

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,800

Open access books available

142,000

International authors and editors

180M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Recent Applications of RNA Sequencing in Food and Agriculture

*Venkateswara R. Sripathi, Varsha C. Anche,
Zachary B. Gossett and Lloyd T. Walker*

Abstract

RNA sequencing (RNA-Seq) is the leading, routine, high-throughput, and cost-effective next-generation sequencing (NGS) approach for mapping and quantifying transcriptomes, and determining the transcriptional structure. The transcriptome is a complete collection of transcripts found in a cell or tissue or organism at a given time point or specific developmental or environmental or physiological condition. The emergence and evolution of RNA-Seq chemistries have changed the landscape and the pace of transcriptome research in life sciences over a decade. This chapter introduces RNA-Seq and surveys its recent food and agriculture applications, ranging from differential gene expression, variants calling and detection, allele-specific expression, alternative splicing, alternative polyadenylation site usage, microRNA profiling, circular RNAs, single-cell RNA-Seq, metatranscriptomics, and systems biology. A few popular RNA-Seq databases and analysis tools are also presented for each application. We began to witness the broader impacts of RNA-Seq in addressing complex biological questions in food and agriculture.

Keywords: RNA-Seq, transcriptome, transcripts, genes, variants, gene expression, analysis, applications, databases, and tools

1. Introduction

Transcriptome broadly refers to a collection of RNA transcripts within a particular context that includes combinations of spatial and temporal factors: biological level of organization, from organelle to organism; and phase of growth, differentiation, or development, from zygote through adult. Additionally, one can investigate transcriptomes under more experimental contexts by controlling or varying the factors mentioned above, along with combinations of environmental, genetic, and physiological conditions. All of these factors influence the constituents of a transcriptome, an array of RNA types that traditionally fall into two categories: coding, the messenger RNAs (mRNAs); and non-coding (ncRNAs), such as ribosomal (rRNAs), transfer (tRNA), small interfering (siRNAs), micro (miRNAs), tRNA-derived small (tsRNA), Piwi-interacting (piRNAs), short hairpin (shRNAs), small nuclear (snRNAs), small nucleolar (snoRNAs), long non-coding (lncRNAs), and circular RNAs (circRNAs) [1, 2]. Interestingly, studies have questioned this sharp distinction between coding and non-coding RNAs, paving the way for more research into multifunctional RNA

types that transcend this traditional dichotomy [3, 4]. Given the complex definitions of transcriptome and its constituent RNAs, keen attention is required in understanding and managing the context within which a transcriptome is generated and analyzed throughout the experimental procedure and downstream analysis.

Thus far, RNA research efforts have concentrated on a few major types of RNAs: mRNAs, rRNAs, tRNAs, and miRNAs. Accounting for 3-4% of the total RNA in a cell [5], mRNAs are products of transcription and, in eukaryotes, multiple processing steps that usually involve the addition of adenosine monophosphates to form a poly(A) tail via polyadenylation [6]. This coding mRNA is then translated into an amino acid (AA) chain by the ribosome, in a process incorporating ribosomal proteins, AAs, and non-coding RNAs, such as rRNAs and tRNAs. About 60% of the ribosome's mass [7] and up to 95% of the total RNA in a cell [8] can consist of rRNAs, which facilitate mRNA and tRNA binding while catalyzing the transfer of an AA from the tRNA to the growing AA chain. Many processes that comprise gene expression, including the steps mentioned above, can be regulated by miRNAs [9]. These short (17-22 bp), single-stranded, non-coding RNAs are exclusive to eukaryotes and typically bind to complementary sequences on mRNA molecules, thereby inducing degradation or inefficient translation of the target transcript [10].

These four major types of RNA and the multitude of minor types can be selectively isolated and analyzed using various wet lab and dry lab techniques, depending on the specific applications and biological questions under investigation. In the case of transcriptome profiling for coding RNAs in a eukaryotic organism, the ratio of mRNA to rRNAs can be increased: first during library preparation through poly(A) selection, ribosomal depletion, and size selection strategies; and again during the bioinformatic analysis by rRNA filtering during the initial quality control (QC) step in the pipeline. Especially for capturing miRNAs, in addition to rRNA decontamination steps, size selection strategies are used for selective isolation of small RNA [11]. Many bioinformatics tools are available customized for short sequence alignments [12], and a few can evaluate the thermodynamics of miRNA secondary structures [13]. The molecular biology of RNA transcription, processing, transportation, and translation can be drastically different between phylogenetically distant organisms, and hence the taxonomy of the species being studied is often considered. A variety of wet lab and dry lab techniques have been developed to account for the biological differences in mRNA structure and processing throughout the phylogenetic tree of life.

Transcriptome analysis evolved steadily from nucleic acid detection methods (e.g., northern blots), to hybridization-based methods (e.g., microarrays), through a multitude of sequencing-based methods (e.g., RNA-Seq). RNA-Seq has been the most widely used approach for analyzing transcriptomes obtained from phylogenetically diverse organisms [14]. The swift advancements in RNA-Seq research are being driven by the continual improvements in sequencing technologies (first, second, and third generation), which have steadily provided higher throughput, lower cost, and more accurate sequencing for transcriptome analyses. Despite the availability of many sequencing technologies, the Illumina short-read method remains the most widely used platform for transcriptome sequencing, and many consider it as the gold-standard sequencing for single-nucleotide resolution transcriptome analysis with an accuracy of 99.99% and minimal biases [15]. This method has evolved from 35 bp to 350 bp fragment sequencing in the past decade, and it offers multiple library preparation options, including single-end, mate-pair, and paired-end. Library preparation can yield either stranded sequences, where the sense and/or antisense orientation of the output reads is known, or unstranded sequences, where the read orientation is unknown. Stranded RNA-Seq enables the resolution

of both sense and antisense transcription for genes overlapping on opposite strands [16], and it remains the standard for most RNA-Seq applications.

A thorough conceptual understanding of the prospective RNA-Seq experiment is required to overcome the plethora of potential biases, errors, misinterpretations, and other various challenges common in RNA-Seq experiments [17, 18]; researchers ought to precisely monitor and engineer each phase of the entire process, wet lab through the dry lab, from beginning to end and in all steps between: experimental design, sample collection, RNA isolation, RNA-QC, adapter ligation, multiplexing, library preparation, library-QC, sequencing, data collection, demultiplexing, pre-processing, data-QC, analyses, and interpretation. The experimental design is the first fundamental process in RNA-Seq analysis. When the goal is to detect statistically significant, differentially expressed genes (DEGs), increasing the number of replicates usually has a more positive effect than increasing the sequencing depth, especially when sequencing over 2 million reads per sample [19, 20]. For most RNA-Seq experiments, six or more biological replicates are recommended, and at least three biological replicates are necessary. If one aims to identify DEGs, then pooling biological replicates before multiplexing is discouraged, but such pooling might be pragmatic when one only attempts to assemble a comprehensive transcriptome. Contrary to biological replicates, technical replicates are unnecessary for RNA-Seq on modern sequencing platforms [19], and resources can be better utilized by increasing the number of biological replicates and minimizing batch effects from unintended influences, such as variance in personnel, in the laboratory environment, and in the selection and usage of materials and methods. A thorough review of the expansive RNA-Seq landscape is available, and to confine our discussion to the scope of this chapter, we will be highlighting the most popular and current RNA-Seq applications in food and agriculture.

2. RNA sequencing (RNA-Seq) applications

2.1 Differential gene expression (DGE)

As previously mentioned, transcriptomes are spatially and temporally dynamic, and they evolve in response to changing environmental, genetic, and physiological conditions. For instance, the transcriptome of one cell type can be significantly different from another cell type, even within the same tissue, and similarly, the transcriptome of a particular cell can vary drastically, as it transitions through the cell cycle, differentiates, acclimates to environmental factors, adapts to the introduction of particular treatments, or changes during disease progression. RNA-Seq can detect such changes in gene expression levels between samples and, in DGE studies, between two or more experimental groups [21, 22]. DGE analysis seeks to identify statistically significant genes that are expressed differently between groups, which are generated through careful attention to experimental design [23]. DGE studies can elucidate functional elements of the genome by identifying gene-level relationships between transcript abundance and experimental conditions, thereby illuminating the mechanisms of associated physiological processes and expanding our understanding of the links between genotype and phenotype [24].

While DGE analysis focuses on quantifying and comparing the complete collection of all transcript isoforms for a gene to identify differentially expressed genes (DEGs), differential isoform expression (DIE) analysis focuses on quantifying and comparing each individual isoform in a collection of transcripts associated with a particular gene, to identify differentially expressed isoforms (DEIs)

Databases			Tools		
S.No	Database	Citation	S.No	Tool	Citation
1	SPEED2	[26]	1	Ballgown	[27]
2	ImaGEO	[28]	2	limma	[29]
3	GXD	[30]	3	NOISeq	[31]
4	RED	[32]	4	DESeq2	[33]
5	Omnibus database	[34]	5	edgeR	[35]

Table 1.
Some popular databases and tools in finding DEGs in RNA-Seq data.

between experimental groups [25]. The materials and methods for analyzing DEGs differ from those used for DEIs. The decision to find differential genes or isoforms is crucial and determines the downstream analysis, and it is ideally taken at the beginning of the experiment. Given these differences, we discussed the methods most relevant to DGE analysis, since it has been more deeply studied and widely applied. Some methods being applied to investigate DEGs include northern blot, western blot, quantitative real-time PCR (qPCR), expressed sequence tags (ESTs), microarrays, and RNA-Seq. Most bioinformatics pipelines for DGE analysis of RNA-Seq data include five main stages: QC, alignment, quantification, normalization, and DGE calculation, which usually assumes either a negative binomial, log-normal, or nonparametric statistical distribution. Many databases and bioinformatics tools are available for all these stages and downstream analyses, and a few popular, reliable databases and DGE calculation tools are presented below (**Table 1**). Often each program will output slightly different collections of statistically significant DEGs [21], so many investigators use multiple tools, assign higher confidence to intersectional DEGs, and then continue by piping these results through various downstream functional analyses, which will be discussed later in this chapter.

RNA-Seq followed by DGE analysis has been extensively used in the agriculture and food industry. Poultry scientists have applied RNA-Seq analysis to identify DEGs associated with the eggshell formation in the shell gland at different time-points in laying hens [36]. A dairy research group identified significant enrichment of DEGs associated with mammary gland development, milk protein formation, lipid metabolism, and other biological processes linked with milk production traits in lactating cows [37]. Interestingly, the possible roles of DEGs involved in pathogenesis-related pathways in response to peanut allergy have been examined by comparing the transcriptome profiles of high-risk and risk-free infants, facilitating early detection of food allergies in infants [38]. The symbiotic association between rhizobium bacteria and root nodules in leguminous plants is important in agriculture and soil metagenomics, as this interaction improves soil fertility by nitrogen fixation and increases crop production. Differences in nodulation phenotypes have been observed by comparing two diverse symbiotic systems at different time-points using RNA-Seq [39]. Furthermore, these researchers identified DEGs in response to specific strains of rhizobia in soybean roots, and the majority of these DEGs were involved in plant-pathogen interactions and flavonoids biosynthesis [39]. By studying global transcriptome profiles in strawberry fruits, plant scientists have elucidated the influence of red and blue light on the differential expression of genes associated with anthocyanin biosynthesis and accumulation [40].

2.2 Variants calling and detection

The genetic variations in the coding region may or may not alter the amino acid sequence, resulting in asynonymous or synonymous variants, respectively; characterizing such variants is important for associating the genomic locations with a trait or phenotype [41]. RNA-Seq can be used to identify variations in the coding sequences, including single-nucleotide variants (SNVs), short insertions/deletions (indels <50 bp), and structural variants (SVs). SNVs result from a single nucleotide substitution at a particular coordinate and single-nucleotide polymorphism (SNP) refers to a frequent SNV, generally present in at least 1% of the subject population [42]. SNPs are ubiquitous throughout the coding, non-coding, and regulatory regions of the genome. In comparison, a haplotype is a set of genes, alleles, or SNPs, which are inherited together. Copy number variations (CNVs) are a type of SV where regions in the genome are repeated, and the number of these repeats varies among individuals due to duplication or deletion events. The percentage of CNVs detected in diverse organisms varied significantly. Over 80% and > 15% of the detected SNPs and CNVs were associated with gene expression in the mammalian system, respectively [43].

Many experimental methods have been developed to detect genetic variants in the genomes of plants and animals, and a few routinely used techniques include rhAmp (RNase H2-dependent amplification assay), Kompetitive Allele-Specific PCR (KASP), TaqMan, Fluidigm, AmpliSeq, Fluorescence In Situ hybridization (FISH), qRT-PCR, microarray, and RNA-Seq. When generating RNA-Seq data for the downstream bioinformatics analysis, sequencing depth is a major consideration, given its influence on not only the overall results but also the cost of experimentation; and after analyzing variants for mutated myeloid genes, researchers suggested 30-40 million paired-end reads per sample was sufficient [44]. Additionally, highly variable coverage between different genes can hinder variant calling and annotation of RNA-Seq data. To identify variants (SNPs and short indels) in RNA-Seq reads, a typical bioinformatics pipeline involves three phases: data clean-up, variant discovery and filtering, and evaluation. A selection of databases and programs for variant analysis is presented below (**Table 2**).

The application of RNA-Seq in genome-wide screening for genetic variants is imperative to accelerate the usage of genome-based breeding approaches for selecting agriculturally desirable traits in plants [55] and animals [41, 56]. Functional SNPs associated with quality traits (e.g., plant color, flowering, fruit color, size, and ripening) and/or quantitative traits (e.g., grain yield, abiotic, and biotic stress tolerance) may result in phenotypic diversity among individuals. Previous studies have used RNA-Seq analysis to identify SNPs in relatively smaller genomes, such as barley [57], and larger genomes, such as wheat [58]. One of the main

Databases			Tools		
S.No	Database	Citation	S.No	Tool	Citation
1	AWESOME	[45]	1	AthCNV	[46]
2	KoVariome	[47]	2	GATK workflow	[48]
3	lncRNASNP2	[49]	3	SQUID	[50]
4	SNP2TFBS	[51]	4	DeepVariant	[52]
5	rSNPBase	[53]	5	VarDict	[54]

Table 2.
A few databases and tools in finding structural variants in RNA-Seq.

goals of livestock germplasm improvement is identifying the genetic variation associated with phenotypic traits of economic importance. By screening 15 duck transcriptomes, SNPs in genes related to fat metabolism and digestion were found in genomic regions that have undergone selective pressures [41]. In a similar study, SNPs associated with the fat deposition in sheep have been identified, potentially leading to breeding programs that reduce tail size in fat-tailed phenotypes [59]. While comparing RNA-Seq variant analysis methodologies for investigating beef production in Nellore steers, researchers recently identified SNPs in genes related to feed efficiency, an economically important trait in cattle [60].

2.3 Allele-specific expression (ASE)

RNA-Seq data can be used to investigate allele-specific expressions (ASEs), which denotes a differential expression of two or more alleles in a diploid or a polyploid organism, sometimes may result in multiple traits and phenotypes. Heterozygous SNPs may lead to ASE, and this phenomenon is conserved in most higher organisms, including those in plant and animal kingdoms. Due to the intrinsic potential of heterozygous SNPs, ASE can be a sensitive marker for detecting cis-regulatory variation and reducing background noise in an individual [61]. Heterozygous variants have been identified in coding regions of mRNA, possibly leading to a variant polypeptide or a truncated protein [62]; non-coding regions (splice site, 5'-UTR, or 3'-UTR), possibly influencing mRNA processing and degradation [63]; and non-coding regulatory regions (promoter, enhancer, or silencer), possibly affecting the binding of transcription and epigenetic factors [64]. Genetic and epigenetic factors regulate transcriptional activity and contribute to ASE, and an imbalanced expression via heterozygous SNP loci in a non-haploid genome may lead to a diseased or abnormal condition [65]. Using whole genome sequencing (WGS) alone, variants throughout the entire genome can be identified. However, by combining WGS and RNA-Seq analyses, ASE and allele silencing information can also be obtained.

Of the many bioinformatics tools and databases created to explore ASE, a few are listed here (**Table 3**). However, despite the recent developments in ASE bioinformatics analysis, significant challenges in applying these tools include: 1) required family tree information, i.e., sequencing data from the individual under investigation and their respective parents, which is more laborious and costly; 2) required phased genotype information, i.e., the haplotype of the individual must be known in order to use the source file as input; 3) commonly required genomic and transcriptomic data to obtain ASE, but MBASED (**Table 3**) requires only RNA-Seq data; 4) common usage of short-read data (100-250 bp) due to the low error rate, which is incapable of covering multiple SNVs and subject to read bias at the exon-intron

Databases			Tools		
S.No	Database	Citation	S.No	Tool	Citation
1	dbGaP	[66]	1	EMASE	[67]
2	Genotype-Tissue Expression, GTEx	[68]	2	IDP-ASE	[69]
3	AD ASTRA	[70]	3	QuSAR	[71]
4	dbNSFP	[72]	4	ASEQ	[73]
5	Genevar	[74]	5	MBASED	[75]

Table 3.
Some widely used databases and tools in finding ASE in RNA-Seq data.

junctions; and 5) lack of advanced statistical methods. Long read (1-100 kb) data allows the detection of multiple SNVs, but it is prone to high error rates and low throughput, which is not ideal for downstream ASE quantification. Therefore, researchers can use a hybrid sequencing approach that combines both short and long reads. IDP-ASE (**Table 3**) can utilize such hybrid data to simultaneously phase haplotype and quantify the ASE at both gene and transcript/isoform levels. More sophisticated tools are required to identify ASE associated with multiple phenotypes and complex traits in comprehensive datasets.

Using genome-wide analysis, the underlying genetic and molecular mechanisms associated with ASE in heterosis have been determined in hybrid rice [76]. ASE of *Dof* genes in response to plant hormone signaling and abiotic stresses is likely mediated through cis-regulatory elements that could be useful for sugarcane crop improvement [77]. Genome-wide expression quantitative trait loci (eQTL) and ASE analyses helped identify candidate genes that determine the meat quality traits in pigs [78]. Similarly, ASE is a widespread phenomenon in the bovine genome, and its effects on the meat quality and production traits in Nellore steers have been studied by combining genotyping and RNA-Seq data from skeletal muscle tissue [79]. With RNA-Seq data from three different tissues (liver, fat, and breast muscle) in commercial broiler chickens, researchers examined the biological mechanisms of ASE variants and their associated meat traits in poultry production by using recently developed bioinformatics software, Variant Call Format (VCF) ASE Detection Tool (VADT) [80].

2.4 Alternative splicing (AS)

During the canonical splicing process in eukaryotes, introns are removed as lariats, and the flanking exons are rejoined to form a processed mRNA, with sequences in the RNA determining where splicing occurs. Usually, exons of the same mRNA are spliced, but sometimes exons from different mRNAs can be combined by trans-splicing [81]. The RNA splicing machinery is a complex of proteins called the spliceosome, its major components being small nuclear Ribo-Nuclear Proteins (snRNPs). The three main types of spliceosome complexes are GU-AG spliceosome (major spliceosome), AU-AC spliceosome, and trans-spliceosome [82]. In general, three main classes of RNA splicing are found: pre-mRNA splicing, Group II introns self-splicing, and Group I introns self-splicing. A single gene can produce multiple products by alternative splicing (AS). In addition to normal, canonical splicing, the primary AS events identified in eukaryotes are exon skipping (ES), mutually exclusive exons (EE), alternative 5' donor sites (A5), alternative 3' acceptor sites (A3), alternative promoters (AP), intron retention (IR), and alternative polyadenylation (APA) [83]. Of these, the later three events gained attention recently with the advancements in RNA-Seq. AS is often regulated by activator and repressor proteins, and it can lead to premature termination of translation due to the interaction of exon junction complexes (EJC) with release factors, triggering the Nonsense-Mediated mRNA Decay (NMD) pathway [84].

RNA-Seq data can be assembled into full-length isoforms from the raw reads associated with AS of the same gene, and then the corresponding AS events can be identified and characterized. Mate-pair and paired-end sequences have performed better than single-end short-reads for detecting AS patterns [85]. Among the contemporary approaches, long-read sequencing (PacBio/Oxford Nanopore) is an ideal solution for generating full-length transcript sequences and detecting AS events and isoforms [86]. Full-length isoforms can be assembled with or without a reference, and each approach requires specific bioinformatics software. Some of these AS tools and databases are presented here (**Table 4**). Many AS tools can be

Databases			Tools		
S.No	Database	Citation	S.No	Tool	Citation
1	DIGGER	[87]	1	SplicingFactory	[88]
2	MeDAS	[89]	2	ASpli	[90]
3	ASlive	[91]	3	ASGAL	[92]
4	CuAS	[93]	4	MAJIQ	[94]
5	SpliceDisease	[95]	5	rMATS	[96]

Table 4.
A few popular databases and tools in finding AS events in RNA-Seq data.

used to analyze these AS events genome-wide and/or for a single gene. For example, the ASGAL pipeline (**Table 4**) begins by building a splice graph from a reference genome and an annotation file. Then, the RNA-Seq reads are aligned to the splice graph. Finally, these splice graph alignments are used to detect novel AS events.

Emerging functional roles of AS in generating transcriptomic and proteomic diversity have been evident in diverse biological processes [97]. In the tea leaves of a *Camellia sinensis* cultivar, approximately 64% of genes underwent an AS event, and many of these events were influenced by heat, drought, and their combined stresses [98]. Naturally occurring splice variants in the population have been used in detecting genotype-specific AS events, and in turn, these events have served as biomarkers for genome-wide association studies (GWAS) in rice subjected to salt stress [99]. Comparative transcriptome analyses of fruit, seedling, and flower tissues in tomatoes revealed more AS events in fruits. About 60% of the tomato's multi-exon genes undergo AS events, among which IR is prevalent. Also, the gene expression is preferentially regulated at the isoform level during early fruit development [100].

2.5 Alternative polyadenylation (APA) site usage

During post-transcriptional processing at the 3'UTR region of pre-mRNA, differential usage of polyadenylation sites can lead to a diverse set of transcript isoforms with different 3'UTR lengths and sequences, as part of a ubiquitous regulatory mechanism called Alternative Polyadenylation (APA). Most eukaryotic genes have multiple APA sites (APAs) that are often found in a coding region (CR-APA) or 3'UTR (UTR-APA) [101]. APAs found in internal intronic and exonic regions account for a small proportion of identified APAs, but these predominantly disrupt the coding regions and can result in variable protein isoforms or NMD decay [102]. In contrast, APAs found in the terminal exon and 3'UTR regions account for a significant proportion of identified APAs, and though such APAs usually do not disrupt the coding regions, they may result in transcript isoforms with variable lengths. A poly(A) tail in the 3'UTR region of an mRNA transcript generally provides mRNA stability, localization, and translational efficiency, so these factors are subject to APA-mediated regulation [103]. Since the 3'UTR region can have hotspots for the binding of miRNAs and RNA-binding proteins (RBPs), any modifications in this region may lead to new RNA species interactions or the formation of novel secondary structures, thereby affecting translational efficiency [101, 103]. APAs likely play a role in many processes involved in gene expression, including nuclear export, localization, stability, degradation, repression, translation, and protein diversification [104]. Additionally, APAs associated with differentiation, proliferation, and tissue-specific expression have been reported [105].

Databases			Tools		
S.No	Database	Citation	S.No	Tool	Citation
1	TREND-DB	[106]	1	Deerect-apa	[107]
2	Animal-APAdb	[108]	2	APalyzer	[109]
3	PlantAPAdb	[110]	3	scDAPA	[111]
4	APAAtlas	[112]	4	DeepPASTA	[113]
5	APADB	[114]	5	TAPAS	[115]

Table 5.
 Some popular databases and tools in finding APAs in RNA-Seq data.

APAs at the gene-level can be discovered using EST, microarray, RNA-Seq, 3' RNA-Seq, and qRT-PCR methodologies. However, genome-wide screening for APAs can be achieved through NGS based approaches, such as Whole Transcriptome Termini Site sequencing (WTTS-Seq), poly(A) site sequencing (PAS-Seq), direct RNA sequencing (DRS), poly(A) single-molecule sequencing, as well as 3' region extraction and deep sequencing (3' READS). Moreover, researchers can engage in cell type-specific APA profiling by preprocessing the samples with specialized wet-lab methods, such as cell sorting, crosslinking immunoprecipitation and green fluorescent protein (GFP)-tagging, and cellular and molecular barcoding. All these methods utilize total RNA or mRNA as their starting material, but they diverge in their usage of polyA enrichment, library preparation, and sequencing strategies. Usually, NGS data analysis for APAs includes preprocessing, size selection, QC, mapping/assembly, normalized expression value assessment for the poly(A) enriched 3'UTRs or transcripts, DGE, functional annotation, motif analysis, and pathway analysis. A few tools that use most of these steps and databases for APA analysis are presented (Table 5).

APA processing has been associated with around 70% of human genes, with the longest resulting isoform for each usually observed to be the most abundant [102, 116]. Recent studies have proposed a role for APAs in leaf development and stress response in the two dominant rice (*Oryza sativa* L.) subspecies, indica and japonica, possibly accounting for significant differences in their phylogenetic divergence [117]. They also demonstrated that variations in 3'UTR length from APA resulted in DEGs associated with many important agronomic traits related to rice yield [117]. The possible role of APA in remodeling root-associated transcriptomes has been observed in Sorghum [118], Bamboo [119], and Arabidopsis [120] in response to diverse abiotic stresses. Currently, APA is underexplored and offers many opportunities for significant contributions to the food and agriculture sectors.

2.6 microRNA (miRNA) profiling

RNA-Seq can identify and characterize diverse classes of small (17-200 bp) ncRNAs, including miRNAs, siRNAs, piRNAs, tsRNAs, snoRNAs, and snRNAs. Almost all types of RNAs crosstalk, and especially miRNAs, the abundant class of sRNAs act as mediator molecules in regulating and deregulation of genes via complementary binding to miRNA response elements (MREs) on target transcripts [121]. Moreover, co-localization and co-expression of ncRNA and mRNA and their interactions are well established [122]. MiRNA genes can be found in exonic, intronic, and intergenic regions of the genome, and they are predominantly localized, form clusters, and generally transcribed together as a single transcriptional unit. The various miRNAs can positively and/or negatively regulate

gene expression post-transcriptionally or by translational repression [123]. While competing endogenous RNA, ceRNAs (e.g., lncRNAs and circRNAs) contain MREs and can regulate gene expression by acting as “miRNA sponges”, thus reducing the availability of one or more miRNAs for other potential targets [121]. A nascent miRNA transcript undergoes post-transcriptional processing and nuclear export during the canonical regulation, eventually being loaded into the RNA-induced silencing complex (RISC) [124]. After the incorporated miRNA binds to a target mRNA at MREs often located in the 3'-UTR, RISC mediates gene expression by post-transcriptional gene silencing (PTGS) or by mRNA cleavage or mRNA degradation [124]. However, the presence of ceRNAs challenges the canonical miRNA regulation of gene targets, and the mechanisms and functions of miRNA sponges are still unclear [121].

Though several wet lab and computational methods have been evolved in the past two decades for genome-wide screening of miRNAs, in silico approaches, continue to be more widely used due to the ease in exploring the properties of miRNAs. MiRNAs are highly conserved, and the thermodynamics of miRNA secondary structures and target binding have been elucidated; identification of conserved and novel miRNAs and their targets can be performed using readily available bioinformatics tools. A few frequently accessed databases and tools used are listed here (**Table 6**). Most studies have applied homology-based approaches in identifying conserved miRNAs, and miRNA precursors can be identified by conducting secondary structure analysis using RNAfold [140] or mfold [141]. The properties of miRNAs, such as cooperativity and multiplicity, can also predict miRNAs and their targets computationally [123].

Since the first reported miRNAs in *C. elegans*, different miRNAs have been identified in numerous organisms across multiple kingdoms [123]. Several studies have demonstrated their involvement in various biological processes and their potential to alter key agronomic traits [142]. Using RNA-Seq, the functional roles of miRNAs in various stresses (heat, drought, and salinity) have been reported in Arabidopsis [143] and Cotton [144]. Also, many conserved and novel miRNAs and their putative gene targets were identified in Upland cotton and its closest progenitor species using RNA-Seq, and the majority of these targets were transcription factors that were involved in the regulation of fiber growth and development and stress responses [123]. The role of miRNAs in various diseases has been established over two decades, but, recently some naturally occurring food-derived compounds and exogenous diet-derived miRNAs have been implicated in determining the human gut-associated miRNA expression and their profiles, which contributes to human health and well-being of an individual [145].

S.No	Databases		miRNA gene prediction tools		miRNA target prediction tools	
	Database	Citation	Tool	Citation	Tool	Citation
1	Rfam	[125]	UEA sRNA workbench	[126]	miRWalk	[127]
2	deepBase	[128]	Mirnovor	[129]	mirDIP	[130]
3	miRDB	[131]	miReader	[132]	psRNATarget	[133]
4	miRbase	[134]	miRDeep-star	[135]	TargetScan	[136]
5	Noncode	[137]	miRNAkey	[138]	mirSOM	[139]

Table 6.
A few popular databases and tools for miRNA analysis using RNA-Seq.

2.7 Circular RNAs

Among the many ncRNAs species, circRNAs are characterized by a stable, closed-loop structure formed through back-splicing via an upstream splice acceptor (SA) site, in contrast to the downstream SA sites of standard linear splicing [146]. CircRNAs span exonic, intronic, intergenic regions, UTR (5' and 3'), and lncRNA loci [147], and they are stable, conserved, non-random, as well as cell-type and tissue-specific [146]. Additionally, circRNAs have been found in all life domains, and, similar to miRNAs, their orthologous expression facilitates discovery, validation, and functional assignments. CircRNAs are transcribed at higher levels than mRNA in specific cells, tissues, or conditions, and they are expressed during chromatin remodeling [146] and in some disease-specific contexts [148]. For example, 14.4% of actively transcribed genes in human fibroblasts produced circRNAs [147], and due to their orthologous, tissue-specific, and spatial expression tendencies, circRNAs may be employed as plausible biomarkers in disease control and treatment [148]. Biological functions for circRNAs continue to be discovered and currently include scaffolding for RNA-binding proteins; formation of regulatory complexes; promotion of translation; regulation of protein function; and target decoys for other regulatory molecules, like miRNAs [149].

Similar to the methods used in experimental validation of linear mRNA, circRNA-forming exons can be determined by RNA-Seq, back-splice junction specific quantitative PCR (qPCR), northern blot, microarrays, RNA fluorescence in situ hybridization (FISH), Chromatin immunoprecipitation (ChIP), RNA immunoprecipitation (RIP), RNA pulldown, mass spectrometry, *in vitro* synthesis, luciferase reporter assays, and denaturing PAGE. RNase-R treated poly(A) mRNA samples and polyadenylated RNA-Seq are ideal for enriching and identifying circRNAs. These circRNAs can also be characterized by utilizing overexpression (*cis/trans*), knockdown (RNAi machinery), or knockout (CRISPR/Cas9 system) strategies. Based on the presence of a back-splice junction spanning locations in the RNA-Seq reads, researchers can characterize various types of circRNAs in their data [150] with a variety of bioinformatics tools and databases available (**Table 7**).

The biogenesis mechanisms and functional roles of plants are different from animals, but their expression-specific patterns are very similar [161]. Plant circRNAs have been implicated in stress-induced (dehydration, chilling, high-light, etc.) expression patterns [162]. Intricate regulatory roles of circRNAs in ripening through ethylene signaling pathway has been investigated using integrated RNA-Seq and bioinformatics analysis in tomato [163]. The role of circRNAs in the fat deposition by regulating adipogenic differentiation and lipid metabolism has been determined by studying subcutaneous adipose tissues of two pig breeds using

Databases			Tools		
S.No	Database	Citation	S.No	Tool	Citation
1	Circbank	[151]	1	circRNAprofiler	[152]
2	exoRBase	[153]	2	CircPlant	[154]
3	PlantcircBase	[155]	3	CircCode	[156]
4	circRNADb	[157]	4	Circ RNA wrap	[158]
5	circBase	[159]	5	CircTools	[160]

Table 7.
 Some databases and tools in finding circular RNAs from RNA-Seq data.

RNA-Seq and bioinformatics and their potential to serve as early diagnostic markers in treating metabolism-related diseases [164]. CircRNAs found on four casein genes in the bovine mammary gland harbor complementary sites for specific miRNAs, suggesting their regulatory role in milk protein synthesis. These circRNAs can be used to fine-tune the gene expression of casein genes, thus producing high-quality milk protein and enhanced milk in dairy cows [165].

2.8 Single-cell RNA-Seq

Cell-specific transcriptome changes are critical for understanding single cells or groups of cells throughout tissues, organs, and organ systems. Single-cell RNA-Seq (scRNA-Seq) can be used to measure individual gene expression in a single cell and the distribution of expression levels across a cell population. It was first developed to undertake the whole-transcriptome analysis of a single mouse blastomere [166] and gained widespread popularity recently due to sequencing chemistry advancements and the steep decline in sequencing costs since 2014. scRNA-Seq can illuminate the complex interplay between intrinsic cellular processes and extrinsic stimuli in cell fate determination [167], and scRNA-Seq can facilitate novel discovery species or regulatory processes, which may serve as tools in biotechnology and medicine [168]. Many scRNA-Seq protocols have been developed, often differing in their methods used for cell isolation [169], but studies continue to be limited by the difficulties of culturing certain cell types and by issues involving accurate and precise viable cell isolation [170].

Different methodologies are available in generating single-cell RNA-Seq data from a biological sample. However, most of these methodologies utilize these steps: 1) digest the tissue, i.e., single-cell dissociation; 2) isolate single cells by plate-based or droplet-based methods; 3) capture intracellular mRNA and prepare the massively multiplexed library with sample-specific cellular barcodes or unique molecular identifiers (UMI); 4) sequence on an NGS platform to generate raw reads. Several different platforms and frameworks (stand-alone, cloud-based, and interactive web-based) are presently available for conducting the bioinformatics analysis of scRNA-Seq data, and a few examples for each platform are listed in **Table 8**. The majority of scRNA-Seq frameworks partially or fully follow these steps: QC; alignment; mapping QC; cell QC; normalization; batch correction; imputation; cell cycle-assignment; feature selection; dimensionality reduction and visualization; pseudotime; cell type annotation; DGE; unsupervised clustering; and network analysis.

scRNA-Seq has been a valuable tool in determining differential gene expression by using gene cluster analyses among heterogeneous cell types and understanding their complex interactions and cellular responses in woody plants [186]. The use

Databases			Web-based scRNA-tools		Cloud-based scRNA-tools	
S.No	Database	Citation	Tool	Citation	Tool	Citation
1	SC2disease	[171]	scMappR	[172]	GranatumX	[173]
2	Curated database	[174]	CHARTS	[175]	Cumulus	[176]
3	PanglaoDB	[177]	alona	[178]	SCelVis	[179]
4	scRNA-tools database	[180]	SingleCellNet	[181]	PscB	[182]
5	scRNASeqDB	[183]	Single Cell Explorer	[184]	Falco	[185]

Table 8.
A few popular databases and tools for single-cell RNA-Seq analysis.

of scRNA-Seq and single-cell gene regulatory networks (scGRN) frameworks in studying complex agronomic traits and resistance to various stresses in crops have been proposed [187]. Gene expression profiles among subcellular populations of the skeletal muscle and its development in chicken have been determined using scRNA-Seq, which are important in producing quantity and quality meat in poultry [188]. In sea urchins, using scRNA-Seq, different cell types commonly seen during the embryo development have been identified by the selective inhibition of Delta/Notch and Wnt responsive pathways [189]. Studying the infant and adult cattle mammary glands (MG) with scRNA-Seq, dairy scientists developed a MG-specific single-cell atlas, determined the cell-type heterogeneity, and identified a novel myofibroblast that can differentiate into luminal epithelial cells, and has potential role in lactation and immunity [190].

2.9 Metatranscriptomics

Metatranscriptome refers to the total RNA sequences (protein-coding and non-coding) collected from a location or source or body, which corresponds to the expression profiles of prokaryotic and eukaryotic species found in natural environments such as soil, sea, space, gut, airways, feces, and skin [191]. Metagenomics focuses on the overall genetic composition of the microbial community, while metatranscriptomics provides more profound insights about the genes expressed, their abundance, diversity, differential expression, and aims to address the functional, metabolic, and pathway diversity present in a microbial community [192]. Metatranscriptome is a dynamic entity that can detect gene expression variability with time and environmental changes [193]. Metatranscriptomics is a culture-free profiling method that helps understand the structure (i.e., microbial communities and taxonomic analysis), function (DEGs, enrichment, and annotation), and mechanisms (adaptability, selection, and domestication) of complex microbial communities [194]. It also helps in understanding RNA-mediated regulation and in deriving biological signatures associated with microbial communities.

The experimental methods for analyzing RNA, such as northern blot, qRT-PCR, microarrays, cDNA clone-based Sanger sequencing, and RNA-Seq, are also used for studying and analyzing metatranscriptomes. The main challenges in molecular metatranscriptome methods include low total RNA yield commonly found in environmental samples, high rRNA content in total RNA and its removal, and the fidelity of microbial mRNA isolated. Metatranscriptome analysis using RNA-Seq can distinguish and handle metadata [195], whereas the previous transcript analysis approaches failed to: categorize or catalog metadata, understand community-wide gene expression, and determine functional diversity. Most of the metatranscriptome tools utilize one or more steps from the following: 1) preprocessing (QC, trimming, and filtering), 2) Binning, 3) Mapping or *de novo* assembly, 4) taxonomic units, 5) species profiling, 6) DEGs, 7) annotation and function assignment, and 8) pathway or network analysis [193]. The key challenges in metatranscriptome analysis are: the lack of comprehensive datasets from diverse groups of samples and their associated metadata; the scarcity of metagenomic reference data; the small overlap between metagenome and metatranscriptome datasets; rRNA filtering; and the enrichment of low-abundance mRNAs. Some databases and tools routinely used to access or analyze metatranscriptomes are presented here (**Table 9**).

Though several applications have been documented in the recent past, only selected studies from agriculture and food disciplines are presented here. In agriculture, metatranscriptome analysis can help us find beneficial and harmful rhizosphere-associated microbes specific to plant and soil types. Thus, it allows us to enrich associated rhizosphere microbes that promote crop health and yield.

Databases			Tools		
S.No	Database	Citation	S.No	Tool	Citation
1	SILVA	[196]	1	QIIME 2	[197]
2	Greengenes	[198]	2	SAMSA2	[199]
3	eggNOG database	[200]	3	ASaiM	[201]
4	NCBI RefSeq	[202]	4	MG-RAST	[203]
5	SEED Subsystems	[204]	5	MetaTrans	[205]

Table 9.
Some widely used databases and tools in metatranscriptomics analysis.

Metatranscriptomics has been used in deciphering multifunctional genes and enzymes linked with the degradation of contaminants in the crop rhizosphere [206]. Metatranscriptomic profiling helped to determine the variation in the rumen's microbial composition based on the host feed efficiency in beef cattle [207]. In the food industry, metatranscriptomics can be applied to detect food contamination, toxins, and metabolic activities of food-associated microbes and enhance food safety, quality, and function. Metatranscriptomics has been used in finding insights into the core functional microbiota of soy sauce aroma type liquor production in the fermentation process under varied environmental conditions [208]. Metatranscriptome analysis has been used to study the community dynamics of bacteria in fermented foods [209]. Using metatranscriptome sequencing followed by 16S and 18S rRNA analysis, temperature-induced changes in the structural landscape and functional diversity of the mesophilic and thermophilic food web communities respond to two contrasting temperatures in the rice fields have been observed [210].

2.10 Systems biology/biological network analysis

The ultimate goal of RNA-Seq analysis is to understand the underlying biological processes and mechanisms linked with gene expression and regulation. From molecule to biospheres, biological systems can be represented as networks of pairwise relationships between biological entities throughout various levels of organization. The interactions between biomolecules can be: direct, via physical contact, or indirect, via causal chains or mere correlations. Interactomes that are commonly studied include networks between: DNA–RNA; DNA–Protein; RNA–RNA; RNA–Protein; and Protein–Protein. Theoretically, any network of words can be merged with these interactions, as some elements are shared by both, like common gene, transcript, or protein identifiers. The systems biology approach examines the overall structure and function of a cell or an organism, rather than looking at its components as isolated events [211]. The systems biology approach considers gene expression of an organism or an interaction as a sum of individual genes, sets of genes, and other compounding factors [212]. Gene regulatory networks (GRNs) and co-expression analyses are common elements while studying a biological problem as a system rather than as an individual problem [213].

Given the growing avalanche of RNA-Seq data along with the wealth of network analysis (NA) programs, there are tremendous opportunities to find networks within and between their available datasets, guiding them toward valuable insights, future validation experiments, and a more holistic understanding of their research. NA of RNA-Seq data can illuminate the interrelationships and functional associations [214] between several elements: regulators/co-regulators,

Databases			Tools		
S.No	Database	Citation	S.No	Tool	Citation
1	DualSeqDB	[216]	1	pARACNE	[217]
2	KBase	[218]	2	SCENIC	[219]
3	MODOMICS	[220]	3	SERGIO	[221]
4	EcoCyc	[222]	4	GRNBoost2	[223]
5	doRiNA	[224]	5	dynGENIE3	[225]

Table 10.
 A few databases and tools for systems biology analysis using RNA-Seq.

upstream/downstream sequences, and genic features; differentially expressed subnetworks; global connectivity among genes and gene networks. Often combined with the aforementioned biomolecular interactions, a more abstracted view of biological systems can be provided by semantic networks, which involve the relationships between categories of biological meaning, commonly ontological, that have been assigned to the biomolecules. Traditional systems biology relied on mathematical and statistical models. In contrast, modern systems biology depends on computer models that simulate an organism's entire biological systems by considering all components [215]. So, these approaches depend on the constant selection of predictors, building models, and testing. Thus, it allows us to move from descriptive science to data science in providing a holistic answer to the biological question under investigation. Thankfully, the inherent complexity of systems biology is ameliorated by the availability of many open-source tools to reconstruct and visualize networks (a few tools and databases are presented in **Table 10**).

RNA-Seq data from a plant (maize) and a pathogen (*Aspergillus flavus*) interaction has been studied as a system to determine GRNs and co-regulated expression patterns in early processes of infection in imparting resistance to *A. flavus* in maize [226]. Systems biology approach has been utilized in unraveling the complex interactions among transcriptomic, metabolomic, and organoleptic components in tomatoes using MetGenMAP, MapMan, and Cytoscape tools [227]. Also, the role of systems biology in building genome-scale metabolic models (GEMs) for characterizing plant-pathogen (*Phytophthora infestans*) interaction, and disease prevention using cellular localization and network reconstruction tools such as KEGG, LocTree 3, and RAVEN [228]. In the food industry, a systems biology framework, Allergen Peptide Browser that stores and catalogs mass spectrometry data has been used in detecting food allergens such as egg, casein, nuts, gluten, wheat, soy, and fish in food products by employing selected and multiple reaction monitoring approach [229]. Systems biology's role in deciphering underlying common molecular pathways that regulate adipose tissue growth and development in chicken has been determined by examining gene modules, functional enrichment, and network analysis (KEGG, Cytoscape, and WGCNA package) [230].

3. Conclusions

In conclusion, a combination of multi-omic approaches and bioinformatics tools developed to date has unquestionably expanded the scope of RNA-Seq applications and improved our understanding of gene expression data. In addition to the applications discussed in this chapter, fusion gene analysis, RNA editing,

RNA interference, and Epitranscriptomics can also be used to understand novel functions of the gene, complex interactions, and the interplay between coding and non-coding regions during gene regulation. In the near future, we will be able to: sequence transcriptomes from complex environments, study more comprehensive RNA datasets using data science tools, functionally validate predicted genes using gene-editing technologies, which will positively impact the food and agriculture sectors.

Acknowledgements

The authors acknowledge Ms. Shalini P. Etukuri and Dr. Govind Sharma at Alabama A&M University for reviewing this book chapter. Also, authors would like to thank anonymous reviewers and editor for their efforts in improving this book chapter. The authors acknowledge the funding support by the Capacity Building grant #2020-38821-31103 from the USDA National Institute of Food and Agriculture.

Conflict of interest


The authors declare no conflict of interest.

Author details

Venkateswara R. Sripathi*, Varsha C. Anche, Zachary B. Gossett
and Lloyd T. Walker
Center for Molecular Biology, Alabama A&M University, Normal, AL, USA

*Address all correspondence to: v.sripathi@aamu.edu

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Chen H, Shan G. The physiological function of long-noncoding RNAs. *Non-coding RNA research*. 2020 Sep 17. DOI: 10.1016/j.ncrna.2020.09.003.
- [2] Fernandes JC, Acuña SM, Aoki JI, Floeter-Winter LM, Muxel SM. Long non-coding RNAs in the regulation of gene expression: physiology and disease. *Non-coding RNA*. 2019 Mar;5(1):17. DOI: 10.3390/ncrna5010017
- [3] Li J, Liu C. Coding or noncoding, the converging concepts of RNAs. *Frontiers in genetics*. 2019 May 22;10:496. DOI: 10.3389/fgene.2019.00496
- [4] Hubé F, Francastel C. Coding and non-coding RNAs, the frontier has never been so blurred. *Frontiers in genetics*. 2018 Apr 18;9:140. DOI:10.3389/fgene.2018.00140
- [5] Han F, Lillard SJ. In-situ sampling and separation of RNA from individual mammalian cells. *Analytical chemistry*. 2000 Sep 1;72(17):4073-4079. DOI: 10.1021/ac000428g
- [6] Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. *Molecular cell*. 2011 Sep 16;43(6):853- DOI: 10.1016/j.molcel.2011.08.017
- [7] Gutell RR, Lee JC, Cannone JJ. The accuracy of ribosomal RNA comparative structure models. *Current opinion in structural biology*. 2002 Jun 1;12(3):301- DOI: 10.1016/S0959-440X(02)00339-1
- [8] Peano C, Pietrelli A, Consolandi C, Rossi E, Petiti L, Tagliabue L, De Bellis G, Landini P. An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microbial informatics and experimentation*. 2013 Dec;3(1):1-1. DOI: 10.1186/2042-5783-3-1
- [9] Ha M, Kim VN. Regulation of microRNA biogenesis. *Nature reviews Molecular cell biology*. 2014 Aug;15(8):509-524. DOI: 10.1038/nrm3838
- [10] Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM, Zhang GZ. Biological functions of microRNAs: a review. *Journal of physiology and biochemistry*. 2011 Mar 1;67(1):129-139. DOI: 10.1007/s13105-010-0050-6
- [11] Zhao S, Zhang Y, Gamini R, Zhang B, von Schack D. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Scientific reports*. 2018 Mar 19;8(1):1-2. DOI: 10.1038/s41598-018-23226-4
- [12] Chen L, Heikkinen L, Wang C, Yang Y, Sun H, Wong G. Trends in the development of miRNA bioinformatics tools. *Briefings in bioinformatics*. 2019 Sep;20(5):1836-1852. DOI: 10.1093/bib/bby054
- [13] Hertel J, Stadler PF. Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data. *Bioinformatics*. 2006 Jul 15;22(14):e197-e202. DOI: 10.1093/bioinformatics/btl257
- [14] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*. 2009 Jan;10(1):57-63. DOI: 10.1038/nrg2484
- [15] Tan G, Opitz L, Schlapbach R, Rehrauer H. Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific reports*. 2019 Feb 27;9(1):1-7. DOI: 10.1038/s41598-019-39076-7
- [16] Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitch S, Lehrach H, Soldatov A. Transcriptome analysis by

strand-specific sequencing of complementary DNA. *Nucleic acids research*. 2009 Oct 1;37(18):e123-. DOI: 10.1093/nar/gkp596

[17] Oszolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics*. 2011 Feb;12(2):87-98. DOI: 10.1038/nrg2934

[18] Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS computational biology*. 2017 May 18;13(5):e1005457. DOI: 10.1371/journal.pcbi.1005457

[19] Liu Y, Zhou J, White KP. RNA-seq differential expression studies: more sequence or more replication?. *Bioinformatics*. 2014 Feb 1;30(3):301-304. DOI: 10.1093/bioinformatics/btt688

[20] Baccarella A, Williams CR, Parrish JZ, Kim CC. Empirical assessment of the impact of sample number and read depth on RNA-Seq analysis workflow performance. *BMC bioinformatics*. 2018 Dec;19(1):1-2. DOI: 10.1186/s12859-018-2445-2

[21] Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. *PloS one*. 2017 Dec 21;12(12):e0190152. DOI: 10.1371/journal.pone.0190152

[22] de Jong TV, Moshkin YM, Guryev V. Gene expression variability: the other dimension in transcriptome analysis. *Physiological genomics*. 2019 May 1;51(5):145-158. DOI: 10.1152/physiolgenomics.00128.2018

[23] Williams AG, Thomas S, Wyman SK, Holloway AK. RNA-seq data: challenges in and recommendations for experimental design and analysis. *Current protocols in human genetics*. 2014 Oct;83(1):11- DOI: 10.1002/0471142905.hg1113s83

[24] Adriaens ME, Bezzina CR. Genomic approaches for the elucidation of genes and gene networks underlying cardiovascular traits. *Biophysical reviews*. 2018 Aug;10(4):1053-1060. DOI: 10.1007/s12551-018-0435-2

[25] Merino GA, Conesa A, Fernández EA. A benchmarking of workflows for detecting differential splicing and differential expression at isoform level in human RNA-seq studies. *Briefings in bioinformatics*. 2019 Mar;20(2):471-481. DOI: 10.1093/bib/bbx122

[26] Rydenfelt M, Klinger B, Klünemann M, Blüthgen N. SPEED2: inferring upstream pathway activity from differential gene expression. *Nucleic acids research*. 2020 Jul 2;48(W1):W307-W312. DOI: 10.1093/nar/gkaa236

[27] Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature biotechnology*. 2015 Mar;33(3):243-246. DOI: 10.1038/nbt.3172

[28] Toro-Domínguez D, Martorell-Marugán J, López-Domínguez R, García-Moreno A, González-Rumayor V, Alarcón-Riquelme ME, Carmona-Sáez P. ImaGEO: integrative gene expression meta-analysis from GEO database. *Bioinformatics*. 2019 Mar 1;35(5):880-882. DOI: 10.1093/bioinformatics/bty721

[29] Ritchie ME, Phipson B, Wu DI, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015 Apr 20;43(7):e47-. DOI: 10.1093/nar/gkv007

[30] Smith CM, Hayamizu TF, Finger JH, Bello SM, McCright IJ, Xu J,

- Baldarelli RM, Beal JS, Campbell J, Corbani LE, Frost PJ. The mouse gene expression database (GXD): 2019 update. *Nucleic acids research*. 2019 Jan 8;47(D1):D774-D779. DOI: 10.1093/nar/gky922
- [31] Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, Conesa A. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic acids research*. 2015 Dec 2;43(21):e140-. DOI: 10.1093/nar/gkv711
- [32] Xia L, Zou D, Sang J, Xu X, Yin H, Li M, Wu S, Hu S, Hao L, Zhang Z. Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *Journal of Genetics and Genomics*. 2017 May 20;44(5):235-241. DOI: 10.1016/j.jgg.2017.05.003
- [33] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014 Dec;15(12):1-21. DOI: 10.1186/s13059-014-0550-8
- [34] Clough E, Barrett T. The gene expression omnibus database. In *Statistical genomics 2016* (pp. 93-110). Humana Press, New York, NY. DOI: DOI: 10.1007/978-1-4939-3578-9_5.
- [35] Chen Y, Lun AT, Smyth GK. Differential expression analysis of complex RNA-seq experiments using edgeR. *Statistical analysis of next generation sequencing data*. 2014:51-74. DOI: 10.1007/978-3-319-07212-8_3
- [36] Khan S, Wu SB, Roberts J. RNA-sequencing analysis of shell gland shows differences in gene expression profile at two time-points of eggshell formation in laying chickens. *BMC genomics*. 2019 Dec;20(1):1-20. DOI: 10.1186/s12864-019-5460-4
- [37] Yang J, Jiang J, Liu X, Wang H, Guo G, Zhang Q, Jiang L. Differential expression of genes in milk of dairy cattle during lactation. *Animal genetics*. 2016 Apr;47(2):174-180. DOI: 10.1111/age.12394
- [38] Devonshire AL, Gursel DB, Fan H, Erickson KA, Pongracic JA, Singh AM, Kumar R. Differential Gene Expression Among Infants at High-Risk for Peanut Allergy. *Journal of Allergy and Clinical Immunology*. 2019 Feb 1;143(2):AB82. DOI: 10.1016/j.jaci.2018.12.255
- [39] Yuan S, Li R, Chen S, Chen H, Zhang C, Chen L, Hao Q, Shan Z, Yang Z, Qiu D, Zhang X. RNA-Seq analysis of differential gene expression responding to different rhizobium strains in soybean (*Glycine max*) roots. *Frontiers in plant science*. 2016 May 30;7:721. DOI: 10.3389/fpls.2016.00721
- [40] Zhang Y, Jiang L, Li Y, Chen Q, Ye Y, Zhang Y, Luo Y, Sun B, Wang X, Tang H. Effect of red and blue light on anthocyanin accumulation and differential gene expression in strawberry (*Fragaria × ananassa*). *Molecules*. 2018 Apr;23(4):820. DOI: 10.3390/molecules23040820
- [41] Lin R, Du X, Peng S, Yang L, Ma Y, Gong Y, Li S. Discovering all transcriptome single-nucleotide polymorphisms and scanning for selection signatures in ducks (*Anas platyrhynchos*). *Evolutionary Bioinformatics*. 2015 Jan;11:EBO-S21545. DOI: 10.4137/EBO.S21545
- [42] Maughan PJ, Yourstone SM, Byers RL, Smith SM, Udall JA. Single-Nucleotide Polymorphism Genotyping in Mapping Populations via Genomic Reduction and Next-Generation Sequencing: Proof of Concept. *The Plant Genome*. 2010 Nov;3(3). DOI: 10.3835/plantgenome2010.07.0016
- [43] Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C,

- Thorne N, Redon R, Bird CP, De Grassi A, Lee C, Tyler-Smith C. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007 Feb 9;315(5813):848-853. DOI: 10.1126/science.1136678
- [44] Quaglieri A, Flensburg C, Speed TP, Majewski IJ. Finding a suitable library size to call variants in RNA-seq. *BMC bioinformatics*. 2020 Dec;21(1):1-9. DOI: 10.1186/s12859-020-03860-4
- [45] Yang Y, Peng X, Ying P, Tian J, Li J, Ke J, Zhu Y, Gong Y, Zou D, Yang N, Wang X. AWESOME: a database of SNPs that affect protein post-translational modifications. *Nucleic acids research*. 2019 Jan 8;47(D1):D874-D880. DOI: 10.1093/nar/gky821
- [46] Zmienko A, Marszalek-Zenczak M, Wojciechowski P, Samelak-Czajka A, Luczak M, Kozlowski P, Karlowski WM, Figlerowicz M. AthCNV: A map of DNA copy number variations in the Arabidopsis genome. *The Plant Cell*. 2020 Jun 1;32(6):1797-1819. DOI: 10.1105/tpc.19.00640
- [47] Kim J, Weber JA, Jho S, Jang J, Jun J, Cho YS, Kim HM, Kim H, Kim Y, Chung O, Kim CG. KoVariome: Korean National Standard Reference Variome database of whole genomes with comprehensive SNV, indel, CNV, and SV analyses. *Scientific reports*. 2018 Apr 4;8(1):1-4. DOI: 10.1038/s41598-018-23837-x
- [48] Brouard JS, Schenkel F, Marete A, Bissonnette N. The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *Journal of animal science and biotechnology*. 2019 Dec;10(1):1-6. DOI: 10.1186/s40104019-0359-0
- [49] Miao YR, Liu W, Zhang Q, Guo AY. lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucleic acids research*. 2018 Jan 4;46(D1):D276-D280. DOI: 10.1093/nar/gkx1004
- [50] Ma C, Shao M, Kingsford C. SQUID: transcriptomic structural variation detection from RNA-seq. *Genome biology*. 2018 Dec 1;19(1):52. DOI: 10.1186/s13059-018-1421-5
- [51] Kumar S, Ambrosini G, Bucher P. SNP2TFBS—a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic acids research*. 2017 Jan 4;45(D1):D139-D144. DOI: 10.1093/nar/gkw1064
- [52] Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT, Gross SS. A universal SNP and small-indel variant caller using deep neural networks. *Nature biotechnology*. 2018 Nov;36(10):983-987. DOI: 10.1038/nbt.4235
- [53] Guo L, Du Y, Chang S, Zhang K, Wang J. rSNPBase: a database for curated regulatory SNPs. *Nucleic Acids Research*. 2014 Jan 1;42(D1):D1033-D1039. DOI: 10.1093/nar/gkt1167
- [54] Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, Johnson J, Dougherty B, Barrett JC, Dry JR. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic acids research*. 2016 Jun 20;44(11):e108-. DOI: 10.1093/nar/gkw227
- [55] Morgil H, Gercek YC, Tulum I. Single nucleotide polymorphisms (SNPs) in plant genetics and breeding. *In The Recent Topics in Genetic Polymorphisms 2020* Mar 28. IntechOpen. DOI: 10.5772/intechopen.91886
- [56] Fang L, Sahana G, Su G, Yu Y, Zhang S, Lund MS, Sørensen P.

Integrating sequence-based GWAS and RNA-Seq provides novel insights into the genetic basis of mastitis and milk production in dairy cattle. *Scientific reports*. 2017 Mar 30;7(1):1-6. DOI: 10.1038/srep45560

[57] Tanaka T, Ishikawa G, Ogiso-Tanaka E, Yanagisawa T, Sato K. Development of genome-wide SNP markers for barley via reference-based RNA-Seq analysis. *Frontiers in plant science*. 2019 May 10;10:577. DOI: 10.3389/fpls.2019.00577

[58] Nishijima R, Yoshida K, Motoi Y, Sato K, Takumi S. Genome-wide identification of novel genetic markers from RNA sequencing assembly of diverse *Aegilops tauschii* accessions. *Molecular Genetics and Genomics*. 2016 Aug;291(4):1681-1694. DOI: 10.1007/s00438-016-1211-2

[59] Bakhtiarizadeh MR, Alamouti AA. RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. *Scientific Reports*. 2020 Aug 11;10(1):1-3. DOI: 10.1038/s41598-020-70527-8

[60] Lam S, Zeidan J, Miglior F, Suárez-Vega A, Gómez-Redondo I, Fonseca PA, Guan LL, Waters S, Cánovas A. Development and comparison of RNA-sequencing pipelines for more accurate SNP identification: practical example of functional SNP detection associated with feed efficiency in Nellore beef cattle. *BMC genomics*. 2020 Dec;21(1):1-7. DOI: 10.1186/s12864-020-07107-7

[61] Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nature Reviews Genetics*. 2010 Aug;11(8):533-538. DOI: 10.1038/nrg2815

[62] Kukurba KR, Zhang R, Li X, Smith KS, Knowles DA, Tan MH,

Piskol R, Lek M, Snyder M, MacArthur DG, Li JB. Allelic expression of deleterious protein-coding variants across human tissues. *PLoS Genet*. 2014 May 1;10(5):e1004304. DOI: 10.1371/journal.pgen.1004304

[63] Li G, Bahn JH, Lee JH, Peng G, Chen Z, Nelson SF, Xiao X. Identification of allele-specific alternative mRNA processing via transcriptome sequencing. *Nucleic acids research*. 2012 Jul 1;40(13):e104-. DOI: 10.1093/nar/gks280

[64] Reddy TE, Gertz J, Pauli F, Kucera KS, Varley KE, Newberry KM, Marinov GK, Mortazavi A, Williams BA, Song L, Crawford GE. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome research*. 2012 May 1;22(5):860-869. DOI: 10.1101/gr.131201.111

[65] Berger E, Yorukoglu D, Zhang L, Nyquist SK, Shalek AK, Kellis M, Numanagić I, Berger B. Improved haplotype inference by exploiting long-range linking and allelic imbalance in RNA-seq datasets. *Nature communications*. 2020 Sep 16;11(1):1-9. DOI: 10.1038/s41467-020-18320-z

[66] Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, Lee M, Popova N, Sharopova N, Kimura M, Feolo M. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic acids research*. 2014 Jan 1;42(D1):D975-D979. DOI: 10.1093/nar/gkt1211

[67] Raghupathy N, Choi K, Vincent MJ, Beane GL, Sheppard KS, Munger SC, Korstanje R, Pardo-Manual de Villena F, Churchill GA. Hierarchical analysis of RNA-seq reads improves the accuracy of allele-specific expression. *Bioinformatics*. 2018 Jul 1;34(13):2177-2184. DOI: 10.1093/bioinformatics/bty078

- [68] Stanfill AG, Cao X. Enhancing Research Through the Use of the Genotype-Tissue Expression (GTEx) Database. *Biological Research For Nursing*. 2021 Feb 18;1099800421994186. DOI: 10.1177/1099800421994186
- [69] Deonovic B, Wang Y, Weirather J, Wang XJ, Au KF. IDP-ASE: haplotyping and quantifying allele-specific expression at the gene and gene isoform level by hybrid sequencing. *Nucleic acids research*. 2017 Mar 17;45(5):e32-. DOI: 10.1093/nar/gkw1076
- [70] Abramov S, Baulin E, Makeev VJ, Boytsov A, Yevshin I, Kulakovskiy IV, Bykova D, Kolpakov F. AD ASTRA: the database of Allelic Dosage-corrected Allele-Specific TRAnscription factor binding suggests causal regulatory sequence variants of pathologies. *In Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2020) 2020* (pp. 14-14). DOI: 10.18699/BGRS/SB-2020-001
- [71] Harvey CT, Moyerbrailean GA, Davis GO, Wen X, Luca F, Pique-Regi R. QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics*. 2015 Apr 15;31(8):1235-1242. DOI: 10.1093/bioinformatics/btu802
- [72] Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3. 0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Human mutation*. 2016 Mar;37(3):235-241. DOI: 10.1002/humu.22932
- [73] Romanel A, Lago S, Prandi D, Sboner A, Demichelis F. ASEQ: fast allele-specific studies from next-generation sequencing data. *BMC medical genomics*. 2015 Dec;8(1):1-2. DOI: 10.1186/s12920-015-0084-2
- [74] Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*. 2010 Oct 1;26(19):2474-2476. DOI: 10.1093/bioinformatics/btq452
- [75] Mayba O, Gilbert HN, Liu J, Haverty PM, Jhunjhunwala S, Jiang Z, Watanabe C, Zhang Z. MBASED: allele-specific expression detection in cancer tissues and cell lines. *Genome biology*. 2014 Aug;15(8):1-21. DOI: 10.1186/s13059-014-0405-3
- [76] Shao L, Xing F, Xu C, Zhang Q, Che J, Wang X, Song J, Li X, Xiao J, Chen LL, Ouyang Y. Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proceedings of the National Academy of Sciences*. 2019 Mar 19;116(12):5653-5658. DOI: 10.1073/pnas.1820513116
- [77] Cai M, Lin J, Li Z, Lin Z, Ma Y, Wang Y, Ming R. Allele specific expression of Dof genes responding to hormones and abiotic stresses in sugarcane. *PloS one*. 2020 Jan 16;15(1):e0227716. DOI: 10.1371/journal.pone.0227716
- [78] Liu Y, Liu X, Zheng Z, Ma T, Liu Y, Long H, Cheng H, Fang M, Gong J, Li X, Zhao S. Genome-wide analysis of expression QTL (eQTL) and allele-specific expression (ASE) in pig muscle identifies candidate genes for meat quality traits. *Genetics Selection Evolution*. 2020 Dec;52(1):1-1. DOI: 10.1186/s12711-020-00579-x
- [79] de Souza MM, Zerlotini A, Rocha MI, Bruscardin JJ, da Silva Diniz WJ, Cardoso TF, Cesar AS, Afonso J, Andrade BG, de Alvarenga Mudadu M, Mokry FB. Allele-specific expression is widespread in *Bos indicus* muscle and affects meat quality candidate genes. *Scientific Reports*. 2020 Jun 23;10(1):1-1. DOI: 10.1038/s41598-020-67089-0

- [80] Tomlinson MJ, Polson SW, Qiu J, Lake JA, Lee W, Abasht B. Investigation of allele specific expression in various tissues of broiler chickens using the detection tool VADT. *Scientific reports*. 2021 Feb 17;11(1):1-3. DOI: 10.1038/s41598-021-83459-8
- [81] Reynolds DJ, Hertel KJ. Ultra-deep sequencing reveals pre-mRNA splicing as a sequence driven high-fidelity process. *PloS one*. 2019 Oct 3;14(10):e0223132 DOI: 10.1371/journal.pone.0223132.
- [82] Will CL, Lührmann R. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology*. 2011 Jul 1;3(7):a003707. DOI:10.1101/cshperspect.a003707
- [83] Hu H, Yang W, Zheng Z, Niu Z, Yang Y, Wan D, Liu J, Ma T. Analysis of alternative splicing and alternative polyadenylation in *Populus alba var. pyramidalis* by single-molecular long-read sequencing. *Frontiers in genetics*. 2020 Feb 7;11:48. DOI: 10.3389/fgene.2020.00048
- [84] Karousis ED, Nasif S, Mühlemann O. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdisciplinary Reviews: RNA*. 2016 Sep;7(5):661-682. DOI: 10.1002/wrna.1357
- [85] Rossell D, Attolini CS, Kroiss M, Stöcker A. Quantifying alternative splicing from paired-end RNA-sequencing data. *The annals of applied statistics*. 2014 Mar;8(1):309. DOI: 10.1214/13-aos687
- [86] Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome biology*. 2020 Dec;21(1):1-6. DOI: 10.1186/s13059-020-1935-5
- [87] Louadi Z, Yuan K, Gress A, Tsoy O, Kalinina OV, Baumbach J, Kacprowski T, List M. DIGGER: exploring the functional role of alternative splicing in protein interactions. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D309-D318. DOI: 10.1093/nar/gkaa768
- [88] Szikora P, Pór T, Sebestyén E. SplicingFactory-Splicing diversity analysis for transcriptome data. *bioRxiv*. 2021 Jan 1. DOI: 10.1101/2021.02.03.429568
- [89] Li Z, Zhang Y, Bush SJ, Tang C, Chen L, Zhang D, Urrutia AO, Lin JW, Chen L. MeDAS: a Metazoan Developmental Alternative Splicing database. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D144-D150. DOI: 10.1093/nar/gkaa886
- [90] Estefania M, Andres R, Javier I, Marcelo Y, Ariel C. ASpli: Integrative analysis of splicing landscapes through RNA-Seq assays. *Bioinformatics*. 2021 Mar 2. DOI: 10.1093/bioinformatics/btab141
- [91] Liu J, Tan S, Huang S, Huang W. ASlive: a database for alternative splicing atlas in livestock animals. *BMC genomics*. 2020 Dec;21(1):1-7. DOI: 10.1186/s12864-020-6472-9
- [92] Denti L, Rizzi R, Beretta S, Della Vedova G, Previtali M, Bonizzoni P. ASGAL: aligning RNA-Seq data to a splicing graph to detect novel alternative splicing events. *BMC bioinformatics*. 2018 Dec;19(1):1-21. DOI: 10.1186/s12859-018-2436-3
- [93] Sun Y, Zhang Q, Liu B, Lin K, Zhang Z, Pang E. CuAS: a database of annotated transcripts generated by alternative splicing in cucumbers. *BMC plant biology*. 2020 Dec;20(1):1-7. DOI: 10.1186/s12870-020-2312-y
- [94] Vaquero-Garcia J, Barrera A, Gazzara MR, Gonzalez-Vallinas J, Lahens NF, Hogenesch JB, Lynch KW, Barash Y. A new view of transcriptome complexity and regulation through the

lens of local splicing variations. *elife*. 2016 Feb 1;5:e11752. DOI: 10.7554/eLife.11752

[95] Wang J, Zhang J, Li K, Zhao W, Cui Q. SpliceDisease database: linking RNA splicing and disease. *Nucleic acids research*. 2012 Jan 1;40(D1):D1055-D1059. DOI: 10.1093/nar/gkr1171

[96] Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences*. 2014 Dec 23;111(51):E5593-E5601. DOI: 10.1073/pnas.1419161111

[97] Wang Y, Liu J, Huang BO, Xu YM, Li J, Huang LF, Lin J, Zhang J, Min QH, Yang WM, Wang XZ. Mechanism of alternative splicing and its regulation. *Biomedical reports*. 2015 Mar 1;3(2):152-158. DOI: 10.3892/br.2014.407

[98] Ding Y, Wang Y, Qiu C, Qian W, Xie H, Ding Z. Alternative splicing in tea plants was extensively triggered by drought, heat and their combined stresses. *PeerJ*. 2020 Jan 29;8:e8258. DOI: 10.7717/peerj.8258

[99] Yu H, Du Q, Campbell M, Yu B, Walia H, Zhang C. Genome-wide discovery of natural variation in pre-mRNA splicing and prioritising causal alternative splicing to salt stress response in rice. *New Phytologist*. 2021 Jan 16. DOI: 10.1111/nph.17189

[100] Sun Y, Xiao H. Identification of alternative splicing events by RNA sequencing in early growth tomato fruits. *BMC genomics*. 2015 Dec;16(1):1-3. DOI: 10.1186/s12864-015-2128-6

[101] Chen W, Jia Q, Song Y, Fu H, Wei G, Ni T. Alternative polyadenylation: methods, findings, and

impacts. *Genomics, proteomics & bioinformatics*. 2017 Oct 1;15(5):287-300. DOI: 10.1016/j.gpb.2017.06.001

[102] Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nature reviews Molecular cell biology*. 2017 Jan;18(1):18-30. DOI: 10.1038/nrm.2016.116

[103] Mayr C. Evolution and biological roles of alternative 3' UTRs. *Trends in cell biology*. 2016 Mar 1;26(3):227-237. DOI: 10.1016/j.tcb.2015.10.012

[104] Zhang Y, Liu L, Qiu Q, Zhou Q, Ding J, Lu Y, Liu P. Alternative polyadenylation: methods, mechanism, function, and role in cancer. *Journal of Experimental & Clinical Cancer Research*. 2021 Dec;40(1):1-9. DOI: 10.1186/s13046-021-01852-7

[105] Li Y, Schaefer B, Zou X, Zhang M, Heyd F, Sun W, Zhang B, Li G, Liang W, He Y, Zhou J. Pan-tissue analysis of allelic alternative polyadenylation suggests widespread functional regulation. *Molecular systems biology*. 2020 Apr;16(4):e9367. DOI: 10.15252/msb.20199367

[106] Marini F, Scherzinger D, Danckwardt S. TREND-DB—a transcriptome-wide atlas of the dynamic landscape of alternative polyadenylation. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D243-D253. DOI: 10.1093/nar/gkaa722

[107] Li Z, Li Y, Zhang B, Li Y, Long Y, Zhou J, Zou X, Zhang M, Hu Y, Chen W, Gao X. Deereact-apa: Prediction of alternative polyadenylation site usage through deep learning. *Genomics, Proteomics & Bioinformatics*. 2021 Mar 2. DOI: 10.1016/j.gpb.2020.05.004

[108] Jin W, Zhu Q, Yang Y, Yang W, Wang D, Yang J, Niu X, Yu D, Gong J. Animal-APAdb: a comprehensive animal alternative polyadenylation database. *Nucleic Acids Research*. 2021 Jan

8;49(D1):D47-D54. DOI: 10.1093/nar/gkaa778

[109] Wang R, Tian B. APAlzyer: a bioinformatics package for analysis of alternative polyadenylation isoforms. *Bioinformatics*. 2020 Jun 1;36(12):3907-3909. DOI: 10.1093/bioinformatics/btaa266

[110] Zhu S, Ye W, Ye L, Fu H, Ye C, Xiao X, Ji Y, Lin W, Ji G, Wu X. PlantAPadb: a comprehensive database for alternative polyadenylation sites in plants. *Plant physiology*. 2020 Jan 1;182(1):228-242. DOI: 10.1104/pp.19.00943

[111] Ye C, Zhou Q, Wu X, Yu C, Ji G, Saban DR, Li QQ. scDAPA: detection and visualization of dynamic alternative polyadenylation from single cell RNA-seq data. *Bioinformatics*. 2020 Feb 15;36(4):1262-1264. DOI: 10.1093/bioinformatics/btz701

[112] Hong W, Ruan H, Zhang Z, Ye Y, Liu Y, Li S, Jing Y, Zhang H, Diao L, Liang H, Han L. APAAtlas: decoding alternative polyadenylation across human tissues. *Nucleic acids research*. 2020 Jan 8;48(D1):D34-D39. DOI: 10.1093/nar/gkz876

[113] Arefeen A, Xiao X, Jiang T. DeepPASTA: deep neural network based polyadenylation site analysis. *Bioinformatics*. 2019 Nov 1;35(22):4577-4585. DOI: 10.1093/bioinformatics/btz283

[114] Müller S, Rycak L, Afonso-Grunz F, Winter P, Zawada AM, Damrath E, Scheider J, Schmah J, Koch I, Kahl G, Rotter B. APADB: a database for alternative polyadenylation and microRNA regulation events. *Database*. 2014 Jan 1;2014. DOI: 10.1093/database/bau076

[115] Arefeen A, Liu J, Xiao X, Jiang T. TAPAS: tool for alternative polyadenylation site analysis.

Bioinformatics. 2018 Aug 1;34(15):2521-2529. DOI: 10.1093/bioinformatics/bty110

[116] Derti A, Garrett-Engle P, MacIsaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. A quantitative atlas of polyadenylation in five mammals. *Genome research*. 2012 Jun 1;22(6):1173-1183. DOI: 10.1101/gr.132563.111

[117] Zhou Q, Fu H, Yang D, Ye C, Zhu S, Lin J, Ye W, Ji G, Ye X, Wu X, Li QQ. Differential alternative polyadenylation contributes to the developmental divergence between two rice subspecies, japonica and indica. *The Plant Journal*. 2019 Apr;98(2):260-276. DOI: 10.1111/tpj.14209

[118] Chakrabarti M, de Lorenzo L, Abdel-Ghany SE, Reddy AS, Hunt AG. Wide-ranging transcriptome remodelling mediated by alternative polyadenylation in response to abiotic stresses in Sorghum. *The Plant Journal*. 2020 Jun;102(5):916-930. DOI: 10.1111/tpj.14671

[119] Wang T, Wang H, Cai D, Gao Y, Zhang H, Wang Y, Lin C, Ma L, Gu L. Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). *The Plant Journal*. 2017 Aug;91(4):684-699. DOI: 10.1111/tpj.13597

[120] Cao J, Ye C, Hao G, Dabney-Smith C, Hunt AG, Li QQ. Root hair single cell type specific profiles of gene expression and alternative polyadenylation under cadmium stress. *Frontiers in plant science*. 2019 May 10;10:589. DOI: 10.3389/fpls.2019.00589

[121] Cai Y, Wan J. Competing endogenous RNA regulations in neurodegenerative disorders: current challenges and emerging insights. *Frontiers in molecular neuroscience*. 2018 Oct 5;11:370. DOI: 10.3389/fnmol.2018.00370

- [122] He X, Guo S, Wang Y, Wang L, Shu S, Sun J. Systematic identification and analysis of heat-stress-responsive lncRNAs, circRNAs and miRNAs with associated co-expression and ceRNA networks in cucumber (*Cucumis sativus* L.). *Physiologia plantarum*. 2020 Mar;168(3):736-754. DOI: 10.1111/ppl.12997
- [123] Sripathi VR, Choi Y, Gossett ZB, Stelly DM, Moss EM, Town CD, Walker LT, Sharma GC, Chan AP. Identification of microRNAs and their targets in four *Gossypium* species using RNA sequencing. *Current Plant Biology*. 2018 Sep 1;14:30-40. DOI: 10.1016/j.cpb.2018.09.008
- [124] O'Brien J, Hayder H, Zayed Y, Peng C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in endocrinology*. 2018 Aug 3;9:402. DOI: 10.3389/fendo.2018.00402
- [125] Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z, Rivas E. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D192-D200. DOI: 10.1093/nar/gkaa1047
- [126] Stocks MB, Mohorianu I, Beckers M, Paicu C, Moxon S, Thody J, Dalmay T, Moulton V. The UEA sRNA Workbench (version 4.4): a comprehensive suite of tools for analyzing miRNAs and sRNAs. *Bioinformatics*. 2018 Oct 1;34(19):3382-3384. DOI: 10.1093/bioinformatics/bty338
- [127] Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PloS one*. 2018 Oct 18;13(10):e0206239. DOI: 10.1371/journal.pone.0206239
- [128] Xie F, Liu S, Wang J, Xuan J, Zhang X, Qu L, Zheng L, Yang J. deepBase v3. 0: expression atlas and interactive analysis of ncRNAs from thousands of deep-sequencing data. *Nucleic acids research*. 2021 Jan 8;49(D1):D877-83. DOI: 10.1093/nar/gkaa1039
- [129] Vitsios DM, Kentepozidou E, Quintais L, Benito-Gutiérrez E, van Dongen S, Davis MP, Enright AJ. Mirnovo: genome-free prediction of microRNAs from small RNA sequencing data and single-cells using decision forests. *Nucleic acids research*. 2017 Dec 1;45(21):e177-. DOI: 10.1093/nar/gkx836
- [130] Tokar T, Pastrello C, Rossos AE, Abovsky M, Hauschild AC, Tsay M, Lu R, Jurisica I. mirDIP 4.1—integrative database of human microRNA target predictions. *Nucleic acids research*. 2018 Jan 4;46(D1):D360-D370. DOI: 10.1093/nar/gkx1144
- [131] Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic acids research*. 2020 Jan 8;48(D1):D127-D131. DOI: 10.1093/nar/gkz757
- [132] Jha A, Shankar R. miReader: Discovering novel miRNAs in species without sequenced genome. *PloS one*. 2013 Jun 21;8(6):e66857. DOI: 10.1371/journal.pone.0066857
- [133] Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic acids research*. 2018 Jul 2;46(W1):W49-W54. DOI: 10.1093/nar/gkr319
- [134] Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic acids research*. 2019 Jan 8;47(D1):D155-D162. DOI: 10.1093/nar/gky1141
- [135] An J, Lai J, Lehman ML, Nelson CC. miRDeep*: an integrated application

- tool for miRNA identification from RNA sequencing data. *Nucleic acids research*. 2013 Jan 1;41(2):727-737. DOI: 10.1093/nar/gks1187
- [136] Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *elife*. 2015 Aug 12;4:e05005. DOI: 10.7554/eLife.05005
- [137] Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, Li Z, Bu D, Sun N, Zhang MQ, Chen R. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic acids research*. 2016 Jan 4;44(D1):D203-D208. DOI: 10.1093/nar/gkv1252
- [138] Ronen R, Gan I, Modai S, Sukachev A, Dror G, Halperin E, Shomron N. miRNAkey: a software for microRNA deep sequencing analysis. *Bioinformatics*. 2010 Oct 15;26(20):2615-2616. DOI: 10.1093/bioinformatics/btq493
- [139] Heikkinen L, Kolehmainen M, Wong G. Prediction of microRNA targets in *Caenorhabditis elegans* using a self-organizing map. *Bioinformatics*. 2011 May 1;27(9):1247-1254. DOI: 10.1093/bioinformatics/btr144
- [140] Langdon WB, Petke J, Lorenz R. Evolving better RNAfold structure prediction. In *European Conference on Genetic Programming 2018* Apr 4 (pp. 220-236). Springer, Cham. DOI: 10.1007/978-3-319-77553-1_14
- [141] Zuker Mfold©: RNA modeling program. *GERF Bulletin of Biosciences*. 2010.
- [142] Zhang Z, Teotia S, Tang J, Tang G. Perspectives on microRNAs and phased small interfering RNAs in maize (*Zea mays* L.): functions and big impact on agronomic traits enhancement. *Plants*. 2019 Jun;8(6):170. DOI: 10.3390/plants8060170
- [143] Pegler JL, Oultram JM, Grof CP, Eamens AL. Profiling the abiotic stress responsive microRNA landscape of *Arabidopsis thaliana*. *Plants*. 2019 Mar;8(3):58. DOI: 10.3390/plants8030058
- [144] Ayubov MS, Mirzakhmedov MH, Sripathi VR, Buriev ZT, Ubaydullaeva KA, Usmonov DE, Norboboyeva RB, Emani C, Kumpatla SP, Abdurakhmonov IY. Role of MicroRNAs and small RNAs in regulation of developmental processes and agronomic traits in *Gossypium* species. *Genomics*. 2019 Sep 1;111(5):1018-1025. DOI: 10.1016/j.ygeno.2018.07.012
- [145] Otsuka K, Yamamoto Y, Matsuoka R, Ochiya T. Maintaining good miRNAs in the body keeps the doctor away?: Perspectives on the relationship between food-derived natural products and microRNAs in relation to exosomes/extracellular vesicles. *Molecular nutrition & food research*. 2018 Jan;62(1):1700080. DOI: 10.1002/mnfr.201700080
- [146] Barrett SP, Salzman J. Circular RNAs: analysis, expression and potential functions. *Development*. 2016 Jun 1;143(11):1838-1847. DOI: 10.1242/dev.128074
- [147] Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. *Rna*. 2013 Feb 1;19(2):141-157. DOI: 10.1261/rna.035667.112
- [148] Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, Wu YM, Dhanasekaran SM, Engelke CG, Cao X, Robinson DR. The landscape of circular RNA in cancer. *Cell*. 2019 Feb 7;176(4):869-881. DOI: 10.1016/j.cell.2018.12.021
- [149] Huang A, Zheng H, Wu Z, Chen M, Huang Y. Circular RNA-protein

interactions: functions, mechanisms, and identification. *Theranostics*. 2020;10(8):3503. DOI: 10.7150/thno.42174

[150] Hansen TB, Venø MT, Damgaard CK, Kjems J. Comparison of circular RNA prediction tools. *Nucleic acids research*. 2016 Apr 7;44(6):e58-. DOI: 10.1093/nar/gkv1458

[151] Liu M, Wang Q, Shen J, Yang BB, Ding X. Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA biology*. 2019 Jul 3;16(7):899-905. DOI: 10.1080/15476286.2019.1600395

[152] Aufiero S, Reckman YJ, Tijssen AJ, Pinto YM, Creemers EE. circRNAprofiler: an R-based computational framework for the downstream analysis of circular RNAs. *BMC bioinformatics*. 2020 Dec;21:1-9. DOI: 10.1186/s12859-020-3500-3

[153] Li S, Li Y, Chen B, Zhao J, Yu S, Tang Y, Zheng Q, Li Y, Wang P, He X, Huang S. exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic acids research*. 2018 Jan 4;46(D1):D106-D112. DOI: 10.1093/nar/gkx891

[154] Zhang P, Liu Y, Chen H, Meng X, Xue J, Chen K, Chen M. CircPlant: An Integrated Tool for circRNA Detection and Functional Prediction in Plants. *Genomics, Proteomics & Bioinformatics*. 2020 Jun 1;18(3):352-358. DOI: 10.1016/j.gpb.2020.10.001

[155] Chu Q, Zhang X, Zhu X, Liu C, Mao L, Ye C, Zhu QH, Fan L. PlantcircBase: a database for plant circular RNAs. *Molecular plant*. 2017 Aug 7;10(8):1126-1128. DOI: 10.1016/j.molp.2017.03.003

[156] Sun P, Li G. CircCode: a powerful tool for identifying circRNA coding ability. *Frontiers in genetics*. 2019 Oct 10;10:981. DOI: 10.3389/fgene.2019.00981

[157] Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. *Scientific reports*. 2016 Oct 11;6(1):1-6. DOI: 10.1038/srep34985

[158] Li L, Bu D, Zhao Y. Circ RNA wrap—a flexible pipeline for circ RNA identification, transcript prediction, and abundance estimation. *FEBS letters*. 2019 Jun;593(11):1179-1189. DOI: 10.1002/1873-3468.13423

[159] Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *Rna*. 2014 Nov 1;20(11):1666-1670. DOI: 10.1261/rna.043687.113

[160] Jakobi T, Uvarovskii A, Dieterich C. circTools—a one-stop software solution for circular RNA research. *Bioinformatics*. 2019 Jul 1;35(13):2326-2328. DOI: 10.1093/bioinformatics/bty948

[161] Lu T, Cui L, Zhou Y, Zhu C, Fan D, Gong H, Zhao Q, Zhou C, Zhao Y, Lu D, Luo J. Transcriptome-wide investigation of circular RNAs in rice. *Rna*. 2015 Dec 1;21(12):2076-2087. DOI: 10.1261/rna.052282.115

[162] Zhao W, Chu S, Jiao Y. Present scenario of circular RNAs (circRNAs) in plants. *Frontiers in plant science*. 2019 Apr 2;10:379. DOI: 10.3389/fpls.2019.00379

[163] Wang Y, Wang Q, Gao L, Zhu B, Luo Y, Deng Z, Zuo J. Integrative analysis of circRNAs acting as ceRNAs involved in ethylene pathway in tomato. *Physiologia plantarum*. 2017 Nov;161(3):311-321. DOI: 10.1111/ppl.12600

[164] Li A, Huang W, Zhang X, Xie L, Miao X. Identification and characterization of CircRNAs of two pig breeds as a new biomarker in metabolism-related diseases. *Cellular*

Physiology and Biochemistry.
2018;47(6):2458-2470. DOI:
10.1159/000491619

[165] Zhang C, Wu H, Wang Y, Zhu S, Liu J, Fang X, Chen H. Circular RNA of cattle casein genes are highly expressed in bovine mammary gland. *Journal of dairy science*. 2016 Jun 1;99(6):4750-4760. DOI: 10.3168/jds.2015-10381

[166] Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods*. 2009 May;6(5):377-382. DOI: 10.1038/nmeth.1315

[167] Saliba AE, Westermann AJ, Gorski SA, Vogel J. Single-cell RNA-seq: advances and future challenges. *Nucleic acids research*. 2014 Aug 18;42(14):8845-8860. DOI: 10.1093/nar/gku555

[168] Shalek AK, Benson M. Single-cell analyses to tailor treatments. *Science translational medicine*. 2017 Sep 20;9(408). DOI: 10.1126/scitranslmed.aan4730

[169] Chen G, Ning B, Shi T. Single-cell RNA-Seq technologies and related computational data analysis. *Frontiers in genetics*. 2019 Apr 5;10:317. DOI: 10.3389/fgene.2019.00317

[170] Baran-Gale J, Chandra T, Kirschner K. Experimental design for single-cell RNA sequencing. *Briefings in functional genomics*. 2018 Jul;17(4):233-239. DOI: 10.1093/bfpgp/elix035

[171] Zhao T, Lyu S, Lu G, Juan L, Zeng X, Wei Z, Hao J, Peng J. SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D1413-D1419. DOI: 10.1093/nar/gkaa838

[172] Sokolowski DJ, Faykoo-Martinez M, Erdman L, Hou H,

Chan C, Zhu H, Holmes MM, Goldenberg A, Wilson MD. Single-cell mapper (scMappR): using scRNA-seq to infer the cell-type specificities of differentially expressed genes. *NAR Genomics and Bioinformatics*. 2021 Mar;3(1):lqab011. DOI: 10.1093/nargab/lqab011

[173] Zhu X, Yunits B, Wolfgruber T, Liu Y, Huang Q, Poirion O, Arisdakessian C, Zhao T, Garmire D, Garmire L. GranatumX: A community engaging and flexible software environment for single-cell analysis. *bioRxiv*. 2019 Jan 1:385591. DOI: 10.1101/385591

[174] Svensson V, da Veiga Beltrame E, Pachter L. A curated database reveals trends in single-cell transcriptomics. *Database*. 2020 Jan 1;2020. DOI: 10.1093/database/baaa073

[175] Bernstein MN, Ni Z, Collins M, Burkard ME, Kendzioriski C, Stewart R. CHARTS: a web application for characterizing and comparing tumor subpopulations in publicly available single-cell RNA-seq data sets. *BMC bioinformatics*. 2021 Dec;22(1):1-9. DOI: 10.1186/s12859-021-04021-x

[176] Li B, Gould J, Yang Y, Sarkizova S, Tabaka M, Ashenberg O, Rosen Y, Slyper M, Kowalczyk MS, Villani AC, Tickle T. Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. *Nature Methods*. 2020 Aug;17(8):793-798. DOI: 10.1038/s41592-020-0905-x

[177] Franzén O, Gan LM, Björkegren JL. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*. 2019 Jan 1;2019. DOI: 10.1093/database/baz046

[178] Franzén O, Björkegren JL. alona: a web server for single-cell RNA-seq analysis. *Bioinformatics*. 2020 Jun 1;36(12):3910. DOI: 10.1093/bioinformatics/btaa269

- [179] Obermayer B, Holtgrewe M, Nieminen M, Messerschmidt C, Beule D. SCellVis: exploratory single cell data analysis on the desktop and in the cloud. *PeerJ*. 2020 Feb 19;8:e8607. DOI: 10.7717/peerj.8607
- [180] Zappia L, Phipson B, Oshlack A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS computational biology*. 2018 Jun 25;14(6):e1006245. DOI: 10.1371/journal.pcbi.1006245
- [181] Tan Y, Cahan P. SingleCellNet: a computational tool to classify single cell RNA-Seq data across platforms and across species. *Cell systems*. 2019 Aug 28;9(2):207-213. DOI: 10.1016/j.cels.2019.06.004
- [182] Ma X, Denyer T, Timmermans MC. PscB: A Browser to Explore Plant Single Cell RNA-Sequencing Data Sets. *Plant physiology*. 2020 Jun 1;183(2):464-467. DOI: 10.1104/pp.20.00250
- [183] Cao Y, Zhu J, Jia P, Zhao Z. scRNASeqDB: a database for RNA-Seq based gene expression profiles in human single cells. *Genes*. 2017 Dec;8(12):368. DOI: 10.3390/genes8120368
- [184] Feng D, Whitehurst CE, Shan D, Hill JD, Yue YG. Single Cell Explorer, collaboration-driven tools to leverage large-scale single cell RNA-seq data. *BMC genomics*. 2019 Dec;20(1):1-8. DOI: 10.1186/s12864-019-6053-y
- [185] Yang A, Troup M, Lin P, Ho JW. Falco: a quick and flexible single-cell RNA-seq processing framework on the cloud. *Bioinformatics*. 2017 Mar 1;33(5):767-769. DOI: 10.1093/bioinformatics/btw732
- [186] Tang W, Tang AY. Biological significance of RNA-seq and single-cell genomic research in woody plants. *Journal of Forestry Research*. 2019 Oct;30(5):1555-1568. DOI: 10.1007/s11676-019-00933-w
- [187] Tripathi RK, Wilkins O. Single cell gene regulatory networks in plants: opportunities for enhancing climate change stress resilience. *Plant, Cell & Environment*. 2021 Feb 1. DOI: 10.1111/pce.14012
- [188] Li J, Xing S, Zhao G, Zheng M, Yang X, Sun J, Wen J, Liu R. Identification of diverse cell populations in skeletal muscles and biomarkers for intramuscular fat of chicken by single-cell RNA sequencing. *BMC genomics*. 2020 Dec;21(1):1-1. DOI: 10.1186/s12864-020-07136-2
- [189] Foster S, Teo YV, Neretti N, Oulhen N, Wessel GM. Single cell RNA-seq in the sea urchin embryo show marked cell-type specificity in the Delta/Notch pathway. *Molecular reproduction and development*. 2019 Aug;86(8):931-934. DOI: 10.1002/mrd.23181
- [190] Gu F, Wu J, Zhu S, Valencak TG, Liu JX, Sun HZ. Single-cell RNA-Sequencing Reveals Novel Myofibroblasts with Epithelial Cell-Like Features in the Mammary Gland of Dairy Cattle. 2020. DOI: 10.21203/rs.3.rs-101174/v1
- [191] Aguiar-Pulido V, Huang W, Suarez-Ulloa V, Cickovski T, Mathee K, Narasimhan G. Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis: supplementary issue: bioinformatics methods and applications for big metagenomics data. *Evolutionary Bioinformatics*. 2016 Jan;12:EBO-S36436. DOI: 10.4137/EBO.S36436
- [192] Vanwonterghem I, Jensen PD, Ho DP, Batstone DJ, Tyson GW. Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Current opinion in biotechnology*. 2014 Jun 1;27:55-64. DOI: 10.1016/j.copbio.2013.11.004

- [193] Shakya M, Lo CC, Chain PS. Advances and challenges in metatranscriptomic analysis. *Frontiers in genetics*. 2019 Sep 25;10:904. DOI: 10.3389/fgene.2019.00904
- [194] Jiang Y, Xiong X, Danska J, Parkinson J. Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome*. 2016 Dec;4(1):1-8. DOI: 10.1186/s40168-015-0146-x
- [195] Peimbert M, Alcaraz LD. A hitchhiker's guide to metatranscriptomics. In *Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing 2016* (pp. 313-342). Springer, Cham. DOI: 10.1007/978-3-319-31350-4_13
- [196] Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic acids research*. 2014 Jan 1;42(D1):D643-D648. DOI: 10.1093/nar/gkt1209
- [197] Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature biotechnology*. 2019 Aug;37(8):852-857. DOI: 10.1038/s41587-019-0209-9
- [198] McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal*. 2012 Mar;6(3):610-618. DOI: 10.1038/ismej.2011.139
- [199] Westreich ST, Treiber ML, Mills DA, Korf I, Lemay DG. SAMS2: a standalone metatranscriptome analysis pipeline. *BMC bioinformatics*. 2018 Dec;19(1):1-1. DOI: 10.1186/s12859-018-2189-z
- [200] Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research*. 2019 Jan 8;47(D1):D309-D314. DOI: 10.1093/nar/gky1085
- [201] Batut B, Gravouil K, Defois C, Hiltmann S, Brugère JF, Peyretailade E, Peyret P. ASaiM: a Galaxy-based framework to analyze microbiota data. *GigaScience*. 2018 Jun;7(6):giy057. DOI: 10.1093/gigascience/giy057
- [202] Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic acids research*. 2014 Jan 1;42(D1):D553-D559. DOI: 10.1093/nar/gkt1274
- [203] Wilke A, Bischof J, Gerlach W, Glass E, Harrison T, Keegan KP, Paczian T, Trimble WL, Bagchi S, Grama A, Chatterji S. The MG-RAST metagenomics database and portal in 2015. *Nucleic acids research*. 2016 Jan 4;44(D1):D590-D594. DOI: 10.1093/nar/gkv1322
- [204] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research*. 2014 Jan 1;42(D1):D206-D214. DOI: 10.1093/nar/gkt1226
- [205] Martinez X, Pozuelo M, Pascal V, Campos D, Gut I, Gut M, Azpiroz F,

- Guarner F, Manichanh C. MetaTrans: an open-source pipeline for metatranscriptomics. *Scientific reports*. 2016 May 23;6(1):1-2. DOI: 10.1038/srep26447
- [206] Singh DP, Prabha R, Gupta VK, Verma MK. Metatranscriptome analysis deciphers multifunctional genes and enzymes linked with the degradation of aromatic compounds and pesticides in the wheat rhizosphere. *Frontiers in microbiology*. 2018 Jul 3;9:1331. DOI: 10.3389/fmicb.2018.01331
- [207] Li F. Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. *Applied and environmental microbiology*. 2017 May 1;83(9). DOI: 10.1128/AEM.00061-17
- [208] Song Z, Du H, Zhang Y, Xu Y. Unraveling core functional microbiota in traditional solid-state fermentation by high-throughput amplicons and metatranscriptomics sequencing. *Frontiers in microbiology*. 2017 Jul 14;8:1294. DOI: 10.3389/fmicb.2017.01294
- [209] Weckx S, Van der Meulen R, Allemeersch J, Huys G, Vandamme P, Van Hummelen P, De Vuyst L. Community dynamics of bacteria in sourdough fermentations as revealed by their metatranscriptome. *Applied and environmental microbiology*. 2010 Aug 15;76(16):5402-5408. DOI: 10.1128/AEM.00570-10
- [210] Peng J, Wegner CE, Bei Q, Liu P, Liesack W. Metatranscriptomics reveals a differential temperature effect on the structural and functional organization of the anaerobic food web in rice field soil. *Microbiome*. 2018 Dec;6(1):1-6. DOI: DOI: 10.1186/s40168-018-0546-9
- [211] Kukurba KR, Montgomery SB. RNA sequencing and analysis: Cold Spring Harbor Protocols. 2015;Nov 1;2015(11):pdb-top084970. DOI: 10.1101/pdb.top084970
- [212] Khang TF, Lau CY. Getting the most out of RNA-seq data analysis. *PeerJ*. 2015 Oct 29;3:e1360. DOI: 10.7717/peerj.1360
- [213] Khosravi P, Gazestani VH, Pirhaji L, Law B, Sadeghi M, Goliaei B, Bader GD. Inferring interaction type in gene regulatory networks using co-expression data. *Algorithms for molecular biology*. 2015 Dec;10(1):1-1. DOI: 10.1186/s13015-015-0054-4
- [214] Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced applications of RNA sequencing and challenges. *Bioinformatics and biology insights*. 2015 Jan;9:BBI-S28991. DOI: 10.4137%2FBBI.S28991
- [215] Delgado FM, Gómez-Vela F. Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. *Artificial intelligence in medicine*. 2019 Apr 1;95:133-145. DOI: 10.1016/j.artmed.2018.10.006
- [216] Macho Rendón J, Lang B, Ramos Llorens M, Gaetano Tartaglia G, Torrent Burgas M. DualSeqDB: the host-pathogen dual RNA sequencing database for infection processes. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D687-D693. DOI: 10.1093/nar/gkaa890
- [217] Sebastian S, Ali SA, Das A, Roy S. pARACNE: A Parallel Inference Platform for Gene Regulatory Network Using ARACNe. In *Innovations in Computational Intelligence and Computer Vision 2021* (pp. 85-92). Springer, Singapore. DOI: 10.1007/978-981-15-6067-5_11
- [218] Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW. KBase: the United States department of energy systems biology knowledgebase. *Nature biotechnology*. 2018 Aug;36(7):566-569. DOI: 10.1038/nbt.4163

- [219] Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens W, Cannoodt R, Rouchon Q, Verbeiren T. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols*. 2020 Jul;15(7):2247-2276. DOI: 10.1038/s41596-020-0336-2
- [220] Boccaletto P, Machnicka MA, Purta E, Piątkowski P, Bagiński B, Wirecki TK, de Crécy-Lagard V, Ross R, Limbach PA, Kotter A, Helm M. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic acids research*. 2018 Jan 4;46(D1):D303-D307. DOI: 10.1093/nar/gkx1030
- [221] Dibaeinia P, Sinha S. SERGIO: a single-cell expression simulator guided by gene regulatory networks. *Cell Systems*. 2020 Sep 23;11(3):252-271. DOI: 10.1016/j.cels.2020.08.003
- [222] Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M. EcoCyc: fusing model organism databases with systems biology. *Nucleic acids research*. 2013 Jan 1;41(D1):D605-D612. DOI: 10.1093/nar/gks1027
- [223] Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, Aerts S. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*. 2019 Jun 1;35(12):2159-2161. DOI: 10.1093/bioinformatics/bty916
- [224] Anders G, Mackowiak SD, Jens M, Maaskola J, Kuntzagk A, Rajewsky N, Landthaler M, Dieterich C. doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic acids research*. 2012 Jan 1;40(D1):D180-D186. DOI: 10.1093/nar/gkr1007
- [225] Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Scientific reports*. 2018 Feb 21;8(1):1-2. DOI: 10.1038/s41598-018-21715-0
- [226] Musungu B, Bhatnagar D, Quiniou S, Brown RL, Payne GA, O'Brian G, Fakhoury AM, Geisler M. Use of Dual RNA-seq for Systems Biology Analysis of *Zea mays* and *Aspergillus flavus* interaction. *Frontiers in Microbiology*. 2020 Jun 3;11:853. DOI: 10.3389/fmicb.2020.00853
- [227] D'Esposito D, Ferriello F, Dal Molin A, Diretto G, Sacco A, Minio A, Barone A, Di Monaco R, Cavella S, Tardella L, Giuliano G. Unraveling the complexity of transcriptomic, metabolomic and quality environmental response of tomato fruit. *BMC plant biology*. 2017 Dec;17(1):1-8. DOI: 10.1186/s12870-017-1008-4
- [228] Rodenburg SY, Seidl MF, De Ridder D, Govers F. Genome-wide characterization of *Phytophthora infestans* metabolism: a systems biology approach. *Molecular plant pathology*. 2018 Jun;19(6):1403-1413. DOI: 10.1111/mpp.12623
- [229] Croote D, Quake SR. Food allergen detection by mass spectrometry: the role of systems biology. *NPJ systems biology and applications*. 2016 Sep 29;2(1):1-0. DOI: 10.1038/npjbsa.2016.22
- [230] Gao Z, Ding R, Zhai X, Wang Y, Chen Y, Yang CX, Du ZQ. Common Gene Modules Identified for Chicken Adiposity by Network Construction and Comparison. *Frontiers in genetics*. 2020 May 29;11:537. DOI: 10.3389/fgene.2020.00537