frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 30 June 2023
DOI 10.3389/fpls.2023.1135455

OPEN ACCESS

EDITED BY
Lewis Lukens,
University of Guelph, Canada

REVIEWED BY
Chuang Ma,
Northwest A&F University, China
Xiujun Zhang,
Wuhan Botanical Garden (CAS), China

*CORRESPONDENCE
Jennifer Lachowiec
jennifer.lachowiec@montana.edu

RECEIVED 31 December 2022
ACCEPTED 12 June 2023
PUBLISHED 30 June 2023

CITATION
Upton RN, Correr FH, Lile J, Reynolds GL, Falaschi K, Cook JP and Lachowiec J (2023) Design, execution, and interpretation of plant RNA-seq analyses. *Front. Plant Sci.* 14:1135455. doi: 10.3389/fpls.2023.1135455

Design, execution, and interpretation of plant RNA-seq analyses

Racheal N. Upton, Fernando H. Correr, Jared Lile, Gillian L. Reynolds, Kira Falaschi, Jason P. Cook and Jennifer Lachowiec*

Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT, United States

Genomics has transformed our understanding of the genetic architecture of traits and the genetic variation present in plants. Here, we present a review of how RNA-seq can be performed to tackle research challenges addressed by plant sciences. We discuss the importance of experimental design in RNA-seq, including considerations for sampling and replication, to avoid pitfalls and wasted resources. Approaches for processing RNA-seq data include quality control and counting features, and we describe common approaches and variations. Though differential gene expression analysis is the most common analysis of RNA-seq data, we review multiple methods for assessing gene expression, including detecting allele-specific gene expression and building co-expression networks. With the production of more RNA-seq data, strategies for integrating these data into genetic mapping pipelines is of increased interest. Finally, special considerations for RNA-seq analysis and interpretation in plants are needed, due to the high genome complexity common across plants. By incorporating informed decisions throughout an RNA-seq experiment, we can increase the knowledge gained.

KEYWORDS

differential expression, co-expression networks, allele-specific variation, QTL mapping, experimental design

1 Introduction

The production and analysis of transcriptomic data has become the norm in plant sciences. In 2022, over 5700 articles were identified from the search terms "plant transcriptomics" on the PubMed database, and over 23% of the total data available in the Sequence Read Archive is RNA-seq for algae and land plants (Julca et al., 2022). With thousands of public datasets already available and the on-going generation of new transcriptomes, deep knowledge of plant physiology, biochemistry, development, evolution, and more can be gained through careful analysis.

Technologies are rapidly improving to study transcriptomes and becoming more cost effective to be deployed across diverse plants species. While microarrays and RNA-seq are

the most common approaches to assess transcriptomes, long-read approaches including Iso-Seq (Schaarschmidt et al., 2020) and Nanopore direct RNA sequencing (Liang et al., 2021; Jain et al., 2022) can also be used to examine transcript isoforms and splicing events. Single-cell RNA-seq (scRNA-seq) can be used to analyze the entire mRNA profile of a single cell, allowing for discovery of new cell types, understanding of cell-to-cell variability, and study of rare cell types (Cuperus, 2022). This review focuses on mRNA transcriptomics as it is the most widely used approach for studying the transcriptome of plants.

For RNA-seq findings to be impactful, careful consideration in the design and analysis of the experiments are critical. In this review, we highlight considerations for new RNA-seq experiments emphasizing experimental design and best practices for processing RNA-seq data. Gene expression determined from RNA-seq represents an endophenotype, molecular phenotypes intermediate to genotype and organismal traits (Mackay et al., 2009). Relating endophenotypes to the underlying genetics and more derived phenotypes is a growing area of research (Mackay et al., 2009; Guo et al., 2016). We therefore, lay out various analyses of gene expression and describe approaches that integrate RNA-seq with gene mapping approaches (Guo et al., 2019; Huang et al., 2022; Jiang et al., 2022) with the intent of increasing the information gained.

2 Experimental considerations for successful RNA-seq studies in plants

Thoughtful RNA-seq experimental design is critical for the high-quality data needed to answer complex biological questions. Insufficient experimental design for RNA-seq experiments may lead to makeshift analyses to circumvent inappropriate procedures. These issues have been discussed extensively in human medical literature (Leek et al., 2010; Fang and Cui, 2011; Robles et al., 2012; Williams et al., 2014), but less so in plant sciences. In addition to the standard considerations of experimental design, RNA-seq requires attention to tissue sampling strategy and the impacts of sample processing batch effects. Batch effects refer to the technical artifacts that may be present across a set of samples processed simultaneously. Lack of consistency in sample collection or inattention to batch effects can lead to many complications including lack of statistical power, technical or biological artifacts, and lack of conclusive results. Below we detail considerations that can be instrumental to the success of RNA-seq.

2.1 Experimental design concerns

2.1.1 Replication

The definition of a single replicate can vary across laboratory and field-based studies. The experimental unit can range from a single cellular component to a single plant to a collection of multiple plants within a plot, making the definition of sample replicates highly dependent on the system of interest. Similarly, experimental units in RNA-seq experiments can vary and may differ depending on the experimental goal (Conesa et al., 2016). In experimental design parlance, a factor refers to a single categorical variable manipulated in a study. If there is an experiment with one factor with k levels, then the study has k treatments. When there is a second factor with j levels applied in combination with factor k, there are k x j treatments, and so on (Hoshmand, 2006). When multiple experimental units receive the same treatment, they are considered replicates, and contrasts can be made between varying factor levels or treatments.

With certain experimental goals, like capturing shared patterns of expression, it may be beneficial to more broadly define replicates. For example, if multiple genotypes share the same direction of gene expression change across the levels of another factor, it may be useful to redefine replicates not by a single genotype (Figure 1A), but instead across multiple genotypes (Figure 1B) (Kasirajan et al., 2018; Correr et al., 2020). Kasirajan et al. (2018) assessed differential expression (DE) between upper and lower sugarcane internodes of two groups of genotypes defined by high or low fiber content, and they used different genotypes within the same fiber content class as replicates. Grouping genotypes with shared functional characteristics allowed the identification of shared candidate regulatory mechanisms.

Once the composition of replicate is determined, the number of replicates to examine is also important for a successful study. In general, replicate number has been found to have a stronger impact on differential expression analysis than sequencing depth (Lamarre et al., 2018). When RNA-seq is used for hypothesis generation, small numbers of samples—usually three—can be sufficient to perform the statistical analysis (Van den Berge et al., 2019). An extensive experiment, exploring the replication needed for RNA-seq, advises 6-12 replicates for each treatment, especially if genes with small changes in expression may be relevant (i.e., transcription factors) (Schurch et al., 2016). Because outlying replicates with poor correlations to other samples can be identified during processing, higher replication can support removing these samples with sufficient replication remaining for downstream analyses (Gierliński et al., 2015).

2.1.2 RNA-seq sensitivity to batch effects

Next-generation sequencing data is highly sensitivity to biological variation and technical artifacts, and RNA-seq is no exception. Minimizing potential sources of unwanted variation during sample collection and preparation enables meaningful interpretation (Zhou et al., 2019). Technical processing of RNA-seq is not only sensitive to the facility and instrument used, but also can be sensitive to the reagent lot, date of sequencing and other processing factors (Molania et al., 2022). Simple experimental design approaches to overcome batch effects include randomization of samples at the library preparation step and across sequencing runs (Conesa et al., 2016). If batch effects are unavoidable, additional experimental design elements can be added, and analysis methods for Removing Unwanted Variation (RUV) can detect and remove batch effects in experiments. RUV approaches use known concentrations of spike-in controls (Risso et al., 2014) to account for batch effects. Additionally, RUV for batch effects can be possible without spike-ins based on presumed

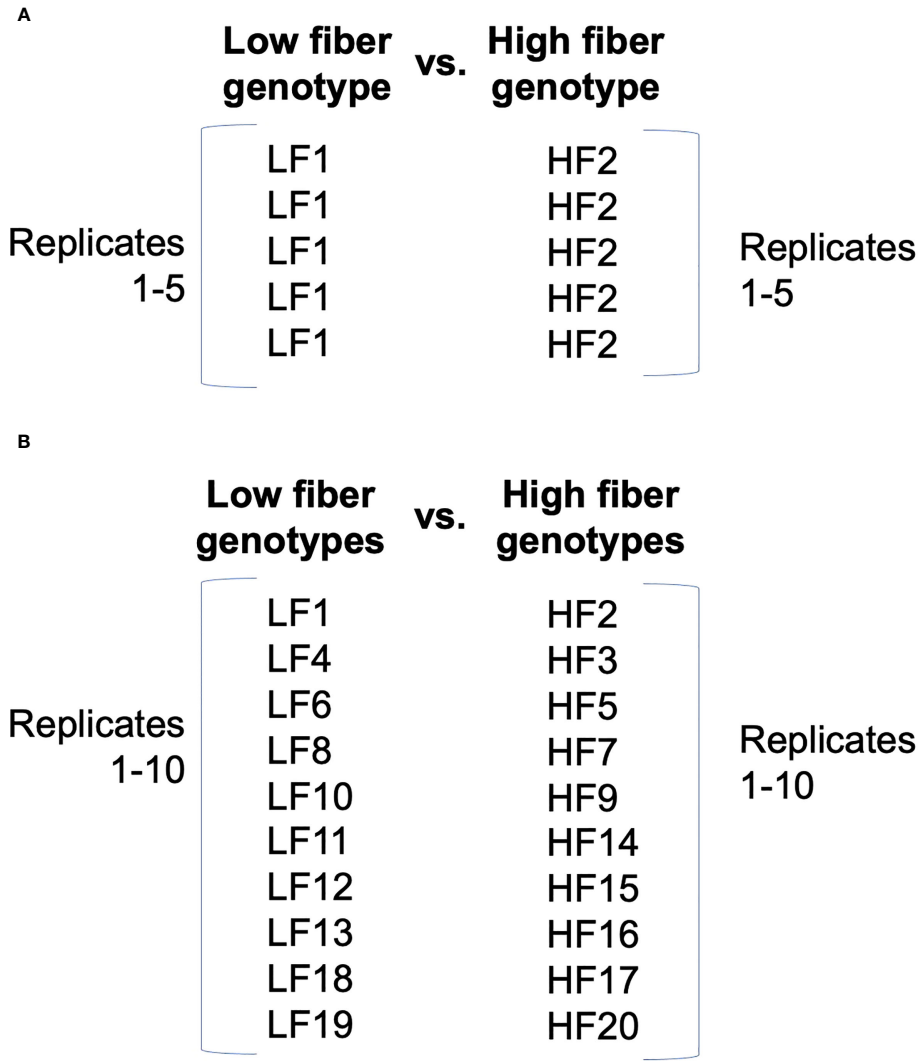

FIGURE 1
Contrasting definitions of replicates in experimental design. The high-fiber (HF) and low-fiber (LF) classes represent two levels of one factor in the experimental design that are of interest to contrast. Unique genotypes are indicted with numbers. **(A)** Typical design in which the same treatment is applied to multiple experimental units to comprise replicates. **(B)** Alternative design in which different genotypes comprise replicates. Design is from Kasirajan et al. (2018).

housekeeping genes (Risso et al., 2014; Zhang et al., 2020), *in silico* pseudoreplication approaches (Molania et al., 2022), or modeling known sample covariates (Risso et al., 2014).

2.2 Sample source considerations

Transcriptomes are highly sensitive, with widespread variability among experimental units exposed to the same treatment detected in even highly controlled environments (Cortijo et al., 2019). Minimizing undesired transcriptional variation can maximize the success of RNA-seq experiments.

2.2.1 Inter-generational impacts on sample sources

It is best practice that all seed needed for an RNA-seq experiment is produced in a common environment the generation prior to the experiment. The growth conditions of the prior generation is an important consideration, as the environments experienced by parental plants can have inter- and even transgenerational effects (Lämke and Bäurle, 2017). A newly formed seed is not a blank slate; the parental environment can affect germ cells and supporting tissues, influence the resulting seed and even the adult plant (Galloway, 2005; Donohue, 2009; Wulff, 2017). For example, parent plants of *Arabidopsis thaliana* grown under higher temperature and radiance produced offspring with higher shoot biomass than parents grown in the lower temperature and radiance condition (Andalo et al., 1999). An RNA-seq experiment designed to compare genotypes could suffer from confounding factors if the environments from which the seed were sourced varied. Another approach to avoid inter-generational effects altogether is to use clones vegetatively propagated from a single individual and apply treatments across clones.

2.2.2 Considerations of composite tissues and bulk RNA-seq

RNA-seq is generally performed on composite samples comprised of multiple tissues or organs, termed bulk RNA-seq. Different tissues and cell types can have varied expression of critical genes that may be masked when measuring pooled transcript levels thus "washing out" the signal of relevant genes in composite samples, which can limit the usefulness of RNA-seq (Johnson et al., 2013). Limitations to biological interpretation with composite samples have been recognized for some time through work focusing on microarray comparisons between cell types and whole tissues. In root cell types identified through reporter gene expression and isolated with fluorescence-activated cell sorting, over 50% of genes with differential expression were repressed in one root cell type and induced in another (Gifford et al., 2008). Pooling of composite cells and tissues may obscure the identification of relevant and causal transcripts that could contribute to improved quality and yield traits but is nearly unavoidable (Schon and Nodine, 2017).

Additional sources of transcriptional variation in composite samples complicate matters further. Varying cell types can simply have different concentrations of total RNA (Baker et al., 1990; Vennapusa et al., 2020; Walsh et al., 2020). For example, within seeds, embryos have high transcriptional activity while the endosperm are less active (Giacomello et al., 2017; Palovaara et al., 2017). Further, RNA is more difficult to extract from some plant tissues and species, such as those rich in polysaccharides or those containing high concentrations of secondary metabolites (Salzman et al., 1999; Gao et al., 2001). Therefore, composite samples with starchy tissues may have transcripts that poorly represent these components due to extraction difficulties. The sources of variation in composite samples can limit the true utility of RNA-seq. Regardless of the challenges of composite tissues studied with bulk RNA-seq, pinpointing the choice of tissue to sample is best informed by the scientific question of interest and hypothesized outcomes supported by independent biological data.

2.2.3 Minimizing diurnal and circadian effects

Throughout the course of a day, temperature, sunlight, and water potential all change, impacting plants (Hotta, 2021). Circadian variation is of particular importance when considering photosynthetic tissues. In plants, changes in the transcriptome due to circadian rhythm can cause a 25% fluctuation in differentially expressed genes (Hayes et al., 2010; Hudson, 2010), and mRNA decay rates in *Arabidopsis thaliana* vary widely from just under four minutes to over 24 hours with median half-lives of 2-4 hours (Narsai et al., 2007; Sorenson et al., 2018). Therefore, the timing of RNA-seq studies must be highly controlled during sample collection. In studies where there are multiple treatments and circadian rhythm is not the focus, samples need to be collected at the same time to avoid confounding variance due to time of day. Similarly, for temporal studies across multiple days, samples need to be taken at the same time of day. For some species, a database of circadian controlled gene expression is available for post-hoc correction (Li et al., 2016), but a conscientious design is a superior approach.

3 From reads to genetic features

After thoughtful RNA-seq experimental design, sample collection, and generation of RNA-seq reads, it can be overwhelming to approach the myriad of different tools available for analysis. In this section we aim to provide a general overview of the different approaches for quantifying the expression of genetic features (genes or transcript isoforms) from raw sequencing reads. The software available are categorized by purpose—quality control, alignment, or quantification—and the genomic information available for the organism. The software described below is commonly used for RNA-seq pipelines; however, this is not a complete list of all software as the tools available are numerous and new software is being developed rapidly (Van den Berge et al., 2019).

3.1 Preprocessing and quality control

Preprocessing comprises the first step for any sequencing analysis. This stage involves the removal of technical artifacts such as adaptors, PhiX sequence, rRNA sequences, assessing sequence quality and if necessary, quality trimming. Commonly, sequencing centers perform the preprocessing steps of the RNA-seq pipeline, but it is best practice to perform a data quality check in-house. A multitude of tools are available for the preprocessing steps. Quality of raw sequences can be checked via tools such as FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016) or fastp (Chen S. et al., 2018). Preprocessing tools such as such as Cutadapt (Martin, 2011), Trimmomatic (Bolger et al., 2014), BBTools (Bushnell, 2022), and fastp (Chen et al., 2018) can be used for contamination removal or quality trimming. rRNA contamination can be identified and removed with tools like BBTools and SortMeRNA (Kopylova et al., 2012).

Preprocessing steps should be implemented with caution, as they impact downstream analyses. A more stringent read trimming shortens reads and thus influences mapping and the estimates of expression levels for genes and their isoforms, ultimately impacting the differential expression analysis (Williams et al., 2016). For *de novo* transcriptome analysis, a stringent trimming strategy has also been observed to produce incomplete transcriptome reconstruction (MacManes, 2014; Mbandi et al., 2014). To ensure the greatest amount of information is retained for analysis, employing less stringent read-trimming, or no read-trimming at all is suggested (Del Fabbro et al., 2013; MacManes, 2014; Mbandi et al., 2014; Williams et al., 2016).

3.2 Alignment and reconstruction: reference-based and *de novo* approaches

The next step in the RNA-seq analysis pipeline is to align reads using a reference genome or *de novo* methods. Similar to preprocessing steps, there are many bioinformatic tools available for aligning reads, and they are categorized by whether they require

a reference genome or a *de novo* transcriptome. In this section we provide a brief overview of the different methods available for both reference and *de novo*-based approaches, with a focus on the types of biological questions each method can address and the limitations of each approach.

Alignment-based reconstruction occurs when a reference genome is available. The approach is very similar to aligning genomic reads, but splicing events need to be considered for RNA-seq data (Mehmood et al., 2020). Some of the software available to perform the reference-based alignment includes HISAT2 (Kim et al., 2019) and STAR (Dobin et al., 2013). Reference-based alignment is split into two parts: reference genome indexing and alignment of reads to the indexed reference genome. Additionally, some alignment tools can incorporate the discovery of novel exon-exon junctions like HISAT2 (Kim et al., 2019) and RsubRead (Liao et al., 2019).

One drawback to reference-based alignments is reference-bias: if there are sequences present in the RNA-seq data that are not present in the reference data, the data will not align and will be lost for downstream analyses unless alternative mining is performed (Yang et al., 2022). A reference-based transcript reconstruction can be performed post-alignment by software like Cufflinks (Trapnell et al., 2010), StringTie2 (Kovaka et al., 2019) and Bookend (Schon et al., 2022). Alternatively, new approaches that align reads to pangenomes are available. Pangenomes store population-level genetic variation into a graph-based structure rather than a single linear genome (Eizenga et al., 2020), allowing for improved read alignment, including for haplotype-aware RNA-seq read alignment (Sibbesen et al., 2023).

When a reference genome or pangenome is not available for your species of interest, researchers can choose to perform a *de novo* assembly transcriptome. *De novo* transcriptome assembly can be performed by tools as Trinity (Grabherr et al., 2011) and TransLiG (Liu et al., 2019). The same tools for a reference genome-based approach are used to align RNA-Seq to a *de novo* transcriptome with minor modifications – e.g., not using the splice-aware function of the aligners.

The next step in the bioinformatics pipeline is counting reads mapped. For reference-based approaches tools such as HT-Seq (Anders et al., 2015) or featureCounts can be used (Liao et al., 2014). Also, transcript quantification using a reference transcriptome can be performed using *alignment-free* methods like Kallisto (Bray et al., 2016) and Salmon (Patro et al., 2017). It is worthwhile mentioning that RNA-Seq strandedness impacts quantification, identification of isoforms and *de novo* transcriptome assemblies. The concept refers to the strategy employed in library construction, wherein stranded library preparations maintain the transcript directionality. Hence, researchers should be aware of the kit utilized for library construction when performed the aforementioned steps of RNA-seq analysis. When the information of strandedness and direction of strandedness is not known, tools like how_are_we_stranded_here (Signal and Kahlke, 2022) can help determine the strandedness of paired-end libraries. Finally, tools like tximport (Soneson et al., 2015) and tximeta (Love et al., 2020) can be used to summarize the quantification of transcript abundances into an expression matrix. For a detailed review about alignment and quantification, we refer the reader to Van den Berge et al. (2019).

4 DEGs and beyond: RNA-seq analysis types

Once RNA-seq transcript abundance has been acquired, a multitude of analyses are available to examine a transcriptome. Most often, treatments are contrasted for the identification of differentially expressed genes (DEGs). We also detail additional analyses that can be performed to further understand transcriptional profiles.

4.1 Differential expression

Measuring plant phenotypic plasticity in extreme environments can aid in knowledge to develop plants for future abiotic and biotic environmental conditions arriving with climate change (Gage et al., 2017; Monforte, 2020). Using various molecular mechanisms, plants respond to changing environmental conditions by altering their physiology and development (Lachowiec et al., 2015). Molecular plant plasticity enables adaptation to climatic shifts and predicts an individual's survival success (Nicotra and Davidson, 2010; Nicotra et al., 2010; Fox et al., 2019; Anderson and Song, 2020; Pazzaglia et al., 2021), warranting further investigation (Brooker et al., 2022).

Increasingly studies use RNA-seq to understand molecular plant plasticity (Kumar et al., 2022; Sreeratree et al., 2022). The primary approach analyzes differential expression (DE), which evaluates the transcriptional abundance across conditions through simultaneous statistical testing for significant changes in expression levels in all detected genes, transcripts, or different usage of transcripts/exons (Soneson et al., 2015; Van den Berge et al., 2019). Software like edgeR (Robinson et al., 2010), DESeq2 (Love et al., 2014), and limma (Ritchie et al., 2015; Stark et al., 2019) all provide robust DE analyses. For reviews of the main aspects in the differential expression analysis we refer the reader to Costa-Silva et al. (2021); Stark et al. (2019) and Van den Berge et al. (2019).

Differential expression analysis incorporates a matrix of features' expression levels and knowledge about the experimental design. Tests for identifying differentially expressed genes (DEGs) rely on contrasting conditions, such as different tissues, genotypes, and conditions. Exploring the up- or downregulation of genes under a stress condition relative to a control condition indicates how a plant combats a stressor and how the stressor harms the plant. For example, in the sorghum lateral root apex, low levels of phosphorus caused major expression changes in the lateral root apices, which correlated with enhanced lateral root growth. Specifically, the low-phosphorus-induced genes encoded proteins with functions in nutrient responses and contribution to phosphorus metabolism (Gladman et al., 2022). Contrasting genotypes with different performances under stress enables identifying potential mechanisms of stress resistance (Yue et al., 2016; Zhao et al., 2021).

4.2 Allele-specific expression

Allele-specific expression (ASE) describes the phenomenon where alleles within a particular genetic feature (e.g. gene, transcript) have significant differences in their expression levels (Castel et al., 2015). The expression of alleles can be assessed by comparing the expression of genes of a certain genotype with its parents or by identifying polymorphisms and quantifying the expression of each allele. ASE is especially informative in understanding hybrid crops where different parental genotypes are combined (Bell et al., 2013; Shao et al., 2019). ASE analysis also reveals the processes of genetic imprinting, tissue- and stress-specific alleles, as well as the evolution of species.

In wheat seeds, expression of homeologs and alleles is differentially controlled and consideration of each copy of a gene is relevant. In the endosperm, genes exhibited subgenome dominance in particular functions (Pfeifer et al., 2014). Further, imprinted genes were identified more frequently in developing endosperm relative to other tissues, and imprinted gene expression patterns were conserved through wheat evolution (Yang et al., 2018). This imbalanced expression of maternal and paternal alleles and subgenome dominance supports proper seed development.

ASE analysis uncovered genes that were targets of selection during domestication with implications for plant sciences. Lemmon et al. (2014) identified that maize and teosinte diverged in gene expression especially due to *cis* regulation, where the expression of maize alleles is favored in F1 hybrids between the species. Genes with *cis* and *cis* plus *trans* divergent regulation were enriched among putative targets of selection. In rice, a F1 hybrid of genotypes representing two major subpopulations exhibit enrichment of genes ASE in genomic regions of signatures for domestication or artificial selection (Shao et al., 2019). Moreover, the limited ASE in sugarcane internodes predominates as genotype-specific phenomenon, favoring high dosage alleles and purging the expression of potentially deleterious alleles (Margarido et al., 2022). Detecting ASE and regulatory mechanisms of ASE can inform our understanding of plant evolutionary history.

4.3 Alternative splicing

Alternative splicing of precursor mRNAs leads to diversification of the functions of a single gene. Transcriptional or isoform switching refers to a shift in the presence or dominance of transcripts in different samples, including across cell types, development, genotypes, and environments. Within plants, the most common form of alternative splicing is intron retention (Chamala et al., 2015), in contrast to animals where exon skipping is most detected (Kim et al., 2007). When the chosen alignment and counting approaches (see Section 3.2) enable distinguishing transcripts, examination of transcript switching is possible by utilizing the differential expression of individual transcripts or using accessible tools for assessing transcriptional switching (Qiu et al., 2021).

Genome-wide surveys of alternative splicing demonstrate the potential impact of alternative splicing events, with tens of thousands of alternative splice forms detected that are evolutionarily dynamic across angiosperms (Chamala et al., 2015). Across a population of over 350 inbred maize lines, variation in alternative splicing was detected, highlighting that connecting genotype to phenotype can be better informed by considering the expression of particular splice forms (Chen Q. et al., 2018).

4.4 Co-expression networks

Co-expression summarizes large-scale transcriptomics to infergene regulatory networks by identifying modules of genes with similar expression patterns across multiple samples. Co-expression analyses can suggest target genes of interest and corroborate GWAS results and expand potential genetic markers from those findings as well. Putative functions then can be assigned to non-annotated genes if the majority of the genes in a module share similar biological functions, the guilt-by-association principle (Serin et al., 2016; Rao and Dixon, 2019). Co-expression network analyses split into two main approaches: (i) non-targeted—a network based on the topological structure using the relationship of all pairs of genes or (ii) targeted—the use of bait genes as prior information to define network connections. Varied inputs are used to build networks, including replicates of multiple treatments, averaging the expression of replicates grown for a treatment, or defining networks separately to single levels of the experimental factor (Cortijo et al., 2020).

A fundamental step for inferring the co-expression of genes from large-scale transcriptomic data is the use of similarity measures, including correlation and mutual information methods (Ma and Wang, 2012; Huang et al., 2017). Correlation methods selection is based on data type, and common coefficients include Pearson's, Spearman's and Gini's (Huang et al., 2017). Gini's correlation, for example, is advantageous for nonnormally distributed RNA-Seq data, robust against outliers and small sample sizes (Ma and Wang, 2012). A common pipeline to construct a co-expression network involves the calculation of a similarity matrix that is then filtered based on a threshold to select gene pairs; then an adjacency matrix can be calculated and subsequently a clustering algorithm is used to group genes into modules (Serin et al., 2016).

Studies use co-expression networks to find hubs—genes with high network connectivity—and understand their role in biological pathways. For example, sugarcane hub genes changed across four stages of development in the networks of 10-month-old compared to 6-month-old apical culms (Hosaka et al., 2021), revealing candidates related to cell wall and stress and three transcription factors (TFs) potentially acting as regulators of those processes. Co-expression networks may also uncover functions for uncharacterized proteins. De Vega and colleagues (De Vega et al., 2021) inferred TF targets in Miscanthus hybrids that were enriched with carbohydrate metabolism, secondary metabolism, and the generation of precursor metabolites. They also found two TF that linked a core and a loop subnetwork, the last composed mostly by TFs linked to uncharacterized genes. Thus, regulatory co-expression networks are

useful tools to identify targets of TFs, which can be important targets for biotechnology and propose functions for poorly understood proteins (Simons et al., 2006; Fröschel et al., 2019).

Time-series expression data impose a challenge for understanding the dynamics of the coregulation of genes. The complex relationships arising due to time can be detected by gene co-expression measures that account for local dependence structures in the expression patterns (Wang et al., 2014). The dynamic network biomarker approach (Chen et al., 2012) aims to find a subnetwork of strongly correlated genes just before a critical transition – identified as a tipping point. With this approach researchers found genes at the tipping point for response to stress or ripening of fruits (Wang and Zhang, 2021; Wang et al., 2022).

Co-expression networks can also provide additional data for determining genes involved in regulatory networks. Co-expression combined with ChIP-seq can also identify the targets of a TF. Cortijo and colleagues (Cortijo et al., 2020) identified novel regulatory targets in the *Arabidopsis thaliana* circadian clock by combining modules with genes co-expressed with known TFs and ChIP-seq data. Targets of PSEUDO-RESPONSE REGULATOR 5—a core component of the circadian clock—were found in a module showing enrichment for photosynthesis. For a complete review of co-expression networks in plant biology, the reader is referred to (Rao and Dixon, 2019) and (Serin et al., 2016).

4.5 Pathway enrichment

Enrichment analysis tests if any functional group is over- or underrepresented by a list of genes of interest – e.g., DEGs, co-expressed genes or ASE genes. Enrichment analysis of genes identified by differential expression or modules in co-expression networks are useful to understand if genes in the module are related to similar functions. It can suggest that non-annotated genes likely participate in the same biological pathways as known genes, which can lead to the identity of causal genes (Serin et al., 2016). Functional pathways are represented by ontologies in different frameworks like Gene Ontology (GO) categories (Ashburner et al., 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (KO) (Kanehisa et al., 2017), and MapMan4 bin categories (Schwacke et al., 2019), which all capture varied functions of genes.

For plants with reference genomes, functional annotation is provided by databases along with the nucleotide sequence and the structural annotation. For *de novo* assemblies this procedure requires comparisons with nucleotide or protein databases, retrieving the functional categories from the hits and associating them to genes/transcripts of the reference. While functional annotation of a reference – genome or transcriptome – in terms of GOs, Enzyme Codes (ECs) or KOs can be performed by tools like OmicsBox (BioBam Bioinformatics, 2019) or Trinotate (Bryant et al., 2017), Mercator4 provides the specific annotation for MapMan (Schwacke et al., 2019). The KO/EC annotation can be linked to molecular networks by KEGG Mapper (Kanehisa and Sato, 2020) to visual representation of the pathways. MapMan4 also has its own visualization of functional pathways where besides the category mapping file, fold change or expression values can be provided as input.

Finally, tests for enrichment of categories—over or underrepresentation—frequently make use of Fisher's Exact test, Hypergeometrical test, permutation, or variation of those methods. There are many tools for functional enrichment analysis: goseq (Young et al., 2010), GSEA (Subramanian et al., 2005), OmicsBox (BioBam Bioinformatics, 2019), topGO (Alexa et al., 2006), clusterProfiler (Yu et al., 2012), and agriGO (Du et al., 2010).

4.6 Viral discovery

Viruses are the most abundant biological entities on earth and are strongly associated with all organisms (Rosani and Gerdol, 2017; Mushegian, 2020). It is very common for viruses to contaminate tissue when preparing samples for sequencing. One meta-analysis found that over half of the 700 sequencing libraries examined had viral contamination (Asplund et al., 2019). RNA-seq is susceptible to viral contamination. However, most viral contamination is filtered out during bioinformatic processing.

Viral contamination in RNA-seq studies can provide informative results as well. Kamitani et al. (2016) looked at both plant-virus and virus-virus interactions in natural environments using RNA-seq. Other studies have used RNA-seq to screen for viral pathogens (Selitsky et al., 2020) and novel viral genome discovery (Rosani and Gerdol, 2017). However, some virus types require alternative rRNA depletion methods to be detected, such as non-polyadenylated genomes (Nagano et al., 2015). To capture the full range of viral genome varieties Nagano et al. (2015) optimized rRNA depletion methods for mRNA focused RNA-seq to detect varying types of viral signatures from natural environments. Viral contamination in transcriptomic data may provide insight to unknown biotic stressors or uncover novel defense mechanisms related to a previously unknown viral pathogen. Accessible software for identifying the species of origin allows for examining unmapped sequencing reads and examining expression patterns for additional species (Chen et al., 2020).

5 RNA-seq: from candidate to causal gene

A major goal in plant genetics is to identify alleles or genes that control a trait by associating genotype with phenotype, using both reductionist and holistic approaches. In quantitative trait locus (QTL) mapping, genetic loci that determine a trait are examined one-by-one, typically focusing on loci of major effect to find genotypes with a desired allele. In contrast, genomic prediction collectively considers the impact of the entire genotype on a focal trait, especially impactful for traits controlled by many loci of small effect (Meuwissen et al., 2001).

Within a plant, the genotype is filtered through many levels of endo- or molecular phenotypes to create phenotypes of interest (Mackay et al., 2009). Integrating molecular phenotypes, such as gene expression-, protein- or metabolite-level information with

genotypes can improve understanding the association between genotype and phenotype in QTL mapping.

RNA-seq supports fine mapping QTL, both by providing additional evidence for the causal gene via differential expression data and by allowing the discovery of new variants in the QTL region, which can be used for marker development (Liu et al., 2016; Habib et al., 2018; Jaganathan et al., 2020). A suitable situation for using RNA-seq in a fine mapping project is one designed in such a way that a differentially expressed gene within the target QTL is the gene of interest. One technique to maximize this possibility is by generating a pair of near-isogenic lines (NILs) which are genetically identical except for the QTL genotype, and by extension, gene expression and phenotype (Szalma et al., 2007). Using NIL pairs will minimize background genetic noise, so that differentially expressed genes between the pair are likely associated with the trait of interest (Keurentjes et al., 2007). Candidate genes are those occurring in the QTL region that are differentially expressed between the NIL pair. Subsequent analysis of each candidate gene can include using molecular techniques such as gene silencing to observe phenotypic changes as well as enrichment analysis to infer the genes' function. In addition to expression analysis, RNA-seq data can be used to identify variants in the QTL region which can be used to develop new markers to further increase the resolution of the region (Paritosh et al., 2013). Taken together, the ability of RNA-seq to both provide evidence for the causal gene through expression data and to allow for the identification of new variants in the QTL region for marker development, means that it can be a powerful tool for fine mapping applications.

Similarly, combining co-expression networks with genome-wide association studies (GWAS) has prioritized candidate genes (Chan et al., 2011). In contrast to biparental QTL mapping, GWAS relies on diverse accessions that show phenotypic variation in a trait of interest. Schaefer et al. (2018) developed a framework to integrate GWAS for maize grain ionome traits and co-expression networks leading the identification of two important genes expressed in roots (Schaefer et al., 2018). The maize *dwarf9* (*d9*) dominant allele *D9-1* had higher abundance of elements like iron, sulfur, and strontium compared to the wild type. *D9-1* did not influence cadmium accumulation, as expected by the location of the cadmium GWAS QTL that contained *d9*. Rather the dwarf allele *–D8-mpl* – of the paralog *d8* identified through co-expression analysis, recapitulated this effect. Co-expression networks can also support extending understanding of the genetic architecture of traits through the examination of epistasis. A maize GWAS for senescence-controlling genes identified putative epistatic interactions among QTL that were independently supported with co-expression data (Sekhon et al., 2019). Co-expression networks integrated with mapping approaches are proving to be powerful for the identification of important genes and genetic architecture (Rao and Dixon, 2019).

6 Genome complexity and RNA-seq in plant sciences

Aspects of plant genome biology complicate the use of RNA-seq data, especially relative to the organisms for which many of the bioinformatics tools were developed, though efforts to produce specific tools for polyploid crops are expanding (Foster et al., 2019). The genomes encoding many plants are polyploid with multiple subgenomes (Meyer et al., 2012). Additionally, plants can harbor massive levels of repetitive sequences (Zimin et al., 2014).

6.1 Polyploidy

One feature of plant genomes that can influence the use of RNA-seq is the prevalence of polyploidy (Fu et al., 2016). A study of 203 modern crops identified that 17% have undergone polyploidization (Meyer et al., 2012). Of these, wheat and sugarcane are among the most cultivated crops globally (Weeks, 2017). The pairing of polyploid chromosomes during recombination varies along a spectrum from allopolyploidy to autopolyploidy, along with the presence of aneuploid chromosomes. Within allopolyploids, often formed through hybridization, chromosome copies behave in a diploid fashion and pair and segregate corresponding to species of origin, within subgenomes (Edger et al., 2018; Kuo et al., 2020). In contrast, homologous autopolyploid chromosomes pair at random, even between subgenomes (Spoelhof et al., 2017). The type of polyploidy can also vary along the length of a single chromosome. Polyploidization impacts the bioinformatic pipelines used to assess RNA-seq data, both at the level of processing reads and subsequent analyses. The presence of polyploidy can be problematic for RNA-seq analysis, from both a computational and biological standpoint.

Polyploidy increases computation complexity due to an increase in genome size, relative to a diploid genome and the increased repeat content. The increase in genome size, resulting from the presence of two or more subgenomes, means more memory is required to index and store the reference genome or transcriptome prior to read alignment. As such, analyzing large, polyploid genomes, such as wheat, may be prohibitive to those without expansive compute resources.

The degree of similarity among polyploid subgenomes is a driving force for how RNA-seq analyses may need to proceed differently compared to diploid genomes (Voshall and Moriyama, 2020). In the cases of subgenomes with low levels of divergence at the nucleotide level, little functional divergence may be expected across homeologs, the genes homologous to one another on each subgenome (Wang et al., 2017; Sigel et al., 2019). However, the phenomenon of subgenome dominance, or the tendency for genes to be expressed from a particular subgenomes, is commonly observed across plants (Schnable et al., 2011; Wang X. et al., 2016; Khan et al., 2020). For example, nonbalanced expression was identified in approximately ~30% of wheat homeologs, primarily through suppression of a single homeolog (Ramírez-González et al., 2018). Isolating the expression of homeologous genes may be particularly useful when using RNA-seq to complement QTL mapping studies (Yang et al., 2014). Therefore, distinguishing homeolog expression enables specific studies.

6.2 Repetitive sequences

High levels of repeat content are also problematic from a computational viewpoint to interpret RNA-seq data. Repeat content can refer to both biological repeats (e.g., transposable elements, short sequence repeats) or genomic regions shared between subgenomes. Significant portions of widely grown cereals are comprised of repetitive sequences, excluding homeologs and duplicated genes: wheat-85% (Wicker et al., 2018), maize-85% (Schnable et al., 2009), barley-80% (Wicker et al., 2017), and rice-41% (Li et al., 2021). In the cases when these regions are expressed (Cavrak et al., 2014), they create issues computationally as there may be no way to confidently align reads to regions that are present more than once, resulting in a high number of multi-mapping reads. Multi-mapping reads are discarded by default by popular read counting programs. As such, the RNA-seq analysis of polyploid genomes may be hampered by the loss of a significant portion of data without explicit intervention.

Data loss due to repetitive sequences is problematic as repetitive elements can have important biological functions. For example, screening *A. thaliana* T-DNA insertion lines at the locations of transposable elements for seedling morphological responses to stresses uncovered functional roles of for over 90% of those tested (Joly-Lopez et al., 2017). Outside of transposable elements, short tandem repeats contribute to the repetitive content of the genome with functional consequences. The length of the short tandem repeat encoding a polyglutamine span in the protein *EARLY FLOWERING 3* influenced flowering time across *A. thaliana* accessions (Undurraga et al., 2012). Inability to assign reads to these repeats means that functional polymorphisms are missing from analyses. Longer read technologies will improve reference genomes (Jiao et al., 2017; Michael et al., 2018; Kamal et al., 2022) and transcriptomes to enable study (Wang B. et al., 2016) of currently poorly characterized sequences.

7 Conclusions

The ever-increasing need to support a growing global population requires increases in plant productivity and access. New crops and varieties are needed to adapt to challenging abiotic and biotic environmental conditions globally. With the growing affordability and accessibility of RNA-seq, this technology can be leveraged to study any plant species of interest. The most impactful RNA-seq studies are carefully designed and control many sources of unwanted variation. During analysis, much research concludes by examining DEGs between conditions; however, we have outlined additional forms of analysis to extend the usefulness of RNA-seq data, even data already publicly available. The intent of this review is to provide plant-focused guidelines, strategies, and examples for supporting new users of RNA-seq and inspiration for new applications by established researchers.

Author contributions

RNU, FHC, JLi, GLR, KF, JPC, and JLa drafted and finalized themanuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by USDA NIFA Award 2020-65114-30768 and the Office of Science (BER), US Department of Energy, grant no. DE-SC0021369.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alexa, A., Rahnenfuhrer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22 (13), 1600–1607. doi: 10.1093/bioinformatics/btl140

Andalo, C., Mazer, S. J., Godelle, B., and Machon, N. (1999). Parental environmental effects on life history traits in arabidopsis thaliana (Brassicaceae). *New Phytol.* 142 (2), 173–184. doi: 10.1046/j.1469-8137.1999.00396.x

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2), 166–169. doi: 10.1093/bioinformatics/btu638

Anderson, J. T., and Song, B.-H. (2020). Plant adaptation to climate change–where are we? *J. Sys. Evol.* 58 (5), 533–545. doi: 10.1111/jse.12649

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556

Asplund, M., Kjartansdóttir, K. R., Mollerup, S., Vinner, L., Fridholm, H., Herrera, J. A.R., et al. (2019). Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clin. Microbiol. Infect.* 25 (10), 1277–1285. doi: 10.1016/j.cmi.2019.04.028

Baker, S. S., Rugh, C. L., and Kamalay, J. C. (1990). RNA And DNA isolation from recalcitrant plant tissues. *Biotechniques* 9 (3), 268–272.

Bell, G. D. M., Kane, N. C., Rieseberg, L. H., and Adams, K. L. (2013). RNA-Seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol. Evol.* 5 (7), 1309–1323. doi: 10.1093/gbe/evt072

BioBam Bioinformatics (2019) *OmicsBox – bioinformatics made easy*. Available at: https://www.biobam.com/omicsbox.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34 (5), 525–527. doi: 10.1038/nbt.3519

Brooker, R., Brown, L. K., George, T. S., Pakeman, R. J., Palmer, S., Ramsay, L., et al. (2022). Active and adaptive plasticity in a changing climate. *Trends Plant Sci.* 27 (7), 717–728. doi: 10.1016/j.tplants.2022.02.004

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl *De novo* transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18 (3), 762–776. doi: 10.1016/j.celrep.2016.12.063

Bushnell, B. (2022) *BBMap short read aligner, and other bioinformatic tools*. Available at: https://sourceforge.net/projects/bbmap/.

Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16 (1), 195. doi: 10.1186/s13059-015-0762-6

Cavrak, V. V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L. M., and Mittelsten Scheid, O., et al. (2014). How a retrotransposon exploits the plant's heat stress response for its activation. *PloS Genet.* 10 (1), e1004115. doi: 10.1371/journal.pgen.1004115

Chamala, S., Feng, G., Chavarro, C., and Barbazuk, W. B.. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front. Bioeng. Biotechnol.* 3. doi: 10.3389/fbioe.2015.00033

Chan, E. K. F., Rowe, H. C., Corwin, J. A., Joseph, B., and Kliebenstein, D. J.. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in arabidopsis thaliana. *PloS Biol.* 9 (8), e1001125. doi: 10.1371/journal.pbio.1001125

Chen, L., Liu, R., Liu, Z.-P., Li, M., and Aihara, K. (2012). Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci. Rep.* 2 (1), 342. doi: 10.1038/srep00342

Chen, Q., Han, Y., Liu, H., Wang, X., Sun, J., Zhao, B., et al. (2018). Genome-wide association analyses reveal the importance of alternative splicing in diversifying gene function and regulating phenotypic variation in maize. *Plant Cell* 30 (7), 1404–1423. doi: 10.1105/tpc.18.00109

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884–i890. doi: 10.1093/bioinformatics/bty560

Chen, S., Ren, C., Zhai, J., Yu, J., Zhao, X., Li, Z., et al. (2020). CAFU: a galaxy framework for exploring unmapped RNA-seq data. *Briefings Bioinf.* 21 (2), 676–686. doi: 10.1093/bib/bbz018

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17 (1), 13. doi: 10.1186/s13059-016-0881-8

Correr, F. H., Hosaka, G. K., Barreto, F. Z., Valadão, I. B., Balsalobre, T. W.A., Furtado, A., et al. (2020). Differential expression in leaves of saccharum genotypes contrasting in biomass production provides evidence of genes involved in carbon partitioning. *BMC Genomics* 21 (1), 673. doi: 10.1186/s12864-020-07091-y

Cortijo, S., Aydin, Z., Ahnert, S., and Locke, J. C. (2019). Widespread inter-individual gene expression variability in arabidopsis thaliana. *Mol. Syst. Biol.* 15 (1), e8591. doi: 10.15252/msb.20188591

Cortijo, S., Bhattarai, M., Locke, J. C.W., and Ahnert, S. E. (2020). Co-Expression networks from gene expression variability between genetically identical seedlings can reveal novel regulatory relationships. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.599464

Costa-Silva, J., Domingues, D. S., Menotti, D., Hungria, M., Lopes, F. M., et al. (2021). Computational methods for differentially expressed gene analysis from RNA-seq: an overview. *arXiv preprint arXiv*, 03625. doi: 10.48550/arXiv.2109.03625

Cuperus, J. T. (2022). Single-cell genomics in plants: current state, future directions, and hurdles to overcome. *Plant Physiol.* 188 (2), 749–755. doi: 10.1093/plphys/kiab478

Del Fabbro, C., Scalabrin, S., Morgante, M., and Giorgi, F. M.. (2013). An extensive evaluation of read trimming effects on illumina NGS data analysis. *PloS One* 8 (12), e85024. doi: 10.1371/journal.pone.0085024

De Vega, J. J., Peel, N., Purdy, S. J., Hawkins, S., Donnison, L., Dyer, S., et al. (2021). Differential expression of starch and sucrose metabolic genes linked to varying biomass yield in miscanthus hybrids. *Biotechnol. Biofuels* 14 (1), 98. doi: 10.1186/s13068-021-01948-4

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi: 10.1093/bioinformatics/bts635

Donohue, K. (2009). Completing the cycle: maternal effects as the missing link in plant life histories. *Philos. Trans. R. Soc. B: Biol. Sci.* 364 (1520), 1059–1074. doi: 10.1098/rstb.2008.0291

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z.. (2010). A'griGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38 (suppl_2), W64–W70. doi: 10.1093/nar/gkq310

Edger, P. P., McKain, M. R., Bird, K. A., and VanBuren, R. (2018). Subgenome assignment in allopolyploids: challenges and future directions. *Curr. Opin. Plant Biol.* 42, 76–80. doi: 10.1016/j.pbi.2018.03.006

Eizenga, J. M., Novak, A. M., Sibbesen, J. A., Heumos, S., Ghaffaari, A., Hickey, G., et al. (2020). Pangenome graphs. *Annu. Rev. Genomics Hum. Genet.* 21 (1), 139–162. doi: 10.1146/annurev-genom-120219-080406

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32 (19), 3047–3048. doi: 10.1093/bioinformatics/btw354

Fang, Z., and Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Briefings Bioinf.* 12 (3), 280–287. doi: 10.1093/bib/bbr004

Foster, T. M., Bassil, N. V., Dossett, M., Leigh Worthington, M., and Graham, J.. (2019). Genetic and genomic resources for rubus breeding: a roadmap for the future. *Horticult. Res.* 6 (1), 1–9. doi: 10.1038/s41438-019-0199-2

Fox, R. J., Donelson, J. M., Schunter, C., Ravasi, T., and Gaitán-Espitia, J. D. (2019). Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. *Philos. Trans. R. Soc. B: Biol. Sci.* 374 (1768), 20180174. doi: 10.1098/rstb.2018.0174

Fröschel, C., Iven, T., Walper, E., Bachmann, V., Weiste, C., Dröge-Laser, W., et al. (2019). A gain-of-Function screen reveals redundant ERF transcription factors providing opportunities for resistance breeding toward the vascular fungal pathogen verticillium longisporum. *Mol. Plant-Microbe Interact.* 32 (9), 1095–1109. doi: 10.1094/MPMI-02-19-0055-R

Fu, D., Mason, A. S., Xiao, M., and Yan, H. (2016). Effects of genome structure variation, homeologous genes and repetitive DNA on polyploid crop research in the age of genomics. *Plant Sci.* 242, 37–46. doi: 10.1016/j.plantsci.2015.09.017

Gage, J. L., Jarquin, D., Romay, C., Lorenz, A., Buckler, E. S., Kaeppler, S., et al. (2017). The effect of artificial selection on phenotypic plasticity in maize. *Nat. Commun.* 8 (1), 1348. doi: 10.1038/s41467-017-01450-2

Galloway, L. F. (2005). Maternal effects provide phenotypic adaptation to local environmental conditions. *New Phytol.* 166 (1), 93–100. doi: 10.1111/j.1469-8137.2004.01314.x

Gao, J., Liu, J., Li, B., and Li, Z. (2001). Isolation and purification of functional total RNA from blue-grained wheat endosperm tissues containing high levels of starches and flavonoids. *Plant Mol. Biol. Rep.* 19 (2), 185–186. doi: 10.1007/BF02772163

Giacomello, S., Salmén, F., Terebieniec, B. K., Vickovic, S., Navarro, J. F., Alexeyenko, A., et al. (2017). Spatially resolved transcriptome profiling in model plant species. *Nat. Plants* 3 (6), 17061. doi: 10.1038/nplants.2017.61

Gierliński, M., Cole, C., Schofield, P., Schurch, N. J., Sherstnev, A., Singh, V., et al. (2015). Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. *Bioinformatics* 31 (22), 3625–3630. doi: 10.1093/bioinformatics/btv425

Gifford, M. L., Dean, A., Gutierrez, R. A., Coruzzi, G. M., and Birnbaum, K. D.. (2008). Cell-specific nitrogen responses mediate developmental plasticity. *Proc. Natl. Acad. Sci.* 105 (2), 803–808. doi: 10.1073/pnas.0709559105

Gladman, N., Hufnagel, B., Regulski, M., Liu, Z., Wang, X., Chougule, K., et al. (2022). Sorghum root epigenetic landscape during limiting phosphorus conditions. *Plant Direct* 6 (5), e393. doi: 10.1002/pld3.393

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29 (7), 644–652. doi: 10.1038/nbt.1883

Guo, Z., Magwire, M. M., Basten, C. J., Xu, Z., and Wang, D. (2016). Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. *Theor. Appl. Genet.* 129 (12), 2413–2427. doi: 10.1007/s00122-016-2780-5

Guo, T., Yang, J., Li, D., Sun, K., Luo, L., Xiao, W., et al. (2019). Integrating GWAS, QTL, mapping and RNA-seq to identify candidate genes for seed vigor in rice (Oryza sativa l.). *Mol. Breed.* 39 (6), 87. doi: 10.1007/s11032-019-0993-4

Habib, A., Powell, J. J., Stiller, J., Liu, M., Shabala, S., Zhou, M., et al. (2018). A multiple near isogenic line (multi-NIL) RNA-seq approach to identify candidate genes underpinning QTL. *Theor. Appl. Genet.* 131 (3), 613–624. doi: 10.1007/s00122-017-3023-0

Hayes, K. R., Beatty, M., Meng, X., Simmons, C. R., Habben, J. E., Danilevskaya, O. N., et al. (2010). Maize global transcriptomics reveals pervasive leaf diurnal rhythms but rhythms in developing ears are largely limited to the core oscillator. *PloS One* 5 (9), e12887. doi: 10.1371/journal.pone.0012887

Hosaka, G. K., Correr, F. H., da Silva, C. C., Sforça, D. A., Barreto, F. Z., Balsalobre, T. W.A., et al. (2021). Temporal gene expression in apical culms shows early changes in cell wall biosynthesis genes in sugarcane. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.736797

Hoshmand, R. (2006). *Design of experiments for agriculture and the natural sciences second edition* (Boca Raton, FL: CRC Press).

Hotta, C. T. (2021). From crops to shops: how agriculture can use circadian clocks. *J. Exp. Bot.* 72 (22), 7668–7679. doi: 10.1093/jxb/erab371

Huang, J., Vendramin, S., Shi, L., and McGinnis, K. M. (2017). Construction and optimization of a Large gene coexpression network in maize using RNA-seq data. *Plant Physiol.* 175 (1), 568–583. doi: 10.1104/pp.17.00825

Huang, Y., Wang, H., Zhu, Y., Huang, X., Li, S., Wu, X., et al. (2022). THP9 enhances seed protein content and nitrogen-use efficiency in maize. *Nature* 612 (7939), 292–300. doi: 10.1038/s41586-022-05441-2

Hudson, K. A. (2010). The circadian clock-controlled transcriptome of developing soybean seeds. *Plant Genome* 3 (1), 25. doi: 10.3835/plantgenome2009.08.0025

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theor. Appl. Genet.* 133 (5), 1791–1810. doi: 10.1007/s00122-020-03560-w

Jain, M., Abu-Shumays, R., Olsen, H. E., and Akeson, M. (2022). Advances in nanopore direct RNA sequencing. *Nat. Methods* 19 (10), 1160–1164. doi: 10.1038/s41592-022-01633-w

Jiang, X., Yang, X., Zhang, F., Yang, T., Yang, C., He, F., et al. (2022). Combining QTL mapping and RNA-seq unravels candidate genes for alfalfa (Medicago sativa l.) leaf development. *BMC Plant Biol.* 22 (1), 485. doi: 10.1186/s12870-022-03864-7

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., et al. (2017). Improved maize reference genome with single-molecule technologies. *Nature* 546 (7659), 524–527. doi: 10.1038/nature22971

Johnson, B. R., Atallah, J., and Plachetzki, D. C. (2013). The importance of tissue specificity for RNA-seq: highlighting the errors of composite structure extractions. *BMC Genomics* 14 (1), 586. doi: 10.1186/1471-2164-14-586

Joly-Lopez, Z., Forczek, E., Vello, E., Hoen, D. R., Tomita, A., Bureau, T. E., et al. (2017). Abiotic stress phenotypes are associated with conserved genes derived from transposable elements. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02027

Julca, I., Tan, Q. W., and Mutwil, M. (2022). Toward kingdom-wide analyses of gene expression. *Trends Plant Sci.* 28 (2), 235–249. doi: 10.1016/j.tplants.2022.09.007

Kamal, N., Lux, T., Jayakodi, M., Haberer, G., Gundlach, H., Mayer, K. F.X., et al. (2022). "The barley and wheat pan-genomes," in *Plant bioinformatics: methods and protocols*. Ed. D. Edwards (New York, NY: Springer US), 147–159. doi: 10.1007/978-1-0716-2067-0_7

Kamitani, M., Nagano, A. J., Honjo, M. N., and Kudoh, H. (2016). RNA-Seq reveals virus–virus and virus–plant interactions in nature. *FEMS Microbiol. Ecol.* 92, 11. doi: 10.1093/femsec/fiw176

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45 (D1), D353–D361. doi: 10.1093/nar/gkw1092

Kanehisa, M., and Sato, Y. (2020). KEGG mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29 (1), 28–35. doi: 10.1002/pro.3711

Kasirajan, L., Hoang, N. V., Furtado, A., Botha, F. C., and Henry, R. J.. (2018). Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content. *Sci. Rep.* 8 (1), 11612. doi: 10.1038/s41598-018-30033-4

Keurentjes, J. J. B., Bentsink, L., Alonso-Blanco, C., Hanhart, C. J., Blankestijn-De Vries, H., Effgen, S., et al. (2007). Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* 175 (2), 891–905. doi: 10.1534/genetics.106.066423

Khan, D., Ziegler, D. J., Kalichuk, J. L., Hoi, V., Hyunh, N., Hajihasani, A., et al. (2020). Gene expression profiling reveals subgenome dominance during brassica napus seed development. *Plant Biol*. doi: 10.1101/2020.04.29.068189

Kim, D., Magen, A., and Ast, G.. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37 (8), 907–915. doi: 10.1038/s41587-019-0201-4

Kim, E., Magen, A., and Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* 35 (1), 125–131. doi: 10.1093/nar/gkl924

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28 (24), 3211–3217. doi: 10.1093/bioinformatics/bts611

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M.. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20 (1), 278. doi: 10.1186/s13059-019-1910-1

Kumar, S., Seem, K., Kumar, S., and Mohapatra, T.. (2022). RNA-Seq analysis reveals the genes/pathways responsible for genetic plasticity of rice to varying environmental conditions on direct-sowing and transplanting. *Sci. Rep.* 12 (1), 2241. doi: 10.1038/s41598-022-06009-w

Kuo, T. C. Y., Hatakeyama, M., Tameshige, T., Shimizu, K. K., and Sese, J.. (2020). Homeolog expression quantification methods for allopolyploids. *Briefings Bioinf.* 21 (2), 395–407. doi: 10.1093/bib/bby121

Lachowiec, J., Queitsch, C., and Kliebenstein, D. J. (2015). Molecular mechanisms governing differential robustness of development and environmental responses in plants. *Ann. Bot.* 117 (5), 795–809. doi: 10.1093/aob/mcv151

Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., et al. (2018). Optimization of an RNA-seq differential gene expression analysis depending on biological replicate number and library size. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00108

Lämke, J., and Bäurle, I. (2017). Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biol.* 18 (1), 124. doi: 10.1186/s13059-017-1263-6

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* 11 (10), 733–739. doi: 10.1038/nrg2825

Lemmon, Z. H., Bukowski, R., Sun, Q., and Doebley, J. F. (2014). The Role of cis Regulatory Evolution in Maize Domestication. *PLOS Genet* 10 (11), e1004745. doi: 10.1371/journal.pgen.1004745

Li, S., Shui, K., Zhang, Y., Lv, Y., Deng, W., Ullah, S., et al. (2016). CGDB: a database of circadian genes in eukaryotes. *Nucleic Acids Res.* 45 (D1), D397–D403. doi: 10.1093/nar/gkw1028

Li, K., Jiang, W., Hui, Y., Kong, M., Feng, L.-Y., Gao, L.-Z, et al. (2021). Gapless indica rice genome reveals synergistic contributions of active transposable elements and segmental duplications to rice genome evolution. *Mol. Plant* 14 (10), 1745–1756. doi: 10.1016/j.molp.2021.06.017

Liang, Y., Gong, Z., Wang, J., Zheng, J., Ma, Y., Min, L., et al. (2021). Nanopore-based comparative transcriptome analysis reveals the potential mechanism of high-temperature tolerance in cotton (Gossypium hirsutum l.). *Plants* 10 (11), 2517. doi: 10.3390/plants10112517

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30 (7), 923–930. doi: 10.1093/bioinformatics/btt656

Liao, Y., Smyth, G. K., and Shi, W. (2019). The r package rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47 (8), e47–e47. doi: 10.1093/nar/gkz114

Liu, D., Zhang, J., Liu, X., Wang, W., Liu, D., Teng, Z., et al. (2016). Fine mapping and RNA-seq unravels candidate genes for a major QTL controlling multiple fiber quality traits at the T1 region in upland cotton. *BMC Genomics* 17 (1), 295. doi: 10.1186/s12864-016-2605-6

Liu, J., Yu, T., Mu, Z., and Li, G.. (2019). TransLiG: a *de novo* transcriptome assembler that uses line graph iteration. *Genome Biol.* 20 (1), 81. doi: 10.1186/s13059-019-1690-7

Love, M. I., Soneson, C., Hickey, P. F., Johnson, L. K., Pierce, N. T., Shepherd, L., et al. (2020). Tximeta: reference sequence checksums for provenance identification in RNA-seq. *PloS Comput. Biol.* 16 (2), e1007664. doi: 10.1371/journal.pcbi.1007664

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Ma, C., and Wang, X. (2012). Application of the gini correlation coefficient to infer regulatory relationships in transcriptome analysis. *Plant Physiol.* 160 (1), 192–203. doi: 10.1104/pp.112.201962

Mackay, T. F. C., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10 (8), 565–577. doi: 10.1038/nrg2612

MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Front. Genet.* 5. doi: 10.3389/fgene.2014.00013

Margarido, G. R. A., Correr, F. H., Furtado, A., Botha, F. C., and Henry, R. J.. (2022). Limited allele-specific gene expression in highly polyploid sugarcane. *Genome Res.* 32 (2), 297–308. doi: 10.1101/gr.275904.121

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17 (1), 10. doi: 10.14806/ej.17.1.200

Mbandi, S. K., Hesse, U., Rees, D. J.G., and Christoffels, A.. (2014). A glance at quality score: implication for *de novo* transcriptome reconstruction of illumina reads. *Front. Genet.* 5. doi: 10.3389/fgene.2014.00017

Mehmood, A., Laiho, A., Venäläinen, M. S., McGlinchey, A. J., Wang, N., and Elo, L. L.. (2020). Systematic evaluation of differential splicing tools for RNA-seq studies. *Briefings Bioinf.* 21 (6), 2052–2065. doi: 10.1093/bib/bbz126

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 (4), 1819–1829. doi: 10.1093/genetics/157.4.1819

Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196 (1), 29–48. doi: 10.1111/j.1469-8137.2012.04253.x

Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., et al. (2018). High contiguity arabidopsis thaliana genome assembly with a single nanopore flow cell. *Nat. Commun.* 9 (1), 541. doi: 10.1038/s41467-018-03016-2

Molania, R., Foroutan, M., Gagnon-Bartsch, J. A., Gandolfo, L. C., Jain, A., Sinha, A., et al. (2022). Removing unwanted variation from large-scale RNA sequencing data with PRPS. *Nat. Biotechnol* 41, 82–95. doi: 10.1038/s41587-022-01440-w

Monforte, A. J. (2020). Time to exploit phenotypic plasticity. *J. Exp. Bot.* 71 (18), 5295–5297. doi: 10.1093/jxb/eraa268

Mushegian, A. R. (2020). Are there 10^{31} virus particles on earth, or more, or fewer? *J. Bacteriol.* 202 (9). doi: 10.1128/JB.00052-20

Nagano, A. J., Honjo, M. N., Mihara, M., Sato, M., and Kudoh, H. (2015). "Detection of plant viruses in natural environments by using RNA-seq," in *Plant virology protocols: new approaches to detect viruses and host responses*. Eds. I. Uyeda and C. Masuta (New York, NY: Springer), 89–98. doi: 10.1007/978-1-4939-1743-3_8

Narsai, R., Howell, K. A., Millar, A. H., O'Toole, N., Small, I., and Whelan, J.. (2007). Genome-wide analysis of mRNA decay rates and their determinants in arabidopsis thaliana. *Plant Cell* 19 (11), 3418–3436. doi: 10.1105/tpc.107.055046

Nicotra, A. B., and Davidson, A. (2010). Adaptive phenotypic plasticity and plant water use. *Funct. Plant Biol.* 37 (2), 117–127. doi: 10.1071/FP09139

Nicotra, A. B.. (2010). Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* 15 (12), 684–692. doi: 10.1016/j.tplants.2010.09.008

Palovaara, J., Saiga, S., Wendrich, J. R., van 't Wout Hofland, N., van Schayck, J. P., Hater, F., et al. (2017). Transcriptome dynamics revealed by a gene expression atlas of the early arabidopsis embryo. *Nat. Plants* 3 (11), 894–904. doi: 10.1038/s41477-017-0035-3

Paritosh, K., Yadava, S. K., Gupta, V., Panjabi-Massand, P., Sodhi, Y. S., Pradhan, A. K., et al. (2013). RNA-Seq based SNPs in some agronomically important oleiferous

lines of brassica rapaand their use for genome-wide linkage mapping and specific-region fine mapping. *BMC Genomics* 14 (1), 463. doi: 10.1186/1471-2164-14-463

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14 (4), 417–419. doi: 10.1038/nmeth.4197

Pazzaglia, J., Reusch, T. B.H., Terlizzi, A., Marín-Guirao, L., and Procaccini, G. (2021). Phenotypic plasticity under rapid global changes: the intrinsic force for future seagrasses survival. *Evolution. Appl.* 14 (5), 1181–1201. doi: 10.1111/eva.13212

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., et al. (2014). Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345 (6194), 1250091. doi: 10.1126/science.1250091

Qiu, Z., Chen, S., Qi, Y., Liu, C., Zhai, J., Xie, S., et al. (2021). Exploring transcriptional switches from pairwise, temporal and population RNA-seq data using deepTS. *Briefings Bioinf.* 22 (3), bbaa137. doi: 10.1093/bib/bbaa137

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361 (6403), eaar6089. doi: 10.1126/science.aar6089

Rao, X., and Dixon, R. A. (2019). Co-Expression networks for plant biology: why and how. *Acta Biochim. Biophys. Sin.* 51 (10), 981–988. doi: 10.1093/abbs/gmz080

Risso, D., Ngai, J., Speed, T. P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* 32 (9), 896–902. doi: 10.1038/nbt.2931

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47–e47. doi: 10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Robles, J. A., Qureshi, S. E., Stephen, S. J., Wilson, S. R., Burden, C. J., and Taylor, J. M. (2012). Efficient experimental design and analysis strategies for the detection of differential expression using RNA-sequencing. *BMC Genomics* 13 (1), 484. doi: 10.1186/1471-2164-13-484

Rosani, U., and Gerdol, M. (2017). A bioinformatics approach reveals seven nearly-complete RNA-virus genomes in bivalve RNA-seq data. *Virus Res.* 239, 33–42. doi: 10.1016/j.virusres.2016.10.009

Salzman, R. A., Fujita, T., Zhu-Salzman, K., Hasegawa, P. M., and Bressan, R. A.. (1999). An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. *Plant Mol. Biol. Rep.* 17 (1), 11–17. doi: 10.1023/A:1007520314478

Schaarschmidt, S., Fischer, A., Lawas, L. M.F., Alam, R., Septiningsih, E. M., Bailey-Serres, J., et al. (2020). Utilizing PacBio iso-seq for novel transcript and gene discovery of abiotic stress responses in oryza sativa l. *Int. J. Mol. Sci.* 21 (21), 8148. doi: 10.3390/ijms21218148

Schaefer, R. J., Michno, J.-M., Jeffers, J., Hoekenga, O., Dilkes, B., Baxter, I., et al. (2018). Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell* 30 (12), 2922–2942. doi: 10.1105/tpc.18.00299

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326 (5956), 1112–1115. doi: 10.1126/science.1178534

Schnable, J. C., Springer, N. M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci.* 108 (10), 4069–4074. doi: 10.1073/pnas.1101368108

Schon, M. A., Lutzmayer, S., Hofmann, F., and Nodine, M. D.. (2022). Bookend: precise transcript reconstruction with end-guided assembly. *Genome Biol.* 23 (1), 143. doi: 10.1186/s13059-022-02700-3

Schon, M. A., and Nodine, M. D. (2017). Widespread contamination of arabidopsis embryo and endosperm transcriptome data sets. *Plant Cell* 29 (4), 608–617. doi: 10.1105/tpc.16.00845

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., et al. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22 (6), 839–851. doi: 10.1261/rna.053959.115

Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., et al. (2019). MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant* 12 (6), 879–892. doi: 10.1016/j.molp.2019.01.003

Sekhon, R. S., Saski, C., Kumar, R., Flinn, B. S., Luo, F., Beissinger, T. M., et al. (2019). Integrated genome-scale analysis identifies novel genes and networks underlying senescence in maize. *Plant Cell* 31 (9), 1968–1989. doi: 10.1105/tpc.18.00930

Selitsky, S. R., Marron, D., Hollern, D., Mose, L. E., Hoadley, K. A., Jones, C., et al. (2020). Virus expression detection reveals RNA-sequencing contamination in TCGA. *BMC Genomics* 21 (1), 79. doi: 10.1186/s12864-020-6483-6

Serin, E. A. R., Nijveen, H., Hilhorst, H. W.M., and Ligterink, W.. (2016). Learning from Co-expression networks: possibilities and challenges. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00444

Shao, L., Xing, F., Xu, C., Zhang, Q., Che, J., Wang, X., et al. (2019). Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc. Natl. Acad. Sci.* 116 (12), 5653–5658. doi: 10.1073/pnas.1820513116

Sibbesen, J. A., Eizenga, J. M., Novak, A. M., Sirén, J., Chang, X., Garrison, E., et al. (2023). Haplotype-aware pantranscriptome analyses using spliced pangenome graphs. *Nat. Methods* 20, 239–247. doi: 10.1038/s41592-022-01731-9

Sigel, E. M., Der, J. P., Windham, M. D., and Pryer, K. M. (2019). Expression level dominance and homeolog expression bias in recurrent origins of the allopolyploid fern polypodium hesperium. *Am. Fern J.* 109 (3), 224. doi: 10.1640/0002-8444-109.3.224

Signal, B., and Kahlke, T. (2022). How_are_we_stranded_here: quick determination of RNA-seq strandedness. *BMC Bioinf.* 23 (1), 49. doi: 10.1186/s12859-022-04572-7

Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y.-S., Gill, B. S., et al. (2006). Molecular characterization of the major wheat domestication gene q. *Genetics* 172 (1), 547–555. doi: 10.1534/genetics.105.044727

Soneson, C., Love, M. I., and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* 4, 1521. doi: 10.12688/f1000research.7563.1

Sorenson, R. S., Deshotel, M. J., Johnson, K., Adler, F. R., and Sieburth, L. E. (2018). Arabidopsis mRNA decay landscape arises from specialized RNA decay substrates, decapping-mediated feedback, and redundancy. *Proc. Natl. Acad. Sci.* 115 (7), E1485–E1494. doi: 10.1073/pnas.1712312115

Spoelhof, J. P., Soltis, P. S., and Soltis, D. E. (2017). Pure polyploidy: closing the gaps in autopolyploid research: pure polyploidy. *J. Sys. Evol.* 55 (4), 340–352. doi: 10.1111/jse.12253

Sreeratree, J., Butsayawarapat, P., Chaisan, T., Somta, P., and Juntawong, P. (2022). RNA-Seq reveals waterlogging-triggered root plasticity in mungbean associated with ethylene and jasmonic acid signal integrators for root regeneration. *Plants* 11 (7), 930. doi: 10.3390/plants11070930

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA Sequencing: the teenage years. *Nat. Rev. Genet.* 20 (11), 631–656. doi: 10.1038/s41576-019-0150-2

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi: 10.1073/pnas.0506580102

Szalma, S. J., Hostert, B. M., LeDeaux, J. R., Stuber, C. W., and Holland, J. B. (2007). QTL mapping with near-isogenic lines in maize. *Theor. Appl. Genet.* 114 (7), 1211–1228. doi: 10.1007/s00122-007-0512-6

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28 (5), 511–515. doi: 10.1038/nbt.1621

Undurraga, S. F., Press, M. O., Legendre, M., Bujdoso, N., Bale, J., Wang, H., et al. (2012). Background-dependent effects of polyglutamine variation in the arabidopsis thaliana gene ELF3. *Proc. Natl. Acad. Sci.* 109 (47), 19363–19367. doi: 10.1073/pnas.1211021109

Van den Berge, K., Hembach, K. M., Soneson, C., Tiberi, S., Clement, L., Love, M. I., et al. (2019). RNA Sequencing data: hitchhiker's guide to expression analysis. *Annu. Rev. Biomed. Data Sci.* 2 (1), 139–173. doi: 10.1146/annurev-biodatasci-072018-021255

Vennapusa, A. R., Somayanda, I. M., Doherty, C. J., and Jagadish, S. V.K.. (2020). A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fiber. *Sci. Rep.* 10 (1), 16887. doi: 10.1038/s41598-020-73958-5

Voshall, A., and Moriyama, E. N. (2020). Next-generation transcriptome assembly and analysis: impact of ploidy. *Methods* 176, 14–24. doi: 10.1016/j.ymeth.2019.06.001

Walsh, J. R., Woodhouse, M. R., Andorf, C. M., and Sen, T. Z. (2020). Tissue-specific gene expression and protein abundance patterns are associated with fractionation bias in maize. *BMC Plant Biol.* 20 (1), 4. doi: 10.1186/s12870-019-2218-8

Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., Jiao, Y., et al. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* 7 (1), 11708. doi: 10.1038/ncomms11708

Wang, X., Zhang, H., Li, Y., Zhang, Z., Li, L., Liu, B., et al. (2016). Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA -sequencing. *New Phytol.* 209 (3), 1264–1277. doi: 10.1111/nph.13678

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., et al. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* 49 (4), 579–587. doi: 10.1038/ng.3807

Wang, T., Peng, H., Cao, Y., Xu, J., Xiong, Y., Liu, K., et al. (2022). Dynamic network biomarker analysis reveals the critical phase transition of fruit ripening in grapevine. *Genes* 13 (10), 1851. doi: 10.3390/genes13101851

Wang, Y. X. R., Waterman, M. S., and Huang, H. (2014). Gene coexpression measures in large heterogeneous samples using count statistics. *Proc. Natl. Acad. Sci.* 111 (46), 16371–16376. doi: 10.1073/pnas.1417128111

Wang, T., and Zhang, X. (2021). Genome-wide dynamic network analysis reveals the potential genes for MeJA-induced growth-to-defense transition. *BMC Plant Biol.* 21 (1), 450. doi: 10.1186/s12870-021-03185-1

Weeks, D. P. (2017). "Gene editing in polyploid crops: wheat, camelina, canola, potato, cotton, peanut, sugar cane, and citrus," in *Progress in molecular biology and translational science* (Elsevier), 65–80. doi: 10.1016/bs.pmbts.2017.05.002

Wicker, T., Schulman, A. H., Tanskanen, J., Spannagl, M., Twardziok, S., Mascher, M., et al. (2017). The repetitive landscape of the 5100 mbp barley genome. *Mobile DNA* 8 (1), 22. doi: 10.1186/s13100-017-0102-3

Wicker, T., Gundlach, H., Spannagl, M., Uauy, C., Borrill, P., Ramírez-González, R. H., et al. (2018). Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biol.* 19 (1), 103. doi: 10.1186/s13059-018-1479-0

Williams, A. G., Thomas, S., Wyman, S. K., and Holloway, A. K (2014). RNA-Seq data: challenges in and recommendations for experimental design and analysis. *Curr. Protoc. Hum. Genet.* 83 (1), 11.13.1-11.13.20. doi: 10.1002/0471142905.hg1113s83

Williams, C. R., Baccarella, A., Parrish, J. Z., and Kim, C. C. (2016). Trimming of sequence reads alters RNA-seq gene expression estimates. *BMC Bioinf.* 17 (1), 103. doi: 10.1186/s12859-016-0956-2

Wulff, R. (2017). Environmental maternal effects on seed quality and germination. *Seed Dev. germination*, Eds. by Jaime Kigel, Routledge 491–505. doi: 10.1201/9780203740071-18

Yang, J., Liu, Y., Pu, Z., Zhang, L., Yuan, Z., Chen, G., et al. (2014). Molecular characterization of high pI α-amylase and its expression QTL analysis in synthetic wheat RILs. *Mol. Breed.* 34 (3), 1075–1085. doi: 10.1007/s11032-014-0098-z

Yang, G., Liu, Z., Gao, L., Yu, K., Feng, M., Yao, Y., et al. (2018). Genomic imprinting was evolutionarily conserved during wheat polyploidization. *Plant Cell* 30 (1), 37–47. doi: 10.1105/tpc.17.00837

Yang, A., Tang, J. Y. S., Troup, M., and Ho, J. W. K. (2022). Scavenger: a pipeline for recovery of unaligned reads utilising similarity with aligned reads. *F1000Research* 8, 1587. doi: 10.12688/f1000research.19426.2

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11 (2), R14. doi: 10.1186/gb-2010-11-2-r14

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi: 10.1089/omi.2011.0118

Yue, R., Lu, C., Qi, J., Han, X., Yan, S., Guo, S., et al. (2016). Transcriptome analysis of cadmium-treated roots in maize (Zea mays l.). *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01298

Zhang, Y., Parmigiani, G., and Johnson, W. E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics Bioinf.* 2 (3), lqaa078. doi: 10.1093/nargab/lqaa078

Zhao, N., Cui, S., Li, X., Liu, B., Deng, H., Liu, Y., et al. (2021). Transcriptome and Co-expression network analyses reveal differential gene expression and pathways in response to severe drought stress in peanut (Arachis hypogaea l.). *Front. Genet.* 12. doi: 10.3389/fgene.2021.672884

Zhou, L., Chi-Hau Sue, A., and Bin Goh, W. W. (2019). Examining the practical limits of batch effect-correction algorithms: when should you care about batch effects? *J. Genet. Genomics* 46 (9), 433–443. doi: 10.1016/j.jgg.2019.08.002

Zimin, A., Stevens, K. A., Crepeau, M. W., Holtz-Morris, A., Koriabine, M., Marçais, G., et al. (2014). Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics* 196 (3), 875–890. doi: 10.1534/genetics.113.159715