


# From single- to multi-omics: future research trends in medicinal plants

Lifang Yang , Ye Yang, Luqi Huang, Xiuming Cui and Yuan Liu

Corresponding authors. X. M. Cui, Yunnan Provincial Key Laboratory of *Panax notoginseng*, Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, Yunnan 650500, China. E-mail: sanqi37@vip.sina.com; Y. Liu, Yunnan Provincial Key Laboratory of *Panax notoginseng*, Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, Yunnan 650500, China. E-mail: liuyuan513@kust.edu.cn

## Abstract

Medicinal plants are the main source of natural metabolites with specialised pharmacological activities and have been widely examined by plant researchers. Numerous omics studies of medicinal plants have been performed to identify molecular markers of species and functional genes controlling key biological traits, as well as to understand biosynthetic pathways of bioactive metabolites and the regulatory mechanisms of environmental responses. Omics technologies have been widely applied to medicinal plants, including as taxonomics, transcriptomics, metabolomics, proteomics, genomics, pangenomics, epigenomics and mutagenomics. However, because of the complex biological regulation network, single omics usually fail to explain the specific biological phenomena. In recent years, reports of integrated multi-omics studies of medicinal plants have increased. Until now, there have few assessments of recent developments and upcoming trends in omics studies of medicinal plants. We highlight recent developments in omics research of medicinal plants, summarise the typical bioinformatics resources available for analysing omics datasets, and discuss related future directions and challenges. This information facilitates further studies of medicinal plants, refinement of current approaches and leads to new ideas.

**Keywords:** medicinal plant, single omics, multi-omics analysis, bioinformatics resource

## Introduction

Medicinal plants (MPs) are the main source of natural metabolites such as pigments, condiments, insecticides and medicines. MPs have been used to treat diverse diseases in China, India and Egypt for 5000 years and are still used today, despite the availability of pharmaceuticals [1]. Plant-derived monomers (morphine, artemisinin, taxol, digitali, vinblastine, etc.) are essential for chemical drug development, and mixed secondary metabolites such as total saponins and tanshinones exert strong therapeutic effects [2]. In addition, various well-known MPs, such as *Panax ginseng* and *Panax quiquefolium*, which enhance physical function and improve memory, have been widely used as supplements and in healthcare products [3].

Discovering novel and pharmacologically relevant compounds and determining their biosynthetic pathways in MPs are challenging. The continuous introduction of novel omics concepts and rapid development of sequencing technologies has greatly facilitated the comprehensive dissection of biological processes occurring in plants at the genetic, transcriptional and metabolic levels, leading to the rapid development of omics-based plant studies over the last two decades (Figure 1). Meanwhile, omics studies of MPs are gradually transitioning from single- to

multi-omics, the integrated multi-omics studies are becoming abundant and the number of omics-based MPs studies is increasing rapidly (Figure 2). Most omics studies of MPs have focused on (i) identifying DNA and chemical markers for classifying MPs [4, 5], (ii) locating functional genes controlling specific agronomic traits [6–8], (iii) identifying key metabolic pathways involved in the biosynthesis of active compounds [9–11] and (iv) determining the molecular mechanisms of stress responses [12–14]. These studies provide a theoretical basis for obtaining large quantities of specific compounds through synthetic biology and can enhance the molecular breeding of MPs.

Here, we comprehensively review recent advances and future trends in omics studies of MPs to promote the development of novel ideas and methods related to integrated multi-omics research.

## Phenotypes and DNA markers are used in taxonomy

Phenotyping is the most intuitive approach for identifying and classifying plants but is time-consuming, laborious and often destructive to plants. High-throughput phenotyping platforms

**Lifang Yang** is a PhD student at Kunming University of Science and Technology, China. She is interested in the bulk RNA-seq, scRNA-seq and metabolomics studies on *Panax* genus.

**Ye Yang** is a professor at Kunming University of Science and Technology, China. His research interests mainly involve the molecular mechanisms of plants response to stresses.

**Luqi Huang**, the academician of the Chinese Academy of Engineering, studies the development of traditional Chinese medicine, Chinese Academy of Chinese Medical Sciences, China.

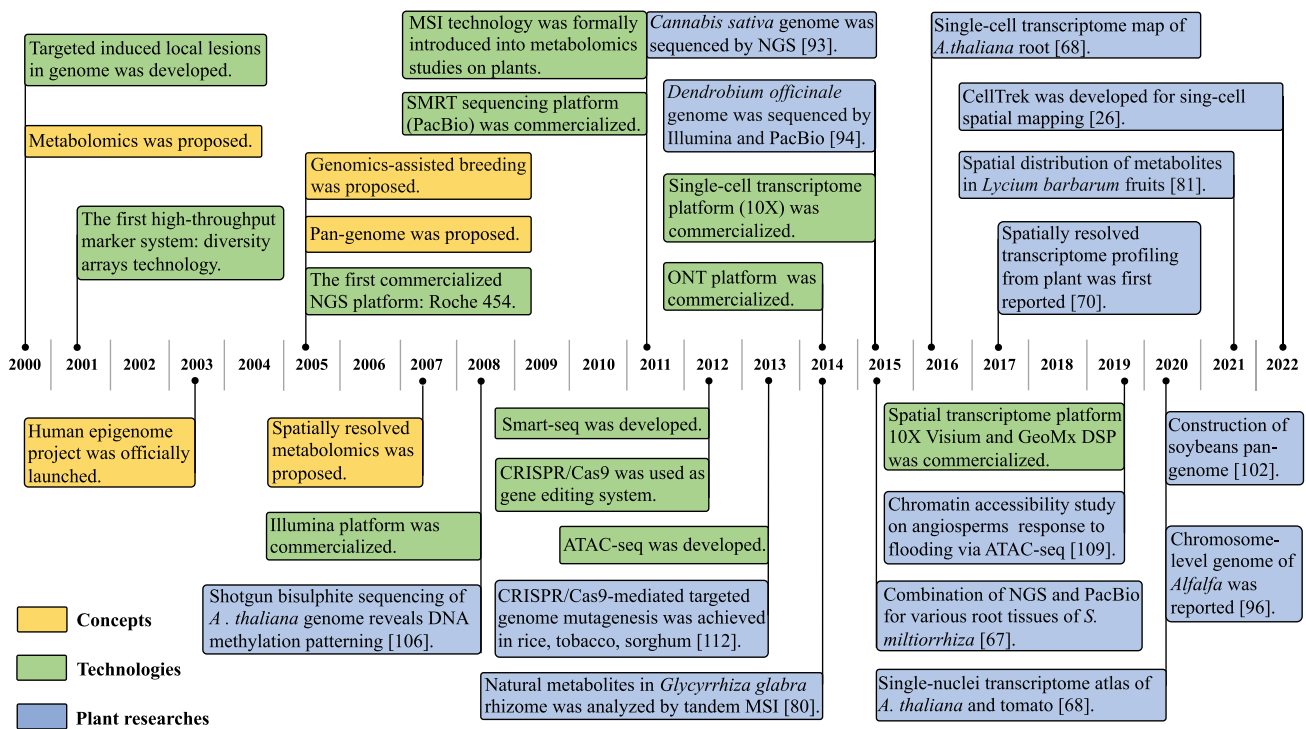
**Xiuming Cui** is a professor at Kunming University of Science and Technology, China. His research interests include sustainable development of germplasm resources of *Panax* genus.

**Yuan Liu** is an associate professor at Kunming University of Science and Technology, China, and is interested in developing bioinformatics analysis software and studying medicinal plant genomes.

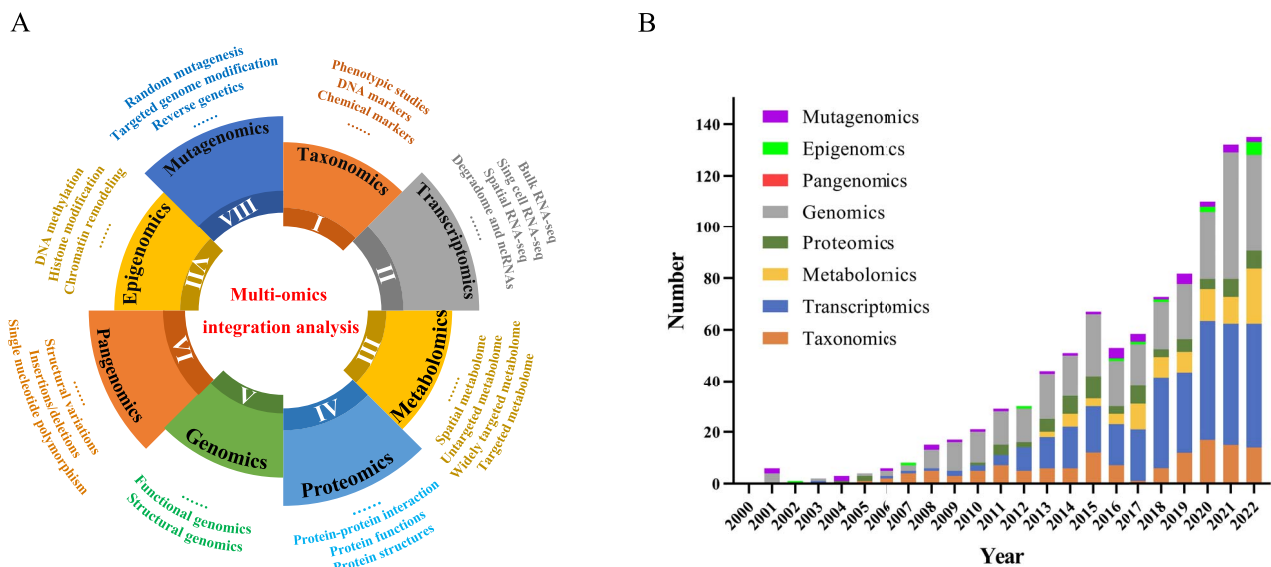
**Received:** August 15, 2022. **Revised:** October 13, 2022. **Accepted:** October 14, 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)



**Figure 1.** Timeline for omics technology development and typical omics-based plant studies over the past two decades. The proposal of omics concepts is shown in yellow, key events related to the development of omics technologies are indicated in green boxes and typical omics-based plant studies were illustrated by blue. Abbreviations: NGS, next-generation sequencing; SMRT, single molecule real-time; MSI, mass spectrometry imaging; Smart-seq, switching mechanism at 5' end of the RNA transcript sequencing; ATAC-seq, assay for transposase-accessible chromatin with high-throughput sequencing; CRISPR/Cas9, clustered regularly interspaced short palindromic repeats/CRISPR-associated9; ONT, Oxford Nanopore Technology.



**Figure 2.** Summary of research pattern and bibliometrics of omics studies on medicinal plants. (A) The pattern of omics studies on medicinal plants: (I) Taxonomics mainly involves identification and classification of medicinal plants based on phenotyping, DNA markers and chemical markers. (II) Transcriptomics studies contain bulk RNA-seq, single-cell RNA-seq (scRNA-seq), spatial RNA-seq (spRNA-seq), as well as degradome and ncRNAs. (III) Metabolomics mainly involves targeted, widely targeted, untargeted metabolome and spatial metabolome studies on secondary metabolites. (IV) Proteomics focuses on structures, functions and protein-protein interaction of protein molecules. (V) Genomics can be divided into structural and functional genomics studies. (VI) Pangenomics lays particular emphasis on the effects of SNPs, indels and SVs. (VII) Epigenomics mainly involves three aspects: DNA methylation, histone modification and chromatin remodeling. (B) The number of articles of omics studies on medicinal plants published from 2000 to 2022 from PubMed database. Keywords of medicinal plant taxonomic, transcriptome, metabolomic, proteomic, genomic, pangenome, DNA methylation and mutagenesis are searched, under the Title/Abstract term in the query box.

with high-resolution, advanced sensors and fully automatic data collection systems are promising advancements in plant phenotyping [15]. Bioinformatics tools and image databases have also been developed for handling the massive amounts of phenotypic data and plant images collected using high-throughput phenotyping platforms (Table 1; [16, 17]).

However, delimitation of certain taxa derived from congeneric species is difficult because of the existence of morphological intermediates. Therefore, many DNA barcodes such as 5S ribosomal RNA, 18S ribosomal RNA, internal transcribed spacer, *matK*, *rbcL*, *trnH-psbA* and *trnL-F* have been widely applied since 2008 for analysing the taxonomy of MPs [4]. In addition, specific types of DNA markers, such as single-nucleotide polymorphisms (SNPs) and simple sequence repeats, can be used to identify MPs. Currently, an interactive database of DNA barcodes from medicinal materials is regularly updated to support medicinal material identification and MP taxonomy studies [18]. Combining DNA barcodes with metabolomics data has been recommended for more accurately taxonomizing MPs and identifying their subspecies or varieties [12,59].

The availability of bioinformatics resources for taxonomic studies of MPs remains limited; thus, it is necessary to construct a standardised taxonomic system that combines phenotypic images with DNA markers and specific metabolites. Accurate taxonomic classification of MP species can not only confirm the authenticity of medicinal raw materials but also ensure the high quality of medicinal products produced from these materials.

## Transcriptomics is the most widely used approach for studying gene expression

Transcriptomics can be divided into microarrays based on hybridization and RNA sequencing (RNA-seq) based on sequencing methods. The major difference between these approaches is that microarray can only detect the expression levels of known genes in samples, whereas RNA-seq can detect the expression information of all genes. In microarray analysis, the roles of specific mRNAs and microRNAs (miRNAs) can be determined under given stress conditions and identify molecular markers of specific compositions in plants [60,61]. RNA-seq can provide a dynamic genetic map of the spatiotemporal expression patterns of genes in different parts and developmental stages of plants. The transcriptomes of MPs with multiple medicinal parts have been sequenced using next-generation sequencing (NGS) platforms to investigate the organ- and tissue-specific expression patterns of genes [62–64]. Dynamic transcriptional changes in MPs under different stress conditions [65] and at different developmental stages [66] have been extensively studied. Compared with NGS, long-read sequencing technologies, such as PacBio and Oxford Nanopore Technologies, can reveal the complexity of transcriptomes, including post-transcriptional modifications, alternative splicing and fusion transcripts; thus, combining NGS and PacBio platforms can provide a finer transcriptome landscape of complex gene expression [67]. Two mainstream methods for transcriptome assembly, the combination of HISAT and StringTie [19] and Trinity [20], are applicable to the availability and non-availability of reference genomes, respectively. Currently, two databases of plant transcriptome data, PPRD and ARS, have important reference value for studying MP transcriptomes (Table 1; [21,22]).

Novel advancements in technology have improved the resolution of transcriptomic research from bulk RNA-seq to single-cell RNA-seq (scRNA-seq). Although limited by reference genomes

and current technologies, scRNA-seq has been applied in plants such as *Zea mays*, *Oryza sativa*, *Solanum lycopersicum* and *Arabidopsis thaliana*, and a single-nucleus transcriptome atlas of *S. lycopersicum* and *A. thaliana* was reported [68]. Application of scRNA-seq and target genome-editing techniques has been proposed for supporting precise crop breeding, as clustered regularly interspaced short palindromic repeat droplet sequencing (CRISPR-seq) depends on a guide RNA vector with a unique barcode that can be detected using scRNA-seq [69]. Moreover, scRNA-seq can be combined with transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) for multi-omics co-labelling, which can simultaneously capture information on transcripts and open chromatin from each cell. A pre-trained deep learning-based method, scDeepSort, can be used to annotate cell types in single-cell transcriptomic datasets [23]. The online tool Plant Single Cell Transcriptome Hub and continuously updated PlantscRNAdb were developed for plant scRNA-seq research [24,25]. These resources promoted scRNA-seq studies on MPs and pave the way for the combined application of scRNA-seq with other omics or techniques.

Spatial transcriptome sequencing (spRNA-seq) can compensate for the loss of spatial location information of cells evaluating using scRNA-seq. The first spatially resolved transcriptome profile of plant fields was obtained from *A. thaliana* in 2017 to determine the processes involved in plant development and evolution [70]. Subsequent spatial transcriptome studies of peanut tissue suggested that the spatial information of cells, independent of marker genes, is more useful for non-model species [71]. A spatiotemporal atlas of organogenesis of orchid flowers revealed that floral organ development is co-regulated by numerous specialised genes that function in different tissues and developmental stages [72]. Two spRNA-seq platforms (10X Visium, 10X Genomics, Pleasanton, CA, USA; and GeoMx DSP, NanoString Technologies, Seattle, WA, USA) have been commercially available since 2019; however, these platforms cannot achieve single-cell resolution. Subsequently, an excellent computational method, CellTrek, was developed that combines two datasets to perform single-cell spatial mapping [26]. Thus, a combination of spRNA-seq and scRNA-seq can accurately depict the spatiotemporal developmental trajectory and biological functions of certain cells of interest in MPs. In addition, a database for spatially resolved transcriptomes, SpatialDB, provides a repository for researchers studying the spatial cellular structure of tissues and the cellular microenvironment [27].

Degradome and non-coding RNAs (ncRNAs) sequencing, another direction for transcriptome data research, provides abundant information on RNA degradation, miRNAs and long ncRNAs and contributes to the identification of miRNA-mediated cleavage of target genes and functional studies of ncRNAs [73]. Combined analysis of degradome sequencing and miRNA profiles has been widely applied in MP research [14,74]. Corresponding bioinformatics tools and databases, such as psRNATarget, PLncPRO, PcircRNA\_finder, PAREameters and MepmiRDB, have been developed to identify and determine the functions of novel ncRNAs in plants (Table 1; [28–32]).

## Metabolomics defines end-products of gene expression

The metabolome is a direct determinant of the authenticity and quality of MPs. Currently, studies based on metabolomics are focused in targeted, widely targeted and untargeted directions. The targeted metabolome is a suitable choice for distinguishing

**Table 1.** The list of typical bioinformatics resources available for omics studies on medicinal plants

Omics	Tool/database name	Brief description	URL	References
Taxonomics	Image Harvest	An open-source software for high-throughput plant image processing and analysis	<a href="http://cropstressgenomics.org/">http://cropstressgenomics.org/</a>	[16]
	SpaTemHTP	A pipeline for analysing spatial-temporal high-throughput phenotyping data	<a href="https://github.com/CRISAT-GEMIS/SpaTemHTP">https://github.com/CRISAT-GEMIS/SpaTemHTP</a>	[17]
	MPID	Medicinal plant images database	<a href="https://library.hkbu.edu.hk/electronic/libdbs/mpd/">https://library.hkbu.edu.hk/electronic/libdbs/mpd/</a>	Null
	PlantCLEF 2019	Image-based identification database for plant species	<a href="https://www.imageclef.org">https://www.imageclef.org</a>	Null
	MMDBD	Medicinal materials DNA barcode database	<a href="http://www.cuhk.edu.hk/icm/mmdbd.htm">www.cuhk.edu.hk/icm/mmdbd.htm</a>	[18]
	HISAT+StringTie	A combination approach for reference genome-based RNA-seq read alignment	Null	[19]
	Trinity	A <i>de novo</i> transcriptome assembler of RNA-seq data without reference genome	<a href="https://github.com/trinityrnaseq/trinityrnaseq/wiki">https://github.com/trinityrnaseq/trinityrnaseq/wiki</a>	[20]
	PPRD	A comprehensive online database for data mining and expression analysis	<a href="http://ipf.sustech.edu.cn/pub/plantrna/">http://ipf.sustech.edu.cn/pub/plantrna/</a>	[21]
	ARS	An online database for exploring public <i>Arabidopsis</i> RNA-seq libraries	<a href="http://ipf.sustech.edu.cn/pub/athma/">http://ipf.sustech.edu.cn/pub/athma/</a>	[22]
	scDeepSort	A pre-trained cell-type annotation approach for single-cell transcriptomics based on deep learning	<a href="https://github.com/ZJUFanLab/scDeepSort">https://github.com/ZJUFanLab/scDeepSort</a>	[23]
Transcriptomics	PscTH	An integrated online tool for exploring plant single-cell transcriptome landscape	<a href="https://jinlab.hzau.edu.cn/PscTH/">https://jinlab.hzau.edu.cn/PscTH/</a>	[24]
	PlantscRNAdb	A database dedicated to plant single-cell RNA analysis	<a href="http://ibi.zju.edu.cn/plantscrnadb/">http://ibi.zju.edu.cn/plantscrnadb/</a>	[25]
	CellTrek	A computational toolkit that can achieve single-cell spatial mapping	Null	[26]
	SpatialDB	A database for spatially resolved transcriptomes	<a href="https://www.spatialomics.org/SpatialDB">https://www.spatialomics.org/SpatialDB</a>	[27]
	psRNATarget	A small RNA target analysis server for plants	<a href="http://plantgrn.noble.org/psRNATarget/">http://plantgrn.noble.org/psRNATarget/</a>	[28]
	PLncPRO	Predicting lncRNAs in plants	<a href="http://ccb.jnu.ac.cn/plncpro/">http://ccb.jnu.ac.cn/plncpro/</a>	[29]
	PcircRNA_finder	Predicting circRNAs in plants	<a href="http://ibi.zju.edu.cn/bioinplant/tools/manual.htm">http://ibi.zju.edu.cn/bioinplant/tools/manual.htm</a>	[30]
	PAREameters	A tool for inferring miRNA targeting criteria in plants	<a href="http://srna-workbench.cmp.uea.ac.uk/">http://srna-workbench.cmp.uea.ac.uk/</a>	[31]
	MepmiRDB	Medicinal plant miRNA and degradome-seq database	<a href="http://mepmirdb.cn/mepmirdb/index.html">http://mepmirdb.cn/mepmirdb/index.html</a>	[32]
	CRISP	A deep learning framework for identifying, simulating and analysing contour regions of interest in metabolomics map	<a href="https://github.com/vivekmathema/GcXGC-CRISP">https://github.com/vivekmathema/GcXGC-CRISP</a>	[33]
Metabolomics	MAPPS	A web-based tool for metabolic pathway prediction and network analysis	<a href="https://mapps.lums.edu.pk">https://mapps.lums.edu.pk</a>	[34]
	MetaboAnalyst 5.0A	Web-based platform for metabolomics data analysis and interpretation	<a href="https://www.metaboanalyst.ca">https://www.metaboanalyst.ca</a>	[35]
	METLIN	A highly annotated database with over 850 000 molecular standards	<a href="http://metlin.scripps.edu">http://metlin.scripps.edu</a>	[36]
	Prosit	Proteome-wide prediction of peptide tandem mass spectra by deep learning	<a href="https://github.com/kusterlab/prosit">https://github.com/kusterlab/prosit</a>	[37]
	pINET	A web platform for downstream analysis and visualization of proteomics data	<a href="http://pinet-server.org">http://pinet-server.org</a>	[38]
	PRIDE	A hub for mass spectrometry-based proteomics evidence	<a href="https://www.ebi.ac.uk/pride/">https://www.ebi.ac.uk/pride/</a>	[39]
	PPDB	The plant proteomics database	<a href="http://ppdb.tc.cornell.edu">http://ppdb.tc.cornell.edu</a>	[40]
	AlphaFold v2.0	A 3D high-accuracy protein-structure prediction database	<a href="https://alphafold.ebi.ac.uk">https://alphafold.ebi.ac.uk</a>	[41]
	STRING v11	Database for providing association networks of protein-protein interactions	<a href="http://string-db.org">http://string-db.org</a>	[42]
	BiGRID	Database for storage of protein, genetic and chemical interactions from humans and major model species	<a href="https://thebiogrid.org">https://thebiogrid.org</a>	[43]
Genomics	SVision	A deep learning approach to resolve complex structural variants in genome	<a href="https://github.com/xjtu-omics/SVission">https://github.com/xjtu-omics/SVission</a>	[44]
	MetalLogo	A heterogeneity-aware sequence logo generator used to display conservations and variations in a batch of DNA or protein sequences	<a href="http://metalogo.omicsnet.org">http://metalogo.omicsnet.org</a>	[45]
Pangenomics	TCMPG	Traditional Chinese medicine plant genome database	<a href="http://cpcb.cdutcm.edu.cn/TCMPG/">http://cpcb.cdutcm.edu.cn/TCMPG/</a>	[46]
	MPGR	Medicinal plants genomics resource	<a href="http://medicinalplantgenomics.msu.edu/">http://medicinalplantgenomics.msu.edu/</a>	Null
	PATO	A pangenome analysis toolkit	<a href="https://github.com/trycisBioinfo/PATO">https://github.com/trycisBioinfo/PATO</a>	[47]
	Panache	A viewer based on web browser for linearized pan-genome	<a href="https://github.com/SouthGreenPlatform/panache">https://github.com/SouthGreenPlatform/panache</a>	[48]
	GreenPhyloDB v5	A comparative plant pan-genomics database	<a href="https://www.greenphylo.org">https://www.greenphylo.org</a>	[49]
Epigenomics	ChINN	A machine learning-based method for predicting chromatin interactions from DNA sequences	<a href="https://github.com/mjflab/chinn">https://github.com/mjflab/chinn</a>	[50]
	PlantPan3.0	A resource for reconstruction of transcriptional regulatory networks from plant ChIP-seq experiments	<a href="http://PlantPAN.itps.ncku.edu.tw/">http://PlantPAN.itps.ncku.edu.tw/</a>	[51]
Mutagenomics Integrated multi-omics	CRISPRIdentify	Identification of CRISPR arrays based on machine learning approach	<a href="https://github.com/BackofenLab/CRISPRIdentify">https://github.com/BackofenLab/CRISPRIdentify</a>	[52]
	multiSLIDE	A web tool for interactive heatmap-based exploration and visualization of multi-omics datasets	<a href="https://github.com/soumitag/multiSLIDE">https://github.com/soumitag/multiSLIDE</a>	[53]
	PaintOmics 4	A web tool for integrating and visualizing multi-omics datasets based on biological pathway maps	<a href="https://paintomics.org/">https://paintomics.org/</a>	[54]
	OmicsAnalyst	A web-based platform for analysis and results visualization of multi-omics datasets	<a href="https://www.omicsanalyst.ca">https://www.omicsanalyst.ca</a>	[55]
	OmicsNet 2.0	A web-based tool for multi-omics integration and network visual analytics	<a href="http://www.omicsnet.ca">http://www.omicsnet.ca</a>	[56]
	MPOD	Integrated multi-omics database for medicinal plants	<a href="http://medicinalplants.ynau.edu.cn/">http://medicinalplants.ynau.edu.cn/</a>	[57]
	1 K-MPGD	An integrated database combining genome and metabolites of medicinal plants	<a href="http://www.herbgenome.com/">http://www.herbgenome.com/</a>	[58]

URL, uniform resource locator; Null represents no URL or Reference.

crude medicinal materials from congeneric species in compound preparations of traditional Chinese medicines [75]. A widely targeted metabolome study of *Pueraria lobata* and its varieties suggested that differences in the nutritional value among these species can be explained by changes in nutrient abundance, whereas medicinal quality can be assessed according to the contents of secondary metabolites [76]. Sixteen key metabolites useful for distinguishing different *Ficus deltoidea* varieties were identified in untargeted metabolome analysis and were stable regardless of the growth environment and geographical origin [77]. Therefore, chemical markers are important factors for MP authentication, whereas the contents of specific metabolites can be used to evaluate the quality of medicinal raw materials. Metabolite profiling of mutagenic lines with loss- or gain-of-function genes reveal specific metabolites that are synthesized under the control of target genes, thereby bridging the gap between genes and metabolites. In addition, a metabolomics-oriented reverse genetic approach can be used to further explore the genes responsible for the chemical structure diversity of secondary metabolites [78]. Therefore, analysis of biosynthetic regulation cascades involved in active metabolite production in MPs, as the first step toward molecular breeding and synthetic biology, has been largely driven by metabolomics-based analyses.

A deep learning framework, CRISP, was developed to identify, simulate and analyse contour regions of interest in metabolomic maps [33]. MAPPS is useful for metabolic network analysis and pathway prediction, whereas MetaboAnalyst 5.0 is a user-friendly platform for analysing raw metabolomics data and exploring metabolite functions [34,35]. The enormous structural diversity of plant-derived compounds suggests that medicinally relevant compounds can still be discovered in plants. METLIN, a highly annotated database containing over 850 000 molecular standards, is useful for screening plant-derived bioactive compounds [36].

Spatial metabolomics overcomes the limitations of bulk metabolomics and can accurately determine the types, contents and spatial distributions of metabolites, and then characterise the chemical makeup of a tissue or organ at spatial resolution [79]. Thus, spatial metabolomics can provide abundant spatial distribution albums of metabolites and achieve 'real-time reporting' of the metabolome in organisms. The *in situ* presentation and spatiotemporal transformation of metabolites can simplify various biological problems in MPs, such as the biosynthetic pathways of natural metabolites [80] and fruit development [81]. Combining spatial metabolomics with spRNA-seq is an exciting approach for investigating biological processes in specialised cell types of MPs.

## Proteomics: a hub linking the transcriptome and metabolome

As proteins are directly involved in performing and controlling almost all biological processes, proteomics is essential for understanding the regulatory mechanisms responsible for the development and secondary metabolism of MPs [82]. iTRAQ quantitative proteomics of *Rehmannia glutinosa* roots revealed that many prenyltransferase present higher expression level at the expansion and maturation stage than the elongation stage [83]. Label-free quantitative proteomic study on *P. ginseng* leaves under heat revealed the molecular mechanism of stress and the influences of ginsenoside production at proteins level [84]. Proteins expressed in *Chrysobalanus icaco*, *Bauhinia variegata* and *Bauhinia forficata* have also been characterised and differentiated to determine the

differences in their medicinal properties [85]. Notably, a recent study suggested that plant odorant-binding proteins bind specific metabolites, leading to changes in transcription activation, gene expression, protein function and metabolism, and play important roles in plant communication and defensive responses, which inspires researchers to further think about that whether the production and accumulation of desired metabolites can be induced by changing the expression and function of specific odourant-binding proteins [86]. The biological functions of a protein depend not only on the linear arrangement of the amino acid sequence but also on its spatial structure; post-translational modifications also have diverse effects on the activity and function of protein molecules [87].

Prosit is a proteome-wide prediction network based on deep learning that can enable larger numbers of identifications at >10x lower false discovery rates [37]. PiNET—a versatile web platform—is used for downstream analysis of proteomic data and visualisation of the results [38]. To date, there is no protein database specific for MPs; however, comprehensive protein databases, such as the continuously updated PRIDE and PPDB, are available for functional studies of proteins in MPs (Table 1; [39,40]). A breakthrough in protein-structure prediction, the AlphaFold protein-structure database, an artificial intelligence (AI) system developed by DeepMind, enables state-of-the-art predictions of protein structures based on their amino acid sequences, allowing biomedical researchers to obtain 3D structural models for almost any protein sequence [41]. In addition, protein-protein interaction networks are useful for functional studies of proteins, in which protein functions can be inferred based on interactions between known and unknown proteins [88]. Information on protein-protein interactions in plants has been deposited in the STRING and BioGRID databases, which are open to the public for MP investigations (Table 1; [42,43]).

## Structural and functional genomics

Structural genomics relies on molecular markers that are available for tagging and mapping of candidate genes related to species traits. Currently, quantitative trait locus (QTLs) mapping and genome-wide association studies (GWAS) are the two most important approaches for studying traits in plants. QTLs has been widely applied in MPs to link complex phenotypes of interest to specific regions on chromosomes and then identifying the number, locations, interactions and functions of these regions [7,89]. GWAS focus on detecting genetic variations in multiple individuals from a population to determine genotypes, followed by statistical analyses between genotypes and phenotypes at the population level to screen genetic variations most likely to affect traits of interest. This method has been applied to evaluate the genes controlling the stem thickness and dry root weight of *P. notogensing* [8], amorpho-4,11-diene synthase gene expansion and ultimately results in higher artemisinin content [90] and high  $\alpha$ -linolenic acid content in the seed oil of *Perilla* [91]. Studies of the relationship between the traits and genotypes of MPs based on GWAS and QTLs have contributed to subsequent utilisation of functional genomics in molecular breeding and genetic improvement.

After plant genome resources became available, a combination of genomics and breeding techniques resulted in development of the novel concept of 'genomics-assisted breeding' for crop improvement in 2005 [92]. The advent of NGS has greatly improved the throughput of genome sequencing, and the introduction of long-read sequencing and Hi-C has enabled

chromosome-level genome assembly and research. The genome of *Cannabis sativa* was sequenced on Roche/454 (Basel, Switzerland) and Illumina (San Diego, CA, USA) platforms in 2011 [93], and that of *Dendrobium officinale* was sequenced on Illumina and PacBio (Menlo Park, CA, USA) platforms in 2015 [94]. Specifically, the number of chromosome-level genomes from various MPs, such as *P. notoginseng* [9], *Artemisia annua* [90], *Opium poppy* [95], *Medicago sativa* [96] and *Bletilla striata* [97], has sharply increased in the last few years. These studies suggest that chromosome-level genomes are important for delineating biological processes occurring in MPs, as they can be used to reduce the negative effects caused by false and incomplete genome assembly. Notably, gene duplication, rearrangement, introgression and fusion events may have directly relationship with the specialised secondary metabolites [95]. Thus, functional genomics is a prerequisite for the precise molecular breeding of MPs to improve their medicinal traits [97]. In addition, some pivotal transcription factors are indispensable for regulating the biosynthesis of active compounds in MPs [98].

SVision was developed to resolve complex structural variations (SVs) in the genome [44], and online bioinformatics tools and continually updated genome databases [45,46] have provided important support for genomic studies of MPs (Table 1).

## Pangenomics focuses on the dynamic genome

With the increasing of genomic studies, researchers realized that a single reference genome is insufficient to represent the genetic diversity of a species. Notably, a comparative genomic study of four *Panax* species illustrated how reshuffling of the ancestral core-eudicot genome results in a highly dynamic genome and causes metabolic diversification of extant eudicot plants [99]. Thus, a new era of pangenomic studies of MPs has emerged. The concept of the pangenome was initially proposed in 2005 and applied to bacteria to account for intraspecific variability. Pangenome refers to collection of all genes in a specific species, these genes can be divided into the core genes shared by all individuals and the dispensable genes present in a specific individual. Currently, pangenome studies of crops such as rice, maize, tomato, cucumber, wheat and soybean have demonstrated that dispensable genes are vital for maintaining the genetic diversity of species, because dispensable genes exhibit higher variability compared with core genes and contain higher-density SNPs and indels [100,101]. Large-scale structural variations (SVs), including copy number variants and presence/absence variants (PAVs) at the population level, are currently the most important focus of crop pangenome studies [102]. SVs directly affect dispensable genes in the pangenome of a species; these genes tend to be responsible for specific plant traits such as fruit traits, flowering time and seed size, environmental adaptation and disease resistance [103]. Moreover, SVs can be used to study gene expression divergence and quantitative trait variations, whereas PAVs can be used as markers in GWAS studies. Bioinformatics tools have also been developed for pangenome analysis (Table 1; [47,48]). In addition, a comparative pangenomics database, GreenPhyIDB v5, was constructed for investigating gene families and homologous relationships among plant genomes [49].

Assembly of the plant genome and pangenome is challenging because of the occurrence of general polyploidization and presence of large number of repetitive sequences. However, long-read sequencing technologies are powerful for pangenome construction in plants with large genome sizes and can span

complex repetitive regions in the genome to identify large-scale SVs. Notably, by combining differential gene identification and CRISPR/Cas9, enables gene functions can be comprehensively dissected and validated. Pangenomic studies of crops have provided valuable references for constructing MP pangenomes. Pangenomes are expected to gradually replace single reference genomes and become a new standard for studying evolutionary clades and genetic variations in plants and MPs.

## Epigenomics is an important supplement to genomics

Epigenetics involves changes in heritable traits caused by DNA methylation, histone modification and chromatin remodeling. Studies of epigenetic phenomena can be carried out on a genome-wide scale; thus, a new omics, epigenomics, combining epigenetics with genomics, has been developed [104]. Epigenomic studies have been performed to analyse epigenetic events occurring during the growth and development of plants, and to evaluate abnormalities caused by stress [105]. In addition, divergence in epigenetic regulation during polyploidization has led to high biochemical diversity among secondary metabolites in the *Panax* genus [99]. Since the DNA methylation pattern of the *A. thaliana* genome was reported in 2008 [106], DNA methylation studies have gradually become universally conducted to evaluate MPs. The pleiotropic roles of DNA methylation in MPs have been reviewed in detail [107]. Chromatin immunoprecipitation sequencing (ChIP-seq) can reveal information on histone modifications in studies of plant development and environmental memory [108], and ATAC-seq can be used to analyse genome-wide chromatin accessibility to explore the possible mechanisms of plant environmental adaptability [109]. Therefore, ChIP-seq and ATAC-seq are complementary methods that show highly consistent results [110]. Furthermore, ATAC-seq and RNA-seq can be combined to study differentially regulated transcription factors in key biological processes in plants [111]. The machine learning-based method chromatin interaction neural network (ChINN) is useful for predicting chromatin interactions based on DNA sequences, and PlantPan3.0 can be used to analyse the results of ChIP-seq experiments on MPs [50,51].

Currently, epigenomics is widely used to study epigenetic phenomena and the underlying epigenetic modification events in MPs. Several studies suggested that epigenetic modifications are closely related to the phenotypic traits of MPs and biosynthetic processes of secondary metabolites. These findings are expected to be applied in epigenetic engineering.

## Mutagenomics for obtaining plant species with desired variations

Mutagenesis is one of the most effective approaches for obtaining species with desired variations and primarily involves random mutagenesis and targeted genome modifications. Random mutagenesis can produce many mutant individuals with diverse traits but requires large-scale screening, which is typically time-consuming and laborious because of the randomness of mutations. In the last two decades, several breakthroughs have been made in the genome-editing field, particularly in the CRISPR/Cas9 system, which is a site-directed mutagenesis technology for introducing targeted genome modifications. Using this system, targeted genome modifications were made in rice, tobacco

and sorghum as early as 2013 [112]. Subsequently, this system was implemented in MPs (*S. miltiorrhiza*, *O. poppy*, *Camelina sativa* and *Dioscorea zingiberensis*) to produce pharmacologically bioactive metabolites through fine-scale targeted mutagenesis [113]. Transgenic herbal raw materials cannot be commercialised at present because of the specific nature of MPs (transgene introgression may lead to unpredictable changes in components and properties of herbal materials); thus, transgene-free genome editing may be important for avoiding transgene incorporation [114]. Transgene-free genome editing based on CRISPR/cas9 may be an optimal choice for improving the quality and yield of valuable MPs and achieving commercialisation. Notably, a machine learning-based approach, CRISPRidentify, can detect and differentiate true from false CRISPR arrays, greatly facilitating the application of CRISPR/Cas9 [52].

For genes with known functions, targeted genome modification is an excellent approach for rapidly and accurately obtaining a desired species. For genes with unknown or uncertain functions produced using genome sequencing and random mutagenesis, reverse genetics technologies can reveal associations between the differential genes and their functions and subsequently verify the functions of candidate genes. Integrated application of functional genomics and mutagenomics is currently the best approaches for improving species traits. Although mutagenomics has not been as widely used in MPs as in crops, its use in MP species is expected to increase with continuous improvements in MP genome resources and rapid development of suitable transformation and regeneration approaches.

## Multi-omics studies of medicinal plants are the future development trend

Rapid development of omics technologies is a prerequisite for successfully performing advanced omics studies. However, each omics technology, such as transcriptomics (including microarray technology, bulk RNA-seq, scRNA-seq and spRNA-seq), metabolomics (including bulk metabolomics and spatial metabolomics), proteomics (including iTRAQ quantitative and label-free quantitative technology) and genomics (including NGS and long-read sequencing technologies) has inherent advantages and disadvantages (Table 2). Therefore, integrated analysis of multi-omics datasets, such as the integration of scRNA-seq and spRNA-seq, spRNA-seq and spatial metabolomics, bulk RNA-seq and metabolomics, and RNA-seq and proteomics, can compensate for the limitations of other methods when comprehensively studying biological processes. Currently, omics studies of MPs are gradually transitioning from single- to multi-omics, which has provided more comprehensive insights into biological processes of interest. Integrated multi-omics studies of MPs have mainly focused on four factors (Figure 3). First, combined analysis of phenotypes, DNA markers and metabolomic data enables the accurate identification of MPs and processed medicinal materials [59,115]. Second, functional genes controlling the key agronomic traits of MPs can be located by linking extrinsic phenotypes to intrinsic genotype control [6,7,116]. Combining GWAS with other omics techniques may contribute to the identification of functional genes regulating complex traits [117]. Third, multi-omics integration can reveal the biosynthetic pathways of secondary metabolites in MPs [9–11,65]. Notably, integration of omics with gene editing tools is useful for the development of precision plant breeding [117]. Finally, multi-omics integration can explain the regulatory mechanisms involved in the responses of MPs to stress [12,13,118]. With the increasing diversity of

omics technologies, researchers often obtain different types of omics datasets derived from the same or different samples, providing highly scientific and reliable access to specific biological processes in MPs. However, these findings also create challenges for the integrated and associated analysis of multiple omics data types.

Several bioinformatics tools for integrating and analysing multi-omics datasets were recently developed [53–56]. MPOD and 1 K-MPGD are specific for multi-omics studies of MPs, and will be continuously updated to provide long-term support for combined analysis of multi-omics datasets (Table 1) [57,58]. Furthermore, data obtained using integrated multi-omics approaches can complement and validate each other when investigating changes in certain biological processes, making the analytical results more comprehensive and credible. Integrated multi-omics approaches will be widely applied in MP research to understand specific biological processes.

## Conclusion

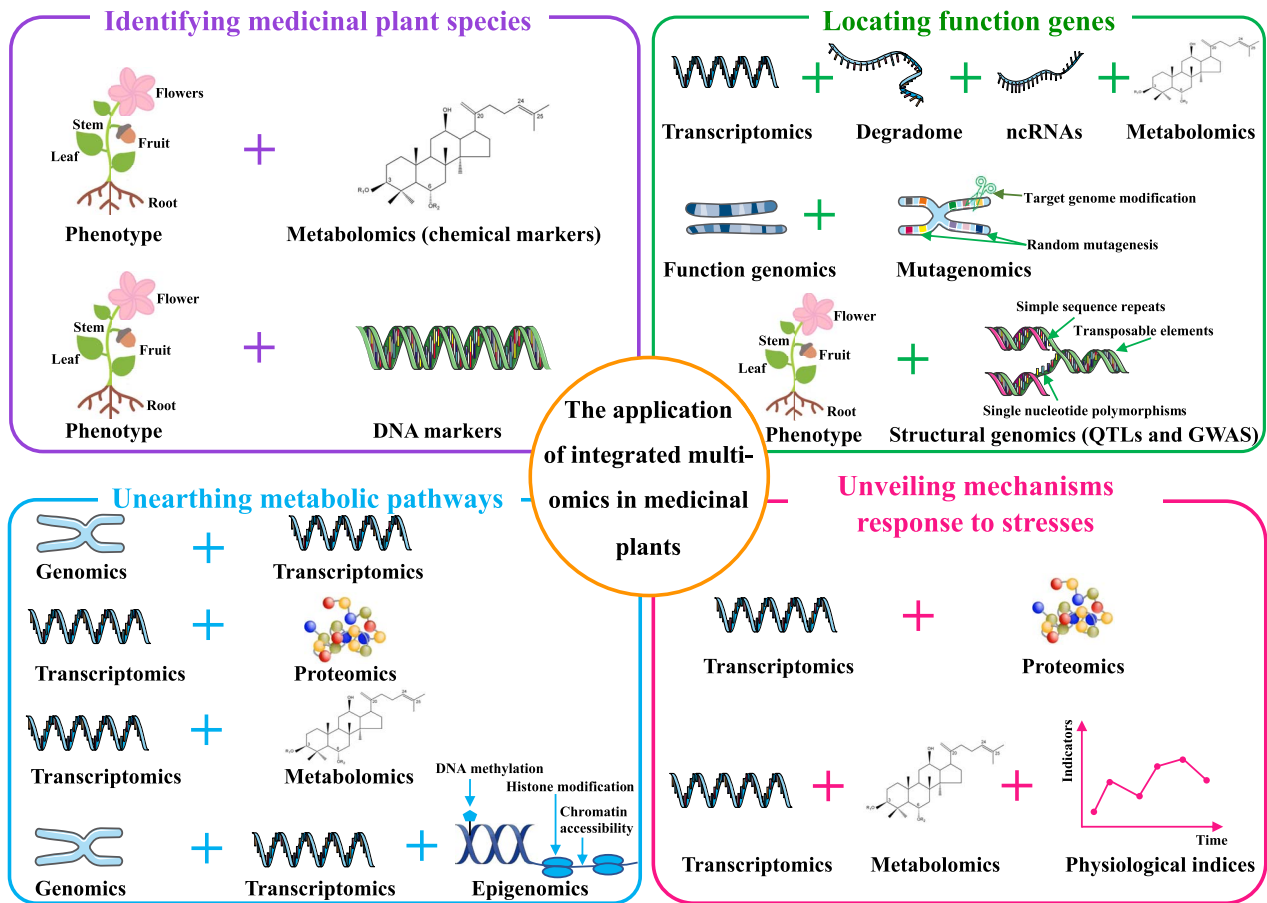
Recent developments in diverse omics technologies have provided an unprecedented opportunity for plant researchers to obtain considerable biological knowledge through integrated analysis of multiple omics datasets. Genomes, transcriptomes, proteomes, metabolomes and other omics datasets derived from various MPs have been reported, and corresponding bioinformatic tools and databases have been developed. Integrated analysis of multi-omics datasets is highly comprehensive for investigating MPs. Results based on multi-omics datasets not only provide a foundation for obtaining MP species with high yield, good quality and disease resistance through molecular breeding but also provide a theoretical basis for achieving steady biotransformation of desired secondary metabolites through synthetic biology. Notably, it is now feasible to identify functional genes controlling key biological traits and determine the catalytic mechanisms of key enzymes involved in biosynthetic pathways of active compounds by performing multi-omics and bioinformatic studies. However, there are many unsettled issues in genome editing and the knock-out or overexpression of functional genes for MPs because of the lack of suitable transformation and regeneration approaches. Synthetic biology involves strain improvement, microbial system development and reconstruction and optimisation of metabolic models suitable for specific metabolite types, which are very challenging.

Although MPs have been widely examined in omics studies, further detailed examination is required. There have been few scRNA-seq and spRNA-seq studies of MPs because of the limitations of reference genomes and technologies. Furthermore, transgene-free genome modifications based on the CRISPR/Cas9 system have not been widely applied to MPs, as suitable transformation and regeneration approaches are lacking. Increasing evidence has shown that epigenetic modifications have non-negligible effects on gene expression; however, there are fewer epigenomic studies of MPs than of crops. In addition, ncRNAs play important roles in regulating gene expression; however, there is only one miRNA database specific for MPs, and no database exists for circRNAs and long ncRNAs in MPs. Finally, it remains challenging to integrate different results from multiple omics research, establish correlations between results and provide reasonable explanations for causalities because of differences in the representation of different omics datasets, particularly for more than three omics data types. The lack of bioinformatic tools and omics databases limits the interpretation of specific

Table 2. Advantages and disadvantages of the leading technologies for omics

Omics	Technologies	Advantages	Disadvantages	Main application in plant fields	Representative research
Transcriptomics	Microarray	(i) The fidelity of gene expression is high because amplification is not required. (ii) It is very suitable for research with high requirement for quantification of genes expression.	(i) The template needs to be designed prior to the experiment. (ii) Novel transcripts cannot be detected. (iii) The detection result is inaccurate when the gene expression level is too low or high. (iv) The number of genes detected at one time is limited.	Identifying molecular markers for specific composition, revealing the regulatory mechanisms of stress response.	[60,61]
	Bulk RNA-seq	(i) All genes expressed at a certain time node or development stage of an organism can be detected simultaneously. (ii) Novel transcripts and splice isoforms or even genes can be found.	(i) The imbalance of PCR amplification processes will destroy the true concentration proportion of different fragments in the samples, resulting in errors in the calculation of gene expression. (ii) It only represents the average gene expression level of a sample.	Studying plant growth and development, stress response and the regulatory mechanisms of genes involved in accumulation and distribution of secondary metabolites.	[62–66]
	ScRNA-seq	(i) Revealing the genes expression state of single cell and reflecting the heterogeneity between cells. (ii) Discovering novel and rare cell types. (iii) Exploring regulatory mechanisms of genes expression during cell development and differentiation. (iv) It is very suitable for rare sample or sample with a small number of cells. In addition to the advantages of scRNA-seq, snRNA-seq is also applicable to (i) frozen tissue. (ii) samples that are difficult to dissociate, (iii) cells with irregular shape, such as nerve cells and muscle cells.	(i) The process of dissociating tissues may cause changes in gene expression, especially in plant cells. (ii) Different types of cells may have different degrees of difficulty in dissociation, which may lead to rare cell types cannot be captured, and the proportion of cell types finally obtained is biased. (iii) The sample must be fresh living tissue. (iv) It loses RNAs in cytoplasm, may resulting in the loss of transcripts information with important biological significances. (v) It is usually necessary to use a flow cytometer to sort the nucleus, which leads to longer experimental time and more transcription changes without biological significance. Single-cell resolution has not been achieved yet due to technology limitation.	Analyzing cell differentiation trajectory, inferring the development process of the tissue and finding the heterogeneity between distinct cell types in the tissue.	[68]
	SnRNA-seq			Similar to scRNA-seq studies.	[68]
	SpRNA-seq	Combining spatial location information with genes expression to display genes transcription information at different locations in tissues or samples.	It only represents the average level of metabolites.	Identifying key genes and regulatory pathways responsible for development process of tissues and organs.	[70–72]
Metabolomics	Bulk metabolomics	It can provide all metabolites (including types and content) of any sample at specific developmental stage or environmental condition.	It only represents the average level of metabolites.	Studying the content of nutrient and secondary metabolites, identifying the quality and authenticity of raw medicinal materials.	[75–77]
	Spatial metabolomics	It can accurately characterize types, contents and spatial distributions of metabolites, and achieved 'real-time reporting' of metabolome in organisms.	Mass spectrometry imaging (MSI) technology mostly belongs to solid sampling, its sensitivity and detection limit often differ from that of bulk metabolomics.	Studying biosynthesis and distribution of natural metabolites, fruit development and maturity.	[80,81]
Proteomics	LC-MS/MS-based iTRAQ	(i) Wide analysis range and good separation effect. (ii) Reliable qualitative and accuracy quantitative results.	(i) It can only detect differential expressed proteins. (ii) It is easy to introduce errors in sample processing.	Study differential expressed proteins under different growth condition or development stage.	[83]
	Quantitative Proteomics				
	LC-MS/MS-based Label-free	(i) The number of samples is not limited and applicable to large sample size. (ii) The detection range of peptide fragments is wide and conducive to detection of low abundance proteins. (iii) It can identify whether proteins exist.	(i) Complex data processing. (ii) High dependence on the stability of mass spectrometry results.	Study differential expressed proteins under different growth condition or development stage.	[84]
	Quantitative Proteomics				
Genomics	NGS-based	(i) High sequencing throughput. (ii) Low cost. (iii) Low sequencing error rate.	(i) Short reads. (ii) High assembly error rate. (iii) Unable to obtain high-quality reference genome. (iv) High sequencing error rate. (v) High cost.	It was used for genome sequencing in the early stage, but now it is mainly used for transcriptome sequencing.	[62–66,93]
	Long-read sequencing-based	(i) Long reads, combing with Hi-C can provide chromosome-level genome. (ii) It can span complex repetitive regions in the genome to discover larger-scale structural variations.		Obtaining chromosome-level genome.	[94–97]





**Figure 3.** The application summary of integrated multi-omics approaches in medicinal plants. It is mainly involved in four aspects: (i) identifying medicinal plants species by integration of phenotype and DNA markers or chemical markers (purple box); (ii) locating function genes by combination of transcriptomics with degradome and ncRNAs, function genomics with mutagenomics, phenotype with structural genomics (green box); (iii) unearthing metabolic pathways by the integration of transcriptomics and genomics, proteomics, metabolomics, as well as the combination of genomics with transcriptomics and epigenomics (blue box) and (iv) unveiling regulation mechanisms response to stress by integration of transcriptomics and metabolomics, and physiological indices (red box).

phenomena, inhibiting the understanding of certain biological processes. Therefore, more comprehensive bioinformatics sources for integrated analysis and visualisation of different omics datasets are urgently needed. Although a wide range of integrative bioinformatics tools have been proposed for analysing multi-omics datasets, biological interpretation is difficult because of the limitations of the tools themselves. Notably, machine learning and artificial intelligence are promising approaches for integrating and analysing multi-omics datasets based on their predictive performance, flexibility and capability to capture hierarchical and nonlinear features [119].

An increasing number of studies of MPs will lead to further omics databases and bioinformatics tools, enabling research to progress from single- to multi-omics. Integrated multi-omics studies on MPs are expected to expand and facilitate the development of molecular breeding of MPs as well as synthetic biology approaches.

#### Key Points

- We summarise research advances and future trends in current mainstream omics approaches in medicinal plants, including taxonomics, transcriptomics, metabolomics, proteomics, genomics, pangenomics, epigenomics and mutagenomics.

- We review typical bioinformatics tools and databases available for omics dataset analysis of medicinal plants.
- We highlight the integrated patterns of multi-omics studies of medicinal plants and discuss associated prospects and challenges.
- Omics studies of medicinal plants are gradually transitioning from single to multi-omics because of large advantages in integrated multi-omics.

#### Authors' contributions

L.Y. wrote the manuscript; L.H., Y.Y. and X.C. provided valuable advice for the manuscript and Y.L. conceived the initial idea and reviewed the manuscript.

#### Funding

This work was supported by the Yunnan Major Scientific and Technological Projects [grant number KKAN20222025]; National Natural Science Foundation of China [grant number 31960134]; Establishment of Sustainable Use for Valuable Chinese Medicine Resources [grant number 2060302] and Major Science and Technology Special Project of Yunnan Province [grant number 202102AA310034].

## References

- Jamshidi-Kia F, Lorigooini Z, Amini-Khoei H. Medicinal plants: past history and future perspective. *J HerbMed Pharmacol* 2018;**7**: 1–7.
- Chopra B, Dhingra AK. Natural products: a lead for drug discovery and development. *Phytother Res* 2021;**35**:4660–702.
- Liu L, Xu FR, Wang YZ. Traditional uses, chemical diversity and biological activities of *Panax* L. (Araliaceae): a review. *J Ethnopharmacol* 2020;**263**:112792.
- Yu J, Wu X, Liu C, et al. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicol Environ Saf* 2021;**208**:111691.
- Liu FJ, Jiang Y, Li P, et al. Untargeted metabolomics coupled with chemometric analysis reveals species-specific steroidal alkaloids for the authentication of medicinal *Fritillariae* Bulbus and relevant products. *J Chromatogr A* 2020;**1612**:460630.
- Luo M, Li AX, Wang FQ, et al. Integrative analysis of multiple metabolomes and transcriptome revealed color expression mechanism in red skin root syndrome of *panax ginseng*. *Ind Crop Prod* 2022;**177**:114491.
- Cui Y, Fan BL, Xu X, et al. A high-density genetic map enables genome synteny and QTL mapping of vegetative growth and leaf traits in *Gardenia*. *Front Genet* 2021;**12**:802738.
- Fan GY, Liu XC, Sun S, et al. The chromosome level genome and genome-wide association study for the agronomic traits of *Panax notoginseng*. *iScience* 2020;**23**:101538.
- Yang ZJ, Liu GZ, Zhang GH, et al. The chromosome-scale high-quality genome assembly of *Panax notoginseng* provides insight into dencichine biosynthesis. *Plant Biotechnol J* 2021;**19**: 869–71.
- Kim J, Kang SH, Park SG, et al. Whole-genome, transcriptome, and methylome analyses provide insights into the evolution of platycoside biosynthesis in *Platycodon grandiflorus*, a medicinal plant. *Hortic Res* 2020;**7**:112.
- Zhan CS, Li XH, Zhao ZY, et al. Comprehensive analysis of the triterpenoid saponins biosynthetic pathway in *Anemone flaccida* by transcriptome and proteome profiling. *Front Plant Sci* 2016;**7**:1094.
- Zheng H, Yu MY, Han Y, et al. Comparative transcriptomics and metabolites analysis of two closely related *Euphorbia* species reveal environmental adaptation mechanism and active ingredients difference. *Front Plant Sci* 2022;**13**:905275.
- Jiang CH, Bi YK, Mo JB, et al. Proteome and transcriptome reveal the involvement of heat shock proteins and antioxidant system in thermotolerance of *Clematis florida*. *Sci Rep* 2020;**10**: 8883.
- Wang YJ, Dai J, Chen R, et al. miRNA-based drought regulation in the important medicinal plant *Dendrobium huoshanense*. *J Plant Growth Regul* 2022;**41**:1099–108.
- Li DL, Quan CQ, Song ZY, et al. High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. *Front Bioeng Biotechnol* 2020;**8**:623705.
- Knecht AC, Campbell MT, Caprez A, et al. Image Harvest: an open-source platform for high-throughput plant image processing and analysis. *J Exp Bot* 2016;**67**:3587–99.
- Kar S, Garin V, Kholova J, et al. SpaTemHTP: a data analysis pipeline for efficient processing and utilization of temporal High-throughput phenotyping data. *Front Plant Sci* 2020;**11**: 552509.
- Wong TH, But GWC, Wu HY, et al. Medicinal Materials DNA Barcode Database (MMDBD) version 1.5-one-stop solution for storage, BLAST, alignment and primer design. *Database (Oxford)* 2018;**2018**:bay112.
- Perteua M, Kim D, Perteua GM, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 2016;**11**:1650–67.
- Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 2013;**8**: 1494–512.
- Yu YM, Zhang H, Long YP, et al. Plant Public RNA-seq Database: a comprehensive online database for expression analysis of ~45 000 plant public RNA-Seq libraries. *Plant Biotechnol J* 2022;**20**:806–8.
- Zhang H, Zhang F, Yu YM, et al. A comprehensive online database for exploring ~20,000 public Arabidopsis RNA-Seq libraries. *Mol Plant* 2020;**13**:1231–3.
- Shao X, Yang HH, Zhuang X, et al. scDeepSort: a pre-trained cell-type annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network. *Nucleic Acids Res* 2021;**49**:e122.
- Xu ZP, Wang QQ, Zhu XQ, et al. Plant single cell transcriptome Hub (PscTH): an integrated online tool to explore the plant single-cell transcriptome landscape. *Plant Biotechnol J* 2022;**20**: 10–2.
- Chen HY, Yin XX, Guo LB, et al. PlantscRNAdb: a database for plant single-cell RNA analysis. *Mol Plant* 2021;**14**:855–7.
- Wei RM, He SY, Bai SS, et al. Spatial charting of single-cell transcriptomes in tissues. *Nat Biotechnol* 2022;**40**:1190–9.
- Fan Z, Chen RS, Chen XW. SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res* 2020;**48**: D233–7.
- Dai XB, Zhuang ZH, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res* 2018;**46**:W49–54.
- Singh U, Khemka N, Rajkumar MS, et al. PLncPRO for prediction of long non-coding RNAs (lncRNAs) in plants and its application for discovery of abiotic stress-responsive lncRNAs in rice and chickpea. *Nucleic Acids Res* 2017;**45**:e183.
- Chen L, Yu YY, Zhang XC, et al. PcircRNA\_finder: a software for circRNA prediction in plants. *Bioinformatics* 2016;**32**:3528–9.
- Thody J, Moulton V, Mohorianu I. PAREameters: a tool for computational inference of plant miRNA-mRNA targeting rules using small RNA and degradome sequencing data. *Nucleic Acids Res* 2020;**48**:2258–70.
- Yu DL, Lu JJ, Shao WS, et al. MepmiRDB: a medicinal plant microRNA database. *Database (Oxford)* 2019;**2019**:baz070.
- Mathema VB, Duangkumpha K, Wanichthanarak K, et al. CRISP: a deep learning architecture for GC × GC–TOFMS contour ROI identification, simulation and analysis in imaging metabolomics. *Brief Bioinform* 2022;**23**:bbab550.
- Riaz MR, Preston GM, Mithani A. MAPPS: a web-based tool for metabolic pathway prediction and network analysis in the postgenomic era. *ACS Synth Biol* 2020;**9**:1069–82.
- Pang ZQ, Chong J, Zhou GY, et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res* 2021;**49**:W388–96.
- Xue JC, Guijas C, Benton HP, et al. METLIN MS2 molecular standards database: a broad chemical and biological resource. *Nat Methods* 2020;**17**:953–4.
- Gessulat S, Schmidt T, Zolg DP, et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods* 2019;**16**:509–18.

38. Shamsaei B, Chojnacki S, Pilarczyk M, et al. piNET: a versatile web platform for downstream analysis and visualization of proteomics data. *Nucleic Acids Res* 2020;**48**:W85–93.
39. Perez-Riverol Y, Bai JW, Bandla C, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 2022;**50**:D543–52.
40. Sun Q, Zybailov B, Majeran W, et al. PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Res* 2009;**37**:D969–74.
41. Varadi M, Anyango S, Deshpande M, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022;**50**:D439–44.
42. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;**47**:D607–13.
43. Oughtred R, Stark C, Breitkreutz BJ, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res* 2019;**47**:D529–41.
44. Lin JD, Wang SB, Audano PA, et al. SVision: a deep learning approach to resolve complex structural variants. *Nat Methods* 2022. <https://doi.org/10.1038/s41592-022-01609-w>.
45. Chen YW, He Z, Men YH, et al. MetaLogo: a heterogeneity-aware sequence logo generator and aligner. *Brief Bioinform* 2022;**23**:bbab591.
46. Meng FB, Tang Q, Chu TZ, et al. TCMPEG: an integrative database for traditional Chinese medicine plant genomes. *Hortic Res* 2022;**9**:uhac060.
47. Fernández-de-Bobadilla MD, Talavera-Rodríguez A, Chacon L, et al. PATO: pangenome analysis toolkit. *Bioinformatics* 2021;**37**:4564–6.
48. Durant E, Sabot F, Conte M, et al. Panache: a web browser-based viewer for linearized pangenomes. *Bioinformatics* 2021;**37**:4556–8.
49. Guignon V, Toure A, Droc G, et al. GreenPhylDB v5: a comparative pangenomic database for plant genomes. *Nucleic Acids Res* 2021;**49**:D1464–71.
50. Cao F, Zhang Y, Cai YC, et al. Chromatin interaction neural network (ChINN): a machine learning-based method for predicting chromatin interactions from DNA sequences. *Genome Biol* 2021;**22**:226.
51. Chow CN, Lee TY, Hung YC, et al. PlantPAN3.0: a new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. *Nucleic Acids Res* 2019;**47**:D1155–63.
52. Mitrofanov A, Alkhnbashi OS, Shmakov SA, et al. CRISPRidentifiy: identification of CRISPR arrays using machine learning approach. *Nucleic Acids Res* 2021;**49**:e20.
53. Ghosh S, Datta A, Choi H. multiSLIDE is a web server for exploring connected elements of biological pathways in multi-omics data. *Nat Commun* 2021;**12**:2279.
54. Liu TY, Salguero P, Petek M, et al. PaintOmics 4: new tools for integrative analysis of multi-omics datasets supported by multiple pathway databases. *Nucleic Acids Res* 2022;**50**:W551–9.
55. Zhou GY, Ewald J, Xia JG. OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res* 2021;**49**:W476–82.
56. Zhou GY, Pang ZQ, Lu Y, et al. OmicsNet 2.0: a web-based platform for multi-omics integration and network visual analytics. *Nucleic Acids Res* 2022;**50**:W527–33.
57. He SM, Yang L, Ye S, et al. MPOD: applications of integrated multi-omics database for medicinal plants. *Plant Biotechnol J* 2022;**20**:797–9.
58. Su XJ, Yang LL, Wang DL, et al. 1 K Medicinal Plant Genome Database: an integrated database combining genomes and metabolites of medicinal plants. *Hortic Res* 2022;**9**:uhac075.
59. Mishra P, Kumar A, Nagireddy A, et al. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnol J* 2016;**14**:8–21.
60. Li H, Wang Y, Wang Z, et al. Microarray and genetic analysis reveals that csa-miR159b plays a critical role in abscisic acid-mediated heat tolerance in grafted cucumber plants. *Plant Cell Environ* 2016;**39**:1790–804.
61. Wood IP, Pearson BM, Garcia-Gutierrez E, et al. Carbohydrate microarrays and their use for the identification of molecular markers for plant cell wall composition. *PNAS* 2017;**114**:6860–5.
62. Zheng LL, Zhou C, Li TH, et al. Global transcriptome analysis reveals dynamic gene expression profiling and provides insights into biosynthesis of resveratrol and anthraquinones in a medicinal plant *Polygonum cuspidatum*. *Ind Crop Prod* 2021;**171**:113919.
63. Tian M, Zhang X, Zhu Y, et al. Global transcriptome analyses reveal differentially expressed genes of six organs and putative genes involved in (iso) flavonoid biosynthesis in *Belamcanda chinensis*. *Front Plant Sci* 2018;**9**:1160.
64. Dhiman N, Kumar A, Kumar D, et al. De novo transcriptome analysis of the critically endangered alpine Himalayan herb *Nardostachys jatamansi* reveals the biosynthesis pathway genes of tissue-specific secondary metabolites. *Sci Rep* 2020;**10**:17186.
65. Zhang JY, Cun Z, Wu HM, et al. Integrated analysis on biochemical profiling and transcriptome revealed nitrogen-driven difference in accumulation of saponins in a medicinal plant *Panax notoginseng*. *Plant Physiol Biochem* 2020;**154**:564–80.
66. Lei HY, Niu TZ, Song HF, et al. Comparative transcriptome profiling reveals differentially expressed genes involved in flavonoid biosynthesis between biennial and triennial *Sophora flavescens*. *Ind Crop Prod* 2021;**161**:113217.
67. Xu ZC, Peters RJ, Weirather J, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J* 2015;**82**:951–61.
68. Seyfferth C, Renema J, Wendrich JR, et al. Advances and opportunities in single-cell transcriptomics for plant research. *Annu Rev Plant Biol* 2021;**72**:847–66.
69. Datlinger P, Rendeiro AF, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* 2017;**14**:297–301.
70. Giacomello S, Salmen F, Terebieniec BK, et al. Spatially resolved transcriptome profiling in model plant species. *Nat Plants* 2017;**3**:17061.
71. Liu YY, Li CH, Han Y, et al. Spatial transcriptome analysis on peanut tissues shed light on cell heterogeneity of the peg. *Plant Biotechnol J* 2022;**20**:1648–50.
72. Liu C, Leng J, Li YL, et al. A spatiotemporal atlas of organogenesis in the development of orchid flowers. *Nucleic Acids Res* 2022;gkac771.
73. Lin SS, Chen YH, Lu MYJ, et al. Degradome sequencing in plants. *Methods Mol Biol* 2019;**1932**:197–213.
74. Zheng Y, Chen K, Xu ZN, et al. Small RNA profiles from *Panax notoginseng* roots differing in sizes reveal correlation between miR156 abundances and root biomass levels. *Sci Rep* 2017;**7**:9418.

75. Yang WZ, Qiao X, Li K, et al. Identification and differentiation of *Panax ginseng*, *Panax quinquefolium*, and *Panax notoginseng* by monitoring multiple diagnostic chemical markers. *Acta Pharm Sin B* 2016;**6**:568–75.
76. Shang XH, Huang D, Wang Y, et al. Identification of nutritional ingredients and medicinal components of *Pueraria lobata* and its varieties using UPLC-MS/MS-based metabolomics. *Molecules* 2021;**26**:6587.
77. Afzan A, Kasim N, Ismail NH, et al. Differentiation of *Ficus deltoidea* varieties and chemical marker determination by UHPLC-TOFMS metabolomics for establishing quality control criteria of this popular Malaysian medicinal herb. *Metabolomics* 2019;**15**:35.
78. Rai A, Saito K, Yamazaki M. Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J* 2017;**90**:764–87.
79. Fox BW, Schroeder FC. Toward spatially resolved metabolomics. *Nat Chem Biol* 2020;**16**:1039–40.
80. Li B, Bhandari DR, Janfelt C, et al. Natural products in *Glycyrrhiza glabra* (licorice) rhizome imaged at the cellular level by atmospheric pressure matrix-assisted laser desorption/ionization tandem mass spectrometry imaging. *Plant J* 2014;**80**:161–71.
81. Zhao WH, Zhang YD, Shi YP. Visualizing the spatial distribution of endogenous molecules in wolfberry fruit at different development stages by matrix-assisted laser desorption/ionization mass spectrometry imaging. *Talanta* 2021;**234**:122687.
82. Mergner J, Kuster B. Plant proteome dynamics. *Annu Rev Plant Biol* 2022;**73**:67–92.
83. Chen PL, Wei XY, Qi QT, et al. Study of terpenoid synthesis and prenyltransferase in roots of *Rehmannia glutinosa* based on iTRAQ quantitative proteomics. *Front Plant Sci* 2021;**12**:693758.
84. Kim SW, Gupta R, Min CW, et al. Label-free quantitative proteomic analysis of *Panax ginseng* leaves upon exposure to heat stress. *J Ginseng Res* 2019;**43**:143–53.
85. Pedrete TA, Hauser-Davis RA, Moreira JC. Proteomic characterization of medicinal plants used in the treatment of diabetes. *Int J Biol Macromol* 2019;**140**:294–302.
86. Giordano D, Facchiano A, D'Auria S, et al. A hypothesis on the capacity of plant odorant-binding proteins to bind volatile isoprenoids based on in silico evidences. *Elife* 2021;**10**:e66741.
87. Liu KD, Yuan CC, Li HL, et al. A qualitative proteome-wide lysine crotonylation profiling of papaya (*Carica papaya* L.). *Sci Rep* 2018;**8**:8230.
88. Elhabashy H, Merino F, Alva V, et al. Exploring protein-protein interactions at the proteome level. *Structure* 2022;**30**:462–75.
89. Rehman F, Gong HG, Li Z, et al. Identification of fruit size associated quantitative trait loci featuring SLAF based high-density linkage map of goji berry (*Lycium spp.*). *BMC Plant Biol* 2020;**20**:474.
90. Liao B, Shen X, Xiang L, et al. Allele-aware chromosomal-level genome assembly of *Artemisia annua* reveals the correlation between ADS expansion and artemisinin yield. *Mol Plant* 2022;**15**:1310–28.
91. Zhang YJ, Shen Q, Leng L, et al. Incipient diploidization of the medicinal plant *Perilla* within 10,000 years. *Nat Commun* 2021;**12**:5508.
92. Varshney RK, Graner A, Sorrells ME, et al. Genomics-assisted breeding for crop improvement. *Trends Plant Sci* 2005;**10**:621–30.
93. Van Bakel H, Stout JM, Cote AG, et al. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol* 2011;**12**:R102.
94. Yan L, Wang X, Liu H, et al. The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. *Mol Plant* 2015;**8**:922–34.
95. Guo L, Winzer T, Yang XF, et al. The opium poppy genome and morphinan production. *Science* 2018;**362**:343–7.
96. Shen C, Du HL, Chen Z, et al. The chromosome-level genome sequence of the autotetraploid *Alfalfa* and resequencing of core germplasms provide genomic resources for *Alfalfa* research. *Mol Plant* 2020;**13**:1250–61.
97. Jiang L, Lin MF, Wang H, et al. Haplotype-resolved genome assembly of *Bletilla striata* (Thunb.) Reichb. f. to elucidate medicinal values. *Plant J* 2022;**111**:1340–53.
98. Wang C, Hao XL, Wang Y, et al. Identification of WRKY transcription factors involved in regulating the biosynthesis of the anti-cancer drug camptothecin in *Ophiorrhiza pumila*. *Hortic Res* 2022;**9**:uhac099.
99. Wang ZH, Wang XF, Lu TY, et al. Reshuffling of the ancestral core-eudicot genome shaped chromatin topology and epigenetic modification in *Panax*. *Nat Commun* 2022;**13**:1902.
100. Li W, Liu JN, Zhang HY, et al. Plant pan-genomics: recent advances, new challenges, and roads ahead. *J Genet Genomics* 2022;**49**:833–46.
101. Yang YD, Saand MA, Huang LY, et al. Applications of multi-omics technologies for crop improvement. *Front Plant Sci* 2021;**12**:563953.
102. Liu YC, Du HL, Li PC, et al. Pan-genome of wild and cultivated soybeans. *Cell* 2020;**182**:162–76.e13.
103. Yuan YX, Bayer PE, Batley J, et al. Current status of structural variation studies in plants. *Plant Biotechnol J* 2021;**19**:2153–63.
104. Novik KL, Nimmrich I, Genc B, et al. Epigenomics: genome-wide study of methylation phenomena. *Curr Issues Mol Biol* 2002;**4**:111–28.
105. Muthamilarasan M, Singh NK, Prasad M, et al. Multi-omics approaches for strategic improvement of stress tolerance in underutilized crop species: a climate change perspective. *Adv Genet* 2019;**103**:1–38.
106. Cokus SJ, Feng SH, Zhang XY, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 2008;**452**:215–9.
107. Guo WQ, Ma H, Wang CZ, et al. Epigenetic studies of Chinese herbal medicine: pleiotropic role of DNA methylation. *Front Pharmacol* 2021;**12**:790321.
108. Zhao T, Zhan ZP, Jiang DH. Histone modifications and their regulatory roles in plant development and environmental memory. *J Genet Genomics* 2019;**46**:467–76.
109. Reynoso MA, Kajala K, Bajic M, et al. Evolutionary flexibility in flooding response circuitry in angiosperms. *Science* 2019;**365**:1291–5.
110. Luo LH, Gribskov M, Wang SF. Bibliometric review of ATAC-seq and its application in gene expression. *Brief Bioinform* 2022;**23**:bbac061.
111. Cai JH, Wu ZL, Song ZY, et al. ATAC-seq and RNA-seq reveal the role of AGL18 in regulating fruit ripening via ethylene-auxin crosstalk in papaya. *Postharvest Biol Technol* 2022;**191**:111984.
112. Jiang WZ, Zhou HB, Bi HH, et al. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Res* 2013;**41**:e188.
113. Dey A. CRISPR/Cas genome editing to optimize pharmacologically active plant natural products. *Pharmacol Res* 2021;**164**:105359.
114. Gu XY, Liu LJ, Zhang HW, et al. Transgene-free Genome Editing in Plants. *Front Genome Editing* 2021;**3**:805317.

115. Zhou MM, Yang GQ, Sun GL, et al. Resolving complicated relationships of the *Panax bipinnatifidus* complex in south-western China by RAD-seq data. *Mol Phylogenet Evol* 2020;**149**:106851.
116. Zhang YT, Cui JB, Hu HL, et al. Integrated four comparative-omics reveals the mechanism of the terpenoid biosynthesis in two different overwintering *Cryptomeria fortunei* phenotypes. *Front Plant Sci* 2021;**12**:740755.
117. Weckwerth W, Ghatak A, Bellaire A, et al. Panomics meets germplasm. *Plant Biotechnol J* 2020;**18**:1507–25.
118. Peng Z, Wang Y, Zuo WT, et al. Integration of metabolome and transcriptome studies reveals flavonoids, abscisic acid, and nitric oxide comodulating the freezing tolerance in *Liriope spicata*. *Front Plant Sci* 2021;**12**:764625.
119. Li RF, Li LX, Xu YG, et al. Machine learning meets omics: applications and perspectives. *Brief Bioinform* 2022;**23**:bbab460.