the DEGs, including LMW-GS, Alpha-gliadin and Gibberellin. The LMW-GS and gliadins (alfa/beta, omega, and gamma) are the main storage proteins and the major components of the gluten polymer. The accumulation of these storage proteins affects wheat quality formation<sup>17</sup>. The relative fold changes in gene expression measured by qRT-PCR were found to be consistent with their expression levels determined by RNA-Seq. Both techniques yielded similar patterns of changes, validating the results obtained from RNA-Seq (Fig. 7).

### Conclusions

The transcriptomics analysis is a powerful tool to study molecular processes taking place during grain development. In this study, differential expressed genes of two breed wheat genotypes “O-64.1.10” and “Omid” were investigated at 15 DPA by transcriptomics approaches. a total number of 733 DEGs were identified between each genotype, as well as the DEGs analysis suggested that up-regulated genes in “O-64.1.10” were enriched on A-genome, whereas down-regulated genes in “O-64.1.10” were overexpression on B-genome. Our recent studies have identified several crucial factors that regulate the baking quality of wheat, including grain storage proteins (gluten), TFs, TRs, PKs, plant hormones, carbohydrates (starch and sucrose), lipids (unsaturated fatty acids), and vitamins. Using the KEGG pathway enrichment analysis of DEGs comparing “O-64.1.10” with “Omid” showed that various pathways of metabolic pathways, protein processing in endoplasmic reticulum, biosynthesis of secondary metabolites, spliceosome, ribosome, plant hormone signal transduction, and biosynthesis of amino acids were enriched in the “O-64.1.10” genotype with the highest number of the genes, whereas the pathways associated with metabolic pathways, biosynthesis of secondary metabolites, and ribosome were frequent in the “Omid” genotype. Yu et al.<sup>17</sup> reported that metabolic pathway network analysis that major and minor metabolic pathways regulate one another to ensure regular seed development and nutritive reserve accumulation. Yan et al.<sup>61</sup> showed that carbohydrate and hormone metabolism has play important roles in the grain size and weight in wheat.

Our results indicated that TraesCS2A01G483600 (encodes Elongation factor 1-alpha) expressed only in “O-64.1.10” genotypes. Paul et al.<sup>48</sup> suggested that Elongation factor 1-alpha gene play an important role in cell expansion in early developmental wheat grain. The validation of transcripts via qRT-PCR presented in this study holds significant academic value. These transcripts have been found to have a positive correlation with the enhancement of the baking quality of wheat grains. Lastly, in addition, we also identified 127 RNA-seq SSR



**Figure 7.** Quantitative Reverse Transcription PCR validation of RNA-seq results for selected DEGs.

markers among all genes with differential expression in “O-64.1.10” and its wild type genotype. The repeat pattern of identified SSRs indicated an abundance of dinucleotides followed by tri, and hexanucleotide repeats.

In conclusion, results from our study provides valuable shed light into the genetic mechanisms underlying baking quality in wheat flour production, and could have significant implications for the baking industry.

### Data availability

The datasets generated and/or analysed during the current study are available in the SRA NCBI repository, <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA979921>. BioSample accessions: <https://www.ncbi.nlm.nih.gov/biosample/SAMN35615225>. <https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN35615224>.

Received: 23 May 2023; Accepted: 7 May 2024

Published online: 13 May 2024

### References

- Duan, J., Xia, C., Zhao, G., Jia, J. & Kong, X. Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. *BMC Genomics* **13**, 1–2 (2012).
- Singh, A. *et al.* Genome-wide transcriptome study in wheat identified candidate genes related to processing quality, majority of them showing interaction (quality x development) and having temporal and spatial distributions. *BMC Genomics* **15**, 1–9 (2014).
- Takač, V. *et al.* Differences in processing quality traits, protein content and composition between spelt and bread wheat genotypes grown under conventional and organic production. *Foods* **10**(1), 156 (2021).
- Xue, C., Matros, A., Mock, H. P. & Mühling, K. H. Protein composition and baking quality of wheat flour as affected by split nitrogen application. *Front. Plant Sci.* **10**, 642. <https://doi.org/10.3389/fpls.2019.00642> (2019).
- Peng, Y. *et al.* Wheat quality formation and its regulatory mechanism. *Front. Plant Sci.* **13**, 834654. <https://doi.org/10.3389/fpls.2022.834654> (2022).
- Schuster, C., Huen, J. & Scherf, K. A. Comprehensive study on gluten composition and baking quality of winter wheat. *Cereal Chem.* **100**, 142–155 (2023).
- Xue, G. P., Sadat, S., Drenth, J. & Lynne, M. L. The heat shock factor family from *Triticum aestivum* in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *J. Exp. Bot.* **65**(2), 539–557 (2012).
- Oyeyinka, S. A. & Basse, I. A. Composition, functionality, and baking quality of flour from four brands of wheat flour. *J. Culinary Sci. Technol.* <https://doi.org/10.1080/15428052.2023.21918745> (2023).
- Salmanowicz, B. P. *et al.* The relationship between grain hardness, dough mixing parameters and bread-making quality in winter wheat. *Int. J. Mol. Sci.* **13**(4), 4186–4201 (2012).
- Jańczak-Pieniżek, M. *et al.* Comparative assessment of the baking quality of hybrid and population wheat cultivars. *Appl. Sci.* **10**, 7104. <https://doi.org/10.3390/app10207104> (2020).
- Michel, S. *et al.* Improving the baking quality of bread wheat by genomic selection in early generations. *Theor. Appl. Genet.* **131**, 477–493 (2018).
- Rahemi, M. R., Yamchi, A., Navabpour, S., Soltanloo, H. & Roepstorff, P. Evaluation of gamma ray effect on wheat bakery properties in Omid, Roshan and Tabasai cultivars by artificial neural network. *JonSat.* **40**(4), 119–127 (2020).
- Rangan, P., Furtado, A. & Henry, R. 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. <https://doi.org/10.1186/s12864-017-4154-z> (2017).
- Furtado, A. *et al.* A novel highly differentially expressed gene in wheat endosperm associated with bread quality. *Sci. Rep.* **5**, 10446. <https://doi.org/10.1038/srep10446> (2015).
- Chi, Q. *et al.* Global transcriptome analysis uncovers the gene co-expression regulation network and key genes involved in grain development of wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics* **19**, 853–866 (2019).
- Guan, J. *et al.* Transcriptome analysis of developing wheat grains at rapid expanding phase reveals dynamic gene expression patterns. *Biology* **11**(2), 281 (2022).
- Yu, Y. *et al.* Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development. *Crop J.* **4**(2), 92–106 (2016).
- Henry, R. J., Furtado, A. & Rangan, P. Wheat seed transcriptome reveals genes controlling key traits for human preference and crop adaptation. *Curr. Opin. Plant Biol.* **1**(45), 231–236 (2018).
- Paul, S. *et al.* RNA-Seq analysis of developing grains of wheat to intrigue into the complex molecular mechanism of the heat stress response. *Front. Plant Sci.* **13**, 904392 (2022).
- Curtis, T. Y. *et al.* Contrasting gene expression patterns in grain of high and low asparagine wheat genotypes in response to sulphur supply. *BMC Genomics* **20**, 628 (2019).
- Pfeifer, M. *et al.* International wheat genome sequencing consortium genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091 (2014).
- Shewry, P. R. *et al.* An integrated study of grain development of wheat (cv. Hereward). *J. Cereal Sci.* **56**(1), 21–30 (2012).
- Saulnier, L., Guillon, F. & Chateigner-Boutin, A. L. Cell wall deposition and metabolism in wheat grain. *J. Cereal Sci.* **56**(1), 91–108 (2012).
- Zamani, M. J. *et al.* Marker-assisted selection for recognizing wheat mutant genotypes carrying HMW glutenin alleles related to baking quality. *Sci. World J.* **2014**, 1–5. <https://doi.org/10.1155/2014/387912> (2014).
- Izadi-Darbandi, A., Yazdi-Samadi, B., Shanejat-Boushehri, A. A. & Mohammadi, M. Allelic variations in Glu-1 and Glu-3 loci of historical and modern Iranian bread wheat (*Triticum aestivum* L.) cultivars. *J. Genet.* **89**, 193–9 (2010).
- Payne, P. I. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Plant Biol.* **38**, 141–153 (1987).
- Shewry, P. R., Halford, N. G. & Tatham, A. S. High molecular weight subunits of wheat glutenin. *Cereal Sci.* **15**, 115–119 (1992).
- Ahmad, M. Molecular marker-assisted selection of HMW glutenin alleles related to wheat bread quality by PCR-generated DNA markers. *Theor. Appl. Genet.* **101**, 892–896 (2000).
- Tabiki, T., Ikeguschi, S. & Ikeda, T. M. Effects of High-molecular-weight and Low-molecular-weight glutenin subunit alleles on common wheat flour quality. *Breed. Sci.* **56**, 131–136 (2006).
- Shahnejat-Bushehri, A. A., Gomarian, M. & Yazdi-Samadi, B. The high molecular weight glutenin subunit composition in old and modern bread wheat's cultivated in Iran. *Aust. J. Agric. Res.* **57**, 1109–1114 (2006).
- Kocourková, Z. *et al.* Wheat breeding for the improved bread-making quality using PCR based markers of glutenins. *Czech J. Genet. Plant Breed.* **44**(3), 105–113 (2008).
- Gianibelli, M. C., Echaide, M., Larroque, O. R., Carrillo, J. M. & Dubcovsky, J. Biochemical and molecular characterization of Glu-1 loci in Argentinean wheat cultivars. *Euphytica* **128**, 61–73 (2002).

33. Locascio, A., Roig-Villanova, I., Bernardi, J. & Varotto, S. Current perspectives on the hormonal control of seed development in Arabidopsis and maize: A focus on auxin. *Front. Plant Sci.* **5**, 412 (2014).
34. Xue, L. J., Zhang, J. J. & Xue, H. W. Genome-wide analysis of the complex transcriptional networks of rice developing seeds. *PLoS One* **7**(2), e31081. <https://doi.org/10.1371/journal.pone.0031081> (2012).
35. Wei, L. *et al.* Transcriptome analysis of wheat grain using RNA-Seq. *Front. Agric. Sci. Eng.* **1**(3), 214–222 (2015).
36. Iwata, Y. & Koizumi, N. An Arabidopsis transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. *Proc Natl Acad Sci USA* **102**, 5280–5285 (2005).
37. Baena-González, E., Rolland, F., Thevelein, J. M. & Sheen, J. A central integrator of transcription networks in plant stress and energy signalling. *Nature* **448**(7156), 938–942 (2007).
38. Finkelstein, R. R. & Lynch, T. J. Abscisic acid inhibition of radicle emergence but not seedling growth is suppressed by sugars. *Plant Physiol.* **122**, 1179–1186 (2000).
39. Gao, Y. *et al.* The endosperm-specific transcription factor TaNAC019 regulates glutenin and starch accumulation and its elite allele improves wheat grain quality. *Plant Cell* **33**, 603–622 (2021).
40. Rahemi, M. R., Yamchi, A., Navabpour, S., Soltanloo, H. & Roepstorff, P. Evaluation of gamma ray effect on wheat bakery properties in Omid, Roshan and Tabasai cultivars by artificial neural network. *J. Nuclear Sci. Technol.* **40**(4), 119–127 (2020).
41. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**(4), 357–360 (2015).
42. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**(5), 511–515 (2010).
43. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33**(16), 2583–2585 (2017).
44. Thiel, T., Michalek, W., Varshney, R. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–22 (2003).
45. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for Biologist Programmers. In *Bioinformatics Methods and Protocols* (eds Misener, S. & Krawetz, S. A.) 365–386 (Humana Press, 1999). <https://doi.org/10.1385/1-59259-192-2:365>.
46. Pfaffl, M. W., Horgan, G. W. & Dempfle, L. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**(9), e36 (2002).
47. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>ΔΔCT method. *Methods* **25**(4), 402–408. <https://doi.org/10.1006/meth.2001.1262> (2001).
48. Paul, S. *et al.* RNA-Seq analysis of developing grains of wheat to intrigue into the complex molecular mechanism of the heat stress response. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2022.904392> (2022).
49. Lara, P. *et al.* Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. *J. Biol. Chem.* **278**(23), 21003–21011 (2003).
50. Guedes Corrêa, L. G. *et al.* The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PLoS One* **3**(8), e2944. <https://doi.org/10.1371/journal.pone.0002944> (2008).
51. Skylas, D. J. *et al.* Heat shock of wheat during grain filling: Proteins associated with heat-tolerance. *J. Cereal Sci.* **35**(2), 175–188 (2002).
52. Reňák, D., Gibalova, A., Šolcová, K. & Honys, D. A new link between stress response and nucleolar function during pollen development in Arabidopsis mediated by AtREN1 protein. *Plant Cell Environ.* **37**(3), 670–83 (2014).
53. Kotak, S., Vierling, E., Baumlein, H. & Koskull-Döring, P. V. A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis. *Plant Cell.* **19**(1), 182–195 (2007).
54. Xue, G. P., Sadat, S., Drenth, J. & McIntyre, C. L. The heat shock factor family from *Triticum aestivum* in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *J. Exp. Botany* **65**(2), 539–557 (2014).
55. Bozaykut, P., Ozer, N. K. & Karademir, B. Regulation of protein turnover by heat shock proteins. *Free Radic. Biol. Med.* **1**(77), 195–209 (2014).
56. Hasrak, S. *et al.* A study to assess the role of gluten encoded genes and their regulatory elements in bread making quality of wheat. *Iran. J. Biotech.* **17**(4), e2164 (2019).
57. Yang, Y. *et al.* Multi-locus GWAS of quality traits in bread wheat: Mining more candidate genes and possible regulatory network. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2020.01091> (2020).
58. Butelli, E. *et al.* Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell.* **24**, 1242–1255 (2012).
59. Kumar, A. & Bennetzen, J. L. Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532 (1999).
60. Moden, Y. *et al.* Comprehensive survey of transposon mPing insertion sites and transcriptome analysis for identifying candidate genes controlling high protein content of rice. *Front. Plant Sci.* **13**, 969582 (2022).
61. Yan, L. *et al.* Transcriptome analysis reveals potential mechanisms for different grain size between natural and resynthesized allohexaploid wheats with near-identical AABB genomes. *BMC Plant Biol.* **18**, 28 (2018).

## Acknowledgements

The authors would like to acknowledge the reviewers for their helpful and constructive comments on this manuscript.

## Author contributions

All authors reviewed the manuscript Hossein Ahmadi-Ochtapeh wrote the main manuscript and Hassan Soltanloo edited Advisors, Seyyede Sanaz Ramezanpour, Ahad Yamchi and Vahid Shariati respectively.

## Funding

This study was supported by a grant from Gorgan University of Agricultural and Natural Resources.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-61528-y>.

**Correspondence** and requests for materials should be addressed to H.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024