The Wheat Transcriptome and Discovery of Functional Gene Networks

5

Tayyaba Andleeb, James Milson and Philippa Borrill

Abstract

Gene expression patterns have been a widely applied source of information to start understanding gene function in multiple plant species. In wheat, the advent of increasingly accurate and complete gene annotations now enables transcriptomic studies to be carried out on a routine basis and studies by groups around the world have compared gene expression changes under an array of environmental and developmental stages. However, associating data from differentially expressed genes to understanding the biological role of these genes and their applications for breeding is a major challenge. Recently, the first steps to apply network-based approaches to characterise gene expression have been taken in wheat and these networks have enabled the prediction of gene functions in wheat but only for a handful of traits. Combining advanced analysis methods with better sequencing technology will increase our capacity to place gene expression in wheat in the context of functions of genes that influence agronomically important traits.

5.1 Gene Function Through Gene Expression

In order to understand gene function, one of the first things researchers would like to do is measure gene expression—when, where and how much of a gene's transcript is present? Measuring the expression level of a single gene through quantitative PCR can reveal insight into a specific gene and its potential biological role. However, to explore the integrated nature of gene expression and how entire biological processes work at the transcriptional level, it is desirable to measure the expression level of multiple genes simultaneously using transcriptomics. In model species, transcriptomics has shed insight into the regulation of developmental processes, responses to the environment and

Tayyaba Andleeb and James Milson have contributed equally to this work.

T. Andleeb
Department of Plant Sciences, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan

T. Andleeb · J. Milson
School of Biological Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

J. Milson · P. Borrill (✉)
Department of Crop Genetics, John Innes Centre, Norwich Research, Norwich NR4 7UH, UK
e-mail: philippa.borrill@jic.ac.uk

genotype-specific responses, all of which would be highly advantageous to understand for wheat improvement. Therefore, transcriptomics has been widely applied in wheat biology.

Initially, transcriptomics largely relied upon microarray approaches. These were useful in determining gene expression patterns, but microarrays in wheat were limited because of the incomplete gene model annotations available when microarrays were designed, therefore many genes were missing from the arrays. The advent of RNA-seq to measure gene expression enabled more accurate measurement of the wheat transcriptome. Transcriptomics could be applied even before high-quality genome assemblies were available because de novo transcriptome assemblies could be generated to answer specific biological questions using individual datasets. However, to get the highest quality and most comprehensive results in a transcriptomic experiment, having a reference transcriptome is valuable and also removes the requirement to carry out a de novo assembly for each new project. Furthermore, the availability of a reference transcriptome facilitates the identification of homoeolog-specific transcripts and therefore allows gene expression to be quantified in a homoeolog-specific manner.

5.2 Measuring Homoeolog-Specific Gene Expression

As consequence of the polyploid nature of wheat, >50% of genes in the wheat genome are present as triads of related homoeologous genes on the A, B and D subgenomes (IWGSC et al. 2018). Studies on a gene-by-gene basis have revealed that each homoeolog in wheat can have different expression levels. For example, the calcium-dependent protein kinase *TaCPK2* has differential responses to stress between homeologs with the A homeolog upregulated in response to powdery mildew infection and the D homoeolog upregulated in response to cold stress (Geng et al. 2013). However, to analyse homoeolog-specific expression using qPCR is labour-intensive and requires the

design of homoeolog-specific primers for each gene of interest. The use of transcriptomics allows quicker and easier homoeolog-specific gene expression measurements. Several different ways to quantify homoeolog-specific gene expression in allopolyploids have been implemented including alignment to the individual subgenomes and read classification according to mismatches or inter-homoeolog SNPs (Kuo et al. 2020), alignment to the whole genome sequence using a standard aligner and selecting only uniquely mapping reads (e.g. He et al. 2022) or pseudoalignment to the transcriptome using kallisto which has been demonstrated to assign reads to appropriate homoeolog using nullitetrasomic lines (Borrill et al. 2016; Ramírez-González et al. 2018). Homoeolog-specific gene analysis has been used to study multiple biological questions and has for example revealed homoeolog-specific gene expression responses to stress conditions (e.g. Clavijo et al. 2017) and developmental stage and tissue-specific homoeolog expression (Ramírez-González et al. 2018). In order to maximise information gained from applying transcriptomic approaches, it is necessary to define which genes are present within the genome and have accurate gene annotations to capture the complexities of gene expression in this polyploid species.

5.3 Building Transcriptome Annotations in Wheat

5.3.1 Expressed Sequence Tags and Full-Length cDNAs

The large size of the wheat genome made sequencing the entire wheat genome and the genes within it a difficult prospect in the 1990s and 2000s due to the high cost and sequencing technology limitations (see also Chap. 1). However, the importance and usefulness of having gene sequence information was clear. An alternative way to obtain gene sequence focussed on expressed sequence tags (ESTs), which provided a quicker way to determine
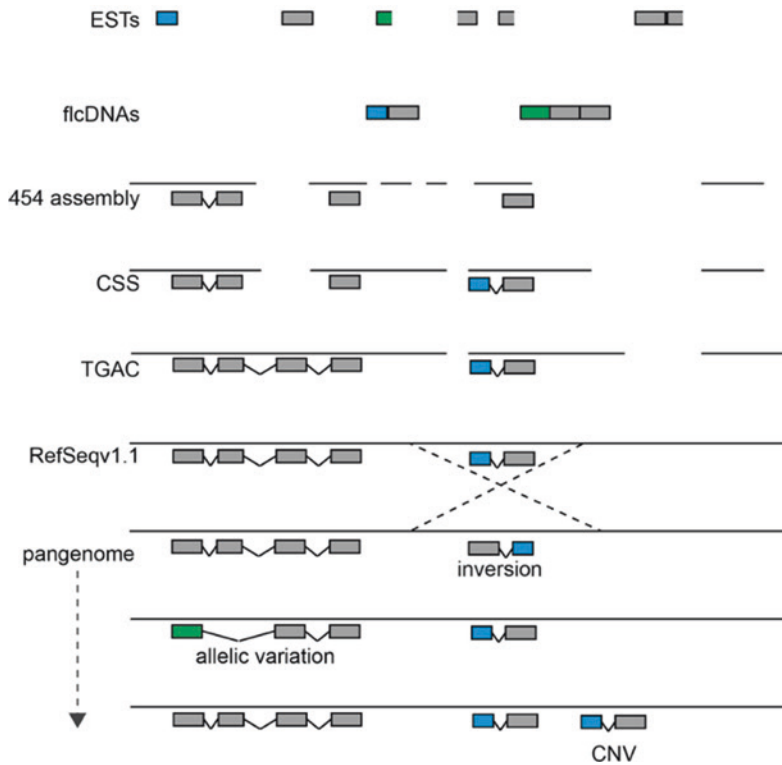
Fig. 5.1 Improvements in transcriptome assemblies in the last 20 years. Transcriptome sequences have progressed from expressed sequence tags (EST) which had unknown chromosomal positions and were often partial sequences, through full-length cDNAs (flcDNAs) to the initial genome assemblies (454 assembly) which often lacked annotation, through to fragmented assemblies with gene model predictions such as the CHINESE SPRING Survey (CSS) and The Genome Analysis Centre (TGAC) assembly, to highly complete transcriptome assemblies on contiguous chromosome-scale scaffolds (RefSeqv1.1). Sequencing and assembly of transcriptomes for multiple wheat cultivars will reveal the pan-transcriptome and variation therein including copy number variation (CNV)

gene sequences and expression information (Fig. 5.1). ESTs were generated by extracting RNA from a tissue or tissues of interest and building a cDNA library in *E. coli*. Plasmids from the *E. coli* library were extracted and sequenced through Sanger sequencing before bioinformatic analysis to group sequences into contigs containing related sequences. ESTs were generated from multiple wheat tissues (Ogihara et al. 2003; Manickavelu et al. 2012) and samples grown under stress conditions (Chao et al. 2006; Mochida et al. 2006) resulting in the identification of over 1 million EST sequences grouped into tens of thousands of contigs. By filtering these contigs for sequences containing both start and stop codons, it was possible to identify full-length cDNA representing entire coding sequences, although the numbers were significantly lower than the number of ESTs. For example, the 1 million EST generated by Manickavelu et al. (2012) were classified into 37,138 contigs of which ~7000 were full length. Significant efforts were made to obtain a good representation of full-length cDNAs, and the resulting sequences (~ 20,000 full-length cDNAs) were gathered into databases (Kawaura et al. 2009; Mochida et al. 2009).

5.3.2 Integrating Gene Annotation into Genome Assemblies

In parallel with the development of flcDNA libraries, many groups embarked upon

projects to sequence the wheat genome. The first sequence of a wheat genome with associated gene annotations was published in 2012 using the cultivar CHINESE SPRING (Brenchley et al. 2012). The low sequencing coverage (5x) using 454 technology meant that the assembly was highly fragmented (over 5 million scaffolds), yet it was extremely useful to researchers offering the first extensive set of genomic sequences. Approximately 95,000 genes were annotated using orthologs to flcDNAs from rice, sorghum, Brachypodium and barley. Two-thirds of these genes were assigned to the A, B or D subgenome but it was not possible to assign genes to individual chromosomes. This data provided larger number of gene annotations than were available from flcDNAs, although not all flcDNAs were represented and many of the gene models were fragmented (Fig. 5.1). Nonetheless, this assembly illustrated that whole genome sequencing of wheat was possible and could make major contributions to generating a complete set of gene models.

The next major improvement in gene models was achieved by applying flow-sorting technology to separate individual chromosome arms prior to sequencing (see Chap. 3). This allowed gene models to be assigned to individual chromosome arms, identifying homoeologous genes with confidence, and positional information was added through the use of synteny and genetic mapping approaches. In total 124,201 genes were annotated and assigned to individual chromosomes, and 75,183 had positional information. These genes were located across a total 10.2 Gb assembly of CHINESE SPRING (the CHINESE SPRING Survey; CSS; Fig. 5.1; IWGSC et al. 2014). However, the fragmented nature of this assembly with only 70% of the assembly in contigs longer than 1 kb, meant that although the number of genes identified was high, many genes were not full length for example due to a gene model being truncated at the end of a contig (Brinton et al. 2018).

Improvements to assembling complete gene models came largely through improved contiguity in genome assemblies. The use of varying sized mate-pair libraries and a new assembly algorithm produced a new CHINESE SPRING assembly (Clavijo et al. 2017) with a longer contig size with over 80% of the assembly having contigs larger than 32 kb. In total 104,091 gene models were annotated, which is ~ 20,000 genes fewer than in the CSS assembly (IWGSC et al. 2014), but these new gene models were generally more complete because the higher assembly contiguity meant it was much less likely that a gene model was truncated at the end of a contig (Fig. 5.1). An additional CHINESE SPRING assembly (Triticum3.1) achieved much-increased contiguity by combining Illumina short reads with PacBio long reads, with over 50% of the assembly having contigs larger than 232 kb (Zimin et al. 2017), but this assembly lacked gene annotations.

The next step change came with the publication of the RefSeqv1.0 CHINESE SPRING genome assembly (IWGSC et al. 2018). This pseudomolecule-level 14.5 Gb assembly used a de novo assembly approach, an improved assembly method and additional layers of genetic, physical and sequencing data to generate a long-range ordered assembly with accurate assignment of homoeologs. In total 107,891 high-confidence genes were annotated by combining the outputs of two prediction pipelines. These gene models represented a higher proportion of conserved BUSCO single-copy genes than previous assemblies with 90% of BUSCO genes present as three complete copies in the RefSeq assembly, compared to 70% in the TGAC assembly and 25% in the CSS assembly. Approximately, 2,000 gene models were manually refined, resulting in the RefSeqv1.1 gene model set (Fig. 5.1).

Although highly complete, further improvements have been made to these gene models. By combining the long-read-based Triticum_aestivum_3.1 genome assembly with information from the RefSeqv1.0 assembly to improve scaffolding and annotation, a more complete (15.1 GB) annotated CHINESE SPRING assembly was obtained: *Triticum_aestivum_4.0* (Alonge et al. 2020). The use of long reads enabled many repeat regions to be expanded in this assembly, including regions containing

thousands of additional gene copies. This gave a total of 108,639 genes localised to individual chromosomes. In parallel, further refinements were made to the RefSeqv1.0 by incorporating optical maps and PacBio long reads to generate RefSeqv2.1 (Zhu et al. 2021). Although the total assembly size did not change much (14.6 GB in RefSeqv2.1 vs. 14.5 GB in RefSeqv1.0), positions and orientations of scaffolds were corrected for 10% of the genome and gaps were filled. In total 106,913 high-confidence genes were annotated by aligning gene annotations from the RefSeqv1.1 and community annotations.

5.3.3 Remaining Challenges to Improve the Accuracy and Completeness of the Gene Model Set

Discrepancies remain between the *Triticum_aestivum*_4.0 and RefSeqv2.1 assemblies in some regions, and integration of new data types will be required to resolve localised gaps or errors, and to assign all scaffolds to accurate positions. Gene annotations may also be inaccurate in a minority of regions due to remaining gaps or inaccuracies. Both these assemblies rely on the transfer of gene models from RefSeqv1.1, so there may be value in re-annotating these genomes from de novo predictions and RNA-seq data to take advantage of these more accurate sequences. A final consequence of relying largely on the RefSeqv1.1 gene models is that alternative spliced isoforms may not be fully represented with only 15.7% of high-confidence genes having alternative isoforms (IWGSC et al. 2018), due to conservative parameters used during the transcriptome assembly.

Although technical challenges remain to perfect the CHINESE SPRING gene models, a more pressing challenge will be to identify variation between gene models in different wheat cultivars. Work by Montenegro et al. (2017) showed that gene content was variable between 18 wheat cultivars, with ~ 81,000 genes shared

between all cultivars and an additional 60,000 genes detected in at least one cultivar. The large average number of genes detected in each cultivar in this study (128,656) may be an artefact of basing gene model discovery on the fragmented CSS assembly; nonetheless, the variation in gene models is likely to have significant consequences to understanding wheat biology (see Chap. 4). More recently whole genome sequencing of 15 cultivars in additional to CHINESE SPRING revealed extensive structural and haplotype divergence between wheat cultivars (Fig. 5.1; Walkowiak et al. 2020). Significant differences were found in gene content between cultivars with ~ 12% of genes showing presence–absence variation, although this was based on projecting gene annotations from CHINESE SPRING, rather than de novo genome annotation tailored to each cultivar. Individual genome annotations for each of these high-quality genome sequences will be a valuable resource for biologists and breeders alike and is likely to identify genes absent from CHINESE SPRING.

Beyond increasing the number of cultivars, it will also be important to increase the accuracy of gene models beyond the coding region, which is so far the most accurate portion of wheat gene models. The 5′ and 3′ untranslated regions are annotated in many genes, but their accuracy is not known and specialised next-generation sequencing approaches could be used, such as CAGE-seq to identify transcription start sites and PolyA-seq to identify transcription end sites, as has been done in cotton to generate accurate untranslated region annotations (Wang et al. 2019). The use of PacBio Iso-seq long reads in conjunction with Illumina short reads and stringent filtering can also increase the accuracy of transcript start and end sites, as well as providing information about splice junctions. This has been achieved in wheat's close relative barley (Coulter et al. 2021). This approach identified that 73% of multi-exonic barley genes had two or more transcript isoforms, suggesting that the current wheat annotations may be missing transcript isoforms in many multi-exonic genes.

5.4 Methods of Measuring Gene Expression at the Genome-Wide Level

The availability of high-quality gene models now facilitates the accurate measurements of gene expression using RNA-seq. The most common type of RNA-seq is the enrichment and subsequent sequencing of polyadenylated RNA to study mRNA levels. Reduced representation sequencing can also be applied to reduce costs. For example, 3′ end sequencing can be used for investigating the expression profile of genes at a lower cost due to reduced sequencing requirements and targeted RNA-seq can be used to sequence-specific targets, primarily those with low expression profiles. More recently, low input RNA-seq methods from small tissues to single-cell approaches have been developed. These enable the measurement of gene expression in different cell types and determine co-expression and gene regulation in single cells, although their application in wheat remains limited.

5.5 Diverse Biological Questions Can Be Answered with Transcriptomics

Transcriptomics approaches have been applied in many different types of studies in wheat. These include observing changes in the transcriptome over a developmental time course, studying gene expression responses to different stresses or investigating the effect of a specific gene on downstream molecular pathways (Fig. 5.2).

5.6 Elucidating Genetic Control of Developmental Processes

Transcriptomic approaches can help build understanding of developmental processes by studying gene expression throughout a time course or by focussing on the transcriptional changes induced by manipulating a gene regulating development, for example through mutants or overexpression. Here we will discuss typical approaches which use RNA-seq to understand developmental processes in wheat.

5.6.1 Studying Gene Expression During Time Courses

Grain development is an important process which influences final yield and quality in all cereal crops and has therefore been examined at the transcriptomic level by several groups. For example, using the CHINESE SPRING Survey (CSS) sequence annotation, Pfeifer et al. (2014) identified cell-type and homoeolog-specific gene expression during grain development at three timepoints. Building upon this work Chi et al. (2019) studied gene expression across four timepoints in grain development, although they did not dissect grains into individual cell types. Differentially expressed genes were clustered into groups based on developmental stages and assigned putative functions based on gene ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) enrichment analyses. Many more differentially expressed genes were identified than was possible using previous microarray-based approaches and the more accurate and complete gene models facilitated the analysis (Yu et al. 2016). A similar approach was used to investigate wheat spike development at four different stages (Feng et al. 2017). Clustering analysis of genes differentially expressed over the time course identified dynamically expressed transcription factors which the authors hypothesise may regulate spikelet initiation and floral organ patterning, inferred from their times of expression and orthologs in model plants. The putative functions of the differentially expressed genes found in this study were assigned using GO enrichment analysis, giving an insight into the functions of individual genes as well as temporal dynamics of expression (Feng et al. 2017).
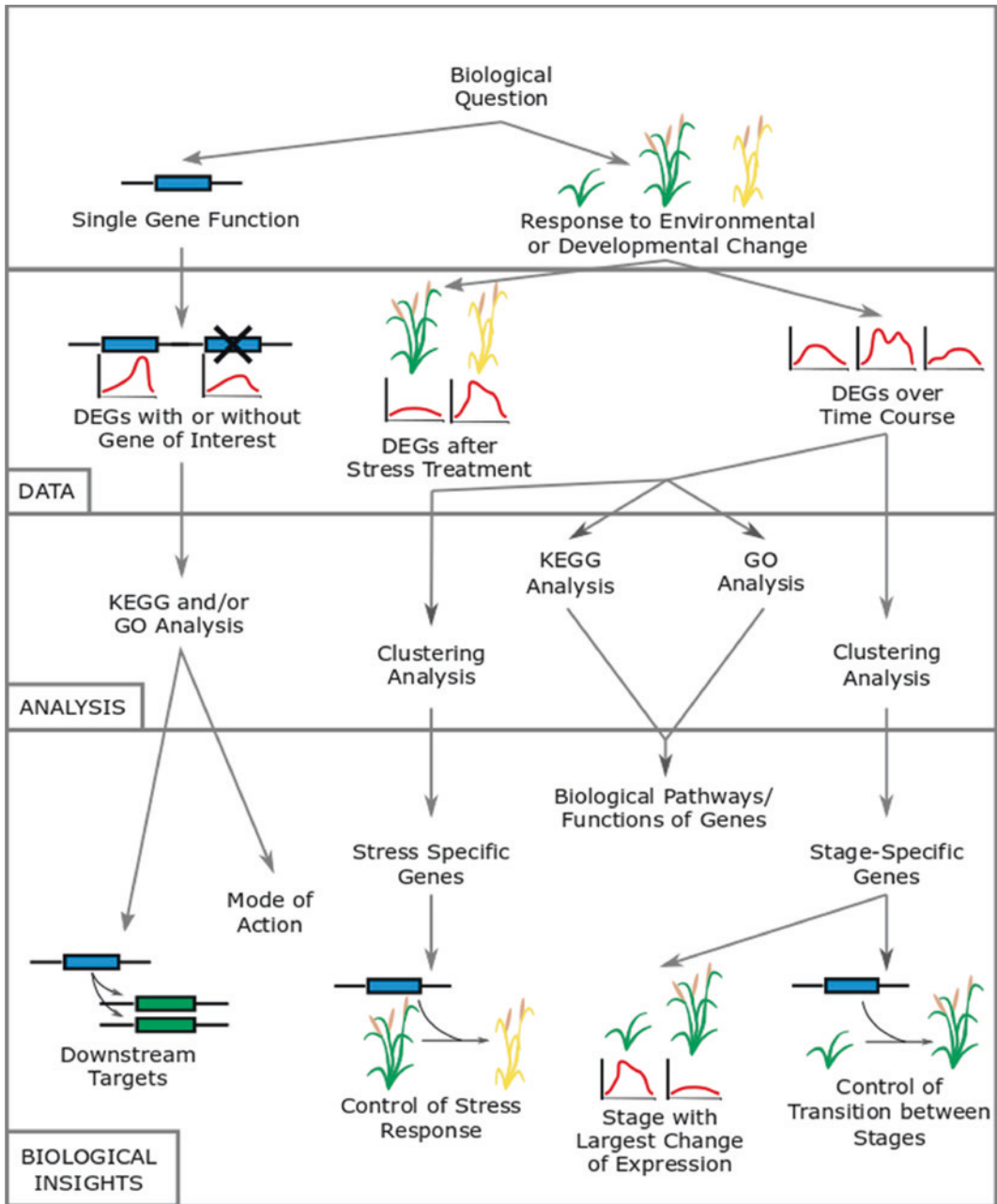
Fig. 5.2 RNA-seq is frequently used to assess the effects of altering a single gene or environmental/developmental change on gene expression. The data collected is used to identify differentially expressed genes (DEGs) which can then be analysed through methods including Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway or gene ontology (GO) analysis, or by clustering gene expression profiles. Specific exploring of differentially expressed genes, pathway and clustering information can uncover the biological pathways and mechanisms through which a gene or environmental/developmental response operates

5.6.2 Understanding the Influence of Individual Genetic Components on a Developmental Process

Understanding general expression changes during development is important, but many geneticists aim to characterise the precise effects of individual genes and RNA-seq can contribute to this goal. Flowering time is one of the best-characterised processes in wheat with many important genes identified. Transcriptomic approaches have deepened our understanding of flowering time pathways by comparing the expression profiles of wild type and plants mutated in or overexpressing key floral regulators (see also Chap. 11). For example, Pearce et al. (2016) studied the phytochrome light receptors using RNA-seq-based methods to better understand how they regulate the developmental transitions controlled by changes in light levels. Under long-day conditions, *PHYB* was found to regulate approximately six times more genes than *PHYC* and only a small number of genes were under transcriptional control of both phytochrome genes. Similarly, under short-day conditions *PHYB* influenced the transcription of approximately five times more genes than *PHYC* (Kippes et al. 2020). Surprisingly in *phyB* and *phyC* mutants flowering was accelerated under short-day conditions, which is unexpected in a long-day plant like wheat. Transcriptomic analysis revealed this may be mediated through flowering promoting genes *VRN-A1* and *PPD-B1*. This work shows that these RNA-seq transcriptome methods can uncover the functions of genes in a developmental process as well as identify downstream targets of these genes.

5.6.3 Atlases of Gene Expression

Beyond individual studies of gene expression, collating gene expression data for future analysis via gene expression atlases allows researchers to address a range of biological questions without the need to carry out more RNA-sequencing. Several different atlases have been built for wheat including the expVIP gene expression atlas which contains RNA-seq data from >1,000 RNA-seq samples, including diverse tissue types, developmental stages, cultivars and environmental conditions (Borrill et al. 2016; Ramírez-González et al. 2018). A pictorial representation of gene expression across 70 different tissue-developmental stages is also available through the wheat eFP browser which provides a powerful tool for intuitive gene expression exploration (Winter et al. 2007; Ramírez-González et al. 2018).

5.7 Response to Environmental Stress

Transcriptome analyses are also a powerful tool to understand how wheat responds to different environmental stresses, including both abiotic and biotic stresses. Genome-wide scale changes in the transcriptome can be investigated by examining the transcriptome changes after the application of the stress or differences between plants with susceptible or resistant genotypes. The effect of single genes on the response can be investigated by comparing lines with precise genetic differences such as near-isogenic lines, overexpression or mutant lines.

5.7.1 Genome-Wide Transcriptional Responses to Stress Conditions

RNA-seq has been used to characterise gene expression changes in response to a wide range of environmental stresses from pathogen infection (e.g. Zhang et al. 2014; Dobon et al. 2016) through to abiotic stresses including drought, heat, salinity and cold (e.g. Liu et al. 2015; Xiong et al. 2017; Li et al. 2018; Gálvez et al. 2019). The effects of yellow rust infection on gene expression is one of the best studied pathogen infections in wheat, at the transcriptional level. Here we will explore insights that have been gained using RNA-seq to study rust infection, which may be widely applicable to other pathosystems and to other environmental interactions.

Early studies using RNA-seq examined temporal changes in gene expression in wheat (Zhang et al. 2014), or in both wheat and the fungal pathogen itself revealing temporal interactions between host and pathogen (Dobon et al. 2016). Comparisons between susceptible and resistant lines have also proved fruitful. Infection with a mixture of powdery mildew and leaf rust revealed that a specific set of genes were downregulated only in the susceptible line. These genes had functions related to programmed cell death and response to cellular damage, indicating that the two fungal pathogens evade the wheat defense system by inducing transcriptional level changes (Poretti et al. 2021). This agrees with earlier results which examined a time course of RNA-seq in wheat plants infected with yellow rust. Immune response regulators were rapidly upregulated after yellow rust infection, but this upregulation was suppressed in subsequent timepoints. Only in resistant interactions was this suppression alleviated, while in susceptible reactions the immune response regulators continued to be suppressed (Dobon et al. 2016). This parallels the findings of Poretti et al. (2021) that specific suppression is required in susceptible wheat lines for successful infection.

Transcriptomics studies are also now leading to the identification and functional characterisation of genes involved in pathogen resistance and susceptibility. Corredor-Moreno et al. (2021) used data from 68 pathogen-infected wheat varieties to investigate genes which influence wheat rust susceptibility. Since samples were collected from different varieties, growth conditions and developmental stages, the authors clustered gene expression profiles to identify genes linked to yellow rust susceptibility. This reduced the amount of background differentially expressed genes which are not involved in the infection response, but instead are linked to variety, growth condition or developmental stage. By focussing on clusters which showed strong expression differences between the most and least susceptible cultivars, susceptibility-associated genes were identified. These susceptibility-associated genes were enriched for branched-chain amino acid (BCAA) biosynthetic genes. Comparison with publicly available data highlighted the gene *branched-chain aminotransferase 1* (*TaBCAT1*) as a candidate gene, which was ultimately validated as a susceptibility gene using mutant lines. This study highlights a new way of identifying genes with roles in infection response and shows the potential genetic variation we can find beyond the pairwise comparisons of lines with different susceptibilities, which is the more routine approach.

5.7.2 Elucidating Biological Mechanisms of Stress-Associated Genes Using Transcriptomics

It is becoming increasingly routine to characterise lines with phenotypic alternations in stress responses using RNA-seq. This can provide insight into the molecular pathways through which a gene involved in stress responses operates and identify future breeding targets downstream in the process.

Taking drought stress as an example, several studies have recently associated NAC transcription factors with drought tolerance and studied the pathways through which they act. The first NAC gene (*TaSNAC8-6A*) improved seedling stage drought tolerance (Mao et al. 2020). RNA-seq analysis in roots showed that even under well-watered conditions, genes with GO terms associated with drought, auxin and ABA responses were upregulated in lines overexpressing this gene. Under drought conditions, more genes associated with drought, auxin and ABA response were upregulated, in the overexpression line than in well-watered conditions. The authors hypothesise that these changes enhance root development and increase water use efficiency, leading to increased drought tolerance. The second NAC (*TaNAC071-A*) increased yield under drought conditions by increasing water use efficiency (Mao et al. 2022). RNA-seq in leaves revealed that stress-responsive pathways such as response to abscisic acid and response to osmotic stress were upregulated in

lines overexpressing this NAC. Furthermore, orthologs of well-established drought-inducible genes were upregulated in the overexpression lines including genes involved in stomatal closure, suggesting that *TaNAC071-A* may increase drought tolerance by more quickly closing the stomata and reducing the transpiration rate. Interestingly, a separate study revealed through RNA-seq that increasing stomatal closure under drought is a common mechanism controlled by NAC transcription factors in wheat Ma et al. (2022).

5.8 Limitations of Current Transcriptomic Studies

A common limitation in many species is that RNA-seq has generally been carried out on pooled tissue which results in the loss of a large amount of potential information from single cells or individual tissue types. For example, by sampling a whole leaf and grinding it up prior to RNA extraction, the generated expression profiles are an average across many cell types. Therefore, any spatial differences expression within a tissue cannot be observed. Until recently, large quantities of RNA were needed for RNA-seq; therefore in order to study specific cell/tissue types, labour-intensive methods had to be used to gather large quantities of material such as aleurone and endosperm from developing grain (Pfeifer et al. 2014) and developing meiocytes (Martín et al. 2018). However, the development of low input RNA-seq methods now allows gene expression studies with much reduced sample collection requirements and enables studies on very small tissue samples which were not feasible before. Low input methods were used by Backhaus et al. (2022) to investigate the gene expression patterns in different regions of the developing spike. The developing spike was dissected at double ridge and glume primordia stage into three sections (apical, central, basal) for sequencing, without any pooling of different samples required. Surprisingly Backhaus et al. (2022) found that the largest differences in the transcriptome were between the basal and apical sections, rather than between different consecutive timepoints of development. The discovery that position has a stronger effect than the developmental time point could not have been made by doing bulk-RNA-seq of the whole spike, as has been done by previous studies (e.g. Feng et al. 2017), uncovering the unique and powerful information available using this low input approach.

While the ability to sequence small samples is a major step forwards, resolution at the single-cell level is now being applied in other plant species such as *Arabidopsis* (Thibivilliers et al. 2020). However, single-cell RNA-seq (scRNA-seq) still has limitations including the complexity of the method itself, mainly the capture of single cells (Chen et al. 2019) and the risk of overamplification based on the small amount of RNA provided from a single or small number of cells (Hrdlickova et al. 2017). However, the main issue for scRNA-seq in plant transcriptomics is the need to degrade the cell wall, with the different compositions and types meaning different protocols are required (Thibivilliers et al. 2020). The application of scRNA-seq will present new opportunities for wheat research, and success in applying this method to monocots such as rice and maize (e.g. Xu et al. 2021; Zhang et al. 2021) lay the groundwork for future studies.

A second key limitation of many studies to date has been the use of glasshouse and controlled environment conditions, to minimise variations in transcriptome changes due to factors other than what is being experimentally manipulated. However, this is not necessarily indicative of gene expression during development or responses to stress in the field environment. It is becoming increasingly important to understand gene expression in real-world fluctuating environments, and field-based studies are becoming more common (e.g. Quijano et al. 2015; Li et al. 2018; Corredor-Moreno et al. 2021). Field-based studies can develop increased insight into biological pathways and provide important information for breeding. For example, a field-based experiment revealed that multiple interactive pathways that influence cold tolerance to

prepare for over-winter stress, and these complex interactions may have been missed in controlled environment conditions where changes are often abrupt (Li et al. 2018). However, variability in gene expression caused by environmental influence can be strong and make analysing changes due to a single gene difficult, as was found for the powdery mildew resistance allele *Pm3b* (Quijano et al. 2015). Therefore, researchers will need to assess the relative benefits of the realistic nature of gene expression under field conditions against the potential pitfalls for each experiment.

5.9   Constructing Gene Networks for Hypothesis Generation and Candidate Gene Identification

Although comparisons of gene expression between samples at different timepoints or in different environmental conditions can be informative, applying network approaches to understand gene interactions and pathway-level responses to environmental and developmental changes is a complementary and powerful approach. Networks can integrate a wide range of information from gene expression and co-expression through to protein-level interactions and scientific literature links (Hassani-Pak et al. 2016), but here we will focus on gene networks built mainly from gene expression measurements.

5.9.1   Co-expression Networks

Co-expression networks can be built from thousands of genes using the similarity in their expression patterns across multiple conditions to determine which genes are grouped (Fig. 5.3a). Based on "guilt-by-association" genes that belong to the same co-expression group are often considered to be co-regulated, for example by shared transcription factors, and to be part of the same biological process.

An important application of gene co-expression networks is the functional annotation of uncharacterised genes (Serin et al. 2016). The development of a high-quality reference sequence for wheat enabled the generation of detailed co-expression networks focussing on specific wheat tissues (leaf, grain, root and spike) and stress conditions (abiotic and biotic) (Ramírez-González et al. 2018). A comparison of the four tissue-specific networks revealed modules of genes which were uniquely co-expressed in the root including several genes whose orthologs regulate root development in Arabidopsis. The other genes present in these root-specific modules represent novel genes that according to "guilt-by-association" may play roles in root development. Additional studies have used co-expression networks to identify candidate genes involved in meiosis, grain development and flowering time pathways (IWGSC et al. 2018; Alabdullah et al. 2019; Chi et al. 2019).

While these studies showed the potential of co-expression networks to identify candidate genes associated with a biological process of interest, functional validation of newly identified genes was lacking. The value of these predictions has been illustrated in wheat using the disease-related network generated by Ramírez-González et al. (2018). Polturak et al. (2022) revealed that the top pathogen-induced modules contained multiple clusters of physically adjacent genes that correspond to six pathogen-induced biosynthetic pathways. Heterologous expression of these co-expressed genes in *Nicotiana benthamiana* produced flavonoids and terpenes that may play a role in defence signalling or as phytoalexins. This study shows the power of co-expression to assign functions to previously uncharacterised genes.

Several online tools have been developed which allow wheat researchers to identify genes that are co-expressed. WheatOmics allows users to search for genes co-expressed with a gene of interest in either grain or multi-tissue co-expression networks (Ma et al. 2021) and KnetMiner integrates information about co-expression
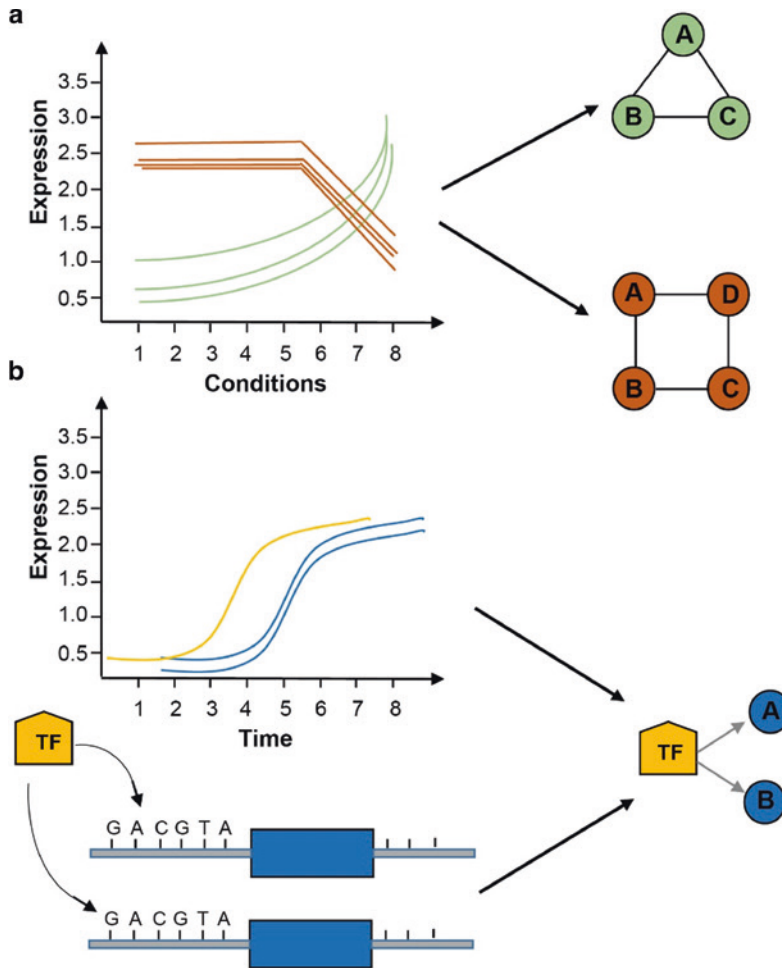
Fig. 5.3 Graphical representation of gene networks. **a** Gene co-expression networks group genes with similar expression patterns across multiple conditions. Interactions between genes (circles) can be direct or indirect. **b** Gene regulatory networks represent direct interactions between genes with directionality. In the example here, a transcription factor (TF; yellow pentagon) is expressed earlier in time and binds to the promoter sites of two downstream genes (blue); the regulatory network on the right shows the directionality of these interactions (arrowheads)

from a network built using 850 wheat RNA-seq samples with a meiosis-specific co-expression, network (IWGSC et al. 2018; Alabdullah et al. 2019; Hassani-Pak et al. 2021). Online tools are also available to construct co-expression networks using custom datasets, such as unpublished RNA-seq data including CoExpNetViz (Tzfadia et al. 2016) and Gene Network Construction Tool Kit (GeNeCK) (Zhang et al. 2019).

5.9.2   Gene Regulatory Networks

In contrast to co-expression networks, the links within gene regulatory networks (GRNs) represent direct gene interactions rather than the association of expression patterns (Fig. 5.3b). GRNs can be built using transcriptome data alone, or they can incorporate additional data types for transcription factor-DNA interactions which inform the network structure (reviewed in

Ko and Brandizzi 2020). GRNs typically have a scale-free network architecture with a few hub genes with multiple connections to other genes and many poorly connected nodes (Barabasi and Oltvai 2004). The hub genes act as master regulators of a GRN and play important roles in biological systems and therefore identifying and manipulating hub genes may enable the manipulation of a biological process of interest.

GRNs in wheat have been used to generate hypotheses about gene function and to identify hub genes which have a strong influence on a biological process. A large GRN was built using 850 RNA-seq samples to predict transcription factor-target interactions using the machine learning-based GENIE3 algorithm (Huynh-Thu et al. 2010). To test the validity of the transcription factor targets identified by GENIE3, Harrington et al. (2020) compared the target genes of the senescence-regulating transcription factor *NAM-A1* to genes differentially expressed in *nam-a1* mutant lines compared to wild type. The *NAM-A1* target genes predicted by GENIE3 overlapped considerably with the differentially expressed genes in lines with reduced *NAM-A1* expression, indicating that GENIE3 can provide biologically relevant predictions. Furthermore, additional senescence-associated transcription factors were identified by combining GENIE3 target information with independent senescence-related expression data. Similarly, combining the GENIE3 network with co-expression networks enabled the identification of candidate genes involved in root development and stress responses (Ramírez-González et al. 2018).

While the GENIE3 approach relies upon diverse RNA-seq samples from different tissues and conditions, GRNs have also proved valuable to understand developmental timeseries in wheat. A ten-timepoint time course of flag leaf senescence was sampled and the resulting RNA-seq data was used to construct a GRN using the time-aware causal structure inference algorithm (Penfold and Wild 2011; Borrill et al. 2019). Filtering the GRN for highly connected and central hub genes identified known senescence regulator *NAM-A1* amongst the 36 top-ranked

genes, indicating that this approach identified biologically relevant genes. Functional validation of *NAM-A2*, another top-ranked gene and an uncharacterised paralog of *NAM-A1*, showed the power of this approach to identify genes regulating senescence.

5.9.3   Limitations of Gene Networks

The first attempts to use gene networks in wheat have focussed on hypothesis generation and identifying candidate genes involved in a biological process of interest. While useful insights have been gained, there is still more work to be done to fully leverage the power of gene networks. To date, most gene networks in wheat have been built using gene expression data, although some other types of information are incorporated into tools such as Knetminer and inetbio (Lee et al. 2017; Hassani-Pak et al. 2021). In other species, the accuracy of networks has been improved by incorporating additional data sources such as transcription factor binding sites, open chromatin regions and protein–protein interactions (reviewed in Haque et al. 2019; Ko and Brandizzi 2020). In wheat, these types of data are becoming available, for example with the publication of accessible chromatin regions identified by ATAC-seq (Concia et al. 2020) and this information could be incorporated into future networks to improve the predictive ability.

A second challenge is the validation of gene networks in wheat. In model systems comparison to "gold standard" networks allows the accuracy of different network construction methods to be determined (Marbach et al. 2012). However, in wheat, we know little about the true topology of gene networks so validation using this approach is not possible. Instead, network predictions can be validated on an individual gene basis by examining mutant or gene-edited lines for predicted phenotypic effects (Borrill et al. 2019). Alternatively, gene interactions in the network could be tested using molecular biology approaches. Another promising

approach is to integrate several different network construction approaches which can boost the breadth and accuracy of gene interactions in biological networks (Marbach et al. 2012).

A final issue which affects wheat gene networks is that having a large polyploid genome with >110,000 genes presents practical challenges for some GRN construction techniques. Although co-expression can be carried out on thousands of genes simultaneously (e.g. IWGSC et al. 2018; Ramírez-González et al. 2018), some widely used GRN approaches only permit tens to hundreds of genes due to computational constraints. One method to circumvent this limitation is to filter genes likely to be of interest before entering them into the GRN to reduce the number of genes (e.g. Borrill et al. 2019). Alternatively, some algorithms such as GENIE3 can use tens of thousands of genes as input, although the computational steps take several weeks on a high-performance computing cluster, therefore this approach will not be accessible to all.

5.10 Conclusions and Future Outlook

The use of transcriptomics has greatly increased in wheat over the past few years, benefitting from a high-quality genome annotation and decreasing sequencing costs. Accurate gene models now simplify the analysis of transcriptomic data and increase the value of the biological information gained. While traditional studies have focussed on understanding changes in gene expression in response to environmental stresses or developmental changes, there are an increasingly varied applications of RNA-seq from identifying candidate genes by surveying genetically diverse populations through to building gene regulatory networks for hypothesis generation. Rapid developments in technologies for transcriptomics will enable us to deepen our understanding of wheat biology for example uncovering high-resolution gene expression patterns.

References

Alabdullah AK, Borrill P, Martin AC, Ramirez-Gonzalez RH, Hassani-Pak K, Uauy C, Shaw P, Moore G (2019) A co-expression network in hexaploid wheat reveals mostly balanced expression and lack of significant gene loss of homeologous meiotic genes upon polyploidization. Front Plant Sci 1325

Alonge M, Shumate A, Puiu D, Zimin AV, Salzberg SL (2020) Chromosome-scale assembly of the bread wheat genome reveals thousands of additional gene copies. Genetics 216:599–608

Backhaus AE, Lister A, Tomkins M, Adamski NM, Simmonds J, Macaulay I, Morris RJ, Haerty W, Uauy C (2022) High expression of the MADS-box gene *VRT2* increases the number of rudimentary basal spikelets in wheat. Plant Physiol 189: 1536–1552

Barabasi A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113

Borrill P, Ramirez-Gonzalez R, Uauy C (2016) expVIP: a customizable RNA-seq data analysis and visualization platform. Plant Physiol 170:2172–2186

Borrill P, Harrington SA, Simmonds J, Uauy C (2019) Identification of transcription factors regulating senescence in wheat through gene regulatory network modelling. Plant Physiol 180:1740–1755

Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo M-C, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KFX, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705–710

Brinton J, Simmonds J, Uauy C (2018) Ubiquitin-related genes are differentially expressed in isogenic lines contrasting for pericarp cell size and grain weight in hexaploid wheat. BMC Plant Biol 18:22–22

Chao S, Lazo GR, You F, Crossman CC, Hummel DD, Lui N, Laudencia-Chingcuanco D, Anderson JA, Close TJ, Dubcovsky J, Gill BS, Gill KS, Gustafson JP, Kianian SF, Lapitan NLV, Nguyen HT, Sorrells ME, McGuire PE, Qualset CO, Anderson OD (2006) Use of a large-scale Triticeae expressed sequence tag resource to reveal gene expression profiles in hexaploid wheat (*Triticum aestivum* L.). Genome 49:531–544

Chen G, Ning B, Shi T (2019) Single-cell RNA-Seq technologies and related computational data analysis. Front Genet 10

Chi Q, Guo L, Ma M, Zhang L, Mao H, Wu B, Liu X, Ramirez-Gonzalez RH, Uauy C, Appels R, Zhao H (2019) Global transcriptome analysis uncovers the gene co-expression regulation network and key genes involved in grain development of wheat (*Triticum aestivum* L.). Funct Integr Genomics 19:853–866

Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome Res 27:885–896

Concia L, Veluchamy A, Ramirez-Prado JS, Martin-Ramirez A, Huang Y, Perez M, Domenichini S, Rodriguez Granados NY, Kim S, Blein T, Duncan S, Pichot C, Manza-Mianza D, Juery C, Paux E, Moore G, Hirt H, Bergounioux C, Crespi M, Mahfouz MM, Bendahmane A, Liu C, Hall A, Raynaud C, Latrasse D, Benhamed M (2020) Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biol 21:104

Corredor-Moreno P, Minter F, Davey PE, Wegel E, Kular B, Brett P, Lewis CM, Morgan YML, Macías Pérez LA, Korolev AV, Hill L, Saunders DGO (2021) The branched-chain amino acid aminotransferase TaBCAT1 modulates amino acid metabolism and positively regulates wheat rust susceptibility. Plant Cell 33:1728–1747

Coulter M, Entizne JC, Guo W, Bayer M, Wonneberger R, Milne L, Schreiber M, Haaning A, Muehlbauer G, McCallum N, Fuller J, Simpson C, Stein N, Brown JWS, Waugh R, Zhang R (2021) BaRTv2: a highly resolved barley reference transcriptome for accurate transcript-specific RNA-seq quantification. bioRxiv: 2021.009.010.459729

Dobon A, Bunting DCE, Cabrera-Quio LE, Uauy C, Saunders DGO (2016) The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. BMC Genomics 17:380

Feng N, Song G, Guan J, Chen K, Jia M, Huang D, Wu J, Zhang L, Kong X, Geng S, Liu J, Li A, Mao L (2017) Transcriptome profiling of wheat inflorescence development from spikelet initiation to floral patterning identified stage-specific regulatory genes. Plant Physiol 174:1779–1794

Gálvez S, Mérida-García R, Camino C, Borrill P, Abrouk M, Ramírez-González RH, Biyiklioglu S, Amil-Ruiz F, Dorado G, Budak H, Gonzalez-Dugo V, Zarco-Tejada PJ, Appels R, Uauy C, Hernandez P, The I (2019) Hotspots in the genomic architecture of field drought responses in wheat as breeding targets. Funct Integr Genomics 19:295–309

Geng S, Li A, Tang L, Yin L, Wu L, Lei C, Guo X, Zhang X, Jiang G, Zhai W, Wei Y, Zheng Y, Lan X, Mao L (2013) TaCPK2-A, a calcium-dependent protein kinase gene that is required for wheat powdery mildew resistance enhances bacterial blight resistance in transgenic rice. J Exp Bot 64:3125–3136

Haque S, Ahmad JS, Clark NM, Williams CM, Sozzani R (2019) Computational prediction of gene regulatory networks in plant growth and development. Curr Opin Plant Biol 47:96–105

Harrington SA, Backhaus AE, Singh A, Hassani-Pak K, Uauy C (2020) The wheat GENIE3 network provides biologically-relevant information in polyploid wheat. G3: Genes Genomes, Genetics 10:3675–3686

Hassani-Pak K, Castellote M, Esch M, Hindle M, Lysenko A, Taubert J, Rawlings C (2016) Developing integrated crop knowledge networks to advance candidate gene discovery. Appl Transl Genomics 11:18–26

Hassani-Pak K, Singh A, Brandizi M, Hearnshaw J, Parsons JD, Amberkar S, Phillips AL, Doonan JH, Rawlings C (2021) KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. Plant Biotechnol J 19:1670–1678

He F, Wang W, Rutter WB, Jordan KW, Ren J, Taagen E, DeWitt N, Sehgal D, Sukumaran S, Dreisigacker S, Reynolds M, Halder J, Sehgal SK, Liu S, Chen J, Fritz A, Cook J, Brown-Guedira G, Pumphrey M, Carter A, Sorrells M, Dubcovsky J, Hayden MJ, Akhunova A, Morrell PL, Szabo J, Rouse M, Akhunov E (2022) Genomic variants affecting homoeologous gene expression dosage contribute to agronomic trait variation in allopolyploid wheat. Nat Commun 13:826

Hrdlickova R, Toloue M, Tian B (2017) RNA-Seq methods for transcriptome analysis. Wires RNA 8:e1364

Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P (2010) Inferring regulatory networks from expression data using tree-based methods. PLoS ONE 5:e12776

IWGSC TIWGSC, Mayer Klaus FX, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski Adam J, Sourdille P, Endo Takashi R, Kubaláková M, Číhalíková J, Dubská Z, Vrána J, Šperková R, Šimková H, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh Nagendra K, Khurana J, Akhunov E, Choulet F, Alberti A, Barbe V, Wincker P, Kanamori H, Kobayashi F, Itoh T, Matsumoto T, Sakai H, Tanaka T, Wu J, Ogihara Y, Handa H, Maclachlan PR, Sharpe A, Klassen D, Edwards D, Batley J, Olsen O-A, Sandve Simen R, Lien S, Steuernagel B, Wulff B, Caccamo M, Ayling S, Ramirez-Gonzalez Ricardo H, Clavijo Bernardo J, Wright J, Pfeifer M, Spannagl M, Martis Mihaela M, Mascher M, Chapman J, Poland Jesse A, Scholz U, Barry K, Waugh R, Rokhsar Daniel S, Muehlbauer Gary J, Stein N, Gundlach H, Zytnicki M, Jamilloux V, Quesneville H, Wicker T, Faccioli P, Colaiacovo M, Stanca Antonio M, Budak H, Cattivelli L, Glover N, Pingault L, Paux E, Sharma S, Appels R, Bellgard M, Chapman B, Nussbaumer T, Bader Kai C, Rimbert H, Wang S, Knox R, Kilian A, Alaux M, Alfama F, Couderc L, Guilhot N, Viseux C, Loaec M, Keller B, Praud S (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

IWGSC, Appels R, Eversole K, Stein N, Feuillet C, Keller B, Rogers J, Pozniak CJ, Choulet F,

Distelfeld A, Poland J, Ronen G, Sharpe AG, Barad O, Baruch K, Keeble-Gagnère G, Mascher M, Ben-Zvi G, Josselin A-A, Himmelbach A, Balfourier F, Gutierrez-Gonzalez J, Hayden M, Koh C, Muehlbauer G, Pasam RK, Paux E, Rigault P, Tibbits J, Tiwari V, Spannagl M, Lang D, Gundlach H, Haberer G, Mayer KFX, Ormanbekova D, Prade V, Šimková H, Wicker T, Swarbreck D, Rimbert H, Felder M, Guilhot N, Kaithakottil G, Keilwagen J, Leroy P, Lux T, Twardziok S, Venturini L, Juhász A, Abrouk M, Fischer I, Uauy C, Borrill P, Ramirez-Gonzalez RH, Arnaud D, Chalabi S, Chalhoub B, Cory A, Datla R, Davey MW, Jacobs J, Robinson SJ, Steuernagel B, Ex Fv, Wulff BBH, Benhamed M, Bendahmane A, Concia L, Latrasse D, Bartoš J, Bellec A, Berges H, Doležel J, Frenkel Z, Gill B, Korol A, Letellier T, Olsen O-A, Singh K, Valárik M, Vossen Evd, Vautrin S, Weining S, Fahima T, Glikson V, Raats D, Číhalíková J, Toegelová H, Vrána J, Sourdille P, Darrier B, Barabaschi D, Cattivelli L, Hernandez P, Galvez S, Budak H, Jones JDG, Witek K, Yu G, Small I, Melonek J, Zhou R, Belova T, Kanyuka K, King R, Nilsen K, Walkowiak S, Cuthbert R, Knox R, Wiebe K, Xiang D, Rohde A, Golds T, Čížková J, Akpinar BA, Biyiklioglu S, Gao L, N'Daiye A, Kubaláková M, Šafář J, Alfama F, Adam-Blondon A-F, Flores R, Guerche C, Loaec M, Quesneville H, Condie J, Ens J, Maclachlan R, Tan Y, Alberti A, Aury J-M, Barbe V, Couloux A, Cruaud C, Labadie K, Mangenot S, Wincker P, Kaur G, Luo M, Sehgal S, Chhuneja P, Gupta OP, Jindal S, Kaur P, Malik P, Sharma P, Yadav B, Singh NK, Khurana JP, Chaudhary C, Khurana P, Kumar V, Mahato A, Mathur S, Sevanthi A, Sharma N, Tomar RS, Holušová K, Plíhal O, Clark MD, Heavens D, Kettleborough G, Wright J, Balcárková B, Hu Y, Salina E, Ravin N, Skryabin K, Beletsky A, Kadnikov V, Mardanov A, Nesterov M, Rakitin A, Sergeeva E, Handa H, Kanamori H, Katagiri S, Kobayashi F, Nasuda S, Tanaka T, Wu J, Cattonaro F, Jiumeng M, Kugler K, Pfeifer M, Sandve S, Xun X, Zhan B, Batley J, Bayer PE, Edwards D, Hayashi S, Tulpová Z, Visendi P, Cui L, Du X, Feng K, Nie X, Tong W, Wang L (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361:eaar7191. https://doi.org/10.1126/science.aar7191

Kawaura K, Mochida K, Enju A, Totoki Y, Toyoda A, Sakaki Y, Kai C, Kawai J, Hayashizaki Y, Seki M, Shinozaki K, Ogihara Y (2009) Assessment of adaptive evolution between wheat and rice as deduced from full-length common wheat cDNA sequence data and expression patterns. BMC Genomics 10:271–271

Kippes N, VanGessel C, Hamilton J, Akpinar A, Budak H, Dubcovsky J, Pearce S (2020) Effect of phyB and phyC loss-of-function mutations on the wheat transcriptome under short and long day photoperiods. BMC Plant Biol 20:297

Ko DK, Brandizzi F (2020) Network-based approaches for understanding gene regulation and function in plants. Plant J 104:302–317

Kuo TCY, Hatakeyama M, Tameshige T, Shimizu KK, Sese J (2020) Homeolog expression quantification methods for allopolyploids. Brief Bioinform 21:395–407

Lee T, Hwang S, Kim CY, Shim H, Kim H, Ronald PC, Marcotte EM, Lee I (2017) WheatNet: a genome-scale functional network for hexaploid bread wheat, *Triticum aestivum*. Mol Plant 10:1133–1136

Li Q, Byrns B, Badawi MA, Diallo AB, Danyluk J, Sarhan F, Laudencia-Chingcuanco D, Zou J, Fowler DB (2018) Transcriptomic insights into phenological development and cold tolerance of wheat grown in the field. Plant Physiol 176:2376–2394

Liu Z, Xin M, Qin J, Peng H, Ni Z, Yao Y, Sun Q (2015) Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). BMC Plant Biol 15:1–20

Ma S, Wang M, Wu J, Guo W, Chen Y, Li G, Wang Y, Shi W, Xia G, Fu D (2021) WheatOmics: a platform combining multiple omics data to accelerate functional genomics studies in wheat. Mol Plant 14:1965–1968

Ma J, Tang X, Sun B, Wei J, Ma L, Yuan M, Zhang D, Shao Y, Li C, Chen K-M, Jiang L (2022) A NAC transcription factor, TaNAC5D-2, acts as a positive regulator of drought tolerance through regulating water loss in wheat (*Triticum aestivum* L.). Environ Exp Bot 196:104805

Manickavelu A, Kawaura K, Oishi K, Shin-I T, Kohara Y, Yahiaoui N, Keller B, Abe R, Suzuki A, Nagayama T, Yano K, Ogihara Y (2012) Comprehensive functional analyses of expressed sequence tags in common wheat (*Triticum aestivum*). DNA Res 19:165–177

Mao H, Li S, Wang Z, Cheng X, Li F, Mei F, Chen N, Kang Z (2020) Regulatory changes in TaSNAC8-6A are associated with drought tolerance in wheat seedlings. Plant Biotechnol J 18:1078–1092

Mao H, Li S, Chen B, Jian C, Mei F, Zhang Y, Li F, Chen N, Li T, Du L, Ding L, Wang Z, Cheng X, Wang X, Kang Z (2022) Variation in cis-regulation of a NAC transcription factor contributes to drought tolerance in wheat. Mol Plant 15:276–292

Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Aderhold A, Allison KR, Bonneau R, Camacho DM, Chen Y, Collins JJ, Cordero F, Costello JC, Crane M, Dondelinger F, Drton M, Esposito R, Foygel R, de la Fuente A, Gertheiss J, Geurts P, Greenfield A, Grzegorczyk M, Haury A-C, Holmes B, Hothorn T, Husmeier D, Huynh-Thu VA, Irrthum A, Kellis M, Karlebach G, Küffner R, Lèbre S, De Leo V, Madar A, Mani S, Marbach D, Mordelet F, Ostrer H, Ouyang Z, Pandya R, Petri T, Pinna A, Poultney CS, Prill RJ, Rezny S, Ruskin HJ, Saeys Y, Shamir R, Sîrbu A, Song M,

Soranzo N, Statnikov A, Stolovitzky G, Vega N, Vera-Licona P, Vert J-P, Visconti A, Wang H, Wehenkel L, Windhager L, Zhang Y, Zimmer R, Kellis M, Collins JJ, Stolovitzky G, The DC (2012) Wisdom of crowds for robust gene network inference. Nat Methods 9:796–804

Martín AC, Borrill P, Higgins J, Alabdullah A, Ramírez-González RH, Swarbreck D, Uauy C, Shaw P, Moore G (2018) Genome-wide transcription during early wheat meiosis is independent of synapsis, ploidy level, and the Ph1 locus. Front Plant Sci 9

Mochida K, Kawaura K, Shimosaka E, Kawakami N, Shin IT, Kohara Y, Yamazaki Y, Ogihara Y (2006) Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat. Mol Genet Genomics 276:304–312

Mochida K, Yoshida T, Sakurai T, Ogihara Y, Shinozaki K (2009) TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. Plant Physiol 150:1135–1146

Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan C-KK, Visendi P, Lai K, Doležel J, Batley J, Edwards D (2017) The pangenome of hexaploid bread wheat. Plant J 90:1007–1013

Ogihara Y, Mochida K, Nemoto Y, Murai K, Yamazaki Y, Shin-I T, Kohara Y (2003) Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. Plant J 33:1001–1011

Pearce S, Kippes N, Chen A, Debernardi JM, Dubcovsky J (2016) RNA-seq studies using wheat PHYTOCHROME B and PHYTOCHROME C mutants reveal shared and specific functions in the regulation of flowering and shade-avoidance pathways. BMC Plant Biol 16:141

Penfold CA, Wild DL (2011) How to infer gene networks from expression profiles, revisited. Interface Focus 1:857–870

Pfeifer M, Kugler Karl G, Sandve Simen R, Zhan B, Rudi H, Hvidsten Torgeir R, Mayer Klaus FX, Olsen O-A (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345:1250091

Polturak G, Dippe M, Stephenson Michael J, Chandra Misra R, Owen C, Ramirez-Gonzalez Ricardo H, Haidoulis John F, Schoonbeek H-J, Chartrain L, Borrill P, Nelson David R, Brown James KM, Nicholson P, Uauy C, Osbourn A (2022) Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. Proc Natl Acad Sci 119:e2123299119

Poretti M, Sotiropoulos AG, Graf J, Jung E, Bourras S, Krattinger SG, Wicker T (2021) Comparative transcriptome analysis of wheat lines in the field reveals multiple essential biochemical pathways suppressed by obligate pathogens. Front Plant Sci 12

Quijano CD, Brunner S, Keller B, Gruissem W, Sautter C (2015) The environment exerts a greater influence than the transgene on the transcriptome of field-grown wheat expressing the Pm3b allele. Transgenic Res 24:87–97

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, Khedikar Y, Robinson SJ, Cory AT, Florio T, Concia L, Juery C, Schoonbeek H, Steuernagel B, Xiang D, Ridout CJ, Chalhoub B, Mayer KFX, Benhamed M, Latrasse D, Bendahmane A, Wulff BBH, Appels R, Tiwari V, Datla R, Choulet F, Pozniak CJ, Provart NJ, Sharpe AG, Paux E, Spannagl M, Bräutigam A, Uauy C, Korol A, Sharpe Andrew G, Juhász A, Rohde A, Bellec A, Distelfeld A, Akpinar Bala A, Keller B, Darrier B, Gill B, Chalhoub B, Steuernagel B, Feuillet C, Chaudhary C, Uauy C, Pozniak C, Ormanbekova D, Xiang D, Latrasse D, Swarbreck D, Barabaschi D, Raats D, Sergeeva E, Salina E, Paux E, Cattonaro F, Choulet F, Kobayashi F, Keeble-Gagnere G, Kaur G, Muehlbauer G, Kettleborough G, Yu G, Šimková H, Gundlach H, Berges H, Rimbert H, Budak H, Handa H, Small I, Bartoš J, Rogers J, Doležel J, Keilwagen J, Poland J, Melonek J, Jacobs J, Wright J, Jones Jonathan DG, Gutierrez-Gonzalez J, Eversole K, Nilsen K, Mayer Klaus FX, Kanyuka K, Singh K, Gao L, Concia L, Venturini L, Cattivelli L, Spannagl M, Mascher M, Hayden M, Abrouk M, Alaux M, Luo M, Valárik M, Benhamed M, Singh Nagendra K, Sharma N, Guilhot N, Ravin N, Stein N, Olsen O-A, Gupta Om P, Khurana P, Chhuneja P, Bayer Philipp E, Borrill P, Leroy P, Rigault P, Sourdille P, Hernandez P, Flores R, Ramirez-Gonzalez Ricardo H, King R, Knox R, Appels R, Zhou R, Walkowiak S, Galvez S, Biyiklioglu S, Nasuda S, Sandve S, Chalabi S, Weining S, Sehgal S, Jindal S, Belova T, Letellier T, Wicker T, Tanaka T, Fahima T, Barbe V, Tiwari V, Kumar V, Tan Y (2018) The transcriptional landscape of polyploid wheat. Science 361:eaar6089. https://doi.org/10.1126/science.aar6089

Serin EA, Nijveen H, Hilhorst HW, Ligterink W (2016) Learning from co-expression networks: possibilities and challenges. Front Plant Sci 7:444

Thibivilliers S, Anderson D, Libault M (2020) Isolation of plant root nuclei for single cell RNA sequencing. Curr Protoc Plant Biol 5:e20120

Tzfadia O, Diels T, De Meyer S, Vandepoele K, Aharoni A, Van de Peer Y (2016) CoExpNetViz: comparative co-expression networks construction and visualization tool. Front Plant Sci 6:1194

Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D, Klymiuk V, Byrns B, Gundlach H, Bandi V, Siri JN, Nilsen K, Aquino C, Himmelbach A, Copetti D, Ban T, Venturini L, Bevan M, Clavijo B, Koo D-H, Ens J, Wiebe K, N'Diaye A, Fritz AK, Gutwin C, Fiebig A, Fosker C, Fu BX, Accinelli GG, Gardner KA, Fradgley N, Gutierrez-Gonzalez J, Halstead-Nussloch G, Hatakeyama M, Koh CS, Deek J, Costamagna AC, Fobert P, Heavens

D, Kanamori H, Kawaura K, Kobayashi F, Krasileva K, Kuo T, McKenzie N, Murata K, Nabeka Y, Paape T, Padmarasu S, Percival-Alwyn L, Kagale S, Scholz U, Sese J, Juliana P, Singh R, Shimizu-Inatsugi R, Swarbreck D, Cockram J, Budak H, Tameshige T, Tanaka T, Tsuji H, Wright J, Wu J, Steuernagel B, Small I, Cloutier S, Keeble-Gagnère G, Muehlbauer G, Tibbets J, Nasuda S, Melonek J, Hucl PJ, Sharpe AG, Clark M, Legg E, Bharti A, Langridge P, Hall A, Uauy C, Mascher M, Krattinger SG, Handa H, Shimizu KK, Distelfeld A, Chalmers K, Keller B, Mayer KFX, Poland J, Stein N, McCartney CA, Spannagl M, Wicker T, Pozniak CJ (2020) Multiple wheat genomes reveal global variation in modern breeding. Nature 588:277–283

Wang K, Wang D, Zheng X, Qin A, Zhou J, Guo B, Chen Y, Wen X, Ye W, Zhou Y, Zhu Y (2019) Multi-strategic RNA-seq analysis reveals a high-resolution transcriptional landscape in cotton. Nat Commun 10:4714

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An "electronic fluorescent picto-graph" browser for exploring and analyzing large-scale biological data sets. PLoS ONE 2:e718

Xiong H, Guo H, Xie Y, Zhao L, Gu J, Zhao S, Li J, Liu L (2017) RNAseq analysis reveals pathways and candidate genes associated with salinity toler-ance in a spaceflight-induced wheat mutant. Sci Rep 7:2731–2731

Xu X, Crow M, Rice BR, Li F, Harris B, Liu L, Demesa-Arevalo E, Lu Z, Wang L, Fox N, Wang X, Drenkow J, Luo A, Char SN, Yang B, Sylvester AW, Gingeras TR, Schmitz RJ, Ware D, Lipka AE, Gillis J, Jackson D (2021) Single-cell RNA sequencing of develop-ing maize ears facilitates functional analysis and trait candidate gene discovery. Dev Cell 56:557-568.e556

Yu Y, Zhu D, Ma C, Cao H, Wang Y, Xu Y, Zhang W, Yan Y (2016) Transcriptome analysis reveals key differen-tially expressed genes involved in wheat grain devel-opment. Crop J 4:92–106

Zhang H, Yang Y, Wang C, Liu M, Li H, Fu Y, Wang Y, Nie Y, Liu X, Ji W (2014) Large-scale transcriptome comparison reveals distinct gene activations in wheat responding to stripe rust and powdery mildew. BMC Genomics 15:1–14

Zhang M, Li Q, Yu D, Yao B, Guo W, Xie Y, Xiao G (2019) GeNeCK: a web server for gene network con-struction and visualization. BMC Bioinform 20:1–7

Zhang T-Q, Chen Y, Liu Y, Lin W-H, Wang J-W (2021) Single-cell transcriptome atlas and chromatin acces-sibility landscape reveal differentiation trajectories in the rice root. Nat Commun 12:2053

Zhu T, Wang L, Rimbert H, Rodriguez JC, Deal KR, De Oliveira R, Choulet F, Keeble-Gagnère G, Tibbits J, Rogers J, Eversole K, Appels R, Gu YQ, Mascher M, Dvorak J, Luo M-C (2021) Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. Plant J 107:303–314

Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL (2017) The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aesti-vum*. GigaScience 6:1–7