# Chapter 2
# The Revolution of Omics Technology in Plant Science

**Emre Aksoy, Hikmet Yılmaz, and Ceyhun Kayıhan**

## Contents



E. Aksoy (✉)
Department of Biological Sciences, Faculty of Arts and Sciences, Middle East Technical University, Ankara, Turkey
e-mail: emreaks@metu.edu.tr

H. Yılmaz · C. Kayıhan
Department of Molecular Biology and Genetics, Faculty of Sciences and Letters, Başkent University, Ankara, Turkey



C. S. Prakash et al. (eds.), *Principles and Practices of OMICS and Genome Editing for Crop Improvement*, https://doi.org/10.1007/978-3-030-96925-7_2

## 2.1 Introduction

Plants are exposed to many different environmental stressors during their life cycle. Depending on the species or genotype, these biotic and abiotic stress conditions can hinder plant growth and development and lead to yield penalties. Tolerance or resistance mechanisms have been studied extensively for years to characterize individual genes, proteins, and metabolites involved in these mechanisms. The development of DNA sequencing approaches facilitated the characterization of genomic regions leading to the whole-genome sequencing (WGS) of different species. Advancements in nucleic acid sequencing accelerated the WGS studies. At present, more than 1,000 plant genome assemblies are accessible in GenBank, even though many of them have low quality. WGS approach was extended to the sequencing of RNAs, proteins, and metabolites. Therefore, a new scientific discipline was required to study the genes, transcripts, proteins, and metabolites holistically.

The Greek terms "ome" and "omics" are expressions derived from the suffix -ome which implies "whole," "all," or "complete." Genome, transcriptome, proteome, and metabolome are the expressions generated by adding the suffix with the terms of the gene, transcript, protein, and metabolite, respectively. Genomics, transcriptomics, proteomics, and metabolomics/lipidomics are the areas of studies that are referred to as omics. As the collective and high-throughput analyses, omics technologies integrated through robust systems biology, bioinformatics, and computational tools aim to study the mechanism, interaction, and function of cell populations, tissues, organs, and the whole organism at the molecular level (Nalbantoglu and Karadag 2019).

The approach toward omics studies has evolved since next-generation sequencing (NGS) technologies are generated. The outputs of next-generation sequencing brought about brand-new approaches to gene regulation and the data on crop genomes. It serves a potential to be used in plant breeding within metagenomic and agrigenomic researches. Gene regulation mechanisms, genes taking part in the plant defense system against pathogens, and abiotic stress factors in the whole plant or at a cellular scale can be revealed via RNA sequencing. The genotypes of lots of single-nucleotide polymorphisms (SNPs) are also determined with the methods developed within NGS. Additionally, molecular markers required for investigating genetic relationships among breeding materials, detailed genetic mapping of targeted genes, and genome-wide associations are developed with the methods called genotyping-by-sequencing (GBS) and whole-genome resequencing. Determination of the genotypes of the required genetic materials enables improving the selection of individuals that resist abiotic stressors and increases the efficiency in agriculture (Vlk and Řepková 2017).

Nowadays, the omics terminology is adapted to other fields of study, including ionomics that deals with ionic changes, methylomics studying the methylation changes in nucleic acids, and toxicogenomics. Here in this chapter, we first describe the evolution of sequencing techniques and give examples of each omics technology in plant science.

## 2.2 First-Generation Sequencing

The sequencing technologies that give rise to decoding and sequencing the genomes of the organisms are based on the discovery of the DNA, which is a double-helix structure consisting of bases (adenine, thymine, cytosine, and guanine). The first laboratory methods used in the interpretation of the DNA sequences in terms of the letters of A, T, C, G, and N representing an ambiguity were generated by Sanger et al. from Cambridge University in 1977 and Maxam et al. from Harvard University in 1980 (Kchouk et al. 2017).

The first-generation sequencing technique was further improved by the Maxam-Gilbert method, which enables sequencing the DNA with chemical degradation of the fragments at specific bases with reagents such as formic acid, dimethyl sulfate, and hydrazine (Maxam et al. 1977). In this method, the strands of the DNA fragments are denatured, and the phosphate groups at the 5' ends of the denatured DNA strands are removed with phosphatase to identify the fragments on the gel after the radioactive isotopes of phosphorus. The radioactively labeled DNA fragments are exposed to chemical reactions in four different tubes in the presence of distinct base-specific chemical reagents. Each of the reagents results in base modification, removal of the base, and phosphodiester cleavage of the DNA strand at that site. Guanine cleavage is induced by DMS + piperidine, while the cleavage of guanine and adenine requires DMS + formic acid + piperidine. Hydrazine piperidine causes cytosine cleavage, and sodium chloride + hydrazine piperidine facilitates the cleavage of cytosine and thymine. At the end of the reactions in the four distinct tubes, labeled fragments with various sizes are separated by electrophoresis (Saraswathy and Ramalingam 2011). The polyacrylamide gel contains urea which prevents the formation of secondary structures in the single-stranded DNA. Then, the DNA sequence is determined by using autoradiography. This sequencing method does not involve DNA cloning. On the other hand, the development of the Sanger sequencing method is more applicable compared to the Maxam-Gilbert method due to its greater simplicity, higher accuracy, and lower radioactivity (Kulski 2016).

As mentioned above, Sanger sequencing developed by Frederick Sanger in 1977 is expressed as the chain termination, dideoxynucleotide, or the sequencing by synthesis method in which one strand of the DNA is used to identify the sequence (Kchouk et al. 2017). In this technique, dideoxynucleotides (ddNTPs) are used which are the analogs of the monomers of the DNA molecules, deoxyribonucleotides (dNTPs), lacking 3′ hydroxyl groups required for the extension of the DNA strands (Heather 2015). The integration of the ddNTPs to the elongating DNA prevents the process to be terminated successfully as the subsequent base cannot be incorporated into the strand. Thus, the DNA fragments with different sizes and the ddNTP molecules at their ends as the analogs of the related bases are obtained. Chain termination reactions are conducted in four different tubes. Each tube contains a different type of ddNTP and the common reaction components including dNTP mix, template DNA, radiolabeled primer, and DNA polymerase. Radioactive isotopes of the phosphorus ($^{32}$P or $^{33}$P) enable identifying the DNA sequence. The

tubes contain a small percentage of ddNTP (about 1%). The polyacrylamide gel with urea is also used, and the DNA sequence is determined in autoradiography (Sanger et al. 1977). The bands of the DNA fragments separated regarding their sizes on the gel slab are displayed with an imaging system, either of X-ray or UV light. The Sanger sequencing was firstly used to sequence the *phiX174* genome (5374 bp) and the bacteriophage λ genome (48501 bp). The speed and accuracy of the sequencing were improved with the automatic sequencing machine based on capillary electrophoresis developed by Applied Biosystems in 1995. The genetic materials of varying plant species such as *Arabidopsis* (The Arabidopsis Genome Initiative 2000), rice (Goff et al. 2002), and soybean (Schmutz et al. 2010) and human genome were also sequenced with Sanger sequencing. The Sanger sequencing has been used for three decades and is still preferred in single or low-throughput DNA sequencing. On the other hand, Sanger sequencing is considered to be time-consuming and expensive. The limited analysis speed also reduces the efficiency besides the inability to decode the complex genomes with the Sanger sequencing (Kchouk et al. 2017).

## 2.3 Next-Generation Sequencing

Following the domination of Sanger sequencing for 30 years, NGS was developed as a high-throughput DNA sequencing technology considered within the second- and third-generation sequencing methods (Kulski 2016). By this method, a high number of simultaneous sequencing reactions become feasible, and the cost of sequencing is lowered due to the developments in detection systems, microfluidics, and integrating the sequencing reactions to minimized dimensions (Türktaş et al. 2015; Kulski 2016). Increased scalability and speed of generating data paved the way for advanced studies on biological systems besides the decrease in time for obtaining gigabase-sized sequences from years to days or hours via NGS (Noman et al. 2017).

NGS enables carrying out studies on genetic approaches in plant breeding and biotechnology, evolution, discovering genetic markers, gene expression profiling via mRNA sequencing, and de novo draft genome sequences within the relevant method of NGS applications such as WGS, exome sequencing (exome-seq), RNA sequencing (RNA-seq), and methylation sequencing (methyl-seq) (Türktaş et al. 2015; Low et al. 2019). The NGS platforms with 99% accuracy rates may also detect nucleotides with errors. Although the current NGS methods are highly accurate, they are still prone to errors. Even the accuracies of more than 99% may accumulate hundreds of thousands of errors in the sequencing of large genomes since NGS platforms generate high amounts of output. The number of times a nucleotide is sequenced is referred to as "coverage" or "depth" (Sims et al. 2014). Coverage may also be used to refer to the percentage of target bases that have been sequenced a specific number of times. Coverage varies depending on the type of NGS and the research application. More coverage tends to be used when in search for a variant

that is less common (<1%) in a sample. For example, whole-genome sequencing generally requires approximately 30x coverage as this will detect 98% of heterozygous single-nucleotide variants identified in a microarray. The coverage can be calculated by the Lander-Waterman equation (Sims et al. 2014).

## 2.4 Second-Generation Sequencing

To overcome the limitations of the first-generation sequencing tools that were used for three decades such as Sanger sequencing, brand new sequencing methods were developed (Kchouk et al. 2017). Second-generation sequencing methods enable sequencing multiple DNA fragments simultaneously that facilitate assembly and determination of complex genomic regions, methylation detection, and gene isoform detection (Muhammad et al. 2019).

Millions of short fragments are read in parallel, the speed of the sequencing process is increased, electrophoresis is not required for detecting the output and the cost is reduced within the second-generation sequencing methods (Kchouk et al. 2017). Template libraries of randomly fragmented DNA or complementary DNA (cDNA) obtained from reverse transcription are generated with shotgun sequencing by ligating the linker or adapter sequences with the DNA molecules rather than performing cloning via a host cell (Kulski 2016). In second-generation sequencing, the read length of these technologies is shorter than the first generation; therefore, amplification is necessary for signal detection (Kang et al. 2019). A solid surface or beads are used in the library amplification process in the presence of miniaturized emulsion droplets or arrays, while the nucleotides to be sequenced are detected via luminescence or changes in electrical charge (Kulski 2016). These sequencing methods are classified in two, namely, sequencing by ligation (SBL) and sequencing by synthesis (SBS), and the sequencing platforms used are Roche/454 established in 2005, Illumina/Solexa in 2006, and the ABI/SOLiD (Sequencing by Oligonucleotide Ligation and Detection) in 2007 (Kchouk et al. 2017; Meera et al. 2019).

## 2.5 Pyrosequencing Technology

Pyrosequencing also known as 454 technology was the first second-generation technology developed in 2005. In this technology, the main principle is to determine the base with chemical luminescence. The pyrosequencing method is different from the Sanger sequencing since the nucleotide incorporation is performed in the presence of DNA polymerase, ATP sulfurylase, luciferase, and apyrase enzymes which are kinetically well-balanced (Ramon et al. 2003). PCR amplification and pyrosequencing of the query DNA fragments are utilized to carry out real-time sequencing (Rothberg et al. 2008). In the pyrosequencing method, adapter molecules provide

the DNA molecules that have been previously fragmented to bind the agarose beads after attaching the DNA fragments. The agarose beads with DNA fragments are mixed with Taq polymerase and buffer solution before being introduced to an oil-water emulsion to induce emulsion PCR (emPCR). The DNA fragments are then amplified in the presence of dNTP and adapters considered as primers (Saraswathy and Ramalingam 2011). The nucleotides are formed and tested in terms of their inclusion in a DNA template which occurs by the release of pyrophosphate (PPi) proportional to the amount of the nucleotides (Ramon et al. 2003). ATP sulfurylase is the enzyme that uses pyrophosphate in ATP synthesis by converting it to ATP in the presence of adenosine 5' phosphosulfate. Production of oxyluciferin from luciferin is facilitated by luciferase driven by ATP. Light emission from the oxyluciferin formed previously providing chemical luminescence takes place as a result. The number of nucleotides is associated with the amount of light emitted providing the determination of the base sequence. The emitted light is illustrated with peaks having heights proportional to the number of nucleotides in a program after it is spotted with a charge-coupled device (CCD) camera. As the apyrase enzyme degrades the excess ATP and dNTP, another pyrosequencing cycle initiates with the integration of the subsequent dNTP, and the complementary strand of the DNA is constructed. A cyclic nucleotide dispensation order (NDO) is utilized to decode an unknown sequence with pyrosequencing. In this method, one of the dNTPs is recruited to the DNA template where the rest of the dNTPs are degraded in the presence of apyrase after each cycle of dNTP dispensation. Non-cyclic NDOs are also generated with the order of nucleotide dispensation and the heights of the peaks in the program in case the DNA sequence is known (Ramon et al. 2003).

Besides the disadvantages such as high cost and low accuracy of reading, the 454 technology can read long sequences (around 700 bp). In addition, the sequences are expected to be smaller than the outputs of the other second-generation sequencing methods, and homopolymers would be sequenced with lower accuracy (Saraswathy and Ramalingam 2011).

## 2.6 Illumina Technology

After developing Illumina sequencing in 2006, Solexa commercialized it as Illumina/Solexa Genome Analyzer. The platforms developed by this company, namely, MiSeq, NextSeq 500, and HiSeq 2500, can put forward 15 Gb, 120 Gb, and 1000 Gb of sequencing data in each run while their maximum read lengths are 2×300 bp, 2×150 bp, and 2×125 bp, respectively. In addition, the NovaSeq 6000 System is declared to present output up to 6 Tb and 20B reads in less than 2 days. It is also claimed that the Illumina sequencing technology has been used in generating more than 90% of the sequencing data of the world as being the most remarkable technology in the NGS market (Krishna et al. 2019).

As being a sequencing by synthesis (SBS)-based technology, cluster generation involves the fragmentation of DNA molecules and ligation of the fragments with

short adapter oligo at both ends. This aids connection and amplification of fragments on a flow cell where sequencing reactions take place. There are microfluidic channels on the flow cell called lanes. Oligonucleotide sequences are attached in each lane and are complementary to the adapters. These complementary oligos form a cluster that is called polony since the appearance of each PCR-amplified DNA fragment looks like a bacterial colony (Turcatti et al. 2008). The flow cell surface that is used for the immobilization of the templates for sequencing enables increased stability of DNA and accessibility of enzymes to the DNA. It also reduces the nonspecific binding of the fluorescently labeled nucleotides. One thousand copies of a template with a diameter of one micron or less are generated within solid-phase amplification. Single-molecule cluster densities reaching the order of 10 million per square centimeter are obtained by different methods including photolithography and mechanical spotting.

In Illumina technology, the PCR amplification of the DNA fragments is performed using the adapter sequence as a primer, and each type of dNTP is labeled with different types of fluorescent labels. In each sequencing cycle, only a single-labeled dNTP is introduced to the nucleic acid chain, and thus each type of dNTP signal helps the detection of base calling, while the signal length helps the identification of the number of the attached dNTPs. Fluorescently labeled nucleotides with a reversible terminator are used in Illumina sequencing. Therefore, the polymerization terminates in the presence of the nucleotide label. The fluorescent label is screened to determine the base after each dNTP incorporation. The dye is then removed from the 3' end by the enzymes for the subsequent nucleotide to be incorporated, and the next cycle begins. Even though the sequences generated after the process are short, large data can be generated accurately and fast (Turcatti et al. 2008). As a technology displaying an error rate below 1%, Illumina sequencing is claimed to be one of the most accurate NGS technologies. The incorporation bias is reduced by the natural competition in the presence of reversible terminator-bound dNTPs that are single, separate molecules. In each cycle, the measurements of the intensities of the signals induce the base calls which are the reasons behind significantly reduced raw error rates compared to the alternative technologies. Imaging the clusters on the flow cell surface is the most time-consuming step of the process besides the nucleotide incorporation phase facilitated by the enzymes. The substitution of a nucleotide located in a specified position in the genome which is named as single-nucleotide substitution is the error taking place most frequently (Turcatti et al. 2008).

Within the resequencing approaches, the sequences are allowed to be aligned to a reference in the Illumina data collection software. The full range of data collection, processing, and analysis modules to streamline collection and analysis of data with minimal user intervention is enabled with this software that was generated with the help of leading researchers. The open format of the software with simple application program interfaces also provides accessing data at various stages of processing and analysis.

## 2.7 Ion Torrent Technology

Ion Torrent technology is based on an SBS process similar to Illumina technology. DNA fragments are amplified by an emulsion PCR (emPCR) on beads that are washed over a picowell plate, and each nucleotide is added later on to release pyrophosphate (Heather 2015). Each ion chip contains a liquid flow chamber which helps the influx and efflux of nucleotides (Merriman et al. 2012). A complementary metal-oxide-semiconductor technology is used to detect the difference in pH caused by the release of protons ($H^+$ ions) during polymerization (Rothberg et al. 2011). The bottom of each chip is covered with millions of pH microsensors. The pH change is not specific to nucleotide types, and each type of dNTP is released in a fixed order. According to the measurement of pH change, the sequence is determined (Merriman et al. 2012). This technology allows for very rapid sequencing during the actual detection phase (Glen et al. 2011). The error rate of the Ion Torrent technology is higher than the Illumina since the indels are the major error in this technology. This technology cannot detect homopolymer sequences of identical nucleotide stretch such as TTTTTT due to the loss of signal as multiple matching dNTPs incorporate (Loman et al. 2012). If the DNA template has a homopolymeric region, pH change should be proportional to the attached nucleotide number. Instead, as the attached nucleotides increase in a homopolymer, the expected pH change decreases gradually. In addition, the lengths of the sequence read obtained in one experiment of Ion Torrent are various rather than being the same. The sequence reads from both ends of a fragment cannot be obtained with the current generation of Torrent devices (Lahens et al. 2017).

## 2.8 Third-Generation Sequencing

Several biological limitations such as assembly and determination of complex genomic regions, gene isoform detection, and methylation detection are not eliminated by the second-generation sequencing technologies because of the short read lengths, even though they present developments outstripping Sanger sequencing (Rhoads and Au 2015). Third-generation sequencing is presented as a promising technology to eliminate the mentioned limitations. The length of the read is also improved to tens of thousands of bases from tens of bases per read within the third-generation sequencing approaches, besides the decrease in time required for sequencing from days to hours and elimination of sequencing biases resulting from the PCR amplification process (Lu et al. 2016).

Unlike second-generation sequencing, third-generation sequencing technologies do not require the sample amplification step and can sequence a single DNA molecule. Also, they may produce more than 10 Kb reads and thus produce highly precise de novo assemblies and contiguous genome reconstruction even at the regions of high content of repetitive elements. These technologies include Pacific

Biosciences, Helicos System (Helicos single-molecule sequencing), and Oxford Nanopore Technologies (ONT). The first commercial NGS implementation was the Helicos System that utilized single-molecule fluorescent sequencing. However, the Helicos Biosciences company filed for bankruptcy in 2012.

## 2.9 Pacific Biosciences Technology

Pacific Biosciences developed the PacBio RS II sequencer and the single-molecule, real-time (SMRT) sequencing system based on the properties of zero-mode waveguides (Schade et al. 2010). PacBio sequencing enables closing the gaps in reference assemblies and determination of structural variation in genomes with the highly contiguous *de novo* assemblies. The mutations that are related to the diseases can be spotted, and extended repetitive regions are sequenced by using relatively long reads. Additionally, isoforms of genes, novel genes, and isoforms of annotated genes can be determined with PacBio transcriptome sequencing as the whole transcripts and relatively long fragments are sequenced. Base modifications such as methylation can also be spotted with PacBio sequencing. Moreover, cost-effective and scalable hybrid sequencing strategies are generated to utilize short reads in relation to long reads (Rhoads et al. 2015).

SMRT sequencing is a method carried out in cells with 150,000 ultra-microwells at a zeptoliter scale. Each well contains a DNA polymerase molecule stabilized at the bottom with a nanostructure including a biotin-streptavidin system called zero-mode waveguides (ZMWs) (Kulski 2016). DNA chains pass through the DNA polymerase, and complementary binding nucleotides promote the detection of the sequences via the signals from fluorescence labels that are attached to the end phosphate groups, which are generated by them (Rhoads et al. 2015). Long reads with high accuracy are obtained with a circulating structure (SMRTbell) constructed by the adaptors. In this technology, first, the sequencing templates are annealed. The complex consisting of template-primer-polymerase is immobilized to the 150,000 ZMWs. After the labeled nucleotides interact with the polymerase, the end phosphate group is cleaved, and the fluorescent signal is detected simultaneously and recorded with a CCD camera. Because the wavelength of the visible light is more than the diameter of a ZMW, the light reflected through the glass bottom reaches the bottom 30 nm of the ZMW. Therefore, the reduction in background noise and the detection of a recruited nucleotide are facilitated with the detection volume.

As the nucleotides are integrated and detected simultaneously rather than the second-generation technologies in which the nucleotides are added in order, the sequencing is completed faster. The accuracy of sequencing 900 bp read length has increased from 99.3% to 99.9% by circularizing the template and sequencing it multiple times by using SMRTbell (Travers et al. 2010; Koren et al. 2013). The drawbacks of SMRT are the high cost, the need for the high amount of DNA samples, and the error rate of 10–15%, which is mostly caused by indels.

## 2.10 Oxford Nanopore Technology

The third-generation sequencing technology developed in 2005 by Oxford Nanopore Technologies Ltd. enables simultaneous analysis of native DNA or RNA sequences at any length in fully scalable formats from pocket to population scale. It uses a nanometer-level channel in a membrane, and it determines the base sequence by the potential difference changing between the membranes passing through a single-stranded DNA (ssDNA). In this technology, the leader and the hairpin adapters are used. Each adapter is ligated to one end of the double-stranded DNA (dsDNA). The leader adapter is denoted as the Y adapter since it has a Y-shaped structure, while the hairpin adapter is called the HP adapter. Sequencing starts at the single-stranded 5' end of the Y adapter, followed by the template strand, then the HP adapter, and the complementary strand. Helicase enzyme translocates along dsDNA to ssDNA, and the hairpin protein makes each base of ssDNA pass through the nanopore at a constant rate, and so base calling may then be performed. Each type of dNTP causes different electrical potential changes that are read, and base sequences are determined (Branton et al. 2010). In this technology, sample preparation is minimal, and long read lengths can be generated in the Kb range compared to the second-generation sequencing technologies. Also, amplification and ligation steps are not required before sequencing. However, the optimization of the speed of DNA translocation through the nanopore should be needed to obtain the accurate measurement of the ionic current changes and to decrease the high error rates of base calling (Stoddart et al. 2009). Thus, the current error rates (roughly around 98%) are very high; therefore it cannot compete with existing sequencing technologies. Moreover, the low depth of coverage obtained with this technology is a possible barrier to accurate eukaryotic genome sequencing at the moment.

## 2.11 Genomics

Shotgun sequencing was used for some of the early plant genomes including *Arabidopsis*, soybean, poplar, and papaya (Michael and Van Buren, 2015). The sequence and genetic structure of plant genomes are determined with an extensive sequencing method called whole-genome sequencing (WGS). In early sequencing projects that focus on WGS, the genomes of strawberry (Shulaev et al. 2011) and wheat (Brenchley et al. 2012) were randomly fragmented, and elements with varying sizes are obtained. The reads obtained from the sequencing process were assembled with bioinformatic tools after the bacterial artificial chromosome (BAC)-end sequencing is performed. De novo projects also utilize WGS besides the resequencing attempts. Preparing a draft of unknown plant genomes is managed with the whole DNA or mRNA de novo sequencing even though the process is time-consuming (Türktaş et al. 2015). Despite the possibility of determining locations of

the contigs or scaffolds with low accuracy, and missing several genes while generating draft genomes, the presence of genome information enables analyses with high throughputs and characterization of genes (Sarethy and Saharan 2021). Later, WGS approach was used to generate the draft genomes of einkorn (Ling et al. 2013), wheat, and *A. tauschii* (Jia et al. 2013). Moreover, resequencing is considered to be useful in transcriptome profiling and detecting SNPs to generate molecular markers. For instance, WGS enabled the construction of the reference genome of potato and discovering SNPs to compare a homozygous doubled-monoploid line with its heterozygous diploid line (The Potato Genome Sequencing Consortium 2011). WGS of many different crop and vegetable species has been completed in the last decade. Although the second-generation sequencing resulted in many lower-quality assemblies, a massive extension WGS of different plant species, especially of the crops, leads to a revolution in plant genomics.

One thousand one hundred twenty-two plant genome assemblies are deposited in GenBank, representing 631 land plant species. The advancements in the long-read sequencing markedly improved the NGS data quality; therefore the number of plant genome assemblies has increased dramatically in the past 20 years. Almost 60% of the plant genome assemblies have been sequenced in the last 3 years alone. Model plants and some crops were the first species whose genomes were fully sequenced. But now, any plant species can be sequenced due to a steady decline in sequencing costs.

The exons considered to be the coding region for the protein synthesis in the genome of an organism are called the exome. Even though they involve the total of the sequences inducing the generation of proteins taking part in phenotypic regulation, they are insufficient to decode fully the mechanisms behind the gene regulation. To enlighten the molecular background of the diseases and phenotypic traits, exome sequencing was introduced as an essential genetic tool. Exome sequencing helps with the identification of genes (whole exome, genes responsible for a disease, or class of genes), determination of phenotypic traits, identification of exome SNPs, and further computational and statistical applications to identify the signals of diseases (Hashmi et al. 2015).

WGS of different populations of the same plant species showed a high degree of genomic variation within the species; therefore it was obvious that single reference genomes no longer can represent the diversity within a species. This observation led to the advancement of the pan-genome concept, which was first developed in bacteria in 2015 (Tettelin et al. 2005). Pan-genomes can distinguish the primary genes that are present in all individuals and variable genes that are found in some individuals but absent in others. Hence, it symbolizes the genomic diversity within the species. Pan-genomes can be curated by three different methods, each with its benefits and disadvantages over the others (Bayer et al. 2020). The first pan-genome study in plants was a comparison of WGS of wild soybean relatives (Li et al. 2014). Another study in rice compared the genomes of three accessions (Schatz et al. 2014). At present, more than 8,000 studies reported pan-genome comparisons in plants (Bayer et al. 2020). These studies have impacts on understanding the biological

significance of genotypic variances at loci linked with tolerance and resistance, developmental processes, and yield enhancement.

The transposable elements in the plant genome are high in copy number since their segmental or tandem duplication takes place frequently. Therefore, an extended amount of repetitive elements is found in plant genomes. Autopolyploid or allopolyploid character of the genome or the age of ploidization affects the progress of the sequencing as ploidy is considered a challenge. To eliminate the obstacles caused by the complexity of the genome, library sequencing of fragmented genome elements is executed by using restriction enzymes or obtaining the sequences without using enzymes (Vlk and Řepková 2017). Variations or significant polymorphisms in the genome are considered to be useful in pre-breeding attempts with resequencing projects. The reference genomes of the desired plant are also intended to be generated within various projects. They are considered as providing information about the structure and function of the genome and the genome assembly patterns of the related species together with molecular markers and candidate genes that can be used in further studies (Vlk and Řepková 2017).

Epigenetic changes such as chromatin modifications, transposable element inactivation, paramutation, transgene silencing, and co-suppression are investigated with the sequencing approaches in detail in various plant species. The changes in gene expression and chromatin-based expressional responses generated against environmental stimuli prove the importance of epigenetic studies in plants (Köhler and Springer 2017). Traditional methods used in epigenetic studies involve bisulfite conversion, methylation-sensitive restriction enzymes, and antibodies specific to 5-methylcytosine. Microarray-based methods were also started to be combined with these methods to carry out a genome-wide analysis of DNA methylation (Buck and Lieb 2004). Recently, NGS technologies paved the way for epigenetic studies (Vlk and Řepková 2017). Therefore, the studies of applied epigenetics cause new opportunities for crop improvement. It has been suggested that varietal selection of crops is associated with variability caused by epigenetic mechanisms (Rodríguez López and Wilkinson 2015; Crisp et al. 2016; Fortes and Gallusci 2017; Gallusci et al. 2017). The potential to develop crop performance and energy use efficiency was shown in *Brassica napus* via an epigenetic selection of isogenic lines (Hauben et al. 2009). Organ-specific epigenetic modifications were determined in maize by Illumina sequencing technology (Wang et al. 2009a). The expression levels of genes are regulated by epigenetic mechanisms in response to plant development and biotic and abiotic stresses, and this affects the phenotype of plants (Kumar 2018).

DNA methylation, histone modifications, and small RNA molecules are the major epigenetic mechanisms affecting the expression levels of genes (Rodríguez López and Wilkinson 2015). DNA methylation is an important chromatin modification that can be inherited in animals and plants. It has been recently suggested that methylation of the promoter and the gene coding region has different effects on gene expression (Wang et al. 2015a, b). The methylation of the promoter region of a gene is related to the repression of transcription (Kass et al. 1997). On the other hand, the methylation of the gene coding region is found with an intermediate expression level in plants. It was shown that it can be involved in reducing erroneous

transcription by reducing intron retention by single-cell transcriptome sequencing data from *Arabidopsis* root quiescent center cells (Horvath et al. 2019). Furthermore, it can enhance the gene expression in certain gene families (Dubin et al. 2015; Anastasiadi et al. 2018). DNA methylation which targets cytosines in varying sequence patterns such as CG, CHG, and CHH can be revealed efficiently with NGS after treating the DNA with sodium bisulfite. Even though mostly the transposons are methylated as being primary targets for epigenetic silencing, the relation between the transposon polymorphism and DNA methylation variation is not easily described because they are highly repetitive and result in large insertion/deletion polymorphisms in the genome. The connection between transposon methylation and transposon insertions was studied using whole-genome bisulfite sequencing data sets by Daron and Slotkin (2017). Also, bisulfite conversion and Illumina sequencing were used together for the identification of the methylated genomic regions in tomato, and it was suggested the ripening of tomato fruits was under the control of epigenetic regulation along with hormonal control (Zhong et al. 2013).

## 2.12 Functional Genomics

Biological investigations were focused on genes and proteins in vitro during the early 1990s. However, as technologies improved and evolved, the approach shifted to research on different molecular aspects, viz., structural genomics, transcriptomics analysis, proteomics, and metabolomics. For instance, a multidisciplinary approach involving integrative analysis is crucial to study the complexity of plant-microorganism interactions (Sarethy and Saharan 2021).

## 2.13 Transcriptomics

The complete set of transcripts in a cell, and their quantity, for a specific developmental stage or condition, is called transcriptome. It is essential for understanding the functional elements of the genome and the molecular regulations of cells and tissues and also for revealing disease and development. The ultimate goals of transcriptomics are to determine all species of transcript such as mRNAs, small RNAs, and non-coding RNAs for revealing the transcriptional structure of genes and the changes in expression levels of each transcript during development and under different conditions (Wang et al. 2009). Several technologies have been developed for transcriptomics, including hybridization- or sequence-based approaches. Commercial high-density oligo microarrays and custom-made microarrays with fluorescently labeled cDNA are the important techniques used in hybridization-based approaches. Furthermore, specialized microarrays have been used for some specific purposes such as the detection of spliced isoforms.

They are high-throughput, relatively inexpensive, and high sensitivity by lowering the detection threshold of the transcriptional level of the less represented genes of a mixture, thus facilitating the analysis of thousands of genes in the same reaction (Kerr et al. 2000). However, they have several limitations such as existing knowledge about genome sequence, high background levels due to cross-hybridization, and complicated normalization methods. Microarrays have been widely used to produce global expression profiles under abiotic stresses in plant species (Kayıhan and Eyidoğan 2019). For instance, the AtH1 *Arabidopsis* GeneChip from Affymetrix has been employed to study transcriptome changes in *Arabidopsis* under salt stress. Accordingly, approximately 35% of the genome (~8000 genes) exhibited expression changes under salt or other abiotic stresses (El Ouakfaoui and Miki 2005). Changes in gene expression caused by drought stress by using microarrays have been suggested by several research groups. For the first time in the literature, Ozturk et al. (2002) found that genes encoding jasmonate-responsive, late embryogenesis abundant, and ABA-responsive proteins were upregulated in barley seedlings exposed to drought. Also, it was found by microarray that changes in the expressions of 300 genes were revealed in spring and winter wheat under cold stress (Gulik et al. 2005). Furthermore, microarray technology provided comprehensive data for $K^+$ deficiency in plants, and this showed a more integrative point of view considering all aspects of $K^+$ management in plants (Kayıhan and Eyidoğan 2019). Kayihan et al. (2017) and Öz et al. (2009) performed the microarray experiments in wheat and barley cultivars exposed to excess boron, respectively. They suggest that WRKY transcription factors, genes related to jasmonate biosynthesis, glutathione S transferase, and NIP4;1 can have a role in boron tolerance mechanisms in cereals. Also, global gene expression analyses were performed in *Arabidopsis thaliana* exposed to high B and low B conditions (Kasajima and Fujiwara 2007). They identified novel high B-induced genes including heat shock protein and multidrug and toxic compound extrusion (MATE) family transporter genes. On the other hand, microarrays have been widely used for transgenic plants such as maize, canola, cotton, tomato, and soybean events (Leimanis et al. 2006; Xu et al. 2007; Schmidt et al. 2008; Zhou 2008; Kim 2010; Feng 2013).

The transcript levels of genes depending on their changes under different conditions are important information, and they can reflect the functions and transcriptional regulation relationships of genes. Modern omics technologies play an important role in better understanding gene expression. The best approach for the characterization of candidate transcripts that are responsible for many biological functions is transcriptome study. NGS technology provides us a powerful tool to reveal the transcriptional landscape of investigated tissue(s) at special developmental stage(s) because it can easily obtain transcriptome data from different plant tissue(s) and developmental stage(s). RNA-seq approach that uses NGS techniques is used for analysis and study of the entire transcriptome, and this approach provides an insight on the expression level of transcripts. Genes expressed within a defined period of time from a particular tissue or cell can be found by RNA-seq. There are some universal steps for this approach. RNA fragments are converted to a cDNA library by reverse transcriptase, and from both ends of cDNA fragments, cDNA

library fragments are ligated to adapter molecules. Then adaptor attached library fragments are sequenced. Through cDNA sequencing, transcriptomes are studied deeply and efficiently. For plant transcriptomes, Illumina technology has generally better coverage. Reference genome and de novo assembling are two types of the assembly methods. For large NGS data of complex genomes without a reference genome, de novo assembly is useful (Wang et al. 2009). De novo transcriptomes are provided by some bioinformatic tools such as TRAPID (Van Bel et al. 2003) and Trinity (Brain et al. 2013). RNA-seq data is used for the development of molecular markers (Trick et al. 2009) and gene characterization (Dassanayake et al. 2009).

RNA-seq has successfully assisted in identifying several genes responsible for biotic and abiotic stress responses in various plant species. A large number of genes related to developmental stages were identified by RNA-seq in cucumber via 454 pyrosequencing (Ando et al. 2012). A combination of microarray and Roche technology was used to identify genes that were linked to the quality of cotton fibers (Nigam et al. 2014). To find genes associated with drought tolerance, RNA-seq analysis was performed in *Populus euphratica* Oliv. grown in arid or semi-arid regions using the Roche 454-GS FLX System (Tang et al. 2012). Likewise, sequencing red clover with Illumina technology discovered genes related to drought tolerance and determined the increase in metabolites such as pinitol, proline, and malate in leaves (Yates et al. 2014). The transcriptome of soybean (Fan et al. 2012), cotton (Xu et al. 2013a, b), and halophyte grass (Yamamoto et al. 2015) was sequenced to explore a molecular mechanism of salt tolerance in these plants. In addition, a whole-genome study was performed in soybean using Illumina technology, which examines the function of the plant-specific family of NAC transcription factors during development and dehydration stress (Le et al. 2011). Ion Torrent technology has been used in transcriptome analysis of finger millet, a hardy grain known for its tolerance to salinity, drought, and disease (Rahman et al. 2014). Transcriptome profiling of jatropha roots was carried out to elucidate molecular reactions to waterlogging (Juntawong et al. 2014). As the third generation, Pacific Biosciences' SMRT technology was used to investigate the interaction of *Xanthomonas oryzae* pv. *oryzicola* and its host, *Oryza sativa* L., by whole-genome sequencing of the pathogen and RNA-seq of the host under attack (Wilkins et al. 2015). Illumina sequencing was used to obtain responsible herbicide resistance genes for *Lolium rigidum* Gaudin (Gaines et al. 2014) and for copper tolerance (Wang et al. 2015a, b). In sweet potatoes (*Ipomoea batatas* L.), biotic stress resistance analyses of catalase genes were performed using NGS technologies, and it was found that a positive response to IbCAT2 may play an important role in stress responses (Yong et al. 2017). In tomatoes, an abiotic stress tolerance identification study was conducted to understand the plant responses and genetic regulatory networks involved in abiotic stress responses (Chaudhary et al. 2019). In plants, RNA-seq technology has been used to determine the patterns of differentially expressed genes between hybrids and their parents to understand the genetic basis of heterosis (Zhai et al. 2013; Hansey et al. 2012; Sexane et al. 2014). Accordingly, gene expression for allopolyploid heterosis was predominant in the emerging hexaploid wheat dominance (Swanson-Wagner et al. 2006), but over-dominance was the key element for

nicotine biosynthesis in tobacco (Tian et al. 2018). Dominance and over-dominance effects were shown by heterotic genes in connection with ear development earlier in maize inforescence (Ding et al. 2014). In the chrysanthemum, two characteristics of flowering – the initial flowering time and the flowering duration – are regulated by the presence of two pairs of main genes (Zhang et al. 2011).

MicroRNAs (miRNAs) are the key regulators at the post-transcriptional level in eukaryotic organisms. They regulate the expression levels of genes in response to development and various stress responses in plants. They are complementary with the target mRNAs and are highly conserved. Up till now, several technical approaches have been used to identify and verify the miRNAs. These are in silico prediction based on conserved sequences, to create miRNA libraries and to follow this with cloning and sequencing and finally the sequencing of miRNAs. In silico prediction was applied in rice (Bonnet et al. 2004). Cadmium-responsive miRNAs and their target genes in *Raphanus sativus* L. roots were identified by Illumina sequencing technology (Xu et al. 2013a, b). Also, circular RNAs were identified by transcriptome analysis by means of SMRT technology by Pacific Biosciences, and it was found that they had an important role in the function of miRNA and transcriptional control (Lu et al. 2015). On the other hand, long non-coding RNAs (lncRNAs), which are longer than 200 bp and do not encode any protein product, are another important regulatory mechanism associated with gene silencing, flowering time regulation, and abiotic stress responses (Wang et al. 2014; Zhang et al. 2014). These molecules were identified in crops, such as wheat (Xin et al. 2011), rape mustard (Yu et al. 2013), apple (Celton et al. 2014), and poplar (Shuai et al. 2014) by tiling array, EST analyses, and RNA-seq.

## 2.14 Proteomics

Proteomics is one of the growing fields of biological research with an immersive positive impact on plant science. Proteomics is a term that refers to the comprehensive identification and quantitative study of protein expression in an organism, cell, tissue, or organelle at a certain time and under specific conditions (Tan and Chen et al. 2011). Understanding proteome profiles provides a direct connection between genomic and transcriptomic regulation and phenotype. Since the first plant proteomic study in maize (Touzet et al. 1996), exponential progress has been made in different crop species although the full potential of plant proteomics has yet to be realized. Recent advances in new or improved technologies, protocols, or workflows have opened up new possibilities for high-throughput proteome analysis and reduced protein assessment errors.

Two-dimensional polyacrylamide gel electrophoresis and differential in-gel electrophoresis (DIGE) have been used in many early proteomic studies to separate the proteins. However, its resolution is not enough to ensure reproducibility and sensitivity (Rabilloud et al. 2010). Therefore, chromatographic

separation followed by mass spectrometry (MS) is now routinely employed in proteomic studies. There are some deviations of chromatographic separation techniques such as high-pressure liquid chromatography (HPLC) and gas chromatography (GC). After proper separation of protein mixtures, they can be identified by single or double MS systems. Sometimes samples can be ionized by matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) before identification in MS. This technique uses a laser energy-absorbing matrix to create ions from large molecules with minimal fragmentation (Jurinke et al. 2004).

Genomics and proteomics have developed separately into two different disciplines, thus limiting the cross talk between scientists in the two fields, limiting the integration of useful information from both fields into a single data modality. However, depending on the encoded genomic variants, mutations, or post-transcriptional modifications at the nucleotide level, the final expressed sequence of a protein may vary. NGS can be used to capture and correctly decipher these variants. Single-nucleotide polymorphisms (SNPs) and small insertion-deletion (INDELs) can be characterized using NGS, and these sequence variants can be easily translated in silico into different proteoforms that can be added to existing protein databases (Hernandez et al. 2014). As a result of the merging of genomics and proteomics, a new field known as proteogenomics has emerged (Jaffe et al. 2004; Nesvizhskii et al. 2014; Low et al. 2016; Sheynkman et al. 2016; Ruggles et al. 2017). The expression of a gene, for example, can be determined at the level of mRNAs and proteins in each allelic form using proteogenomics (Wingo et al. 2017). Exon-exon splice junctions, on the other hand, allow for the analysis of alternatively spliced proteomes. Moreover, proteogenomics has been increasingly used to understand the adaptive diversification of plant species and populations (Voelckel et al. 2017).

Tens of studies have been completed on proteomic analysis of various plant species under different developmental stages, abiotic or biotic stress conditions, at different tissues, organs, and cells (Reviewed by Tan et al. 2017; Mustafa and Komatsu et al. 2021; Smythers et al. 2021). Recently the Arabidopsis PeptideAtlas (www.peptideatlas.org/builds/arabidopsis/) was released to solve critical questions about the *Arabidopsis thaliana* proteome (van Wijk et al. 2021). It includes around 0.5 million unique peptides and 17,858 unique proteins at the highest confidence level.

## 2.15 Metabolomics

Metabolomics is the large-scale study of small molecules, also known as metabolites, in cells, biofluids, tissues, or organisms. The metabolome refers to these small molecules and their interactions within a biological system. Metabolomics is a powerful approach because, unlike other omics approaches, metabolites and their concentrations directly reflect the underlying biochemical activity and state of cells and tissues. As a result, metabolomics is the most accurate representation

of the phenotype. Advancements in chromatographic separation and MS allowed for unbiased, high-throughput screening and characterization of the metabolites to study the metabolic pathways and phytochemicals to complement the other omics approaches (Lee et al. 2012; Kang et al. 2013; Kin et al. 2013; Lee et al. 2015). Because of the metabolome complexity, functional characterization of metabolites is a challenging strategy in plants (Chen et al. 2013; Lee et al. 2019). Moreover, plants within the same family generally produce the same or similar metabolites since the metabolic pathways are highly conserved in plant families, which make it easier to study the metabolites in the same family in different species (Ntie-Kang et al. 2013). Plant metabolomics studies can explain the spatio-temporal differences of some essential metabolites in different plant species, which are affected by environmental factors together with genetic determinants (Lee and Lee et al. 2015; Son et al. 2016). In general, genetic factors, nutritional status, and geo-climatic conditions all influence the chemical composition of different plant parts (Dias et al. 2016).

Currently, MS or nuclear magnetic resonance (NMR) spectrometry is used in many metabolomics studies. Some studies use gas chromatography (GC)-MS for the separation and analysis of volatile compounds. However, studying all metabolites is a big challenge since the combination of multiple metabolomics methods is required for this purpose. Many metabolomics studies have been completed in different crop species (Reviewed by Kumar et al. 2017; Sharma et al. 2018; Fernandez et al. 2020). Recent efforts in plant metabolomics science focus on natural variations of metabolites (Reviewed by Sun et al. 2021). These efforts determined the type of natural variations reflecting the metabolomics changes in a given plant family or taxon (Hu et al. 2014; Kusano et al. 2015; Albrecht et al. 2016; Zhen et al. 2016; Yang et al. 2018a, b; Fang et al. 2019). Later, these natural variations were used to select for the genotypes with superior metabolic profiles (e.g., Zhen et al. 2016) and link a specific metabolite or metabolic pathway to a genomic region via the identification of metabolite QTLs (mQTLs) (Chen et al. 2018a, b; Shi et al. 2020; Jamaloddin et al. 2021) or metabolome-based GWAS (mGWAS) (Luo 2015; Fang and Lou 2019; Chen et al. 2020; Wei et al. 2021).

Similar studies have recently been employed in the determination of ionic changes in different plant species (Yang et al. 2018a, b; Pita-Barbosa et al. 2019; Ali et al. 2021; Singh et al. 2022). Comparative metabolomics and ionomics studies revealed the evolutionary divergence of metabolic pathways and how they are conserved in some species or genotype for enhancing the adaptation to a specific condition (Dos Santos et al. 2017; Mawalagedera et al. 2019; Deng et al. 2020; Rastogi et al. 2020). We are now at the beginning of a new phase in plant metabolism research, in which integrative genomics and metabolomics approaches are used (Rai et al. 2017). The supremacy of genomics and transcriptomics should be integrated with metabolomics and proteomics studies to identify novel genes controlling the metabolism.

## 2.16 Multi-omics

Transcriptomics, proteomics, and metabolomics studies can represent the overall changes in transcripts, proteins, and metabolites, respectively (Aizat et al. 2018); however, a more diverse overall approach is needed to combine and compare large data sets to understand the complex biological systems such as the interactome. Multi-omics data generation and acquisition have become an essential part of modern molecular biology and biotechnology to study the biological pathways under different conditions because of recent advancements in NGS, proteomics, and metabolomics technologies as well as computational and statistical tools (Fondi and Liò 2015; Fabregat et al. 2018). Advancements in systems biology, the computational and mathematical analysis, and modeling of complex biological systems led to a more accurate understanding of complex biological systems. Systematic multi-omics integration (MOI) is essential for systems biology in plant science. MOI in plants has been a difficult task since the genomes of many non-model plant species are not well-annotated, the metabolic processes are diverse, and the interactome is massive (Jamil et al. 2020).

The earliest examples of MOI studies were very successful to demonstrate the power of the integrative omics approach to identify potential candidate genes, proteins, or metabolites for further functional characterization. For example, correlation analysis of transcriptomic and metabolomic data from the potato tubes led to the identification of novel transcript-metabolite pairs that can be further characterized in the future (Urbanczyk-Wochniak et al. 2003). In another study, transcriptomic and metabolomic data were integrated to understand the interactions of sulfur and nitrogen metabolisms and the involvement of secondary metabolites in *Arabidopsis thaliana* (Hirai et al. 2004). Since then, the MOI has been extensively used by plant scientists for functional characterization of unknown genes and to understand the behavior of complex systems under different conditions. Several different online software have been developed to integrate multi-omics data, such as MapMan (Thimm et al. 2004), and reviewed by Fondi and Lio (2015). The systems biology approach has been integrated extensively in different plant species (Rai et al. 2019).

In contrast to these advancements, some hurdles have slowed the utilization of the systems biology approach, particularly in crop species. These include the incomplete transcriptome, proteome, and metabolome data sets or their total unavailability. Current software is not designed to integrate different omics data sets to describe the phenome. Machine learning and artificial intelligence should yet to be incorporated into this software. The metabolome or ionome can be easily influenced by the environmental changes so that the extensive data generated by metabolomics and ionomics may not be readily integrated with the date of transcriptomics and genomics. Therefore, the results obtained at the levels of transcriptome and genome may not be fully reflected at the metabolome or even phenotype (do Amaral and Souza 2017). Therefore, there are lots of complex and dynamic processes working in parallel in the cell.

## 2.17 Single-Cell Technologies

The sequencing of a single-cell genome or transcriptome to obtain genomic, transcriptome, or other multi-omics information to show cell population distinctions and cellular evolutionary relationships is referred to as single-cell sequencing technologies (Wen et al. 2018). The molecular insight into tissue and/or time point/developmental groupings using bulk techniques, which average over many cells, has been gained. However, the inherent biases introduced by averaging over different cell populations limit these approaches. Bulk averaging can, in some cases, lead to qualitatively wrong conclusions, a phenomenon known as Simpson's dilemma (Trapnell et al. 2015). Single-cell technologies have the advantages of detecting heterogeneity among individual cells, distinguishing a small number of cells, and outlining cell maps when compared to standard sequencing technology (Wen et al. 2018). Single-cell genomic approaches offer a potent set of tools for identifying cellular heterogeneity, as well as the formation and differentiation of cell types in complex tissues.

Due to its expensive cost, early single-cell sequencing was not widely used (Wen et al. 2018). High-throughput single-cell transcriptomics has become an accessible and powerful tool for unbiased profiling of complex and heterogeneous systems, thanks to recent improvements in cost and throughput (Klein et al. 2015; Zilionis et al. 2017; Macosko et al. 2015) and the availability of fully commercialized workflows (Zheng et al. 2017). These data sets can be utilized in concert with novel computational approaches to uncover cell types and states (Shekhar et al. 2016; Villani et al. 2017a, b), recreate developmental pathways, make destiny decisions (Trapnell et al. 2014; Welch et al. 2016), and spatially model complex tissues (Satija et al. 2015; Achim et al. 2015).

The emergence of omics techniques has quickly revolutionized our perspectives on plant biology, thanks to the advancement of sequencing technologies. The cellular diversity inside a tissue or organism, on the other hand, is far more complex than can be assessed using bulk analysis, which can only produce population-averaged results (Gawad et al. 2016). As sequencing technologies advanced, allowing smaller and smaller samples, eventually allowing single-cell analysis, the traditional consensus from bulk-based omics studies was questioned (Shapiro et al. 2013). Characterizing the single-cell genome is of significant interest because each cell undergoes a unique chain of DNA synthesis and damage repair events. In a single-cell sequencing-based investigation, there are numerous basic processes. The first step is the preparation of a cell lysate. Plant cell isolation and lysis, unlike animal models, are hampered by the natural cell wall, requiring the use of specific techniques. Single-cell whole-genome amplification (WGA) must be performed once the plant cell lysate has been generated. Single-cell genomics and epigenomic technologies are both based on single-cell WGA; however, single-cell epigenomics is more diverse due to the addition of sample preprocessing procedures for capturing various epigenomic features such as bisulfite conversion for DNA methylation (Smallwood et al. 2014) and proximity DNA ligation for chromatin conformation

(Nagano et al. 2013). Single-cell sequencing technologies have been used to investigate the cell heterogeneity that underlies several bulk omics characteristics, such as genomic variation, DNA methylation, and chromatin accessibility, in a variety of animal models (Huang et al. 2015; Kelsey et al. 2017). In recent years, they have been advanced greatly in terms of sensitivity and throughput. These developments have made it possible to profile cell-specific genomic variants and epigenomic characteristics in plant models for the first time, and they hold a great promise for answering a wide range of plant biological problems at the single-cell level (Stuart and Satija et al. 2019). Recently, multiple experimental protocols, including the Assay for Transposase-Accessible Chromatin with High-Throughput Sequencing (ATAC-seq) (Buenrostro et al. 2015), single-cell combinatorial indexing ATAC-seq (sci-ATAC-seq) (Cusanovich et al. 2015), single-cell transposome hypersensitivity site sequencing (scTHS-seq) (Lake et al. 2018), plate-based scATAC-seq protocol (Chen et al. 2018a, b), and droplet-based single-cell combinatorial indexing ATAC-seq (dsci-ATAC-seq) (Lareau et al. 2019), have been developed to profile genome-wide chromatin accessibility in single cells. Very recently, the use of single-nucleus RNA sequencing (sNucRNA-seq) and single-nucleus assay for transposase-accessible chromatin sequencing (sNucATAC-seq) technologies on *Arabidopsis* roots was reported, and it was suggested that the differential chromatin accessibility is a critical mechanism to regulate gene activity at the cell-type level (Farmer et al. 2021). Furthermore, single-cell resolution maps of open chromatin in the *Arabidopsis* root to address the issue of tissue heterogeneity and to detect likely endoreduplication events were provided by single-cell ATAC-seq (Dorrity et al. 2021).

## 2.18 Single-Cell Transcriptomics

Differential gene expression is largely responsible for the development of multiple cell types and cell-specific functions in multicellular organisms. The transcriptome of individual cells is frequently profiled using single-cell RNA sequencing (scRNA-seq). scRNA-seq (single-cell RNA sequencing) is a next-generation sequencing technology that generates gene expression data from thousands of single cells. This large data collection can be used to answer questions like how many different cell kinds are present in a sample and how common each cell type is. The recent development of single-cell RNA sequencing (scRNA-seq) has deepened our understanding of the cell as a functional unit, revealing new populations of cells with distinct gene expression profiles previously hidden within gene expression analyses performed on bulk cells and providing new insights based on gene expression profiles of hundreds to hundreds of thousands of individual cells (Ziegenhain et al. 2017; Macosko et al. 2015).

Single-cell RNA sequencing has been particularly useful in gaining insight into tissue cellular heterogeneity and identifying previously unknown cell types (Artegiani et al. 2017; Villani et al. 2017a, b; Glass et al. 2017). Single-cell technologies can also be used to identify subpopulations within a known cell type by

looking for differences in gene expression patterns within the cell population (Artegiani et al. 2017; Shalet and Satija et al. 2013). Furthermore, these technologies can effectively isolate the signal from rare cell populations, which would otherwise be lost in the output of RNA sequencing on a bulk cell population (Shalet and Satija et al. 2014; Grün et al. 2015; Mahata et al. 2014; Torre et al. 2018). Besides that, the technology can be used to infer potentially useful markers for cell types that lack known markers, such as cell surface proteins. Because single-cell sequencing is driven by cell clustering based on differentially expressed genes, the genes that drive the clustering can be studied as potential unique markers for the cell population of interest (Artegiani et al. 2017; Zhao and Gao et al. 2017). Finally, single-cell sequencing can be used to investigate cell lineage and differentiation regulation. A population of stem cells, for example, can be induced to differentiate, and single-cell sequencing at various time points can provide "snapshots" of the differentiation process. The trajectories that cell flows to reach each terminally differentiated state, as well as the key genes that are differentially regulated at each branch point, can then be inferred using these snapshots (Artegiani et al. 2017; Treutlein et al. 2014; Trapnell et al. 2014; Qiu et al. 2017).

Biological tissue samples are frequently used as an input material for single-cell experiments. In the first phase, a single-cell suspension is created by digesting the tissue in a process known as single-cell dissociation. Cells must be isolated to profile the mRNA in each one separately. Depending on the experimental protocol, single-cell isolation is done differently. Droplet-based approaches focus on catching each cell in its microfluidic droplet, whereas plate-based methods separate cells into wells on a plate. Multiple cells can be captured together (doublets or multiplets), non-viable cells can be captured, or no cell can be captured at all (empty droplets/wells) in both circumstances. Droplet-based approaches rely on a low concentration flow of input cells to manage doublet rates; hence empty droplets are typical. Each well or droplet includes the chemicals required to break down cell membranes and perform library construct. The process of capturing intracellular mRNA, reverse-transcribed to cDNA molecules and amplified, is known as library construction. The mRNA from each cell can be labeled with a well- or droplet-specific cellular barcode, while the cells go through this process in isolation. Moreover, captured molecules are labeled with a unique molecular identifier (UMI) in many experimental protocols. To enhance the probability of being measured, cellular cDNA is amplified before sequencing.

Cellular cDNA libraries are labeled with cellular barcodes and, depending on the protocol, UMIs after library formation. For sequencing, these libraries are pooled together (multiplexed). Read data is generated by sequencing and is subjected to quality control, grouping based on assigned barcodes (demultiplexing), and alignment in reading processing pipelines. Read data can be further demultiplexed for UMI-based methods to produce counts of captured mRNA molecules (count data). However, analyzing and utilizing the large amounts of data created by single-cell RNA sequencing research is difficult and requires knowledge of the experimental and computational pathways that go from the preparation of input cells to the production of interpretable data. Single-cell gene expression analysis was previously

limited to a few select transcripts from a few individual cells. Modern single-cell sequencing platforms like as Fluidigm C1, Drop-Seq, Chromium 10X, SCI-Seq, and many others have been developed during the past decade thanks to high-throughput sequencing and high-yield cell separation approaches. At any given time, these technologies can define the transcriptional profile of hundreds to thousands of single cells. All rely on the use of DNA barcodes to label mRNA molecules during reverse transcription and/or later processes, allowing the transcripts to be indexed back to their individual cells of origin. Despite the fact that each technique has its own manner of separating cells and labeling mRNA molecules, they all use the same computational pipelines to represent transcriptional profiles.

Single-cell gene expression analyses have not been widely used in plants to date, owing to the presence of the plant cell wall, which makes it difficult to separate and acquire individual cells. Although there is recognition of the potential benefit of large-scale single-cell transcriptome studies in plants, single-cell gene expression studies in plants have so far been limited to a small number of cells (Lieckfeldt et al. 2008; Brennecke et al. 2013; Efroni et al. 2015; Frank and Scanlon 2015; Efroni and Birnbaum 2016; Libault et al. 2017). Several groups have recently used single-cell transcriptomics to plants with high throughput (Denyer et al. 2019; Efroni et al. 2015; Efroni et al. 2016; Jean-Baptiste et al. 2019; Kubo et al. 2019; Nelms et al. 2019; Ryu et al. 2019; Shulse et al. 2019; Zhang et al. 2019). Plant studies using single-cell RNA-seq have primarily focused on the well-studied and understood *Arabidopsis* root system (Denyer et al. 2019; Jean-Baptiste et al. 2019; Ryu et al. 2019; Shulse et al. 2019; Zhang et al. 2019). The *Arabidopsis* root, in particular, is a useful organ for scRNA-seq because it has a relatively small number of cells and cell types, and methods for isolating individual cells by protoplasting are available (Birnbaum et al. 2005; Bruex et al. 2012; Efroni et al. 2015; Li et al. 2016). Even in this highly tractable and well-understood system with many known marker genes and cell types, these landmark studies revealed a slew of new and more robust cell-type marker genes and begun to characterize the transition states that give rise to developmental trajectories (Denyer et al. 2019; Jean-Baptiste et al. 2019; Ryu et al. 2019; Shulse et al. 2019; Zhang et al. 2019). Recently, Qing et al. (2020) performed the scRNA-seq on root tips of two agronomically important rice cultivars and identified more than 20,000 single cells. Using integration analysis of two cultivars, most of the major cell types were identified, and novel cell-type-specific marker genes for both cultivars were characterized. In addition, they found well-conserved cell types between the two rice cultivars associated with specific regulatory programs, including phytohormone signaling, biosynthesis, and response. To identify the effects of tissue heterogeneity, Dorrity et al. (2021) applied scATAC-seq and scRNA-seq to *Arabidopsis* roots separately. They identified thousands of differentially accessible sites using scATAC-seq results and the entirety of a cell's regulatory landscape and its transcriptome using scRNA-seq. To define the endoreduplication, cell division, and developmental progression, they integrated the scATAC-seq and scRNA-seq data and characterized cell type-specific motif enrichments of transcription factor family analysis and linked the expression of family members to changing accessibility at specific loci, resolving direct and indirect effects that shape expression (Dorrity et al. 2021).

## 2.19 Conclusion

The omics technologies have been tremendously developed since their first-time introduction in plant science, which was flowed by exponential studies in different plant taxa. At present, at least one research study on plant omics is published every day. These omics studies have generated extensive data such that the pace of software development to analyze this much of data cannot meet the demand. In the past, individual studies of genomics, transcriptomics, or metabolomics were enough to make a judgment about the plant species or genotypes. However, now the focus has shifted from generating high-throughput biological data sets to the integration of these data sets to derive biological meaning out of it. These data sets are valuable for future efforts in establishing models that can describe plant adaptation, cultivation, and production. Novel approaches such as artificial intelligence and machine learning will be required in the near future to get the most out of these data sets and predict the future scenarios, especially under ongoing climate crises.

In the near future, plant biologists will focus on understanding the interactome of different metabolisms in the plant and how these interactions are affecting the phenome. They will utilize integrated omics technologies together with genome editing and speed breeding. Identification of novel genomic, proteomic, or metabolomic markers will be very useful in screening different plant genotypes, wild relatives, or breeding lines to find and develop new cultivars highly adapted to the changing climate with higher yields and nutritional quality.

## References

Achim K, Pettit JB, Saraiva L et al (2015) High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nat Biotechnol 33:503–509

Aizat WM, Goh HH, Baharum SN (2018) Omics Applications for Systems Biology Cham. Springer, Switzerland

Albrecht U, Fiehn O, Bowman KD (2016) Metabolic variations in different citrus rootstock cultivars associated with different responses to Huanglongbing. Plant Physiol Biochem 107:33–44

Ali S, Tyagi A, Bae H (2021) Ionomic Approaches for Discovery of Novel Stress-Resilient Genes in Plants. Int J Mol Sci 22(13):7182

Anastasiadi D, Esteve-Codina A, Piferrer F (2018) Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. Epigenetics Chromatin 11:37

Ando K, Carr KM, Grumet R (2012) Transcriptome analyses of early cucumber fruit growth identifies distinct gene modules associated with phases of development. BMC Genomics 13:518

Artegiani B, Lyubimova A, Muraro M, van Es JH, van Oudenaarden A, Clevers H (2017) A single-cell RNA sequencing study reveals cellular and molecular dynamics of the hippocampal neurogenic niche. Cell Rep 21:3271–3284

Bayer PE, Golicz AA, Scheben A, Batley J, Edwards D (2020) Plant pan-genomes are the new reference. Nat Plant 6(8):914–920

Bayley H (2015) Nanopore sequencing: From imagination to reality. Clin Chem 61:25–31

Birnbaum K, Jung JW, Wang JY, Lambert GM, Hirst JA, Galbraith DW, Benfey PN (2005) Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. Nat Methods 2:615–619

Bonnet E, Wuyts J, Rouzé P, Van de Peer Y (2004) Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. Proc Natl Acad Sci U S A 101:11511–11516

Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X (2010) The potential and challenges of nanopore sequencing. In: Nanoscience and technology: a collection of reviews from nature journals. World Scientific Singapore, pp 261–268

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D et al (2012) Analysis of the bread wheat genome using wholegenome shotgun sequencing. Nature 491:705–710

Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, Heisler MG (2013) Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods 10:1093–1095

Bruex A, Kainkaryam RM, Wieckowski Y, Kang YH, Bernhardt C, Xia Y, Zheng X, Wang JY, Lee MM, Benfey P et al (2012) A gene regulatory network for root epidermis cell differentiation in Arabidopsis. PLoS Genet 8:e1002446

Buck MJ, Lieb JD (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics 84:349–360

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP et al (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523:486–490

Celton JM, Gaillard S, Bruneau M, Pelletier S, Aubourg S, Martin-Magniette ML, Navarro L, Laurens F, Renou JP (2014) Widespread anti-sense transcription in apple is correlated with siRNA production and indicates a large potential for transcriptional and/or posttranscriptional control. New Phytol 203:287–299

Chaudhary J, Khatri P, Singla P, Kumawat S, Kumari A, Vikram A, Jindal SK, Kardile H, Kumar R, Sonah H, Deshmukh R (2019) Advances in omics approaches for abiotic stress tolerance in tomato. Biology 8(4):90

Chen W, Gong L, Guo Z, Wang W, Zhang H, Liu X, Yu S, Xiong L, Luo J (2013) A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. Mol Plant 6:1769–1780

Chen J, Wang J, Chen W, Sun W, Peng M, Yuan Z, Shen S, Xie K, Jin C, Sun Y, Liu X (2018a) Metabolome analysis of multi-connected biparental chromosome segment substitution line populations. Plant Physiol 178(2):612–625

Chen X, Miragaia RJ, Natarajan KN, Teichmann SA (2018b) A rapid and robust method for single cell chromatin accessibility profiling. Nat Commun 9:5345

Chen J, Hu X, Shi T, Yin H, Sun D, Hao Y, Xia X, Luo J, Fernie AR, He Z, Chen W (2020) Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. Plant Biotechnol J 18(8):1722–1735

Crisp PA, Ganguly D, Eichten SR, Borevitz JO, Pogson BJ (2016) Reconsidering plant memory: intersections between stress recovery, RNA turnover, and epigenetics. Sci Adv 2:e1501340

Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL et al (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348:910–914

Daron J, Slotkin RK (2017) EpiTEome: simultaneous detection of transposable element insertion sites and their DNA methylation levels. Genome Biol 18:91

Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM (2009) Shedding light on an extremophile lifestyle through transcriptomics. New Phytol 183:764–775

Deng M, Zhang X, Luo J, Liu H, Wen W, Luo H, Yan J, Xiao Y (2020) Metabolomics analysis reveals differences in evolution between maize and rice. Plant J 103(5):1710–1722

Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MCP (2019) Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. Dev Cell 48:840–852

Dias MI, Sousa MJ, Alves RC, Ferreira IC (2016) Exploring plant tissue culture to improve the production of phenolic compounds: a review. Ind Crop Prod 82:9–22

Ding H et al (2014) Heterosis in early maize ear inflorescence development, a genome–wide transcription analysis for two maize inbred lines and their hybrid. Int J Mol Sci 15:13892–13915

do Amaral MN, Souza GM (2017) The challenge to translate OMICS data to whole plant physiology: the context matters. Front Plant Sci 8:2146–2146

Dorrity MW, Alexandre CM, Hamm MO et al (2021) The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *Nat Commun* 12:3334

Dos Santos VS, Macedo FA, Do Vale JS, Silva DB, Carollo CA (2017) Metabolomics as a tool for understanding the evolution of Tabebuia sensu lato. Metabolomics 13(6):72

Dubin MJ, Zhang P, Meng D, Remigereau MS, Osborne EJ, Paolo Casal F et al (2015) DNA methylation in arabidopsis has a genetic basis and shows evidence of local adaptation. Elife 4:05255

Efroni I, Birnbaum KD (2016) The potential of single-cell profiling in plants. Genome Biol 17:65

Efroni I, Ip PL, Nawy T, Mello A, Birnbaum KD (2015) Quantification of cell identity from single-cell gene expression profiles. Genome Biol 16:9

Efroni I, Mello A, Nawy T, Ip PL, Rahni R, Del Rose N, Powers A, Satija R, Birnbaum KD (2016) Root regeneration triggers an embryo-like sequence guided by hormonal interactions. Cell 165:1721–1733

El Ouakfaoui S, Miki B (2005) The stability of the Arabidopsis transcriptome in transgenic plants expressing the marker genes nptII and uidA. Plant J 41(6):91–800

Fabregat A, Jupe A, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B et al (2018) The Reactome Pathway Knowledgebase. Nucleic Acids Res 46:649–655

Fan XD, Wang JQ, Yang N, Dong YY, Liu L, Wang FW, Wang N, Chen H, Liu WC, Sun YP, Wu JY, Li HY (2012) Gene expression profiling of soybean leaves and roots under salt, saline–alkali and drought stress by high-throughput Illumina sequencing. Gene 512:392–402

Fang C, Luo J (2019) Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. Plant J 97(1):91–100

Fang C, Fernie AR, Luo J (2019) Exploring the diversity of plant metabolism. Trends Plant Sci 24(1):83–98

Farmer A, Thibivilliers S, Ryu KH, Schiefelbein J, Libault M (2021) Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis roots at the single-cell level. Mol Plant 14(3):372–383

Feng J, Liang Y, Wang F, Chen J (2013) Detection of genetically modified tomato using PCR coupled with ìParaflo™ microfluidics microarrays. J Nanosci Nanotechnol 13:8266–8274

Fernandez O, Millet EJ, Rincent R, Prigent S, Pétriacq P, Gibon Y (2020) Plant metabolomics and breeding. In: Advances in botanical research. Academic Press Inc, pp 207–235

Fondi M, Liò P (2015) Multi-omics and metabolic modelling pipelines: challenges and tools for systems microbiology. Microbiol Res 171:52–64. https://doi.org/10.1016/j.micres.2015.01.003

Fortes AM, Gallusci P (2017) Plant stress responses and phenotypic plasticity in the epigenomics era: perspectives on the grapevine scenario, a model for perennial crop plants. Front Plant Sci 8:82

Frank MH, Scanlon MJ (2015) Cell-specific transcriptomic analyses of three-dimensional shoot development in the moss Physcomitrella patens. Plant J 83:743–751

Gaines TA, Lorentz L, Figge A, Herrmann J, Maiwald F, Ott MC, Han H, Busi R, Yu Q, Powles SB, Beffa R (2014) RNA-Seq transcriptome analysis to identify genes involved in metabolism-based diclofop resistance in *Lolium rigidum*. Plant J 78:865–876

Gallusci P, Dai Z, Génard M, Gauffretau A, Leblanc-Fournier N, Richard-Molar C et al (2017) Epigenetics for plant improvement: current knowledge and modeling avenues. Trends Plant Sci 22:610–623

Gawad C et al (2016) Single-cell genome sequencing: current state of the science. Nat Rev Genet 17:175–188
Glass LL, Calero-Nieto FJ, Jawaid W, Larraufie P, Kay RG, Göttgens B, Reimann F, Gribble FM (2017) Single-cell RNA-sequencing reveals a distinct population of proglucagon-expressing cells specific to the mouse upper small intestine. Mol Metab 6:1296–1303
Glenn TC (2011) Field guide to next-generation DNA sequencers. Mol Ecol Resour 11:759–769
Goff SA, Ricke D, Lan TH, Presting G, Wang R et al (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100
Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, van Oudenaarden A (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525:251–255
Gulik PJ, Drouin S, Yu Z et al (2005) Transcriptome comparison of winter and spring wheat responding to low temperature. Genome 48(5):913–923
Hansey CN et al (2012) Maize (*Zea mays* L.) genome diversity as revealed by RNA–sequencing. PLoS One 7:e33071
Hashmi U, Shafqat S, Khan F, Majid M, Hussain H, Kazi AG, John R, Ahmad P (2015) Plant exomics: concepts, applications and methodologies in crop improvement. Plant Signal Behav 10(1):e976152. https://doi.org/10.4161/15592324.2014.976152
Hauben M, Haesendonckx B, Standaert E, Van Der Kelen K, Azmi A, Akpo H et al (2009) Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. Proc Natl Acad Sci U S A 106:20109–20114
Heather JM (2015) T-cell receptor repertoire sequencing in health and disease. Doctoral dissertation, UCL (University College London)
Hernandez C, Waridel P, Quadroni M (2014) Database construction and peptide identification strategies for proteogenomic studies on sequenced genomes. Curr Top Med Chem 14:425–434
Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proc Natl Acad Sci 101(27):10205–10210
Horvath H, Laenen B, Takuno S, Slotte T (2019) Single-cell expression noise and gene-body methylation in *Arabidopsis thaliana*. Heredity 123:81–91
Hu C, Shi J, Quan S, Cui B, Kleessen S, Nikoloski Z, Tohge T, Alexander D, Guo L, Lin H, Wang J (2014) Metabolic variation between japonica and indica rice cultivars as revealed by non-targeted metabolomics. Sci Rep 4(1):1–10
Huang L et al (2015) Single-cell whole-genome amplification and sequencing: methodology and applications. Annu Rev Genomics Hum Genet 16:79–102
Jaffe JD, Berg HC, Church GM (2004) Proteogenomic mapping as a complementary method to perform genome annotation. Proteomics 4:59–77
Jamaloddin M, Maliha A, Gokulan CG, Gaur N, Patel HK (2021) Metabolomics-assisted breeding for crop improvement: an emerging approach. In: Omics technologies for sustainable agriculture and global food security, vol 1. Springer, Singapore, pp 241–279
Jamil IN, Remali J, Azizan KA, Muhammad NAN, Arita M, Goh H-H, Aizat WM (2020) Systematic multi-omics integration (MOI) approach in plant systems biology. Front Plant Sci 11:944
Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus JT (2019) Dynamics of gene expression in single root cells of A. thaliana. Plant Cell 31:993–1011
Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X et al (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95
Juntawong P, Sirikhachornkit A, Pimjan R, Sonthirod C, Sangsrakru D, Yoocha T, Tangphatsornruang S, Srinives P (2014) Elucidation of the molecular responses to waterlogging in Jatropha roots by transcriptome profiling. Front Plant Sci 5:658
Jurinke C, Oeth P, van den Boom D (2004) MALDI-TOF mass spectrometry. Mol Biotechnol 26(2):147–163

Kang D, Son GH, Park HM, Kim J, Choi JN, Kim HY et al (2013) Culture condition dependent metabolite profiling of Aspergillus fumigatus with antifungal activity. Fungal Biol 117:211–219
Kang Y, Kang CS, Kim C (2019) History of nucleotide sequencing technologies: advances in exploring nucleotide sequences from Mendel to the 21st century. Hortic Sci Technol 37:549–558
Kasajima I, Fujiwara T (2007) Identification of novel Arabidopsis thaliana genes which are induced by high levels of boron. Plant Biotechnol 24(3):355–360
Kass SU, Pruss D, Wolffe AP (1997) How does DNA methylation repress transcription? Trends Genet 13:444–449
Kayıhan C, Eyidoğan F (2019) Omics in oxidative stress tolerance in crops. In: Hasanuzzaman M, Fotopoulos V, Nahar K, Fujita M (eds) Reactive oxygen, nitrogen and sulfur species in plants: production, metabolism, signaling and defense mechanisms. Wiley-Blackwell, United States of America
Kayihan C, Öz MT, Eyidoğan F et al (2017) Physiological, biochemical, and transcriptomic responses to boron toxicity in leaf and root tissues of contrasting wheat cultivars. Plant Mol Biol Repr 35(1):97–109
Kchouk M, Gibrat J, Elloumi M (2017) Generations of sequencing technologies: from first to next generation. Biol Med 9:1–8
Kelsey G et al (2017) Single-cell epigenomics: recording the past and predicting the future. Science 358:69–75
Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression microarray data. J Comput Biol 7:819–837
Kim JH, Kim SY, Lee H, Kim YR, Kim HY (2010) An event-specific DNA microarray to identify genetically modified organisms in processed foods. J Agric Food Chem 58:6018–6026
Kim J, Choi JN, Choi JH, Cha YS, Muthaiya MJ, Lee CH (2013) Effect of fermented soybean product (Cheonggukjang) intake on metabolic parameters in mice fed a high-fat diet. Mol Nutr Food Res 57:1886–1891
Klein AM et al (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161:1187–1201
Köhler C, Springer N (2017) Plant epigenomics-eciphering the mechanisms of epigenetic inheritance and plasticity in plants. Genome Biol 18(1). https://doi.org/10.1186/s13059-017-1260-9
Koren S, Harhay GP, Smith TP et al (2013) Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol 14:R101
Krishna BM, Khan MA, Khan ST (2019) Next-generation sequencing (NGS) platforms: an exciting era of genome sequence analysis. Microbial Genomics in Sustainable Agroecosystems, pp 89–109
Kubo M, Nishiyama T, Tamada Y, Sano R, Ishikawa M, Murata T, Imai A, Lang D, Demura T, Reski R et al (2019) Single-cell transcriptome analysis of Physcomitrella leaf cells during reprogramming using microcapillary manipulation. Nucleic Acids Res 47:4539–4553
Kulski JK (2016) Next-generation sequencing — an overview of the history, tools, and "omic" applications. Next generation sequencing - advances, applications and challenges. InTech
Kumar S (2018) Epigenomics of plant responses to environmental stress. Epigenomes 2:1–17
Kumar R, Bohra A, Pandey AK, Pandey MK, Kumar A (2017) Metabolomics for plant improvement: status and prospects. Front Plant Sci 8:1302
Kusano M, Yang Z, Okazaki Y, Nakabayashi R, Fukushima A, Saito K (2015) Using metabolomic approaches to explore chemical diversity in rice. Mol Plant 8(1):58–67
Lahens NF, Ricciotti E, Smirnova O, Toorens E, Kim EJ, Baruzzo G, Hayer KE, Ganguly T, Schug J, Grant GR (2017) A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. BMC Genomics 18(1):602
Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC et al (2018) Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol 36:70–80
Lareau CA, Duarte FM, Chew JG, Kartha VK, Burkett ZD, Kohlway AS et al (2019) Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat Biotechnol 37:916–924

Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS (2011) Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. DNA Res 18:263–276

Lee S, Do SG, Kim SY, Kim J, Jin Y, Lee CH (2012) Mass spectrometry-based metabolite profiling and antioxidant activity of Aloe vera (Aloe barbadensis Miller) in different growth stages. J Agric Food Chem 60:11222–11228

Lee S, Oh DG, Lee S, Kim GR, Lee JS, Son YK et al (2015) Chemotaxonomic metabolite profiling of 62 indigenous plant species and its correlation with bioactivities. Molecules 20:19719–19734

Lee S, Oh DG, Singh D, Lee HJ, KIM GR, Lee S, Lee JS, Lee CH (2019) Untargeted metabolomics toward systematic characterization of antioxidant compounds in Betulaceae family plant extracts. Metabolites 9:186

Leimanis S, Hernandez M, Fernandez S, Boyer F, Burns M, Bruderer S et al (2006) A microarray-based detection system for genetically modified (GM) food ingredients. Plant Mol Biol 61:123–139

Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang SS (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol 32(10):1045–1052

Li S, Yamada M, Han X, Ohler U, Benfey PN (2016) High-resolution expression map of the Arabidopsis root reveals alternative splicing and lincRNA regulation. Dev Cell 39:508–522

Libault M, Pingault L, Zogli P, Schiefelbein J (2017) Plant systems biology at the single-cell level. Trends Plant Sci 22:949–960

Lieckfeldt E, Simon-Rosin U, Kose F, Zoeller D, Schliep M, Fisahn J (2008) Gene expression profiling of single epidermal, basal and trichome cells of Arabidopsis thaliana. J Plant Physiol 165:1530–1544

Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y et al (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87–90

Liu Q, Liang Z, Feng D, Jiang S, Wang Y, Du Z, Li R, Hu G, Zhang P, Ma Y et al (2020) Transcriptional landscape of rice roots at the single cell resolution. Mol Plant 14(3):384–394

Loman NJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30:434–439

Low TY, Heck AJ (2016) Reconciling proteomics with next generation sequencing. Curr Opin Chem Biol 30:14–20

Low TY, Mohtar MA, Ang MY, Jamal R (2019) Connecting proteomics to next-generation sequencing: proteogenomics and its current applications in biology. Proteomics 19(10):e1800235

Lu T, Cui L, Zhou Y, Zhu Q, Fan D, Gong H, Zhao Q, Zhou C, Zhao Y, Lu D, Luo J, Wang Y, Tian Q, Feng Q, Huang T, Han B (2015) Transcriptome-wide investigation of circular RNAs in rice. RNA 21:2076–2087

Lu H, Giordano F, Ning Z (2016) Oxford nanopore minION sequencing and genome assembly. Genomics Proteomics Bioinformatics 14(5):265–279

Luo J (2015) Metabolite-based genome-wide association studies in plants. Curr Opin Plant Biol 24:31–38

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214

Mahata B, Zhang X, Kolodziejczyk AA, Proserpio V, Haim-Vilmovsky L, Taylor AE, Hebenstreit D, Dingler FA, Moignard V, Göttgens B et al (2014) Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis. Cell Rep 7:1130–1142

Mawalagedera SM, Callahan DL, Gaskett AC, Rønsted N, Symonds MR (2019) Combining evolutionary inference and metabolomics to identify plants with medicinal potential. Front Ecol Evol 7:267

Maxam AM, Gilbert WA (1977) A new method for sequencing DNA. Proc Natl Acad Sci 74:560–564

Meera Krishna B, Khan MA, Khan ST (2019) Next-generation sequencing (NGS) platforms: an exciting era of genome sequence analysis. In: Tripathi V, Kumar P, Tripathi P, Kishore A, Kamle M (eds) Microbial genomics in sustainable agroecosystems. Springer, Singapore

Merriman B, Team ITD, Rothberg JM (2012) Progress in ion torrent semiconductor chip based sequencing. Electrophoresis 33:3397–3417

Michael TP, Van Buren R (2015) Progress, challenges and the future of crop genomes. Curr Opin Plant Biol 24:71–81

Muhammad II, Kong SL, Akmar Abdullah SN, Munusamy U (2019) RNA-seq and ChIP-seq as complementary approaches for comprehension of plant transcriptional regulatory mechanism. Int J Mol Sci 21(1):167

Mustafa G, Komatsu S (2021) Plant proteomic research for improvement of food crop under stresses: a review. Molecul Omics

Nagano T et al (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502:59–64

Nalbantoglu S, Karadag A (2019) Introductory chapter: insight into the OMICS technologies and molecular medicine. https://doi.org/10.5772/intechopen.86450

Nelms B, Walbot V (2019) Defining the developmental program leading to meiosis in maize. Science 364:52–56

Nesvizhskii AI (2014) Proteogenomics: concepts, applications and computational strategies. Nat Methods 11:1114–1125

Nigam D, Kavita P, Tripathi RK, Ranjan A, Goel R, Asif M, Shukla A, Singh G, Rana D, Sawant SV (2014) Transcriptome dynamics during fibre development in contrasting genotypes of *Gossypium hirsutum* L. Plant Biotechnol J 12:20418

Noman MS, Rashid M, Khan TA (2017) Next-generation high-throughput sequencing technologies: concept and applications. J Agric & Rural Res 1(1):15–23

Ntie-Kang F, Lifongo LL, Mbaze LMA, Ekwelle N, Owono LCO (2013) Megnassan E. Cameroonian medicinal plants: a bioactivity versus ethnobotanical survey and chemotaxonomic classification. BMC Complement Altern Med 13:147

Öz MT, Yilmaz R, Eyidogan F et al (2009) Microarray analysis of late response to boron toxicity in barley (Hordeum vulgare L.) leaves. Turkish J Agric Forest 33:191–202

Ozturk ZN, Talame V, Deyholos M et al (2002) Monitoring largescale changes in transcript abundance in drought- and salt stressed barley. Plant Mol Biol 48:551–573

Pita-Barbosa A, Ricachenevsky FK, Flis PM (2019) One "OMICS" to integrate them all: ionomics as a result of plant genetics, physiology and evolution. Theor Exp Plant Physiol 31(1):71–89

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C (2017) Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 14:979–982

Rabilloud T, Chevallet M, Luche S, Lelong C (2010) Two-dimensional gel electrophoresis in proteomics: past, present and future. J Proteomics 73(11):2064–2077

Rahman H, Jagadeeshselvam N, Valarmathi R, Sachin B, Sasikala R, Senthil N, Sudhakar D, Robin S, Muthurajan R (2014) Transcriptome analysis of salinity responsiveness in contrasting genotypes of finger millet (*Eleusine coracana* L.) through RNA-sequencing. Plant Mol Biol 85:485–503

Rai A, Saito K, Yamazaki M (2017) Integrated omics analysis of specialized metabolism in medicinal plants. Plant J 90:764–787

Rai A, Yamazaki M, Saito K (2019) A new era in plant functional genomics. Curr Opin Syst Biol 15:58–67

Ramon D, Braden M, Adams S, Marincola FM, Wang L (2003) Pyrosequencing™: a onestep method for high resolution HLA typing. J Transl Med 1(1):9

Rastogi S, Shah S, Kumar R, Kumar A, Shasany AK (2020) Comparative temporal metabolomics studies to investigate interspecies variation in three Ocimum species. Sci Rep 10(1):1–15

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 13(5):278–289

Rodríguez López CM, Wilkinson MJ (2015) Epi-fingerprinting and epiinterventions for improved crop production and food quality. Front Plant Sci 6:397
Rothberg JM (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348–352
Rothberg JM, Leamon JH (2008) The development and impact of 454 sequencing. Nat Biotechnol 26:1117
Ruggles KV, Krug K, Wang X, Clauser KR, Wang J, Payne SH et al (2017) Methods, tools and current perspectives in proteogenomics. Mol Cell Proteomics 16:959–981
Ryu KH, Huang L, Kang HM, Schiefelbein J (2019) Single-cell RNA sequencing resolves molecular relationships among individual plant cells. Plant Physiol 179:1444–1456
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci 74:5463–5467
Saraswathy N, Ramalingam P (2011) Concepts and techniques in genomics and proteomics. Elsevier, pp 57–76
Sarethy IP, Saharan A (2021) Genomics, proteomics and transcriptomics in the biological control of plant pathogens: a review. Indian Phytopathol 74(1):3–12
Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 33:495–502
Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. Brief Funct Genomics 13:296–307
Schatz MC, Maron LG, Stein JC, Wences AH, Gurtowski J, Biggers E, Lee H, Kramer M, Antoniou E, Ghiban E, Wright MH (2014) Whole genome de novo assemblies of three divergent strains of rice, Oryza sativa, document novel gene space of aus and indica. Genome Biol 15(11):1–16
Schmidt AM, Sahota R, Pope DS, Lawrence TS, Belton MP, Rott ME (2008) Detection of genetically modified canola using multiplex PCR coupled with oligonucleotide microarray hybridization. J Agric Food Chem 56:6791–6800
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T et al (2010) Genome sequence of the paleopolyploid soybean. Nature 463:178–183
Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D et al (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498:236–240
Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, Chen P, Gertner RS, Gaublomme JT, Yosef N et al (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature 510:363–369
Shapiro E et al (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet 14:618–630
Sharma K, Sarma S, Bohra A, Mitra A, Sharma NK, Kumar A (2018) Plant metabolomics: an emerging technology for crop improvement. New visions in plant science, 1st edn. IntechOpen, London, pp 65–79
Shekhar K et al (2016) Comprehensive classification of retinal bipolar neurons by singlecell transcriptomics. Cell 166:1308–1323
Sheynkman GM, Shortreed MR, Cesnik AJ, Smith LM (2016) Proteogenomics: integrating next-generation sequencing and mass spectrometry to characterize human proteomic variation. Annu Rev Anal Chem 9:521–545
Shi T, Zhu A, Jia J, Hu X, Chen J, Liu W, Ren X, Sun D, Fernie AR, Cui F, Chen W (2020) Metabolomics analysis and metabolite-agronomic trait associations using kernels of wheat (Triticum aestivum) recombinant inbred lines. Plant J 103(1):279–292
Shuai P, Liang D, Tang S, Zhang Z, Ye CY, Su Y, Xia X, Yin W (2014) Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. J Exp Bot 65:4975–4983
Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP et al (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43:109–116

Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, Gouran M, Turco GM, Zhu Y, O'Malley RC, Brady SM et al (2019) High-throughput single-cell transcriptome profiling of plant cell types. Cell Rep 27:2241–2247

Sims D et al (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15:121–132

Singh A, Jaiswal A, Singh A, Tomar RS, Kumar A (2022) Plant ionomics: toward high-throughput nutrient profiling. In: Plant nutrition and food security in the era of climate change. Academic Press, p 227254

Smallwood SA et al (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 11:817–820

Smythers AL, Hicks LM (2021) Mapping the plant proteome: tools for surveying coordinating pathways. Emerging Topics in Life Sciences 5(2):203–220

Son SY, Kim NK, Lee S, Singh D, Kim GR, Lee JS et al (2016) Metabolite fingerprinting, pathway analyses, and bioactivity correlations for plant species belonging to the Cornaceae, Fabaceae, and Rosaceae families. Plant Cell Rep 35:917–931

Stoddart D, Heron A, Mikhailova E, Maglia G, Bayley H (2009) Single nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc Natl Acad Sci U S A 106:7702–7707

Stuart T, Satija R (2019) Integrative single-cell analysis. Nat Rev Genet 20:257–272

Sun W, Chen Z, Hong J, Shi J (2021) Promoting human nutrition and health through plant metabolomics: current status and challenges. Biology 10(1):20

Swanson-Wagner RA et al (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci US 103:6805–6810

Tan L, Chen S, Wang T, Dai S (2013) Proteomic insights into seed germination in response to environmental factors. Proteomics 13:1850–1870

Tan BC, Lim YS, Lau SE (2017) Proteomics in commercial crops: an overview. J Proteomics 169:176–188

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci 102(39):13950–13955

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37(6):914–939

Tian M et al (2018) Transcriptomic analysis reveals overdominance playing a critical role in nicotine heterosis in Nicotiana tabacum L. BMC Plant Biol 18:48

Torre E, Dueck H, Shaffer S, Gospocic J, Gupte R, Bonasio R, Kim J, Murray J, Raj A (2018) Rare cell detection by single-cell RNA sequencing as guided by single-molecule RNA FISH. Cell Syst 6:171–179

Touzet P, Riccardi F, Damerval MC, Huet JC, Pernollet JC, Zivy M, de Vienne D (1996) The maize two-dimensional gel protein database: towards an integrated genome analysis program. Theor Appl Genet 93:997–1005

Trapnell C (2015) Defining cell types and states with single-cell genomics. Genome Res 25:1491–1498

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32:381–386

Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38(15):e159

Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 509:371–375

Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. Plant Biotechnol J 7:334–346

Turcatti G, Romieu A, Fedurco M, Tairi AP (2008) A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic Acids Res 36:e25

Türktaş M, Kurtoğlu K, Dorado G, Zhang B, Hernandez P, Ünver T (2015) Sequencing of plant genomes – a review. Turkish J Agric Forest 39(3)

Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Rep 4(10):989–993

Van Bel M, Proost S, Van Neste C, Deforce D, Van de Peer Y, Vandepoele K (2013) TRAPID: an efficient online tool for the functional and comparative analysis of de novo RNASeq transcriptomes. Genome Biol 14:R134

van Wijk KJ, Leppert T, Sun Q, Boguraev SS, Sun Z, Mendoza L, Deutsch EW (2021) The Arabidopsis Peptide Atlas: harnessing worldwide proteomics data to create a comprehensive community proteomics resource. The Plant Cell

Villani AC et al (2017a) Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356:eaah4573

Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S et al (2017b) Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356:eaah4573

Vlk D, Řepková J (2017) Application of next-generation sequencing in plant breeding. Czech J Genet Plant Breed 53:89–96

Voelckel C, Gruenheit N, Lockhart P (2017) Evolutionary transcriptomics and proteomics: insight into plant adaptation. Trends Plant Sci 22(6):462–471

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, Xu J, Chua NH (2014) Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. Genome Res 24:444–453

Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ et al (2015a) CG gene body DNA methylation changes and evolution of duplicated genes in cassava. Proc Natl Acad Sci U S A 112:13729–13734

Wang Y, Dong C, Xue Z, Jin Q, Xu Y (2015b) *De novo* transcriptome sequencing and discovery of genes related to copper tolerance in *Paeonia* ostii. Gene 576:126–135

Wei W, Li S, Wang Y, Wang B, Fan G, Zeng Q, Zhao F, Xu C, Zhang X, Tang T, Feng X (2021) Metabolome-based genome-wide association study provides genetic insights into the natural variation of Foxtail Millet. Frontiers in Plant Science, p 1155

Welch JD, Hartemink AJ, Prins JF (2016) SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. Genome Biol 17:106

Wen L, Tang F (2018) Boosting the power of single-cell analysis. Nat Biotechnol 36(5):408–409

Wilkins KE, Booher NJ, Wang L, Bogdanove AJ (2015) TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. oryzicola while strict conservation suggests universal importance of five TAL effectors. Front Plant Sci 6:536

Wingo TS, Duong DM, Zhou M, Dammer EB, Wu H, Cutler DJ et al (2017) Seyfried. J Proteome Res 16:3336–3347

Xin M, Wang Y, Yao Y, Song N, Hu Z, Qin D, Xie C, Peng H, Ni Z, Sun Q (2011) Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. BMC Plant Biol 11:61

Xu J, Zhu S, Miao H, Huang W, Qiu M, Huang Y et al (2007) Eventspecific detection of seven genetically modified soybean and maizes using multiplex-PCR coupled with oligonucleotide microarray. J Agric Food Chem 55:5575–5579

Xu P, Liu Z, Fan X, Gao J, Zhang X, Zhang X, Shen X (2013a) De novo transcriptome sequencing and comparative analysis of differentially expressed genes in *Gossypium aridum* under salt stress. Gene 525:26–34

Xu L, Wang Y, Zhai L, Xu Y, Wang L, Zhu X, Gong Y, Yu R, Limera C, Liu L (2013b) Genome-wide identification and characterization of cadmium-responsive microRNAs and their target genes in radish (*Raphanus sativus* L.) roots. J Exp Bot 64:4271–4287

Yamamoto N, Takano T, Tanaka K, Ishige T, Terashima S, Endo C, Kurusu T, Yajima S, Yano K, Tada Y (2015) Comprehensive analysis of transcriptome response to salinity stress in the halophytic turf grass *Sporobolus virginicus*. Front Plant Sci 6:241

Yang M, Lu K, Zhao FJ, Xie W, Ramakrishna P, Wang G, Du Q, Liang L, Sun C, Zhao H, Zhang Z (2018a) Genome-wide association studies reveal the genetic basis of ionomic variation in rice. Plant Cell 30(11):2720–2740

Yang X, Wei S, Liu B, Guo D, Zheng B, Feng L, Liu Y, Tomás-Barberán FA, Luo L, Huang D (2018b) A novel integrated non-targeted metabolomic analysis reveals significant metabolite variations between different lettuce (Lactuca sativa. L) varieties. Hortic Res 5(1):1–14

Yates SA, Swain MT, Hegarty MJ, Chernukin I, Lowe M, Allison GG, Ruttink T, Abberton MT, Jenkins G, Skøt L (2014) *De novo* assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. BMC Genomics 15:453

Yong B, Wang X, Xu P, Zheng H, Fei X, Hong Z, Ma Q, Miao Y, Yuan X, Jiang Y, Shao H (2017) Isolation and abiotic stress resistance analyses of a catalase gene from Ipomoea batatas (L.) Lam. Biomed Res Int 2017:6847532

Yu X, Yang J, Li X, Liu X, Sun C, Wu F, He Y (2013) Global analysis of cis-natural antisense transcripts and their heat responsive nat-siRNAs in *Brassica rapa*. BMC Plant Biol 13:208

Zhai R et al (2013) Transcriptome analysis of rice root heterosis by RNA–Seq. BMC Genomics 14:19

Zhang F et al (2011) Genetic analysis and associated SRAP markers for flowering traits of chrysanthemum (Chrysanthemum morifolium). Euphytica 177:15–24

Zhang W, Han Z, Guo Q, Liu Y, Zheng Y, Wu F, Jin W (2014) Identification of maize long non-coding RNAs responsive to drought stress. PLoS One 9:e98958

Zhang TQ, Xu ZG, Shang GD, Wang JW (2019) A single-cell RNA sequencing profiles the developmental landscape of Arabidopsis root. Mol Plant 12:648–660

Zhao X, Gao S, Wu Z, Kajigaya S, Feng X, Liu Q, Townsley DM, Cooper J, Chen J, Keyvanfar K et al (2017) Single-cell RNA-seq reveals a distinct transcriptome signature of aneuploid hematopoietic cells. Blood 130:2762–2773

Zhen S, Zhou J, Deng X, Zhu G, Cao H, Wang Z, Yan Y (2016) Metabolite profiling of the response to high-nitrogen fertilizer during grain development of bread wheat (Triticum aestivum L.). J Cereal Sci 69:85–94

Zheng GXY et al (2017) Massively parallel digital transcriptional profiling of single cells. Nat Commun 8:14049

Zhong S, Fei Z, Chen YR, Vrebalov J, Mcquinn R, Gapper N, Giovannoni J (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat Biotechnol 31:154–159

Zhou PP, Zhang JZ, You YH, Wu YN (2008) Detection of genetically modified crops by combination of multiplex PCR and low-density DNA microarray. Biomed Environ Sci 21:53–62

Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W (2017) Comparative analysis of single-cell RNA sequencing methods. Mol Cell 65:631–643

Zilionis R et al (2017) Single-cell barcoding and sequencing using droplet microfluidics. Nat Protoc 12:44–73