

RESEARCH ARTICLE

Open Access



Ubiquitin-related genes are differentially expressed in isogenic lines contrasting for pericarp cell size and grain weight in hexaploid wheat

Jemima Brinton, James Simmonds and Cristobal Uauy* 

Abstract

Background: There is an urgent need to increase global crop production. Identifying and combining specific genes controlling distinct biological processes holds the potential to enhance crop yields. Transcriptomics is a powerful tool to gain insights into the complex gene regulatory networks that underlie such traits, but relies on the availability of a high-quality reference sequence and accurate gene models. Previously, we identified a grain weight QTL on wheat chromosome 5A (5A QTL) which acts during early grain development to increase grain length through cell expansion in the pericarp. In this study, we performed RNA-sequencing on near isogenic lines (NILs) segregating for the 5A QTL and used the latest gene models to identify differentially regulated genes and pathways that potentially influence pericarp cell size and grain weight in wheat.

Results: We sampled grains at 4 and 8 days post anthesis and found that genes associated with metabolism, biosynthesis, proteolysis and the defence response are upregulated during this stage of grain development in both NILs. We identified a specific set of 112 transcripts differentially expressed (DE) between 5A NILs at either time point, including eight potential candidates for the causal 5A gene and its downstream targets. The 112 DE transcripts had functional annotations including non-coding RNA, transposon-associated, cell-cycle control, ubiquitin-related, heat-shock, transcription and histone-related. Many of the genes identified belong to families that have been previously associated with seed/grain development in other species. Notably, we identified DE transcripts at almost all steps of the pathway associated with ubiquitin-mediated protein degradation. In the promoters of a subset of DE transcripts we identified enrichment of binding sites associated with C2H2, MYB/SANT, YABBY, AT-HOOK and Trihelix transcription factor families.

Conclusions: In this study, we identified DE transcripts with a diverse range of predicted biological functions, reflecting the complex nature of the pathways that control early grain development. Few of these are the direct orthologues of grain size genes in other species and none have been previously characterised in wheat. Further functional characterisation of these candidates and how they interact will provide novel insights into the control of grain size in cereals.

Keywords: Wheat, RNA-seq, Ubiquitin, Grain weight, Pericarp, Transcriptomic, Grain size, Near isogenic lines

* Correspondence: cristobal.uauy@jc.ac.uk
John Innes Centre, Norwich, UK

Background

Crop production must increase to meet the demands of a global population estimated to exceed nine billion by 2050 [1]. Indeed, one in nine people currently live under food insecurity [2]. With limited opportunity for agricultural expansion, increasing yields on existing land could significantly reduce the number of people at risk of hunger [3]. It is estimated that at least a 50% increase in crop production is required by 2050 [4, 5], however current rates of yield increase are insufficient to achieve this goal [6]. It is therefore critical and urgent that we identify ways to increase crop yields.

Final crop yield is influenced by the interaction of many genetic and environmental factors. This complexity hinders its study and has meant that the mechanisms controlling this trait are not well understood. Grain weight, however, an important component of final yield, is more stably inherited and is better understood than yield itself [7]. Grain weight is mainly determined by grain size, which itself is controlled by the coordination of cell proliferation and expansion processes. Studies in both crop and model species have shown that these processes are regulated by a wide range of genes and molecular mechanisms (reviewed in [8, 9]). Control at the transcriptional level has been demonstrated, with the rice transcription factor (TF) *OsSPL16* influencing grain size through cell proliferation [10], whilst a WRKY domain TF, *TTG2*, influences cell expansion in the integument of the *Arabidopsis* seed [11]. Important pathways relating to protein turnover have also been identified, for example the E3 ubiquitin-ligase, *GW2*, negatively regulates grain weight and width in rice through the control of cell division [12]. *GW2* orthologues in other species, including *Arabidopsis* and wheat, also act as negative regulators of seed/grain size suggesting that these mechanisms may be conserved across species [13, 14]. Other pathways/mechanisms which affect grain size include microtubule dynamics [15, 16], G-protein signalling [17, 18] and phytohormone biosynthesis and signalling [19–21].

Wheat is a crop of global importance, accounting for approximately 20% of the calories consumed by the human population [22]. However, our understanding of the mechanisms controlling grain size remains limited in wheat, compared to rice and *Arabidopsis*. Comparative genomics approaches have provided some insight [13, 23] and many quantitative trait loci (QTL) associated with grain size and shape components (grain area, length and width) have been identified [24–29]. However, none of these QTL have been cloned and little is understood about the underlying mechanisms. Previously, we identified a QTL associated with increased grain weight on wheat chromosome 5A. Using BC₄ near isogenic lines (NILs) we determined that the QTL acts during the early stages of

grain development to increase grain length through increased cell expansion in the pericarp [29]. This and other studies suggest that the early stages of grain/ovule development are important for determining final grain size/shape in wheat [13, 30, 31].

Transcriptomics is a powerful tool to gain insights into the complex gene regulatory networks that underlie specific traits and biological processes. Several studies have used transcriptomics approaches to look at the genes expressed during grain development in wheat [32–38]. However, these studies have mostly focused on the later stages of grain development, often focusing on starch accumulation in the endosperm. Additionally, many of these studies were performed using microarrays [33, 36, 37], which represent a fraction of the transcriptome and are unable to distinguish between homoeologous gene copies. More recent studies have used RNA-sequencing (RNA-seq) [34, 35], which is an open-ended platform that provides homoeolog specific resolution. However, the accuracy of RNA-seq is dependent on the availability of a high-quality reference sequence and accurate gene models. Until recently, the large (~17 Gb) and highly repetitive nature of the hexaploid wheat genome meant that genomic resources were limited and incomplete. However, this has changed drastically in the last few years with the release of several whole genome sequences and annotations [39–42]. To date, the RNA-seq grain development studies have used either expressed sequence tags (ESTs) [35, 38] or the Chromosome Survey Sequence (CSS) [34] as references. However in hindsight, these annotations are incomplete with respect to the latest gene models [39, 41]. These novel resources provide new opportunities for more detailed and accurate transcriptomic studies in wheat.

A potential drawback of transcriptomic studies is that comparisons across varieties, tissues or time points can result in a large number of transcripts being differentially expressed. While this informs our understanding of the biological mechanisms, it is difficult to prioritise specific genes for downstream analysis. Comparative transcriptomic approaches using more precisely defined genetic material, tissues and developmental time points can aid in this by defining a smaller set of differentially regulated transcripts. For example, a comparison of the flag leaf transcriptomes of wild-type and RNAi knock-down lines of the *Grain Protein Content 1* (*GPC*) genes was used to identify downstream targets of the *GPC* TFs [43]. Similarly, the transcriptomes of NILs segregating for a major grain dormancy QTL on chromosome arm 4AL were compared and specific candidate genes underlying the QTL were identified [44]. To our knowledge, no such experiments have been performed on isogenic lines with a known difference for grain size in wheat.

In this study, we performed RNA-seq on NILs segregating for a major grain weight QTL on chromosome

arm 5AL. Previously, we showed that the QTL acts during early grain development and that NILs carrying the positive 5A allele (5A+ NILs) have significantly increased thousand grain weight (TGW; 7%), grain length (4%) and pericarp cell length (10%) compared to NILs carrying the negative 5A allele (5A- NILs) [29]. The NILs carry an introgressed segment of ~490 Mb and using recombinant inbred lines we fine-mapped the grain length effect to a 75 Mb region on the long arm of chromosome 5A according to the IWGSC RefSeq v1.0. The aim of the present study was to identify biological pathways that potentially influence grain length and pericarp cell size by using RNA-seq to identify genes that are differentially regulated between the 5A- and 5A+ NILs.

Results

RNA-sequencing of 5A near isogenic lines

We performed RNA-seq on whole grains from two 5A NILs which contrast for grain length based on the 2014 growing season [29]. We chose the time point when NILs show the first significant differences in grain length (8 days post anthesis (dpa); T2) and the preceding time point (4 dpa; T1) to capture differences in gene expression occurring during this period (Fig. 1). We hypothesised that although there was no significant difference in the grain length phenotype at T1, phenotypic differences were beginning to emerge and gene expression changes influencing this may already be occurring. We obtained over 362 M reads across all 12 samples (two time points, two NILs, three biological replicates), with individual samples ranging from 15.0 M to 53.6 M reads and an average of 30.2 M reads (standard error \pm 3.5 M reads) per sample (Table 1).

Comparison between Chinese spring reference transcriptomes

We aligned reads to two different transcriptome sequences from the Chinese Spring reference accession, the IWGSC Chromosome Survey Sequence (CSS) [40] and TGACv1 (TGAC) [41] reference. On average across samples, $69.8 \pm 0.3\%$ of reads aligned to the CSS reference, whilst $84.4 \pm 0.2\%$ of reads aligned to the TGAC reference.

We defined a transcript as expressed if it had an average abundance of >0.5 transcripts per million (tpm) in at least one of the four conditions (2 NILs \times 2 time points). This resulted in 64,020 and 101,652 transcripts being expressed in the CSS and TGAC transcriptomes, respectively. We defined differentially expressed (DE) transcripts (q value <0.05) using sleuth [45] and performed four pairwise comparisons: two 'across time' and two 'between NIL' comparisons. The 'across time' analyses consisted of a comparison between T1 and T2 samples of the 5A- NIL (hereafter symbolised as 5A-T1; Fig. 1, orange)

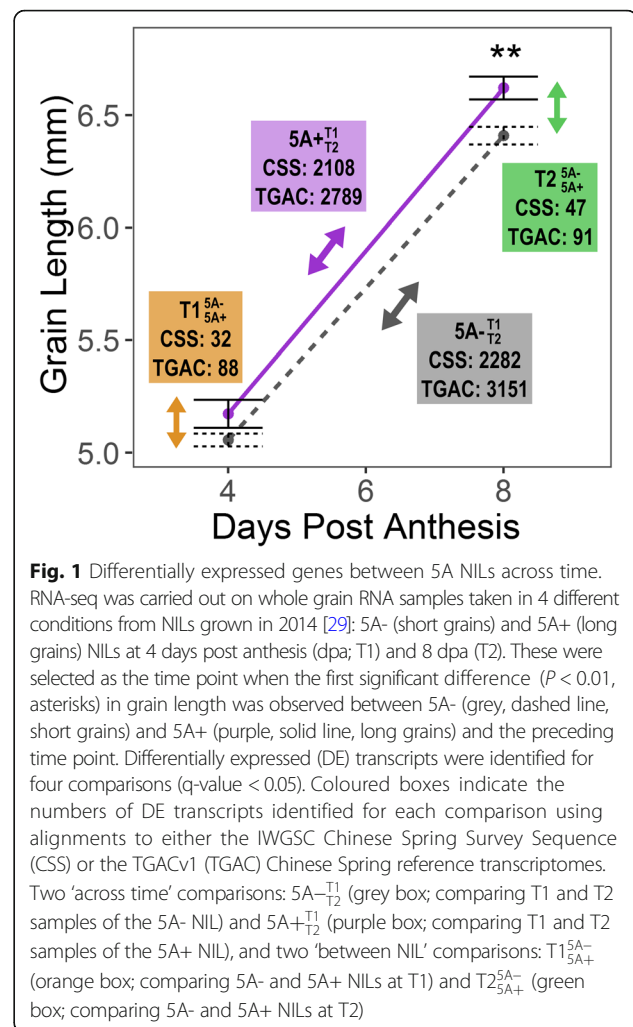


Fig. 1 Differentially expressed genes between 5A NILs across time. RNA-seq was carried out on whole grain RNA samples taken in 4 different conditions from NILs grown in 2014 [29]: 5A- (short grains) and 5A+ (long grains) NILs at 4 days post anthesis (dpa; T1) and 8 dpa (T2). These were selected as the time point when the first significant difference ($P < 0.01$, asterisks) in grain length was observed between 5A- (grey, dashed line, short grains) and 5A+ (purple, solid line, long grains) and the preceding time point. Differentially expressed (DE) transcripts were identified for four comparisons (q-value <0.05). Coloured boxes indicate the numbers of DE transcripts identified for each comparison using alignments to either the IWGSC Chinese Spring Survey Sequence (CSS) or the TGACv1 (TGAC) Chinese Spring reference transcriptomes. Two 'across time' comparisons: 5A-T1 (grey box; comparing T1 and T2 samples of the 5A- NIL) and 5A+T1 (purple box; comparing T1 and T2 samples of the 5A+ NIL), and two 'between NIL' comparisons: T1 5A- vs 5A+ (orange box; comparing 5A- and 5A+ NILs at T1) and T2 5A- vs 5A+ (green box; comparing 5A- and 5A+ NILs at T2)

Fig. 1, grey) and the corresponding comparison for the 5A+ NIL samples (hereafter 5A+T1; Fig. 1, purple). In both 'across time' analyses, transcripts were considered as upregulated or downregulated with respect to T1 i.e. increasing or decreasing in expression across time. The 'between NIL' analyses consisted of a comparison between the 5A- and 5A+ NILs at T1 (hereafter T1 5A- vs 5A+; Fig. 1, orange), and a comparison between the 5A- and 5A+ NILs at T2 (hereafter T2 5A- vs 5A+; Fig. 1, green). In all cases, more DE transcripts were identified in the TGAC compared with the CSS transcriptome, and similar trends were observed for both references across the four comparisons (Fig. 1).

We selected the comparison with the fewest DE transcripts (T1 5A- vs 5A+; 32 and 88 DE transcripts for CSS and TGAC, respectively) to conduct a more in depth analysis of the alignments and references. For all DE transcripts from each alignment we identified the equivalent transcript/gene model in the other reference sequence using *Ensembl* plants release 35 and compared the gene

Table 1 Mapping summary of RNA-seq samples

Genotype	Time point	Replicate	Reads	CSS gene models		TGAC gene models	
				Reads pseudoaligned	% reads pseudoaligned	Reads pseudoaligned	% reads pseudoaligned
5A -	1	1	24,443,658	17,072,939	69.85	20,549,681	84.07
5A -	1	2	34,441,799	23,349,288	67.79	28,483,090	82.70
5A -	1	3	23,462,705	16,220,597	69.13	19,664,859	83.81
5A -	2	1	21,333,672	14,839,724	69.56	18,052,324	84.62
5A -	2	2	14,967,302	10,632,519	71.04	12,803,552	85.54
5A -	2	3	35,522,754	25,491,523	71.76	30,297,336	85.29
5A +	1	1	19,267,564	13,520,181	70.17	16,317,352	84.69
5A +	1	2	22,299,102	15,479,234	69.42	18,780,525	84.22
5A +	1	3	30,531,539	20,789,582	68.09	25,436,453	83.31
5A +	2	1	51,637,607	36,192,489	70.09	43,739,451	84.70
5A +	2	2	53,575,232	37,956,887	70.85	45,497,914	84.92
5A +	2	3	30,553,421	21,604,895	70.71	25,984,674	85.05
		Total	362,036,355	253,149,858	–	305,607,211	–
		Mean	30,169,696	21,095,822	69.87	25,467,268	84.41

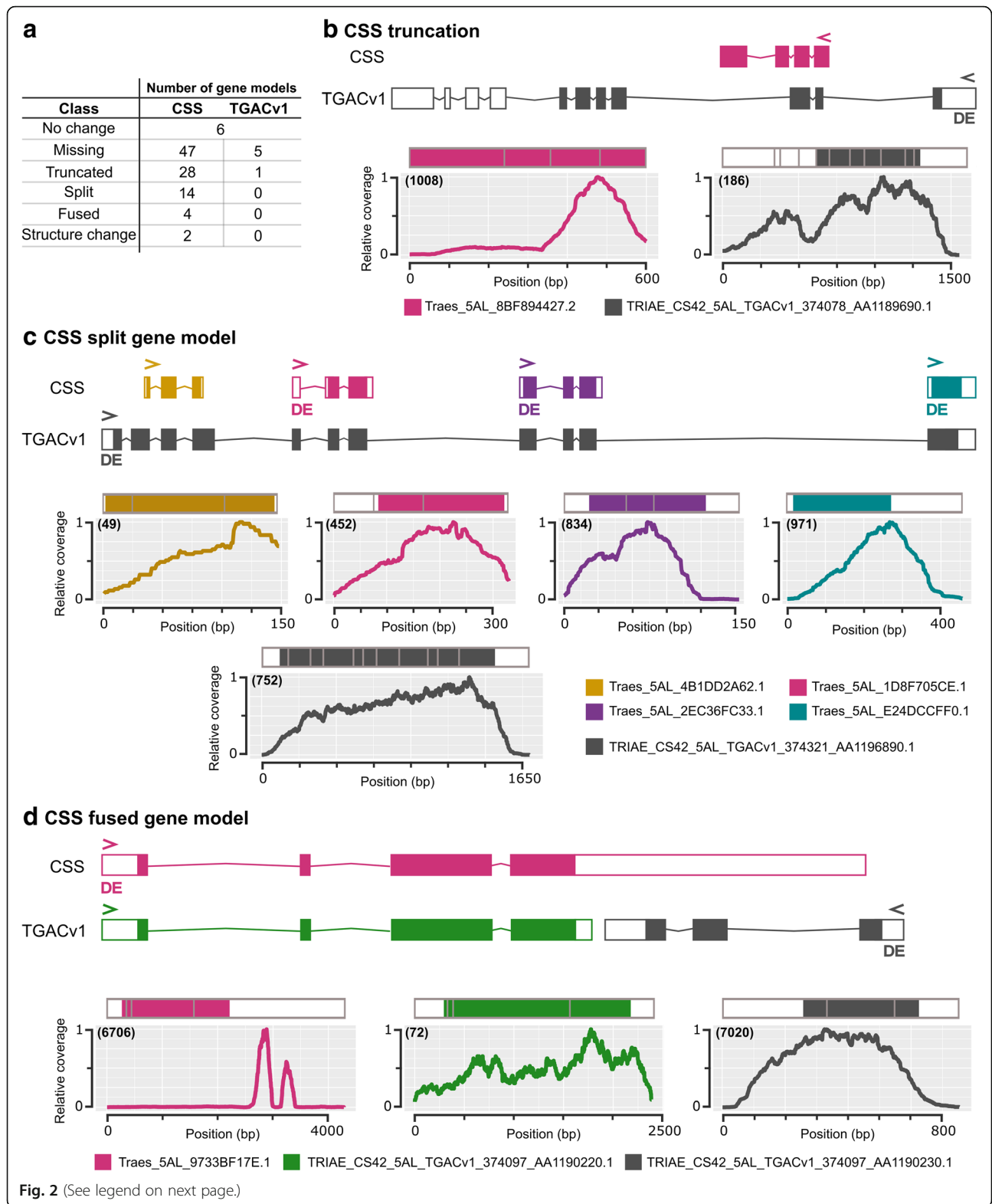
models (Additional file 1). For 64 of the TGAC DE transcripts we did not identify an equivalent CSS DE transcript, either because there was no corresponding CSS gene model (47 transcripts) or the expression change between NILs was non-significant for the CSS transcript. Analogously, eleven CSS DE transcripts did not have an equivalent TGAC gene model DE, five of which were due to there being no corresponding TGAC gene model annotated. Combining both sets, we identified 42 groups of equivalent gene models, 26 of which were differentially expressed in both alignments. Comparing these 42 groups and taking into account fused and split gene models within each dataset, there were a total of 97 gene models across both datasets (50 CSS + 47 TGAC) (Fig. 2a, Additional file 1). Of these, only six were identical between the CSS and TGAC references. All other discrepant gene models fell under categories included truncations in either reference, gene models that were split/fused in one reference sequence, and gene models that differed drastically in their overall structure.

For all discrepant gene models we used transcriptome read mapping and an interspecies comparison to determine which gene model seemed most plausible. Fig. 2b shows an example of the most commonly identified discrepancy where a gene model was truncated in the CSS reference (pink) relative to the TGAC reference (grey). The DE TGAC gene model was supported by our transcriptome data as we observed read coverage across the whole gene model whilst the coverage across the CSS gene model dropped at the position where an intron is predicted in the TGAC model. Another common discrepancy was a single gene model in one reference being split into multiple gene models in the other reference.

Fig. 2c shows an instance where a single DE TGAC gene model comprised four separate CSS gene models. In this case, all five gene models had coverage across the entire gene body, however the single TGAC gene model was more similar to proteins from other species, suggesting that this single gene model was most likely correct. The final example (Fig. 2d) shows two TGAC gene models that were fused into a single CSS gene model. The coverage across the CSS gene model was inconsistent, with most reads concentrated in the 3' untranslated region (UTR). The two TGAC gene models had more consistent coverage across the entire gene models and were both supported by protein alignments with other species. Interestingly, only the shorter TGAC gene model was DE (Fig. 2d, grey), suggesting that differential expression of the CSS gene model was driven by the reads mapping to the putative 3' UTR rather than the coding regions of the transcript (Fig. 2d, pink). Taking together the fact that a higher percentage of reads mapped to the TGAC gene models and that many more of the examined TGAC gene models were supported by interspecies comparison and expression data than the CSS gene models, we decided to continue our analysis using the alignments to the TGAC gene models only.

Many DE transcripts during early grain development are shared between NILs

We identified 3151 and 2789 DE transcripts across early grain development in 5A^{-T1} and 5A^{+T1}, respectively (Fig. 1, Fig. 3a). The DE transcripts were evenly distributed across the 21 chromosomes, showing no overall bias towards any chromosome group or subgenome

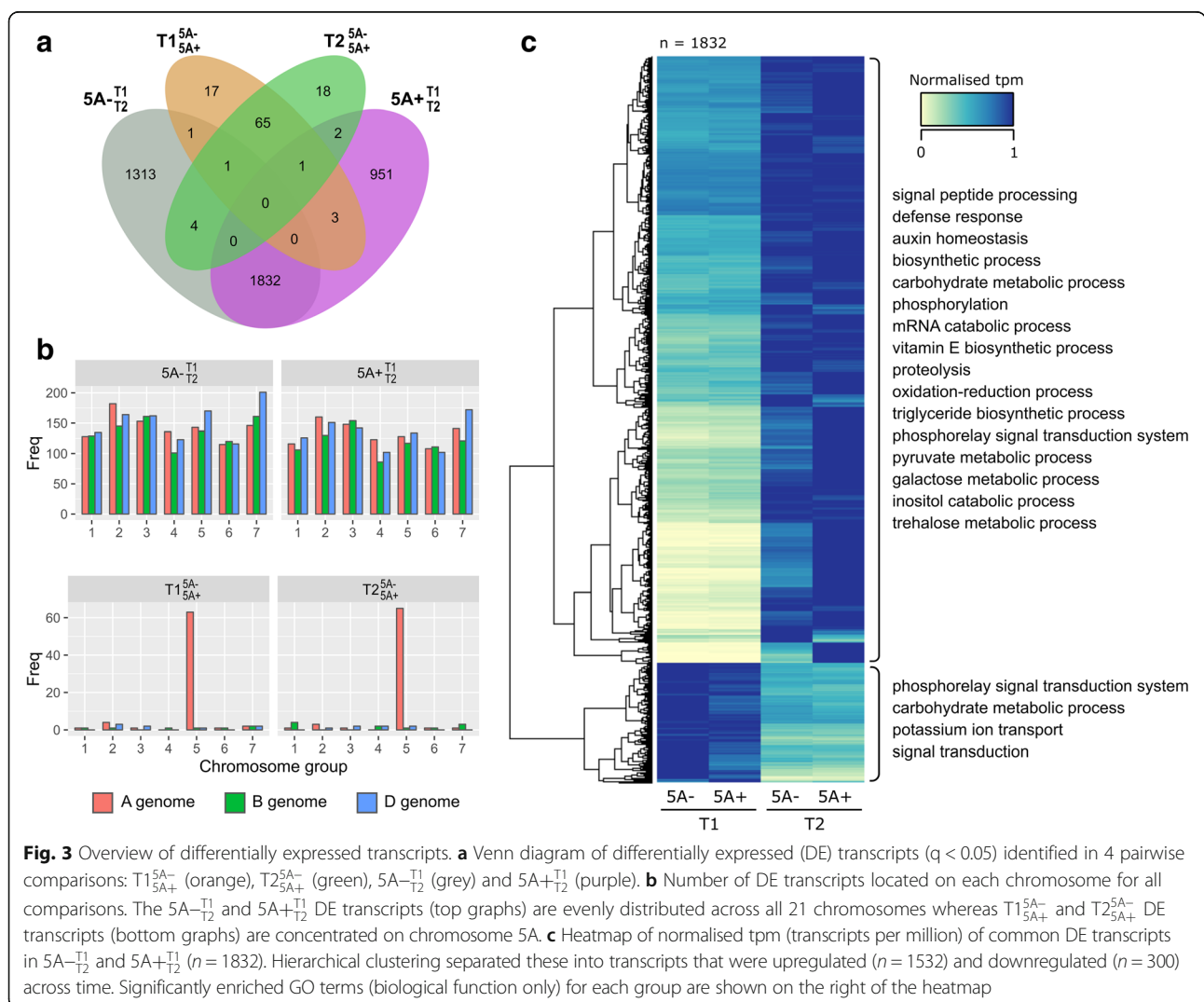


(See figure on previous page.)

Fig. 2 Comparison between CSS and TGACv1 gene models. **a** Discrepancies identified between gene models in the CSS and TGAC reference sequences and the number of gene models falling into categories. Panels **(b)**, **(c)** and **(d)** show specific examples of discrepancies. In each panel, a representation of the unspliced gene model is shown with exons as coloured boxes, untranslated regions as white boxes, and introns as thin lines. Graphs show the relative read coverage across the spliced transcript with the structure represented diagrammatically directly above each graph. The number in brackets shows the maximum absolute read depth for each gene model. > and < in the gene structures indicate the direction of transcription and a 'DE' indicates that the gene model was differentially expressed in $T1_{5A-}^{5A+}$ (q value < 0.05). For each panel transcript names are shown in the coloured legends

(Fig. 3b). Approximately 60% (1832) of the DE transcripts were shared between $5A-T1_{T2}$ and $5A+T1_{T2}$ (Fig. 3a) and 84% (1532) of the shared transcripts were upregulated across time (Fig. 3c). We identified 41 significantly enriched gene ontology (GO) terms in the upregulated transcripts (Additional file 2). Sixteen of the GO terms were associated with biological process and could be grouped under three parent GO terms: metabolic process (GO:0008152), defence response (GO:0006952) and biological regulation (GO:0065007) (Fig. 3b). Within

metabolic process we found terms associated with carbohydrate (GO:0005975) and pyruvate metabolism (GO:0006090), vitamin E (GO:0010189) and triglyceride biosynthesis (GO:0019432), mRNA catabolism (GO:0006402), proteolysis (GO:0006508) and phosphorylation (GO:0016310). Downregulated transcripts (300) were enriched for seven GO terms, four of which were associated with biological process: potassium ion transport (GO:0006813), signal transduction (GO:0007165), phosphorelay signal transduction (GO:0000160) and



carbohydrate metabolism (GO:0005975). The overlap between enriched GO terms in the upregulated and downregulated transcripts (e.g. carbohydrate metabolism) suggests that different aspects of these processes are being differentially regulated during this early grain development stage.

We also identified many transcripts that were only DE across early grain development in one of the two genotypes (i.e. unique to either the $5A-T1$ or $5A+T1$ comparisons). However, many of these transcripts were borderline non-significant in the opposite genotype comparison illustrated by the fact that the distributions of q-values were skewed towards significance (Additional file 3). Additionally, the uniquely DE transcripts were enriched for GO terms similar to the shared transcripts (Additional file 2). Some GO terms, however, were only enriched in the uniquely DE transcripts, for example, cell wall organisation or biosynthesis (GO:0071554) and response to abiotic stimulus (GO:0009628). Overall, these results suggest that although there were some differences between genotypes, broadly similar biological processes were taking place in the grains of both the 5A NILs at the early stages of grain development.

DE transcripts between NILs are concentrated on chromosome 5A

We identified 88 and 91 DE transcripts between the NILs in $T1_{5A+}$ and $T2_{5A+}$, respectively, many fewer than identified in $5A-T1$ or $5A+T1$. This was expected as the NILs are genetically very similar and therefore the difference in developmental stage between the T1 and T2 time points results in greater changes in gene expression. For a subset of the DE transcripts, we confirmed expression changes using qRT-PCR (see [Methods](#) for details). Of the 179 DE transcripts, 67 were common between $T1_{5A+}$ and $T2_{5A+}$, whereas 45 DE transcripts between genotypes were unique and identified only at a single time point (resulting in 112 DE transcripts between NILs at any time point; Fig. 3a). No GO terms were significantly enriched in these groups. Of the 67 common DE transcripts, 54 (80%) were located on chromosome 5A, whilst in both the T1 and T2 unique groups less than 50% were located on chromosome 5A (Fig. 4a). Similar numbers of DE transcripts were more highly expressed in either genotype, with no distinct patterns observed between the unique or common groups.

We looked specifically at the positions of the 74 DE transcripts located on chromosome 5A and found that all were located within the 491 Mbp introgressed region of the NILs (Fig. 4b). Higher numbers of DE transcripts were identified in regions of increased SNP density between the 5A NILs. Previously, we fine-mapped the grain length effect to a 75 Mbp interval on 5AL

(between BS00182017 (317 Mbp) and BA00228977 (392 Mbp; [29])) and eight of the DE transcripts were located within this interval. Three of these transcripts were more highly expressed in the 5A+ NILs ($5A+_{high}$ transcripts), two of which were transcript variants of the same gene (a kinesin-like protein; only .2 variant shown in Fig. 4b). The other $5A+_{high}$ transcript was annotated as a putative retrotransposon protein. One of the five transcripts more highly expressed in the 5A- NIL ($5A-_{high}$ transcript) had no annotation and the remaining four were annotated as a non-coding RNA, a RING/U-box containing protein, a TauE-like protein and a DUF810 family protein.

DE transcripts outside of chromosome 5A are enriched in specific transcription factor binding sites

As all the DE transcripts on chromosome 5A were located within the 491 Mbp introgressed region, it is possible that the differential expression was a direct consequence of sequence variation between the NILs e.g. in the promoter regions. However, the 38 DE transcripts located outside of chromosome 5A have the same nucleotide sequence as they are identical by descent (BC₄ NILs confirmed with 90 k iSelect SNP marker data [29]). These 38 DE transcripts are unlikely to represent false positives (see [Methods](#)), so we hypothesised that these transcripts are downstream targets of DE genes, such as transcription factors (TFs), located within the 5A introgression.

To assess this, we identified transcription factor binding sites (TFBS) present in the promoter regions of these 38 DE transcripts. We identified TFBS associated with 91 distinct TF families present in this group of transcripts (Additional file 4), five of which were enriched relative to all expressed transcripts (Table 2; FDR adjusted $P < 0.05$). The enriched TFBS families were C2H2, Myb/SANT, AT-Hook, YABBY and MADF/Trihelix.

To determine potential candidates for upstream regulators we identified all TFs located within the introgressed region on chromosome 5A [46]. We identified a total of 200 annotated TFs, belonging to 35 TF families. Of these, four families (across 29 genes) overlapped with enriched TFBS families. Four of the 29 TFs were located within the fine-mapped grain length region on chromosome 5A, including C2H2, MYB and MYB_related TFs (Additional file 5).

Functional annotation of DE transcripts

Having analysed DE transcripts between NILs based on chromosome location, we looked at the 112 DE transcripts based on their functional annotations (Additional file 6). We identified multiple categories including transcripts associated with ubiquitin-mediated protein degradation, cell cycle, metabolism, transport,

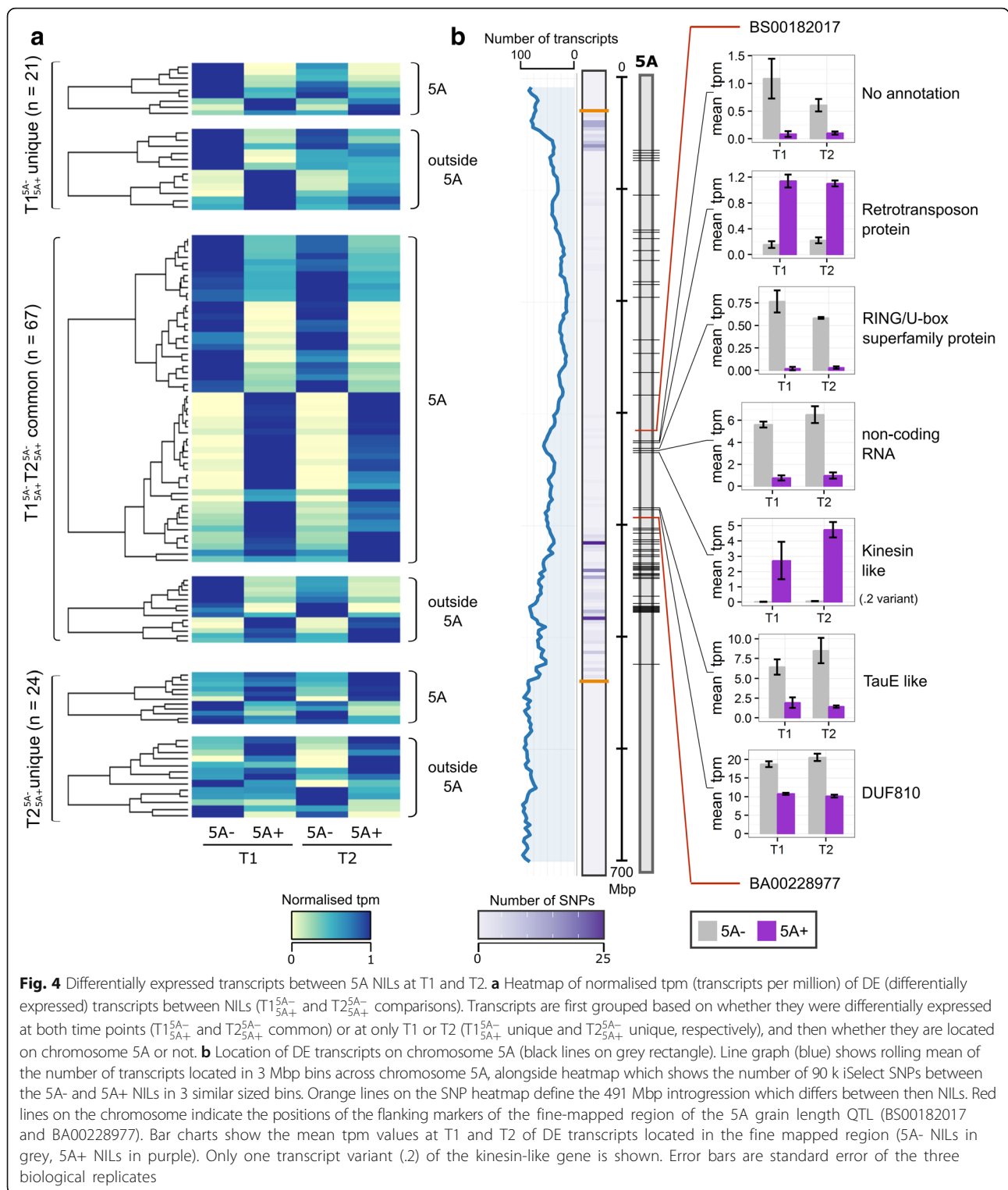


Fig. 4 Differentially expressed transcripts between 5A NILs at T1 and T2. **a** Heatmap of normalised tpm (transcripts per million) of DE (differentially expressed) transcripts between NILs ($T1_{5A-}^{5A+}$ and $T2_{5A-}^{5A+}$ comparisons). Transcripts are first grouped based on whether they were differentially expressed at both time points ($T1_{5A-}^{5A+}$ and $T2_{5A-}^{5A+}$ common) or at only T1 or T2 ($T1_{5A-}^{5A+}$ unique and $T2_{5A-}^{5A+}$ unique, respectively), and then whether they are located on chromosome 5A or not. **b** Location of DE transcripts on chromosome 5A (black lines on grey rectangle). Line graph (blue) shows rolling mean of the number of transcripts located in 3 Mbp bins across chromosome 5A, alongside heatmap which shows the number of 90 k iSelect SNPs between the 5A- and 5A+ NILs in 3 similar sized bins. Orange lines on the SNP heatmap define the 491 Mbp introgression which differs between then NILs. Red lines on the chromosome indicate the positions of the flanking markers of the fine-mapped region of the 5A grain length QTL (BS00182017 and BA00228977). Bar charts show the mean tpm values at T1 and T2 of DE transcripts located in the fine mapped region (5A- NILs in grey, 5A+ NILs in purple). Only one transcript variant (.2) of the kinesin-like gene is shown. Error bars are standard error of the three biological replicates

transposons and non-coding RNAs (Table 3). Few categories were exclusively located on/outside 5A or had exclusively higher expression in the either the 5A- or 5A+ NIL. The category with the most DE transcripts was non-coding RNA (ncRNA, 15 transcripts), although this was

not enriched relative to all expressed transcripts. All ncRNA transcripts were classed as long non-coding RNAs (> 200 bp, [47]) and we found that four of the ncRNAs overlapped with coding transcripts (two in the antisense direction) and one ncRNA was a putative

Table 2 Enriched transcription factor binding sites in promoters of DE transcripts located outside of 5A

TF family	Observed in all expressed transcripts (n = 101,653)	Expected in outside 5A DE transcript (n = 38)	Observed in outside 5A DE transcripts (n = 38)	FDR adjusted <i>p</i> -value
C2H2	77,987	29	36	0.021
Myb/SANT	88,575	33	38	0.021
AT-Hook	90,203	34	38	0.028
YABBY	19,447	7	15	0.034
MADF;Trihelix	16,632	6	13	0.042

Values are the number of transcripts in which binding sites associated with the specified transcription factor (TF) family are present

miRNA precursor (Ta-miR132-3p; [48]). We identified 13 transcripts as putative targets of Ta-miR132-3p in the TGAC reference but none of these target transcripts were differentially expressed in our dataset. The second largest transcript category was transposon-associated (14 transcripts; FDR-adjusted $p = 0.008$), whereas the third largest category was DE transcripts related to ubiquitin and the proteasome (12 transcripts; $p = 0.008$). DE transcripts annotated as homeobox were also enriched (4 transcripts; FDR-adjusted $p = 0.001$). Interestingly, we identified homeodomain TFBS in 27 of the 38 outside 5A DE transcripts although this was not significantly enriched (FDR-adjusted $p = 0.166$, Additional file 4).

Table 3 Categories of DE transcripts between NILs based on predicted function

Category	number of transcripts	Adjusted <i>p</i> -value	5A/not 5A	NIL with higher expression: 5A-/5A+
non-coding RNA	15	0.141	10/5	6/9
transposon-associated	14	0.008	4/10	5/9
ubiquitin	12**	0.008	10/2	8/4
cell cycle	5	-	5/0	2/3
histone-related	5	-	3/2	3/2
heat shock	5	-	3/2	2/3
protease	4	-	3/1	3/1
transport	4	-	3/1	2/2
metabolism	5	-	5/0	4/1
homeobox	4	0.001	3/1	1/3
cell wall	3	-	2/1	2/1
transcription	3	-	2/1	0/3
non-translating	2	-	0/2	1/1
peroxisome	2	-	0/2	0/2
other*	20	-	14/6	11/9
No annotation	8	-	4/4	5/3

Adjusted *p*-values displayed are based on an enrichment test of the functional categories relative to all expressed transcripts. - indicates that an enrichment test was not performed as categories were based on bespoke annotations

*includes transcripts with annotations that could not be grouped by function with other transcripts

**only the 7 transcripts that were annotated as ubiquitin-related in the TGAC annotation were used in the enrichment test (see [methods](#))

The DE transcripts related to ubiquitin were of particular interest as ubiquitin-mediated protein turnover has previously been associated with the control of seed/grain size in wheat [13] and other species including rice and Arabidopsis [12, 14, 49]. The pathway acts through the sequential action of a cascade of enzymes (see Fig. 5a legend) to add multiple copies of the protein ubiquitin (ub) to a substrate protein that is then targeted for degradation by the proteasome. We identified differential expression of transcripts at almost all steps of this pathway (excluding E1): two ubiquitin proteins and one ubiquitin-like protein, one E2 conjugase, six potential E3 ligase components and two putative components of the proteasome (Fig. 5). In addition to these, we also identified four DE transcripts annotated as proteases (Fig. 5), which are known substrates regulated by this pathway [50–52] and that influence organ size through the regulation of cell proliferation. Most of the components of the ubiquitin pathway that were differentially expressed were more highly expressed in the 5A- NIL (11/16, including proteases) (Fig. 5b).

Discussion

In this study, we performed RNA-seq on the grains of 5A NILs with a known difference in pericarp cell size, grain length and final grain weight. We previously determined that the first phenotypic differences between NILs arose during early grain development [29]. We hypothesised that differences in gene expression between NILs during these early stages would allow us to identify specific genes and pathways that affect pericarp cell size and grain size at the transcriptional level.

The importance of a high-quality reference sequence

We initially mapped the RNA-seq data to two different reference transcriptomes: CSS and TGAC. We found that TGAC outperformed the CSS transcriptome both in terms of the number of reads that aligned and in the gene models themselves. This was most likely due to the significant improvement in terms of sequence contiguity of the TGAC reference over the CSS (N50 = 88.8 vs < 10 kb, respectively), allowing more accurate prediction of gene models. Our study highlights the practical

This exemplifies the importance of correctly annotated gene models and improved genome assemblies in gaining a more accurate view of the underlying biology.

Differential expression analysis provides an insight into the biological processes occurring in early grain development

We sampled grains at 4 and 8 dpa to encompass the developmental stage at which the first significant difference in grain length between 5A NILs is observed. During this stage, increases in grain size are largely driven by cell expansion in the pericarp [54, 55], consistent with our previous finding that increased pericarp cell size underlies the difference in final grain length. These time points are also relatively early compared to other grain related RNA-seq studies which have focused on later grain filling processes [34–36]. The ‘across time’ comparisons ($5A-T1_{T2}$ and $5A+T1_{T2}$) identified >2700 DE transcripts in each NIL, and there was a large overlap in the biological processes being differentially regulated. We found that most DE transcripts were upregulated over time and many of these were associated with metabolism and biosynthesis consistent with grains undergoing a period of rapid growth and the start of endosperm cellularisation at this stage of development [32]. Transcripts associated with proteolysis and mRNA catabolism were also upregulated across time consistent with increases in specific proteases and other hydrolytic enzymes at this stage of grain development [56]. These could be indicative of programmed cell death which occurs in both the nucellus and pericarp of the developing grain up to 12 dpa [54]. We also identified an upregulation of transcripts associated with defence response and oxidation-reduction process, consistent with previous reports of accumulation of proteins associated with defence against both pathogens and oxidative stress during the early-mid stages of grain development [57]. Transcriptional studies always have the caveat that changes in gene expression may not translate to changes in protein level [58]. However, proteomic analyses of similar stages of grain development have identified the differential regulation of similar ontologies [57, 59] suggesting that these transcriptional changes are reflective of overall protein status in the grain.

Comparative transcriptomics as a method to identify candidate genes underlying QTL

The relative ease and low-cost of performing RNA-seq on highly isogenic material compared with high resolution genetic mapping makes it an attractive method for identifying candidate genes underlying QTL and better understanding the associated traits. This is particularly true for species with large genomes, such as wheat, for which the cost of whole genome sequencing remains

high. A potential drawback of transcriptomics approaches, however, is the large scale gene expression changes often observed when comparing different varieties or developmental stages. The use of highly isogenic lines can be used to address this by defining very specific chromosome intervals which differ between lines. The use of the 5A NILs in our study exemplified this by allowing the direct comparison of the effect of the 5A introgression on gene expression at two time points during grain development ($T1_{5A+}^{5A-}$ and $T2_{5A+}^{5A-}$). This resulted in a defined set of 112 DE transcripts between genotypes. The majority of $T1_{5A+}^{5A-}$ and $T2_{5A+}^{5A-}$ DE transcripts were located on chromosome 5A and all of these were located within the 5A introgression. This is expected given that the sequence variation in the NILs was restricted to the chromosome 5A region.

DE transcripts located within the fine-mapped interval on chromosome 5A represent good candidates for further characterisation. The kinesin-like gene and RING/U-box superfamily protein are particularly strong candidates based on their functional annotations. Previous studies have demonstrated that Kinesin-like proteins can regulate grain length and cell expansion through involvement with microtubule dynamics [15, 16, 60]. The RING/U-box protein is a putative E3 ligase, a class of enzymes which have been associated with the control of grain size (discussed in more detail later; [12, 13]).

However, caution should be exercised when speculating on the identity of a causal gene(s) based solely on differential expression of transcripts. In the case of the 5A NILs, it is difficult to predict whether DE transcripts in the fine-mapped interval are truly associated with the effect of the 5A QTL or are simply a consequence of sequence variations between the parental cultivars, i.e. ‘guilt by association’. A relevant example was the recent use of transcriptomics to define a candidate gene underlying a grain dormancy QTL (*PM19*) [44]. Subsequent studies showed that a different gene in close physical proximity (*TaMKK3*) [61] was responsible for the natural variation observed [62]. The mis-interpretation of the transcriptomics data was due to complete linkage disequilibrium between the DE *PM19* gene and the causal *TaMKK3* gene in the germplasm used in the original study.

An additional consideration is that the causal gene(s) underlying a QTL may not be differentially expressed. The phenotypic effects of the QTL could be a result of allelic variation that alters the function of the underlying gene independent of expression level. Indeed, coding region sequence polymorphisms can be identified using RNA-seq data [63]. These polymorphism can be converted into genetic markers to further fine map the gene, although we did not pursue this strategy in our study.

It could be argued that resources would be better focussed on fine-mapping and cloning the gene underlying the QTL before performing further studies using methods such as RNA-seq. Indeed, direct manipulation of a causal gene through TILLING or gene editing allows precise interrogation of gene function and how it relates to the trait of interest (reviewed in [64]). However, aside from the identification of the causal gene itself, we would argue that comparative transcriptomics using highly isogenic material is a valuable method for understanding the underlying biology of the trait. This is becoming increasingly commonplace with the relatively low cost of performing RNA-seq (two sequencing lanes in this study), the improved genome assemblies now available in multiple species [65], and the improved software (e.g. [66, 67]) which facilitate downstream analyses of data. Combining insight from transcriptomic studies with prior fine-mapping and phenotypic characterisation allows the selection of candidates for functional characterisation, which can proceed alongside further fine-mapping of the QTL. In the case of this study, genes found to be associated with differences in grain size components would be interesting novel candidates for grain size control in wheat, regardless of whether they are the causal gene underlying the 5A QTL. The case of *PM19* again provides a relevant example, as direct manipulation of *PM19* in transgenic lines resulted in a grain dormancy phenotype despite the fact it was not the causal gene underlying the grain dormancy QTL in question [44, 62].

Ultimately, in the case of the 5A QTL, further fine-mapping of the locus will be required to identify the underlying gene. However, we would consider the use of transcriptomics to be a complementary approach to fine-mapping both by providing additional SNPs and insight into the biological mechanisms underlying the QTL trait.

DE transcripts outside chromosome 5A are candidates for downstream targets of the 5A QTL

We considered DE transcripts outside of chromosome 5A as candidates for downstream targets of genes located in the 5A introgression because the differential expression could not have arisen through sequence variation. These included genes located on all three genomes implying that there is cross-talk at the transcriptional level between the A, B and D genomes. We identified, in the promoters of these genes, enrichment of TF binding sites associated with TF families which have all previously been shown to play diverse roles in the control of organ development [68, 69]. For example YABBY genes, a plant specific family of TFs, play a critical role in patterning and the establishment of organ polarity [70] and fruit size [71]. Another example are the

C2H2 TFs, *NUBBIN* and *JAGGED*, which are involved in determining carpel shape in Arabidopsis [72]. AT-Hook TFs play roles in floral organ development in both maize and rice [73, 74] and modulate cell elongation in the Arabidopsis hypocotyl [75]. Few of these transcription factor families have been characterised in wheat, and although these interactions need to be experimentally validated, they could be potential targets for the manipulation of grain size.

DE transcripts have functions related to the control of seed/organ size

Studies in species such as rice and Arabidopsis have shown that seed size is regulated by a complex network of genes and diverse mechanisms, ultimately through the coordination of cell proliferation and expansion (reviewed in [8, 9]). 5A+ NILs have significantly longer pericarp cells, suggesting that the underlying gene influences cell expansion [29]. Genes that physically modify the cell wall have been shown to directly control cell expansion (reviewed in [76]) and we identified three DE transcripts that have potential roles in cell wall synthesis and remodelling. We also identified a number of DE transcripts associated with the cell cycle and the control of cell proliferation. During seed development, a number of cell cycle types in addition to the typical mitotic cycle are observed. One such alternative cycle type is endoreduplication, characterised by the replication of chromosomes in the absence of cell division, which is associated with cell enlargement (reviewed in [77]). Two of the DE transcripts were the closest wheat orthologues of Arabidopsis genes that have specific roles in organ development: a *GRF*-interacting factor (*GIF*) and *SEUSS* (*SEU*). In Arabidopsis, the *GIF* genes interact with the *GROWTH-REGULATING FACTOR* (*GRF*) TFs and act as transcriptional co-activators to regulate organ size through cell proliferation [78]. Conversely, *SEU* acts a transcriptional co-repressor and interacts with important regulators of development to control many processes, including floral organ development [79].

Seed development requires the coordination of processes across multiple tissues, namely the seed coat, endosperm and embryo. The development and growth of these tissues is inherently interlinked, and it has been proposed that the mechanical constraint imposed by the maternal seed coat/pericarp places an upper limit on the size of the seed/grain [29, 30, 80]. Epigenetic regulation appears to play an important role in the cross talk and coordination of these tissues [81]. The differential expression of 34 non-coding transcripts, transposons and histone-related transcripts between NILs could suggest a difference in epigenetic status associated with the control of pericarp cell size. Additional work to characterise these non-coding RNAs would be warranted to establish their role in grain development.

The ubiquitin-mediated control of seed/grain size has been documented in a number of species (reviewed in [82]), including wheat [13, 83]. DE transcripts associated with the ubiquitin pathway were significantly enriched in the 5A NILs, although only one was located within the fine-mapped interval. The pathway tags substrate proteins with multiple copies of the ubiquitin protein through the sequential action of a cascade of enzymes: E1 (Ub activating), E2 (Ub conjugases) and E3 (Ub ligases). The ubiquitinated substrate proteins are then targeted to the 26S proteasome for degradation [84]. *GW2*, a RING-type E3 ligase, negatively regulates cell division and was identified as the causal gene underlying a QTL for grain width and weight in rice [12]. The Arabidopsis orthologue, *DA2*, acts via the same mechanism to regulate seed size in Arabidopsis [14]. Another E3 ligase, *EOD1/BB* also negatively regulates seed size in Arabidopsis [49]. In general, the E3 ligase determines the specificity for the substrate proteins [84] and *DA2* and *EOD1* may have different substrate targets, however they converge and both target the ubiquitin-activated protease *DA1*. *DA1* also negatively regulates cell proliferation and acts synergistically with both *DA2* and *EOD1*, although it is not clear whether the two E3 ligases act via independent genetic pathways or as part of the same mechanism [14, 50, 85]. *UBP15* (a ubiquitin specific protease) is a downstream target of this pathway and conversely acts as a positive regulator of seed size through the promotion of cell proliferation [51]. Other ubiquitin-associated regulators of organ/grain size have been identified, including components of the 26S proteasome, enzymes with deubiquitinating activity and proteins that have been shown to bind ubiquitin in vitro [52, 86, 87]. The DE transcripts associated with this pathway are not direct homologs of these previously characterised genes. As such the functional characterisation of these putative novel components could provide new insights into the ubiquitin-mediated control of grain size in cereals.

Conclusions

In this study we have both generated candidates for the causal gene underlying the 5A QTL, and have also identified potential downstream pathways controlling grain size. A subset of these candidates is being tested functionally using TILLING mutants [88] and other approaches to provide novel insights into the control of grain size in cereals. Ultimately identifying the individual components and pathways that regulate grain size and understanding how they interact will allow us to more accurately manipulate final grain yields in wheat.

Methods

Plant material

The 5A BC₄ NILs used in this study have been described previously [29]. Briefly, the NILs were generated from a

doubled haploid population between the UK cultivars 'Charger' and 'Badger' using Charger as the recurrent parent. The NILs differ for an approximately 491 Mbp interval on chromosome 5A. We used one genotype each for the 5A- (Charger allele, short grains) and 5A+ NIL (Badger allele, long grains). Plants were grown in 1.1 × 6 m plots (experimental units) during the 2014 growing season in a complete randomised block design with five replications, and spikes were tagged at full ear emergence [29]. The three blocks with the most similar flowering time were used for sampling. Plants were sampled at 4 (time point 1: T1) and 8 (time point 2: T2) days post anthesis (dpa) based on the 2014 developmental time course outlined in [29]. For each genotype, we sampled three grains from three separate spikes from different plants within the experimental unit. Each biological replicate, therefore, consisted of the pooling of nine grains per genotype. Grains were sampled from the outer florets (positions F1 and F2) from the middle section of each of the three spikes. Grains were removed from the spikes in the field, immediately frozen in liquid nitrogen and stored at -80 °C. In total, three biological replicates (from the three blocks in the field) were sampled for each NIL at each time point.

RNA extraction and sequencing

For each biological replicate, the nine grains were pooled and ground together under liquid nitrogen. RNA was extracted in RE buffer (0.1 M Tris pH 8.0, 5 mM EDTA pH 8.0, 0.1 M NaCl, 0.5% SDS, 1% β-mercaptoethanol) with Ambion Plant RNA Isolation Aid (Thermo Fisher Scientific). The supernatant was extracted with 1:1 acidic Phenol (pH 4.3):Chloroform. RNA was precipitated at -80 °C by addition of Isopropanol and 3 M NA Acetate (pH 5.2). The RNA pellet was washed twice in 70% Ethanol and resuspended in RNase free water. RNA was DNase treated and purified using RNeasy Plant Mini kit (Qiagen) according to the manufacturer's instructions. RNA QC, library construction and sequencing were performed by the Earlham Institute, Norwich (formerly The Genome Analysis Centre). Library construction was performed on a PerkinElmer Sciclone using the TruSeq RNA protocol v2 (Illumina 15,026,495 Rev.F). Libraries were pooled (2 pools of 6) and sequenced on 2 lanes of a HiSeq 2500 (Illumina) in High Output mode using 100 bp paired end reads and V3 chemistry. Initial quality assessment of the reads was performed using fastQC [89].

Read alignment and differential expression analysis

Reads were aligned to two reference sequences from the same wheat accession, Chinese Spring: the Chromosome Survey Sequence (CSS; [40] downloaded from *Ensembl* plants release 29) and the TGACv1 reference sequence [41].

The CSS reference consists of 124,201 high confidence transcripts, from which a single transcript was selected based on Ensembl plants release 29 (resulting in 102,503 total transcripts). However, the TGACv1 transcriptome included transcript variants from both high and low confidence gene models (164, 623 and 109,116, respectively), equating to 273,239 total transcripts. We performed read alignment and expression quantification using kallisto-0.42.3 [66] with default parameters, 30 bootstraps ($-b$ 30) and the $-pseudobam$ option. Kallisto has previously been shown to be suitable for the alignment of wheat transcriptome data in a homoeolog specific manner [90].

Differential expression analysis was performed using sleuth-0.28.0 [45] with default parameters. Transcripts with a false-discovery rate (FDR) adjusted p -value (q value) < 0.05 were considered as differentially expressed. Transcripts with a mean abundance of < 0.5 tpm in all four conditions were considered not expressed and were therefore excluded from further analyses. We also randomised samples to determine the expected number of false positives arising from the differential expression analysis. We grouped opposite samples for both genotype and time points and found a mean of 1.83 transcripts differentially expressed under this scenario (standard error = 0.32, $n = 130$).

For each condition the mean tpm of all three biological replicates was calculated. All heatmaps display mean expression values as normalised tpm, on a scale of 0 to 1 with 1 being the highest expression value of the transcript. Read coverage for gene models was obtained using bedtools-2.24.0 genome cov [91] for each pseudobam file and then combined to get a total coverage value of each position. Coverage across a gene model was plotted as relative coverage on a scale of 0 to 1, with 1 being equivalent to the highest level of coverage for the gene model in question.

We also measured the expression of a subset of five DE transcripts from the 'between NIL' comparisons using quantitative reverse transcription PCR (qRT-PCR). Expression values obtained using qRT-PCR were highly consistent with the expression values obtained using RNA-seq (Additional file 7). For qRT-PCR, the same twelve RNA samples used for RNA-seq were used (i.e. three biological replicates per NIL per time point) and cDNA was made using M-MLV Reverse Transcriptase (Life Technologies). Transcript levels were determined using genome specific forward and reverse primers for each transcript (Additional file 8). Primer efficiencies were determined using pooled cDNA from all twelve samples (all $> 88\%$). qRT-PCR reactions were performed using LightCycler 480 SYBR Green I Master Mix with a LightCycler 480 instrument (both Roche Applied Science, UK) and the following conditions: 5 min at 95 °C; 45 cycles of 10 s at 95 °C, 15 s at 60 °C, 30 s at 72 °C.

The specificity of the primers were determined by dissociation curve analysis (from 60 to 95 °C). All reactions were performed with three technical replications. Relative expression levels were calculated relative to actin [92] and linearised values determined using the $(1 + \text{Efficiency})^{-\Delta\Delta C_t}$ method [93].

GO term enrichment

We used the R package Goseq v1.26 [94] to test for enrichment of GO terms in specific groups of DE transcripts. We considered over-represented GO terms with a Benjamini Hochberg FDR adjusted p -value of < 0.05 to be significantly enriched.

Functional annotation

Functional annotations of transcripts were obtained from the TGACv1 annotation [41]. Additionally, for coding transcripts we performed BLASTP against the non-redundant NCBI protein database and conserved domain database, in each case the top hit based on e -value was retained. In cases where all three annotations were in agreement, the TGAC annotation is reported. In cases where the three annotations produced differing results, all annotations are reported. Orthologues in other species such as Arabidopsis and rice were obtained from Ensembl plants release 36. Eight of the 112 DE transcripts had no annotation or protein sequence similarity with other species. We manually categorised the remaining 104 DE transcripts based on their predicted function. Transcripts that fell into a category of size 1 were classed as 'other'. For the non-coding transcripts, we used BLASTN to identify potential miRNA precursors using a set of conserved and wheat specific miRNA sequences obtained from Sun et al., 2014 [48]. We used the $-task$ blastn-short option of BLAST for short sequences and only considered hits of the full length of the miRNA sequence with no mismatches as potential precursors. We used the psRNATarget tool (<http://plantgrn.noble.org/psRNATarget/>) to determine the miRNA targets.

Identification of transcription factor binding sites

We extracted 1000 bp of sequence upstream of the cDNA start site to search for transcription factor binding sites (TFBS). Transcripts with < 1000 bp upstream in the reference sequence were not used in the analysis. We used the FIMO tool from the MEME suite (v 4.11.4; [95]) with a position weight matrix (PWM) obtained from plantPAN 2.0 (<http://plantpan2.itps.ncku.edu.tw/>; [96]). We ran FIMO with a p value threshold of $< 1e-4$ (default), increased the max-stored-scores to 1,000,000 to account for the size of the dataset, and used a $-motifpseudo$ of $1e-8$ as recommended by Peng et al. [97] for use with PWMs. We generated a background model

using the fasta-get-markov command of MEME on all extracted promoter sequences.

Enrichment testing

To test for enrichment of different categories of transcripts relative to all expressed transcripts we performed Fisher's exact test using R-3.2.5. For functional annotation categories, enrichment testing was only performed on categories that could be extracted using GO terms and key words based on their annotation in the TGAC reference. Only DE transcripts that could be extracted using this method were used in the enrichment tests. For example, we identified 12 DE transcripts associated with ubiquitin. The annotation of these transcripts was obtained through a combination of the TGAC annotation and manual annotation. However, only seven of these transcripts could be extracted using GO terms and key words from the whole reference annotation. Therefore, only seven transcripts were used for the enrichment test.

Additional files

Additional file 1: Comparison of CSS and TGAC gene models. (XLSX 21 kb)

Additional file 2: Enriched GO terms in the across time comparisons. (XLSX 20 kb)

Additional file 3: Q-value distributions of uniquely DE transcripts across time. (DOCX 171 kb)

Additional file 4: Transcription factor binding sites identified in outside 5A DE transcripts. (XLSX 28 kb)

Additional file 5: Transcription factors present in the 5A introgression. (XLSX 20 kb)

Additional file 6: Functional annotation of DE transcripts between NILs. (XLSX 18 kb)

Additional file 7: qRT-PCR expression values and graphs. (DOCX 370 kb)

Additional file 8: qRT-PCR primer sequences. (XLSX 9 kb)

Acknowledgements

We thank David Swarbreck and Gemy Kaithakottil (Earlham Institute) for in silico mapping of TGACv1 gene models and to the IWGSC for pre-publication access to RefSeq v.1.0.

Funding

This work was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) Grants BB/P013511/1 and BB/P016855/1. JB was supported by the UK Agriculture and Horticulture Development Board (AHDB) and the John Innes Foundation.

Availability of data and materials

The data sets supporting the results of this article are included within the article and its additional files. Raw sequence reads have been deposited in the NCBI Sequence Read Archive under the Bioproject PRJNA396738.

Authors' contributions

JB designed the research, performed RNA extractions, analysed the data, performed statistical analyses and wrote the manuscript; JS coordinated the field trials and developed the germplasm used in this study; CU designed the research and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

Publisher's Note

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

Received: 15 August 2017 Accepted: 17 January 2018

Published online: 25 January 2018

References

- United Nations. World Population Prospects: the 2015 revision. https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf. Accessed 27 Feb 2017.
- FAO, IFAD, WFP. The State of Food Insecurity in the World 2015. Meeting the 2015 international hunger targets: taking stock of uneven progress. Rome, FAO 2015.
- Rosegrant MW, Tokgoz S, Bhandary P, Msangi S. Looking ahead: Scenarios for the future of food. 2012 Global Food Policy Report. Washington, D.C: International Food Policy Research Institute; 2013. p. 88–101.
- Tilman D, Balzer C, Hill J, Befort BL. Global food demand and the sustainable intensification of agriculture. *Proc Natl Acad Sci*. 2011;108(50):20260–4. <https://doi.org/10.1073/pnas.1116437108>.
- FAO. Global agriculture towards 2050. Rome: FAO; 2009.
- Ray DK, Mueller ND, West PC, Foley JA. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS One*. 2013;8(6):e66428-e. <https://doi.org/10.1371/journal.pone.0066428>.
- Kuchel H, Williams KJ, Langridge P, Eagles HA, Jefferies SP. Genetic dissection of grain yield in bread wheat. I. QTL analysis. *Theor Appl Genet*. 2007;115(8):1029–41. <https://doi.org/10.1007/s00122-007-0629-7>.
- Li N, Li Y. Maternal control of seed size in plants. *J Exp Bot*. 2015;66(4):1087–97. <https://doi.org/10.1093/jxb/eru549>.
- Huang R, Jiang L, Zheng J, Wang T, Wang H, Huang Y, et al. Genetic bases of rice grain shape: so many genes, so little known. *Trends in Plant Science*. 2013;18(4):218–26. <https://doi.org/10.1016/j.tplants.2012.11.001>.
- Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, et al. Control of grain size, shape and quality by *OsSPL16* in rice. *Nature Genetics*. 2012;44(8):950–4. <https://doi.org/10.1038/ng.2327>.
- Garcia D, Gerald JNF, Berger F. Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell*. 2005;17(1):52–60. <https://doi.org/10.1105/tpc.104.027136>.
- Song X-J, Huang W, Shi M, Zhu M-Z, Lin H-X. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet*. 2007;39(5):623–30. <https://doi.org/10.1038/ng2014>.
- Simmonds J, Scott P, Brinton J, Mestre TC, Bush M, Blanco A, et al. A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theor Appl Genet*. 2016;129(6):1099–112. <https://doi.org/10.1007/s00122-016-2686-2>.
- Xia T, Li N, Dumenil J, Li J, Kamenski A, Bevan MW, et al. The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in Arabidopsis. *Plant Cell*. 2013;25(9):3347–59. <https://doi.org/10.1105/tpc.113.115063>.
- Kitagawa K, Kurinami S, Oki K, Abe Y, Ando T, Kono I, et al. A novel kinesin 13 protein regulating Rice seed length. *Plant Cell Physiol*. 2010;51(8):1315–29. <https://doi.org/10.1093/pcp/pcq092>.
- Fujikura U, Elsaesser L, Breuninger H, Sánchez-Rodríguez C, Ivakov A, Laux T, et al. Atkinesin-13A modulates Cell-Wall synthesis and cell expansion in *Arabidopsis thaliana* via the THESEUS1 pathway. *PLoS Genet*. 2014;10(9):e1004627. <https://doi.org/10.1371/journal.pgen.1004627>.
- Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, et al. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor Appl Genet*. 2006;112(6):1164–71. <https://doi.org/10.1007/s00122-006-0218-1>.

18. Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, et al. Natural variation at the DEP1 locus enhances grain yield in rice. *Nat Genet.* 2009;41(4):494–7. <https://doi.org/10.1038/ng.352>.
19. Riefler M, Novak O, Strnad M, Schmülling T. Arabidopsis cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. *Plant Cell.* 2006;18(1):40–54.
20. Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ. The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development.* 2006;133(2):251–61. <https://doi.org/10.1242/dev.02194>.
21. Jiang W-B, Huang H-Y, Hu Y-W, Zhu S-W, Wang Z-Y, Lin W-H. Brassinosteroid regulates seed size and shape in Arabidopsis. *Plant Physiol.* 2013;162(4):1965–77.
22. FAO. Online statistical database: Food balance. FAOSTAT. 2017. <http://www.fao.org/faostat/en/>. Accessed 27 February 2017.
23. Ma M, Zhao H, Li Z, Hu S, Song W, Liu X. TaCYP78A5 regulates seed size in wheat (*Triticum aestivum*). *J Exp Bot.* 2016;67(5):1397–410. <https://doi.org/10.1093/jxb/erv542>.
24. Simmonds J, Scott P, Leverington-Waite M, Turner AS, Brinton J, Korzun V, et al. Identification and independent validation of a stable yield and thousand grain weight QTL on chromosome 6A of hexaploid wheat (*Triticum aestivum* L.). *BMC Plant Biology.* 2014;14(1):191. <https://doi.org/10.1186/s12870-014-0191-9>.
25. Breseghello F, Sorrells ME. QTL analysis of kernel size and shape in two hexaploid wheat mapping populations. *Field Crop Res.* 2007;101(2):172–9. <https://doi.org/10.1016/j.fcr.2006.11.008>.
26. Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, et al. A genetic framework for grain size and shape variation in wheat. *Plant Cell.* 2010;22(4):1046–56. <https://doi.org/10.1105/tpc.110.074153>.
27. Farré A, Sayers L, Leverington-Waite M, Goram R, Orford S, Wingen L, et al. Application of a library of near isogenic lines to understand context dependent expression of QTL for grain yield and adaptive traits in bread wheat. *BMC Plant Biol.* 2016;16 <https://doi.org/10.1186/s12870-016-0849-6>.
28. Kumar A, Mantovani EE, Seetan R, Soltani A, Echeverry-Solarte M, Jain S, et al. Dissection of Genetic Factors underlying Wheat Kernel Shape and Size in an Elite x Nonadapted Cross using a High Density SNP Linkage Map. *Plant Genome-U.S.* 2016;9(1) <https://doi.org/10.3835/plantgenome2015.09.0081>.
29. Brinton J, Simmonds J, Minter F, Leverington-Waite M, Snape J, Uauy C. Increased pericarp cell length underlies a major quantitative trait locus for grain weight in hexaploid wheat. *New Phytol.* 2017;215(3):1026–38. <https://doi.org/10.1111/nph.14624>.
30. Hasan AK, Herrera J, Lizana C, Calderini DF. Cernel weight, grain length and stabilized grain water content are physiological drivers of grain weight determination of wheat. *Field Crop Res.* 2011;123(3):241–7. <https://doi.org/10.1016/j.fcr.2011.05.019>.
31. Guo Z, Chen D, Schnurbusch T. Variance components, heritability and correlation analysis of anther and ovary size during the floral development of bread wheat. *J Exp Bot.* 2015; <https://doi.org/10.1093/jxb/erv117>.
32. Shewry PR, RAC M, Tosi P, Wan Y, Underwood C, Lovegrove A, et al. An integrated study of grain development of wheat (cv. Hereward). *J Cereal Sci.* 2012;56(1):21–30. <https://doi.org/10.1016/j.jcs.2011.11.007>.
33. Wan Y, Poole RL, Huttly AK, Toscano-Underwood C, Feeney K, Welham S, et al. Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics.* 2008;9 <https://doi.org/10.1186/1471-2164-9-121>.
34. Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science.* 2014;345(6194):1250091.
35. Pellny TK, Lovegrove A, Freeman J, Tosi P, Love CG, Knox JP, et al. Cell walls of developing wheat starchy endosperm: comparison of composition and RNA-Seq transcriptome. *Plant Physiol.* 2012;158(2):612–27. <https://doi.org/10.1104/pp.111.189191>.
36. Yu Y, Zhu D, Ma C, Cao H, Wang Y, Xu Y, et al. Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development. *The Crop Journal.* 2016;4(2):92–106. <https://doi.org/10.1016/j.cj.2016.01.006>.
37. Laudencia-Chingcuanco DL, Stamova BS, You FM, Lazo GR, Beckles DM, Anderson OD. Transcriptional profiling of wheat caryopsis development using cDNA microarrays. *Plant Mol Biol.* 2007;63(5):651–68. <https://doi.org/10.1007/s11103-006-9114-y>.
38. Liu W, Wu Z, Zhang Y, Guo D, Xu Y, Chen W, et al. Transcriptome analysis of wheat grain using RNA-Seq. *Frontiers of Agricultural Science and Engineering.* 2014;1(3):214–22. <https://doi.org/10.15302/j-fase-2014024>.
39. IWGSC RefSeq v1.0. <https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies>.
40. IWGSC. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science.* 2014;345(6194):1251788. <https://doi.org/10.1126/science.1251788>.
41. Clavijo BJ, Venturini L, Schudoma C, Garcia Accinelli G, Kaithakottil G, Wright J, et al. 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.* 2017;27:885–96. <https://doi.org/10.1101/gr.217117.116>.
42. Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL. The first near-complete assembly of the hexaploid bread wheat genome. *Triticum aestivum* bioRxiv. 2017; <https://doi.org/10.1101/159111>.
43. Cantu D, Pearce SP, Distelfeld A, Christiansen MW, Uauy C, Akhunov E, et al. Effect of the down-regulation of the high *Grain Protein Content* (GPC) genes on the wheat transcriptome during monocarpic senescence. *BMC Genomics.* 2011;12(1):492. <https://doi.org/10.1186/1471-2164-12-492>.
44. Barrero JM, Cavanagh C, Verbyla KL, Tibbits JFG, Verbyla AP, Huang BE, et al. Transcriptomic analysis of wheat near-isogenic lines identifies *PM19-A1* and *A2* as candidates for a major dormancy QTL. *Genome Biol.* 2015;16(1):93. <https://doi.org/10.1186/s13059-015-0665-6>.
45. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods.* 2017;14(7):687–90. <https://doi.org/10.1038/nmeth.4324>.
46. Borrill P, Harrington SA, Uauy C. Genome-Wide Sequence and Expression Analysis of the NAC Transcription Factor Family in Polyploid Wheat. *G3: Genes|Genomes|Genetics.* 2017;7(9):3019–29. <https://doi.org/10.1534/g3.117.043679>.
47. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature.* 2012;482(7385):339–46.
48. Sun F, Guo G, Du J, Guo W, Peng H, Ni Z, et al. Whole-genome discovery of miRNAs and their targets in wheat (*Triticum aestivum* L.). *BMC Plant Biology.* 2014;14(1):142. <https://doi.org/10.1186/1471-2229-14-142>.
49. Disch S, Anastasiou E, Sharma VK, Laux T, Fletcher JC, Lenhard M. The E3 ubiquitin ligase BIG BROTHER controls *Arabidopsis* organ size in a dosage-dependent manner. *Curr Biol.* 2006;16(3):272–9. <https://doi.org/10.1016/j.cub.2005.12.026>.
50. Dong H, Dumenil J, Lu FH, Na L, Vanhaeren H, Naumann C, et al. Ubiquitylation activates a peptidase that promotes cleavage and destabilization of its activating E3 ligases and diverse growth regulatory proteins to limit cell proliferation in Arabidopsis. *Genes Dev.* 2017;31(2):197–208. <https://doi.org/10.1101/gad.292235.116>.
51. Du L, Li N, Chen L, Xu Y, Li Y, Zhang Y, et al. The ubiquitin receptor DA1 regulates seed and organ size by modulating the stability of the ubiquitin-specific protease UBP15/SOD2 in *Arabidopsis*. *Plant Cell.* 2014;26(2):665–77. <https://doi.org/10.1105/tpc.114.122663>.
52. Huang K, Wang D, Duan P, Zhang B, Xu R, Li N, et al. WIDE AND THICK GRAIN 1, which encodes an otubain-like protease with deubiquitination activity, influences grain size and shape in rice. *Plant J.* 2017;91(5):849–60. <https://doi.org/10.1111/tpj.13613>.
53. Cheng Y, Qin G, Dai X, Zhao Y. NPY1, a BTB-NPH3-like protein, plays a critical role in auxin-regulated organogenesis in *Arabidopsis*. *Proc Natl Acad Sci.* 2007;104(47):18825–9. <https://doi.org/10.1073/pnas.0708506104>.
54. Radchuk V, Weier D, Radchuk R, Weschke W, Weber H. Development of maternal seed tissue in barley is mediated by regulated cell expansion and cell disintegration and coordinated with endosperm growth. *J Exp Bot.* 2011;62(3):1217–27. <https://doi.org/10.1093/jxb/erq348>.
55. Drea S, Leader DJ, Arnold BC, Shaw P, Dolan L, Doonan JH. Systematic spatial analysis of gene expression during wheat caryopsis development. *Plant Cell.* 2005;17(8):2172–85. <https://doi.org/10.1105/tpc.105.034058>.
56. Dominguez F, Cejudo FJ. Characterization of the Endoproteases appearing during wheat grain development. *Plant Physiol.* 1996;112(3):1211–7. <https://doi.org/10.1104/pp.112.3.1211>.
57. Kaspar-Schoenefeld S, Merx K, Jozefowicz AM, Hartmann A, Seiffert U, Weschke W, et al. Label-free proteome profiling reveals developmental-dependent patterns in young barley grains. *Journal of Proteomics.* 2016;143:106–21. <https://doi.org/10.1016/j.jpro.2016.04.007>.
58. Pires JC, Conant GC. Robust yet fragile: expression noise, protein Misfolding, and gene dosage in the evolution of genomes. *Annu Rev Genet.* 2016;50:113–31. <https://doi.org/10.1146/annurev-genet-120215-035400>.
59. Yang M, Gao X, Dong J, Gandhi N, Cai H, von Wettstein DH, et al. Pattern of protein expression in developing wheat grains identified through proteomic analysis. *Front Plant Sci.* 2017;8:962. <https://doi.org/10.3389/fpls.2017.00962>.
60. Li J, Jiang J, Qian Q, Xu Y, Zhang C, Xiao J, et al. Mutation of Rice BC12/GDD1, which encodes a kinesin-like protein that binds to a GA biosynthesis

- gene promoter, leads to dwarfism with impaired cell elongation. *Plant Cell*. 2011;23(2):628–40. <https://doi.org/10.1105/tpc.110.081901>.
61. Torada A, Koike M, Ogawa T, Takenouchi Y, Tadamura K, Wu J, et al. A Causal Gene for Seed Dormancy on Wheat Chromosome 4A Encodes a MAP Kinase Kinase. *Curr Biol*. 2016;26(6):782–7. <https://doi.org/10.1016/j.cub.2016.01.063>.
 62. Shorinola O, Balčárková B, Hyles J, JFG T, Hayden MJ, Holušova K, et al. Haplotype Analysis of the Pre-harvest Sprouting Resistance Locus *Phs-A1* Reveals a Causal Role of *TaMKK3-A* in Global Germplasm. *Front Plant Sci*. 2017;8:1555. <https://doi.org/10.3389/fpls.2017.01555>.
 63. Ramirez-Gonzalez RH, Segovia V, Bird N, Fenwick P, Holdgate S, Berry S, et al. RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnol J*. 2015;13(5):613–24. <https://doi.org/10.1111/pbi.12281>.
 64. Uauy C, Wulff BBH, Dubcovsky J. Combining Traditional Mutagenesis with New High-Throughput Sequencing and Genome Editing to Reveal Hidden Variation in Polyploid Wheat. *Annual Review of Genetics*. 2017;51(1) <https://doi.org/10.1146/annurev-genet-120116-024533>.
 65. Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K, Clark MD. Genomic innovation for crop improvement. *Nature*. 2017;543(7645):346–54. <https://doi.org/10.1038/nature22011>.
 66. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotech*. 2016;34(5):525–7. <https://doi.org/10.1038/nbt.3519>.
 67. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Meth*. 2017;14(4):417–9. <https://doi.org/10.1038/nmeth.4197>.
 68. Kaplan-Levy RN, Brewer PB, Quon T, Smyth DR. The trihelix family of transcription factors – light, stress and development. *Trends in Plant Science*. 2012;17(3):163–71. <https://doi.org/10.1016/j.tplants.2011.12.002>.
 69. Barg R, Sobolev I, Eilon T, Gur A, Chmelnitsky I, Shabtai S, et al. The tomato early fruit specific gene *Lefsm1* defines a novel class of plant-specific SANT/ MYB domain proteins. *Planta*. 2005;221(2):197–211. <https://doi.org/10.1007/s00425-004-1433-0>.
 70. Sarojam R, Sappl PG, Goldschmidt A, Efroni I, Floyd SK, Eshed Y, et al. Differentiating *Arabidopsis* shoots from leaves by combined YABBY activities. *Plant Cell*. 2010;22(7):2113–30. <https://doi.org/10.1105/tpc.110.075853>.
 71. Cong B, Barrero LS, Tanksley SD. Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat Genet*. 2008;40(6):800–4. <https://doi.org/10.1038/ng.144>.
 72. Dinneny JR, Weigel D, Yanofsky MF. NUBBIN and JAGGED define stamen and carpel shape in *Arabidopsis*. *Development*. 2006;133(9):1645–55. <https://doi.org/10.1242/dev.02335>.
 73. Jin Y, Luo Q, Tong H, Wang A, Cheng Z, Tang J, et al. An AT-hook gene is required for palea formation and floral organ number control in rice. *Dev Biol*. 2011;359(2):277–88. <https://doi.org/10.1016/j.ydbio.2011.08.023>.
 74. Gallavotti A, Malcomber S, Gaines C, Stanfield S, Whipple C, Kellogg E, et al. BARREN STALK FASTIGIATE1 is an AT-hook protein required for the formation of maize ears. *Plant Cell*. 2011;23(5):1756–71. <https://doi.org/10.1105/tpc.111.084590>.
 75. Street IH, Shah PK, Smith AM, Avery N, Neff MM. The AT-hook-containing proteins SOB3/AHL29 and ESC/AHL27 are negative modulators of hypocotyl growth in *Arabidopsis*. *Plant J*. 2008;54(1):1–14. <https://doi.org/10.1111/j.1365-3113.2007.03393.x>.
 76. Cosgrove DJ. Growth of the plant cell wall. *Nat Rev Mol Cell Bio*. 2005;6(11):850–61. <https://doi.org/10.1038/nrm1746>.
 77. Dante RA, Larkins BA, Sabelli PA. Cell cycle control and seed development. *Front Plant Sci*. 2014;5:493. <https://doi.org/10.3389/fpls.2014.00493>.
 78. Lee BH, Ko J-H, Lee S, Lee Y, Pak J-H, Kim JH. The *Arabidopsis* GRF-INTERACTING FACTOR gene family performs an overlapping function in determining organ size as well as multiple developmental properties. *Plant Physiol*. 2009;151(2):655–68. <https://doi.org/10.1104/pp.109.141838>.
 79. Bao F, Azhakanandam S, Franks RG. SEUSS and SEUSS-LIKE transcriptional adaptors regulate floral and embryonic development in *Arabidopsis*. *Plant Physiol*. 2010;152(2):821–36. <https://doi.org/10.1104/pp.109.146183>.
 80. Adamski NM, Anastasiou E, Eriksson S, O'Neill CM, Lenhard M. Local maternal control of seed size by KLUH/CYP78A5-dependent growth signaling. *Proc Natl Acad Sci*. 2009;106(47):20115–20. <https://doi.org/10.1073/pnas.0907024106>.
 81. 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*. 2014;5:412. <https://doi.org/10.3389/fpls.2014.00412>.
 82. Li N, Li Y. Ubiquitin-mediated control of seed size in plants. *Front Plant Sci*. 2014;5:332. <https://doi.org/10.3389/fpls.2014.00332>.
 83. Yang Z, Bai Z, Li X, Wang P, Wu Q, Yang L, et al. SNP identification and allelic-specific PCR markers development for TaGW2, a gene linked to wheat kernel weight. *Theor Appl Genet*. 2012;125(5):1057–68. <https://doi.org/10.1007/s00122-012-1895-6>.
 84. Herskko A, Ciechanover A. The ubiquitin system. *Annu Rev Biochem*. 1998;67:425–79. <https://doi.org/10.1146/annurev.biochem.67.1.425>.
 85. Vanhaeren H, Nam Y-J, De Milde L, Chae E, Storme V, Weigel D, et al. Forever young: the role of ubiquitin receptor DA1 and E3 ligase BIG BROTHER in controlling leaf growth and development. *Plant Physiol*. 2016;01410 <https://doi.org/10.1104/pp.16.01410>.
 86. Kurepa J, Wang S, Li Y, Zaitlin D, Pierce AJ, Smalle JA. Loss of 26S proteasome function leads to increased cell size and decreased cell number in *Arabidopsis* shoot organs. *Plant Physiol*. 2009;150(1):178–89. <https://doi.org/10.1104/pp.109.135970>.
 87. Weng J, Gu S, Wan X, Gao H, Guo T, Su N, et al. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res*. 2008;18(12):1199–209.
 88. Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, et al. Uncovering hidden variation in polyploid wheat. *Proc Natl Acad Sci*. 2017;114(6):E913–E21. <https://doi.org/10.1073/pnas.1619268114>.
 89. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Accessed 9 Sept 2015.
 90. Borrill P, Ramirez-Gonzalez R, Uauy C. expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiol*. 2016;170(4):2172–86.
 91. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
 92. Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science*. 2006;314(5803):1298–301.
 93. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc*. 2008;3:1101. <https://doi.org/10.1038/nprot.2008.73>.
 94. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. 2010;11(2):R14. <https://doi.org/10.1186/gb-2010-11-2-r14>.
 95. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27(7):1017–8. <https://doi.org/10.1093/bioinformatics/btr064>.
 96. Chow C-N, Zheng H-Q, Wu N-Y, Chien C-H, Huang H-D, Lee T-Y, et al. PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res*. 2016;44(Database issue):D1154–D160. <https://doi.org/10.1093/nar/gkv1035>.
 97. Peng FY, Hu Z, Yang R-C. Bioinformatic prediction of transcription factor binding sites at promoter regions of genes for photoperiod and vernalization responses in model and temperate cereal plants. *BMC Genomics*. 2016;17:573. <https://doi.org/10.1186/s12864-016-2916-7>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

