**Use of artificial intelligence and bioinformatics for crop improvement to ensure future food security**

Dibyendu Seth1†

Email id: deep032002@gmail.com

**ORCiD**: 0000-0001-9642-1421

(<https://orcid.org/0000-0002-8848-3275>)

Sandip Debnath1†

Email id: sandip.debnath@visva-bharati.ac.in

**ORCiD**: 0000-0002-0234-6633

(<https://orcid.org/0000-0002-0234-6633>)

Sourish Pramanik1†

Email id: sourishpramanik2002@gmail.com

**ORCiD**: 0000-0002-8848-3275

(<https://orcid.org/0000-0002-8848-3275>)

Biswajit Pramanik1†\*

Email id: biswajit1996pramanik@gmail.com

**ORCiD**: 0000-0002-6016-8935 (<https://orcid.org/0000-0002-6016-8935>)

**\*Corresponding author**: biswajit1996pramanik@gmail.com

**†All the authors contributed equally**

**Address:** 1Department of Genetics and Plant Breeding, Palli-Siksha Bhavana (Institute of Agriculture), Visva-Bharati University, Sriniketan, Birbhum, West Bengal, India

**ABSTRACT**

**Background:** In order to feed the world's alarmingly rising population, modern plant breeders confront a steep uphill struggle. At the same time, agricultural crop productivity has been harmed by a variety of factors such as insect pest attack, disease severity, and nutrient insufficiency. Our demand for meeting population needs is growing by the day. It is necessary to engage in interdisciplinary approaches to identify solutions to existing difficulties in order to achieve the same.

**Main body:** We have lately seen a paradigm shift toward using omics knowledge, methodologies, and technology to increase agricultural yield. By providing extensive data on crop genotypes and phenotypes, the omics era has opened several avenues. This will show to be an essential instrument for increasing agricultural productivity and farmer revenue. Recently, Artificial Intelligence (AI) has emerged in the agriculture industry. AI in agriculture is distinguished by its adaptability, high performance, precision, and cost-effectiveness. Understanding plant genomics might facilitate the identification, cloning, and sequencing of genes that help plants withstand damaging environmental influences. Recently, machine learning has been included as a viable interdisciplinary strategy for improving and upgrading the agricultural sector, particularly the food industry, which is quickly changing.

**Conclusion:** This paper investigates plant multi-omics as a solution to future food security challenges by investigating the interaction between bioinformatics and artificial intelligence as a possible method for discovering the genome and its variants in order to genetically modify crops in the future.

**Keywords**—artificial intelligence, bioinformatics, deep learning, genomics, metabolomics, proteomics, transcriptomics,

1. **Background**

Concerning population expansion, environmental degradation, and climate change, sustainable agricultural production and food security are two significant issues [1]. More than two-thirds of daily human energy consumption is derived from crops. The strain is on agriculture to provide greater agricultural yields as the global population grows. Climate change, the dearth of arable land, and water restrictions provide further agricultural challenges. In addition, the recent increase in demand for biofuel crops has created a new market for agricultural goods, therefore exacerbating the problem of food security. In order to overcome these problems, several applications of genetic data have generated various opportunities for combining the rich advantages of sub-systems biology, integrative biology, and large-scale systematic functional genomics programmes. Significant gene sequences and their roles have been uncovered, and the field of plant molecular biology is advancing. Numerous of these sequences are associated with agricultural yields (production), crop quality (protein and carbohydrate content), and resilience to biotic and abiotic threats [2, 3].

Since 2000, scientists have gained access to the whole genome of the mustard plant *Arabidopsis* *thaliana*. (International *Arabidopsis* Genome Initiative, 2000, p. 149) Rice (*Oryza sativa* cv. japonica) has had its whole genome sequenced since 2005 (International Rice Genome Sequencing Project 2005) [4]. The grape genome was the first to use a combination of 454 sequencing and Sanger sequencing, whereas rice was still sequenced using BAC and Sanger sequencing [4]. 40 of the 55 plant genomes sequenced as of 2013 were agricultural [5]. As of April 2022, the Genbank database (<https://www.ncbi.nlm.nih.gov/genbank/statistics/>) has a total of 237520318 sequences.

Obtaining a functional knowledge of plant genomes necessitates the application of bioinformatics, which is necessary for processing and analysing vast amounts of genomic data. Numerous algorithms developed for short reads have difficulty matching long reads for third-generation sequencing data. Complete data may be supplied by genome-wide association studies (GWAS), variant calling, and comparative genomic analysis with the goal of supporting crop development. In many aspects of crop breeding, including genome-wide association studies (GWAS), where genomic sequencing of crop populations may allow gene-level resolution of agronomic variation, quantitative trait locus (QTL) mapping, and more, genomics is gaining importance. This is owing to the ease with which breeders may now access genetic data. Recent advancements in multi-omics have presented a tremendous opportunity for crop improvement. The "omic space," a conceptual model with levels ranging from "genome" to "phenome," has been imagined [6]. This initiates important study on the structure and function of genes in relation to phenotypic changes in plants. As a consequence of advancements in gene expression methods, gene-specific molecular breeding, and the relationship between the genome, proteome, and metabolome, several web-based databases have been created to store the huge volumes of data. Genome sequencing as a tool for genomics is helpful in the amelioration of agronomic characteristics so that genetic potential may be used to increase yield. Due to the declining cost of Deoxyribonucleic acid (DNA) sequencing, crop genome sequencing has increased rapidly over the last decade, presenting breeders with an abundance of options.

Machine learning is an alternative approach for prediction and categorization (ML). Machine learning is an area of computer science that uses mathematical and statistical techniques to progressively train models without direct programming. ML generates several algorithms that progressively acquire knowledge from training data and sample data in order to make predictions. Support vector machines (SVM), boosting, random forests, and Reproducing Kernel Hilbert Space (RKHS) have all been used in a variety of non-parametric machine learning (ML) studies on plants and animals. The fundamental advantage of using ML models for genomic selection (GS) is that they identify patterns from data without any prior assumptions; hence, they account for all variants, their interactions, and environmental variables [7]. A scientist is interested in the effect of each nucleotide in addition to accurately predicting phenotypes from a plant's DNA. Deep learning may produce very accurate predictions, but the models are often difficult to comprehend, making it challenging to examine biological processes via inference. As a consequence, academics have not yet paid much attention to deep learning (DL).

Understanding the flow of biological information underlying complex features requires an alternative systems biology approach comprising the integration of numerous omics data, modelling, and prediction of cellular processes. This method allows a thorough knowledge of the dynamic system in which different levels of biological structure interact with the external environment in order to manifest phenotype. Several omics methodologies (such as genomics, transcriptomics, proteomics, metabolomics, epigenomics, and single-cell omics) are gaining prominence in the area of plant sciences as sequencing costs decrease and knowledge levels increase. Therefore, the sequencing and resequencing information obtained for different crops has allowed the identification of novel alleles from diverse sources conceivable, regardless of the availability of the genome sequence. Therefore, a systematic strategy is necessary to increase sustainable agricultural yields via a greater knowledge of plant network biology (Figure 1). In this paper, the most recent advancements in bioinformatics and artificial intelligence are emphasized in an effort to improve agricultural plants employing genomes, transcriptomics, proteomics, metabolomics, and systems biology, among others.



**Figure 1: Pictorial depiction of genomics, transcriptomics, proteomics, metabolomics and phenomics’ integrated application in crop improvement**

1. **Next Generation Sequencing**

Using Illumina short reads and Sanger sequencing, the cucumber genome was sequenced in 2009 [8], which encouraged the fast use of next-generation sequencing (NGS) [5]. Genome sequence data is used to uncover genes and gene families, as well as coding and non-coding regions, regulatory genes, and repetitive sequences. Since the advent of NGS, genome sequencing and resequencing have become common place in several areas of plant biology. By 2013, 55 plant genomes have been sequenced, including 40 crop genomes [5]. Low-cost high-throughput approaches have also been used to uncover genome-wide molecular phenotypes with several dimensions [9]. Using next-generation sequencing technology and reference genome sequence data, it is possible to identify individual, strain, and/or population differences. Mapping sequence fragments onto a given reference genome data set reliably identifies nucleotide polymorphisms, which is vital for any genetic research.

Despite the fact that plant genome assembly is still challenging due to long repetitive regions, large genome sizes, and frequent polyploidy, advances in sequencing technologies (third generation sequencing technologies) and bioinformatics tools have enabled rapid advancements since the rice genome was sequenced and assembled in 2005 [10]. By utilising reads spanning challenging regions, such as those with high levels of repeated sequence, third-generation sequencing facilitates the construction of high-quality *de novo* assemblies of the whole genome and sheds light on the remaining complex of repeat sequences, including structural variants. In addition, full-length sequenced transcripts (isoform sequencing) obtained by third-generation sequencing methods enable accurate examination of exons, splice sites, and alternatively spliced areas, which aids in genome annotation. Consequently, high-quality crop reference genomes are becoming more widely accessible, enabling downstream methodologies such as comparative genomic analysis, variant calling, and genome-wide association studies (GWAS) to provide complete data for crop development. Third-generation sequencing, such as single-molecule real-time sequencing by Pacific Biosciences (PacBio) and sequencing by Oxford Nanopore Technologies (ONT), has made feasible longer reads and more accurate and contiguous genome assemblies. In recent years, third-generation sequencing technology capable of generating lengthy reads longer than 10 kb have been developed, making crop genome sequencing an additional important tool.

Long-read sequencing, long-range mapping, and chromosomal conformation capture have made highly contiguous chromosome-level plant genome assemblies feasible even for non-model crop species and smaller facilities. Additionally, long-read sequencing successfully finds repetitive sequences. New optical mapping techniques, like as BioNano Genomics, enable rapid labelling of giant DNA molecules exceeding 250 kb, enabling the detection of structural variation and low-cost scaffolding. Chromosome conformation capture sequencing (Hi-C) is a third-generation mapping technique that relies on the physically tight binding of DNA segments. When Hi-C measurements and optical mapping are coupled, chromosomal phasing and scaffolding may be accomplished to a considerably higher degree. Mascher et al. [11] employed short reads, optical, and chromatin interaction mapping data to reconstruct the highly repetitive and polyploid barley genome with a N50 of 1.9 Mb. The enhanced sequence continuity provided by third-generation sequencing may facilitate genomics-based breeding techniques such as trait mapping. Construction of improved, highly contiguous crop genomes is the most successful use of third-generation sequencing to breeding. Although NGS has a number of advantages, one of its significant disadvantages is its inherent biases and imperfect repetitive sequence matching, which results in highly fragmented draught genome assemblies and consequently makes it more difficult to research hidden indels and structural variants.

Historically, crop breeding used phenotypic selection and crossing cycles to generate better genotypes by genetic recombination. Through genomics-based breeding, it is now possible to identify genetic diversity in crop species, which may be used to develop climate-resistant crops [12]. Once genome sequences are available, all genes and genetic variants linked with agronomic characteristics may be identified, and breeding changes can be evaluated at the genotype level. The availability of genomic data to breeders is increasing the significance of genomic data in several aspects of crop breeding, including QTL mapping and GWAS, where genomic sequencing of crop populations may provide gene-level resolution of agronomic variation. The databases listed in Table 1 pertain to genomics research.

**Table 1. Databases in use of plant genomics research**

|  |  |
| --- | --- |
| **Database** | **URL** |
| Phytozome v8.0 | <http://www.phytozome.net/Phytozome_info.php>  |
| Gramene | <http://www.gramene.org/>  |
| Home—BioProject—NCBI | <http://www.ncbi.nlm.nih.gov/sites/entrez?db=bioproject>  |
| BLAST: Basic Local Alignment Search Tool | <http://blast.ncbi.nlm.nih.gov/Blast.cgi>  |
| GrainGenes Class Browser | <http://wheat.pw.usda.gov/cgi-bin/graingenes/browse.cgi?class=marker>  |
| PlantGDB— Resource Plant Comparative Genomics | <http://www.plantgdb.org/>  |
| TreeView | <http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>  |
| GenBank | <https://www.ncbi.nlm.nih.gov/genbank/>  |
| European Molecular BiologicalLaboratory (EMBL) | <https://www.embl.org/>  |
| KnetMiner (Knowledge Network Miner) | <https://knetminer.com/>  |
| LALIGN Server | <http://www.ch.embnet.org/software/LALIGN_form.html>  |
| PopGene | <http://www2.unil.ch/popgen/softwares/fstat.htm>  |
| Arlequin 3.11 | <http://cmpg.unibe.ch/software/arlequin3/>  |
| PRIMER-E | <http://www.primer-e.com/>  |

1. **QTL mapping**

Genomics is the study of an organism's genome, which consists of all of its genes and complete sets of DNA. In contrast to genetics, which focuses on genes and their role in inheritance, genomics examines the collective description and quantification of an organism's genes [13]. The advancement of genomics facilitates the understanding of system biology and the most complex biological systems. Generating genomic resources in specific ways, such as through molecular markers, transcriptome assemblies, biparental mapping populations, genetic linkage maps, comparative genome mapping, and functional genomics, enables the development of new understandings that will aid crop plants in living sustainably.

Multiple methods for high-throughput genotyping have made feasible QTL cloning, the construction of genetic maps, and marker-assisted selections in distinct segregating populations. In addition to the discovery of particular genes associated to complex traits by QTL analysis, genetic markers that span a wide area of a genome provide the study of genetic diversity in relation to natural variations. As genome sequencing and substantial Expressed Sequenced Tag (EST) study in several species have progressed, these sequence data set have become valuable sequence resources for the creation of molecular markers. This is because all SNP marker sets are combined in the predicted model. For several species, including barley, wheat, maize, melon, *Brassica*, common bean, and sunflower, computational identification of EST base single-nucleotide polymorphisms and/or EST-SNP markers for finding sequence-tagged site markers has progressed [14, 15, 16, 17, 18]. QTL analysis facilitates the tracking of trait connections with genetic regions. As additional QTLs are found, a meta-QTL analysis is required to more correctly forecast QTL locations than individuals. The computational tool MetaQTL reduces the QTL's confidence interval in order to accurately predict the correct QTL location and effect. SolQTL and RASQUAL both provide QTL analyses with low bias, the presentation of QTL data, and the linking of QTL data to other genome databases. In maize [19], cotton [20], soybean [21], and wheat [22], Meta-QTL analysis has been used to identify traits associated with crop development and abiotic and biotic responses. However, there are two disadvantages to QTL mapping: (1) inadequate resolution due to coarse mapping makes it difficult to differentiate between pleiotropic and physically adjacent genes; and (2) only the allelic diversity present in the parents of the segregating population can be evaluated.

1. **GWAS & Genomic selection**

GWAS is an alternative to QTL mapping that circumvents its limitations. GWAS relies on wild populations, while QTL analysis use bi-parental populations derived from controlled crosses. This enables greater clarity when finding multiple recombination events and analysing the natural variations associated with phenotypic differences. GWAS, which achieves higher mapping resolution than QTL analysis, reveals marker-trait associations (MTA) that may be related to the level of linkage disequilibrium (LD) between polymorphic markers across a diverse range of genotypes. GWAS is preferred above QTL analysis when a breeder wishes to examine a broad genetic basis for studying many candidate genes to be used in breeding programmes. Before being used to animals and a few model species, GWAS was initially used to analyse complex human characteristics. GWAS has been applied to various crops, including canola, rice, soybean, maize, and wheat during the last decade [23, 24, 25, 26]. However, it may be difficult to identify the breeder when the characteristics are polygenic. On the basis of genomic estimated breeding values (GEBV) in a group of variants, GS gives an advantage in this instance. Using entire SNP marker sets in a projected model may overcome the problem of insufficient QTL translations caused by biparental populations. It has been shown that Lolium perenne GS-based breeding techniques based on computer simulations reduce the four-year breeding cycle. GBS has been used on maize breeding lines to identify 55,000 SNP markers and on elite wheat breeding lines to assess high yield and stem rust resistance [27]. Figure 2 displays a few characteristics that were analysed using GWAS for five key crops.



**Figure 2: Few traits or genes studied via Genome Wide Association Study (GWAS) of 5 major crops**

**(Created in Biorender.com) (**[**https://biorender.com**](https://biorender.com)**)**

1. **Cis-Regulatory Elements (CREs) for Crop Breeding**

Concentration is moving ahead to govern the cis-regulatory elements, the promoters, enhancers, and regulators to control certain gene expressions. CREs are related with chromatin that binds to proteins but are not as expressive as genes, making it more challenging to discover. However, CRE targeting is a good option when the goal is to regulate the gene rather than to completely eliminate it. By using bioinformatic methods like ChIP-seq [28], ATAC-seq [29], DNase I hypersensitivity mapping [30], word-counting [31], and conserved sequence analysis of [32], it is possible to discover regulatory regions in open chromatin.

Despite our poor understanding of CREs, modern methods have made it simpler to identify regulatory areas, while experimental research is still required to establish a specific CRE's contribution to the target gene's expression. Plant CARE's database on plant CREs is called Plant Cis-Acting Regulatory Elements. Numerous cases have been documented up to this point, including the suppression of the gene GRAIN WIDTH 7 due to a mutation in the rice CRE, which produced rice with slender grains despite its negative effect on yield, and the variation in tomato seed compartment numbers caused by the regulation of the WUSCHEL (WUS) and CLAVATA (CLV3) promoters. It is anticipated that CREs linked to desirable features can be found when a mutant library is constructed using the expression data of mutant lines.

1. **Proteomics**

The static nature of the genome prohibits it from having a link with mRNA and proteins due to post-transitional modifications (PTM), function, and localization. To appreciate the function proteins, play in the evolution of plants, it is vital to investigate their structure and interactions. Proteomics is a high-performance technology for detecting and measuring protein performance in a cell or organism at a particular moment. In the majority of proteomics systems, the three fundamental processes are identification or quantification, protein extraction, and separation. We have progressed to the second generation of functional proteomics, which includes quantitative proteomics, subcellular proteomics, various modifications, and protein-protein interactions, as a result of recent, rapid technological advances in proteomics (e.g., advances in mass spectrometry equipment and methodological developments in protein quantification).

Numerous ways are applied to expand the knowledge, resolution, and coverage of the plant proteome. Several aspects, such as the availability of resources, facilities, and applications, such as global or targeted profiling, dictate the research methodology for the proteome. Using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), which combines two-dimensional gel electrophoresis (2-DE) with isoelectric focusing (IEF) as the first dimension and SDS-PAGE as the second, it is possible to separate proteins with high repeatability and resolution. In addition, chromatography-based separation methods, including as gel filtration, ion exchange, and affinity chromatography, may be used to isolate proteins depending on their physicochemical properties. Peptide mass fingerprinting is a prevalent approach for identifying proteins at present. It begins with the breakdown of proteins into peptides, followed by the precise measurement of the mass of the peptides using mass spectrometry (MS). In-gel electrophoresis was created to circumvent the 2D-PAGE limitations of gel-to-gel variation and limited reproducibility (DIGE). DIGE is employed to comprehend how protein expression alters in response to various biotic and abiotic stimuli. Three dimensions are added to two-dimensional gel electrophoresis (2D-PAGE) to avoid co-migration interferences. It provides exceptionally exact identification of proteins and PTMs utilising two separate buffers with varied ion carriers [33].

MS employs a variety of computer algorithms to identify proteins based on peptide mass and fragmentation (MS/MS) data. In all, there are three stages. To transform molecules into gas-phase ions, mass-based ion separation is first done in an electro or magnetic field, followed by measurement of the separated ions with a certain m/z value. Electrospray ionisation (ESI), surface-enhanced laser desorption/ionization (SELDI), and matrix-assisted laser desorption ionisation (MALDI) are ionisation methods. The gel-free techniques, such as quantitative approaches, tag-based labelling, metabolic labelling, and label-free methods, may address the drawbacks of gel-based approaches, such as their inability to separate the whole proteome and poor identification of less abundant proteins.

The discovery of significant proteomic modifications, such as expression, interaction, and modification that are associated with genetic variations and/or observable phenotypic changes also needs a quantitative proteomics approach. Before 2-D electrophoresis in DIGE (Differential Gel Electrophoresis), protein samples are labelled with fluorescent dyes to accurately discriminate between proteins. ICAT (Isotope-Coded Affinity Tagging) utilises *in vitro* isotopic labelling to quantify protein, with labelled tryptic peptides first separated by chromatography and then identified by MS. iTRAQ (Isobaric Tagging for Relative and Absolute Quantification) quantifies proteins using isobaric tags. Crop breeders employ this technique to find markers for biotic and abiotic stresses in order to generate genetically engineered crops. Stable Isotope Labelling by Amino Acid in Cell Culture (SILAC) takes use of *in vivo* labelling of cell populations cultured in N14 or N15 medium and has been shown to be effective in detecting proteome abnormalities caused by post-translational modifications under stressful conditions [34]. MudPIT (Multi-Dimensional Protein Identification Technology) is used for complex multi-dimensional protein analysis. Separation of digested proteins using biphasic or triphasic microcapillary columns is followed by tandem mass spectrometry. Using this method, the systems underlying the regulation of rice tiller numbers have been revealed. Figure 3 shows the methodologies used in proteome investigations, whereas Table 2 includes the major databases utilised in proteomic research.



**Figure 3: Different techniques used in proteomic studies (Created in Biorender.com) (**[**https://biorender.com**](https://biorender.com)**)**

**Table 2: Databases for Proteomic study**

|  |  |
| --- | --- |
| Database | Link |
| Swiss Institute of Bioinformatics’ ExpasySWISS-2DPAGE database | <http://au.expasy.org/ch2d/> |
| Kazusa DNA Research Institute’s Cyano2Dbase | <http://bacteria.kazusa.or.jp/cyano_legacy/Synechocystis/cyano2D/index.html> |
| rice proteome database | <http://gene64.dna.affrc.go.jp/RPD/> |
| Nottingham *Arabidopsis* Stock Centre(NASC) Proteomics database | [http://proteomics.*Arabidopsis*.info/](http://proteomics.arabidopsis.info/) |
| SUB-cellular location database for *Arabidopsis*proteins (SUBA) | <http://suba.plantenergy.uwa.edu.au/> |
| The soybean proteome database | <http://proteome.dc.affrc.go.jp/cgi-bin/2d/2d_view_map.cgi> |
| The *Arabidopsis* Protein PhosphorylationSite Database (PhosPhAt) | <http://phosphat.mpimp-golm.mpg.de/> |
| Protein data bank, PDB | <http://www.pdb.org/pdb/home/home.do> |
| The RIKEN SGPI | <http://www.rsgi.riken.go.jp/rsgi_e/index.html> |
| Genomes TO Protein structures and functions(GTOP) database | <http://spock.genes.nig.ac.jp/~genome/gtop.html> |
| CATH | <http://www.cathdb.info/> |
| Structural Classification of Proteins (SCOP) database | <http://scop.mrc-lmb.cam.ac.uk/scop/> |
| PRoteomics IDEntification database (PRIDE) | <https://www.ebi.ac.uk/pride/> |
| Peptide Atlas | <http://www.peptideatlas.org/>  |
| Mass Spectrometry Interactive Virtual Environment (MassIVE) | <https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>  |
| Plant Proteomics Database(PPDB) | <http://ppdb.tc.cornell.edu/>  |
| 1001 Proteomes (Discontinued)  | <https://www.heazleome.org/tools.html>  |
| GelMap | <https://www.gelmap.de/>  |
| Peptide Atlas SRM ExperimentLibrary (PASSEL) | <http://www.peptideatlas.org/passel/>  |

1. **Transcriptomics**

As previously said, the genome is fixed and cannot reveal how much of it is being expressed. Consequently, transcriptomic approaches are used to quantify the fraction of the genome that is expressed. 1-2 percent of the functional genome is transcription. Transcriptomics is a comprehensive, high-throughput examination of gene expression to find cis-regulatory patterns, predict gene function, and screen candidate genes. Transcriptomics, which investigates how genes are expressed in an organism across many contexts, tissues (spatial transcriptome), and time periods, offers a comprehensive investigation of this expressed genome (temporal transcriptome). It is well knowledge that hybridization-based techniques, such as those used in microarrays and GeneChips, may be used to provide comprehensive gene expression profiles for many species. Deep sequencing of small snippets of expressed RNAs, including sRNAs, is rapidly becoming an efficient method for use with genome-sequenced species. Recent efforts in the field of transcriptomics have enhanced publicly accessible databases, which are a valuable source of secondary applications for co-expression and comparative research.

In the 1970s and 1980s, reverse transcriptase was used to convert cDNA into RNA transcripts in the silk moth [35], and in the 1990s, Sanger sequencing was used to sequence RNA transcripts as expressed sequence tags (ESTs), which are basically used to estimate the gene composition of an organism [36]. After random sequencing in an unbiased cDNA library, ESTs are clustered into groups of transcript sequences using sequence-clustering and/or assembly approaches. Next, the number of ESTs with unique identifiers for each cDNA library and/or sequence cluster is tallied to estimate the quantity of transcripts expressed in each tissue. This concept has also been used in the digital differential display (DDD) tool of the NCBI's UniGene database, which has been utilised in substantial cDNA research for several taxa, including plants. Later, northern blotting and quantitative reverse transcription polymerase chain reaction (qRT-PCR) were utilised to quantify RNA transcripts. Since none of these methods addressed the complete transcriptome, the Sequencing-based Serial Analysis of Gene Expression (SAGE) was developed in 1995 [37]. More than 10 short specific tags (13–15 bp) are concatenated and cloned from each mRNA present in a sample to generate a SAGE library. The sequencing of selected clones from the SAGE library makes the efficient collection of transcript tag sequences feasible. To identify the genes corresponding to each SAGE tag, a dataset of genome sequences or a large collection of expressed sequence tags (ESTs) is required. Several versions of the fundamental protocol (MAGE, SADE, microSAGE, miniSAGE, longSAGE, superSAGE, deepSAGE, 5 ′ SAGE, etc.) have been developed to improve and expand the value of SAGE.

Massive parallel signature sequencing is another sequencing-based technique (MPSS). MPSS uses a 17–20 bp signature sequence near to the 3' end to identify mRNA. Initially, each distinctive sequence is cloned onto microbeads. This approach guarantees that a microbead has just one kind of DNA sequence. For sequencing and measuring, the flow cell comprises an array of microbeads. The signature sequences (MPSS tags) of an MPSS dataset are evaluated, compared to all other signatures, and the number of signatures with similar sequences is counted. Accessible online at http://mpss.udel.edu are databases containing MPSS information on various plant species, including *Arabidopsis*, rice, grapes, and *Magnaporthe grisea* (rice blast fungus). In *Arabidopsis*, high-density TSS mapping was performed utilising the newly published CT-MPSS method for quantitative investigation of the 5 ′ end of transcripts coupled with the cap-trapper strategy for full-length cDNA cloning. The data set of *Arabidopsis* CT-MPSS tags is accessible through the plant promoter database ppdb (http://www.ppdb.gene.nagoya-u.ac.jp), which provides rice and *Arabidopsis* promoter annotation. Many databases for plant transcriptome research are included in Table 3.

**Table 3: Different databases for plant transcriptomic study**

|  |  |
| --- | --- |
| Database | Link |
| Ppdb  | <http://www.ppdb.gene.nagoya-u.ac.jp>  |
| ArrayExpress | <https://www.ebi.ac.uk/arrayexpress/>  |
| ATTED II | <http://atted.jp/>  |
| Genevestigator | <https://www.genevestigator.com/gv/index.jsp>  |
| *Arabidopsis* Gene Expression Database AREX | <http://www.arexdb.org/index.jsp>  |
| RICEATLAS | <http://bioinformatics.med.yale.edu/riceatlas/>  |

1. **Metabolomics**

Metabolomics is defined as the complete and multidimensional study of metabolism that detects metabolites utilising a range of analytical tools and bioinformation. The metabolome of a certain plant is sufficiently complex, but comparing plants is much more challenging. Metabolomic methods are far better to chemical-level phenotyping and diagnostic evaluation because they allow for the simultaneous examination of a large number of metabolites and quantitative analysis of individual metabolites. Using exhaustive metabolic profile data sets, researchers may better comprehend how cells respond to changes in their internal and external surroundings. In addition, chemical phenotypes may be used to identify genes participating in certain metabolic pathways based on changes in metabolic profiles brought about by genetic variations. In recent years, several technical improvements have been achieved in metabolomics instruments. Obtaining metabolic fingerprints using different analytical techniques is the first step in obtaining metabolomics data. In combination with different kinds of MS, sample categorization techniques such as gas chromatography (GC), high-performance or ultra-performance liquid chromatography (LC), and capillary electrophoresis (CE) are used. Since it is a very sensitive approach for isolating and analysing biological components, CEMS is especially effective [38].

In metabolomics, data processing is crucial for assessing biological importance. Classifying samples and/or metabolites often use multivariate statistical approaches such as principal component analysis (PCA), hierarchical clustering analysis (HCA), and self-organization mapping (SOM) [39]. Other 'omics' methodologies are employed in conjunction with the visualisation of metabolic profiles on metabolic maps, such as gene expression profiles of specific genes encoding enzymes involved in certain pathways. Thus, plant metabolomics is a challenging analytical endeavour, but it is essential for comprehending plant growth and development. Metabolomics offers the potential to expand our understanding of plant cellular systems. Because metabolomics may highlight plant cellular processes, it allows us to design molecular breeding to improve plant productivity and functionality in areas such as stress tolerance, pharmaceutical manufacture, functional meals, biomaterials, and energy.

1. **Systems biology**

In recent years, several plant species have generated enormous volumes of data from sources such as the genome, transcriptome, proteome, metabolome, and epigenome. Since they were researched separately, it was not possible to completely appreciate the molecular foundations of complex characteristics and biological networks. To grasp the flow of biological information that underpins complex features, a systems biology strategy comprising the integration of different omics data, modelling, and prediction of cellular processes is required. This method allows a thorough knowledge of the dynamic system in which different levels of biological structure interact with the external environment in order to manifest phenotype. The primary aims of crop biology research are the maximising of yield and the reduction of losses caused by a variety of stress conditions. Because the issue is complex, so too is the answer. The combination of transcriptomics, proteomics, and metabolomics will greatly facilitate the identification and study of complicated plant regulatory networks. Therefore, systems biology evolves as an interesting interdisciplinary field of research that integrates considerable omics data with well-developed mathematical models to test hypotheses and predict biological systems. The processing, scaling, and analysis of multidimensional datasets in order to extract appropriate biological findings remain to be the greatest obstacle in integrating omics data. For the integration and analysis of datasets generated by many platforms, data collection, preparation, appropriate standardisation, and integration into a single matrix are necessary. Afterwards, groupings of genes, proteins, and metabolites with similar patterns were found. Multiple systems are available to aggregate multidimensional omics data, including mixOmics, OnPLS modelling, Integromics, sparse Multi-Block Partial Least Squares, and COVAIN. These techniques allow for the investigation of plant metabolism and the comprehension of the molecular mechanisms behind agronomically relevant plant phenotypes. To find light-specific metabolic and regulatory markers in rice [40], transcriptomics, metabolomics, and genome-scale in silico modelling were used. In 2020, transcriptomics, proteomics, and metabolomics data were analyzed to augment information that had previously shed light on the mechanisms behind the fertility shift in a thermosensitive male sterile line of pigeon peas for use in two-line hybrid breeding [41]. Since phenotypic variation is not just driven by DNA but also by biological regulation in response to the environment, multiomics data are increasingly exploited for phenotypic prediction. The reconstruction of pathways and networks using transcriptome, proteome, and metabolome data may assist in the comprehension of these regulatory networks and their functional interaction with the biological entities. The correct normalization of omics data generates a similarity matrix, which is subsequently transformed into an adjacency matrix and, finally, a directed graph or network abstraction. Global gene co-expression networks are a possible tool for investigating and high-throughput forecasting specialized metabolite pathways. The next step in network biology is dynamic modelling, which gives a comprehensive knowledge of how gene expression regulates protein activity in plants in response to environmental stimuli. By bridging the gap between genotype and phenotype and comprehending the complexity of various characteristics, systems biology provides enormous promise for sustainable agriculture. It is helpful for modelling and assessing multigenic characteristics associated with agricultural output, such as plant architecture, nitrogen use efficiency, water use efficiency, and resilience to biotic and abiotic stress. As a result of recent advancements in high-throughput experimental analysis and computational power, it is now feasible to combine many disciplines to explain any given complex trait. Using well-designed mathematical models based on time series data, one may construct a systems biology-based breeding strategy by identifying key candidate genes for potential usage in breeding programmes.

1. **Machine Learning**

Machine Learning (ML) is a subfield of computer science that uses statistical and mathematical techniques to train models without direct programming [42]. ML develops a variety of algorithms that learn from sample data and train the predictive model. Samuel [43]. ML is the study of programming computers to learn from data. By simplifying functional annotation of genomes and allowing real-time, high-throughput phenotyping of agronomic traits in the greenhouse and field, machine learning helps the discovery of agronomically valuable economic regions. ML is a technique to data analysis that enables computers to learn patterns over time. ML models for GS have the benefit of learning the pattern directly from the data, enabling them to account for all variations, interactions, and environmental factors. For huge, heterogeneous, and formless datasets, such as those produced by optical imaging or sequencing, ML may provide significant benefits over traditional analytic approaches. Crop breeders may use machine learning to rapidly phenotype plants and to examine massive databases for patterns, such as DNA sequence-to-characteristic connections. Machine learning algorithms may employ high-throughput phenotyping and genomic data to automate elements of the gene discovery process that are presently difficult to automate, such as genome labelling and picture interpretation. Figure 4 depicts the fundamental picture interpretation procedure. Although several research have used machine learning (ML) for GS, the subject of deep learning (DL) has yet to be thoroughly investigated.



**Figure 4: Basic workflow of image interpretation**

* 1. **High Throughput Crop Phenotyping**

Important for association studies and crop improvement, plant phenotyping is the measurement of functional or structural features at the cellular to organism level. Due to the fast advancement of genomics research and sequencing technologies, there is a growing need for plant phenotypes to aid in the interpretation of genetic data. Because it is subjective, error-prone, labor-intensive, and time-consuming, traditional phenotyping is typically a bottleneck that restricts the number of characteristics, crops, and ecosystems that may be evaluated. Due to advances in evaluating technologies (high-throughput imaging and automatic sensors) and machine learning, robotic high-throughput phenotyping can be established, which overcomes the limitations of conventional human-based phenotyping by enabling the rapid generation of phenotypical features and features across large populations [2]. Four fundamental characteristics of high-throughput phenotyping are image or sensor detection, phenotypical data categorization, feature quantification, and forecasting based on particular models or algorithms (Figure 3). Jose' Luis Araus and Jill E. Cairns assessed high-throughput phenotyping in the field. In a recent work [44], researchers measured, assessed, and categorised the severity of foliar stressors in Glycine max, such as bacterial and fungal infections, as well as nutritional insufficiency, using a deep ML-based phenotyping approach with an unsupervised identification explanation methodology. Large datasets are necessary for training and model development in machine learning. A limited training set may result in non-significant and problematic predictions, but acquiring a big dataset can be expensive and time-consuming, particularly when crop measurements are frequently performed only once every growing season. Consequently, ML-based phenotyping with high throughput is restricted to a small number of research institutions and businesses. To make ML-based phenotyping generally usable on the farm of the future, further reductions in acquisition and operating expenses are necessary.

* 1. **Machine Learning in Crop Genomics Research**

Several applications of machine learning include genome assembly, recurrent inference of gene regulatory networks, and identification of genuine Single Nucleotide Polymorphisms (SNPs) in polyploid plants. Optimizing polyploid genomic assemblies with complicated redundancy may be achieved via the application of machine learning. Ma et al. [45] provide a detailed review of machine learning algorithms and associated open-source R tools relevant for plant data analysis. A comprehensive genome assembly and annotation provides the basis for monitoring genetic changes within a plant species and for understanding the shape and function of plant genes, both of which are essential steps in the process of agricultural trait discovery. For interactive inference of gene regulatory networks, ML-based methods that can incorporate diverse types of regulatory signals from multiple data sources have gained popularity. Consequently, inferring regulatory element-gene links is a potential field for uncovering unexplored crop improvement opportunities.

GWAS is currently one of the most often utilised approaches for detecting MTA in plant species. Traditional GWASs are excellent for identifying SNP markers with considerable effects on complex traits, but they may ignore a variety of interconnected biological processes and mechanisms that influence the phenotypic of complex traits simultaneously. Variable significance values may be used to identify high-resolution variant-trait associations in ML-mediated GWASs. The implementation of this important genetic strategy in practical plant-breeding programmes may be enhanced by using complicated mathematical approaches such as machine learning (ML) algorithms. (ML-based GWAS for Identifying QTL Underlying Soybean Yield and Its Components)

1. **Deep Learning**

In the genomics era, multifaceted molecular phenotypes involved in information relay, namely the structure, modification, function, and evolution of elements in DNA, RNA, and protein, along with their interactions, are beginning to be revealed at scale and even at lower cost, allowing fine-grained evaluation of information transfer and transformation along Francis Crick's 1957 "central dogma" [46]. In data mining, it has been shown that deep learning models are very effective in predicting molecular phenotypes from upstream molecular phenotypes or directly from genomic DNA sequences.

Deep learning is a subfield of machine learning that focuses on densely connected, artificial neural network-trained networks. Artificial Neural Networks (ANN) are well-known strategies for dealing with machine learning issues that have been studied since the 1940s and are based on the nervous systems of animals [45]. A single artificial neural network (ANN) consists of several hidden layers and one input. Deep Neural Network (DNN) is a new machine learning discipline and a type of artificial neural network. DNNs differ from ANNs in that they contain many more hidden layers; hence, the quantity of data needed increases as the DNN's predictive ability increases. In genomics, transcriptomics, proteomics, metabolomics, and systems biology, deep learning has been used to address complicated biological challenges.

Deep learning, which utilises a high number of neurons and models such as CNN, RNN, and MLP, is applicable to GS [47]. The input layer of these models consists of marker data, whereas the output layer contains responses with several hidden layers. The optimal model performance is determined by hyperparameter selection, which is a time-consuming and computationally costly process. The ability of deep learning models to generate ab initio forecasts on unique, previously unknown sequence data (data not within the training set) is perhaps the most notable characteristic, which has numerous important ramifications, whereas the number of high-capacity and trainable characteristics is the most advantageous. Despite the huge number of genetic variations in a real population, deep learning models can only be trained on a small subset of them to predict the effects of all other variants (i.e., the whole mutation space). Knowledge may move from well-studied species (such as *Arabidopsis*) to closely related but less well-studied species (such as *Arabidopsis*) (such as other species in the Brassicaceae). When many variants within a crucial coding region (such as a QTL for a certain trait) are in tight linkage disequilibrium, we may utilise in silico mutagenesis to transfer them from one haplotype to the next, therefore prioritising causative variants. Such a break in linkage disequilibrium would be labor-intensive and difficult to scale up in wet lab research, and practically impossible in nature. Using a large collection of deep learning models, each targeting a different molecular phenotype, or a multi-task learning model addressing multiple molecular phenotypes simultaneously, it is possible to predict not only the causative mutation underlying a QTL, but also its likely molecular mechanism. Importantly, while using the breeding-by-editing approach, we are no longer restricted to the known beneficial natural variations. Instead, we have unrestricted flexibility to design different beneficial alleles based on the 'knowledge' of the biological processes of interest possessed by our deep learning algorithms. Rodrguez-Leal et al. [48] altered the promoter of the tomato CLAVATA3 gene (SlCLV3) to improve fruit size and inflorescence branching. Utilizing generative models in synthetic biology is another way for producing genetic components with defined functionality. Despite the growing interest in generative models like variational autoencoders and generative adversarial networks, their applications in synthetic biology remain restricted. Using GANs to construct synthetic DNA sequences encoding for antimicrobial peptides is one example.

1. **CONCLUSION**

It is crucial to adapt plant breeding curriculum to the digital age. Researchers and breeders must find a balance between machine-generated suggestions and farmer desires. Developing information for plant breeding is ineffective if researchers lack the capacity to use that knowledge. Information–action strategies, which integrate additional abilities and viewpoints to facilitate the development of knowledge for enhanced breeding and smarter farming, are necessary. Agriculture will depend on Next-Generation AI techniques to make judgments and recommendations based on massive data that is indicative of the environment and the systems biology of a plant. Breeding will be able to perform at greater levels than ever before because to Next-Gen AI's capacity to utilise diverse and complex data in an effective manner. The use of ML and DL has led to significant phenomics and genomics findings. As promising as these discoveries are, they are not yet adequate to contemplate depending only on technology to accelerate the breeding process, which remains a difficult, time-consuming, and costly endeavour. Despite gains in the efficiency of data generation, the plant research community still confronts difficulties with translational procedures. In isolation, genomes, epigenomics, transcriptomics, proteomics, metabolomics, and phenomics continue to be largely distinct fields of study that provide scant insight. To expedite plant development, it is necessary to concurrently use and integrate multi-omics data. Utilizing enormous quantities of genetic data from a variety of sources and formats for crop development is fraught with considerable difficulties in agriculture. To address these problems, novel breeding tactics and bioinformatics technology must be used to turn genetic data into advances in agricultural production and yield stability. Using meta-QTL analysis, GWAS, and genetic screens, researchers may uncover significant gene-trait connections more quickly. While genome editing is an effective method for rapidly introducing beneficial mutations into champion crops, GS enhances selection efficiency without needing knowledge of genetic drivers. ML algorithms may employ high-throughput phenotyping and genomic data to automate difficult-to-automate aspects of the gene discovery process, such as genome annotation and image interpretation. Combining new technologies and methods will allow future plant breeding to achieve the crop growth rate necessary for food security.

**LIST OF ABBREVIATIONS**

|  |  |  |  |
| --- | --- | --- | --- |
| BAC | Bacterial Artificial Chromosomes  | LD | Linkage Disequilibrium |
| GWAS | Genome-Wide Association Studies | GBS | Genomic Based Selection |
| QTL | Quantitative Trait Loci | CREs | Cis-Regulatory Elements |
| DNA | Deoxyribonucleic Acid | MS | Mass Spectrometry |
| RNA | Ribonucleic Acid | cDNA | Complementary DNA |
| ML | Machine Learning | SAGE | Serial Analysis of Gene Expression |
| GS | Genomic Selection | PAGE | polyacrylamide gel electrophoresis |
| DL | Deep Learning | SNP | Single Nucleotide Polymorphism |
| NGS | Next-Generation Sequencing | DNN | Deep Neural Network |
| EST | Expressed Sequence Tag | ANN | Artificial Neural Networks |
| MTA | Marker-Trait Associations | CNN | Convolutional Neural Networks |

**REFERENCES**

1. S. Satpathy, S. Debnath, and A. Mishra, “Study on character association *Lens culinaris* Medik.,” Electron. J. Plant Breed, vol. 12, pp. 58-65, April, 2021.
2. Z. An, C. Wang, B. Raj, S. Eswaran, R. Raseed, S. Debnath, et al., “Application of New Technology of Intelligent Robot Plant Protection in Ecological Agriculture,” J. Food Qual, vol. 1, pp. 1-7, April, 2022.
3. N. Rajan, S. Debnath, A.K. Dutta, B. Pandey, A.K. Singh, R.K. Singh, “Elucidation of Nature of Gene Action and Estimation of Combining Ability Effects for Fruit Yield Improvement and Yield Attributing Traits in Brinjal Landraces,” J. Food Qual, vol. 12, April 2022.
4. R. Velasco, A. Zharkikh, M. Troggio, D.A. Cartwright, A. Cestaro, D. Pruss, et al., “A high quality draft consensus sequence of the genome of a heterozygous grapevine variety,” PloS one, vol. 2, pp. e1326, December 2007.
5. T.P. Michael, and S. Jackson, “The first 50 plant genomes,” Plant Genome, vol. 6, July 2013.
6. T. Toyoda, and A. Wada, “Omic space: coordinate-based integration and analysis of genomic phenomic interactions,” Bioinformatics, vol. 20, pp. 1759-1765, July 2004.
7. J.M. González‐Camacho, L. Ornella, P. Pérez‐Rodríguez, D. Gianola, S. Dreisigacker, and J. Crossa, “Applications of machine learning methods to genomic selection in breeding wheat for rust resistance,” Plant Genome, vol. 11, pp. 170104, July 2018.
8. S. Huang, R. Li, Z. Zhang, L.I. Li, X. Gu, W. Fan, et al., “The genome of the cucumber, *Cucumis sativus* L.,” Nat. genet, vol. 41, pp. 1275-1281, December 2009.
9. H. Wang, E. Cimen, N. Singh, and E. Buckler, “Deep learning for plant genomics and crop improvement,” Curr. Opin. Plant Biol, vol. 24, pp. 34-41, April 2020.
10. T. Sasaki, “The map-based sequence of the rice genome,” Nature, vol. 436, pp. 793-800, August 2005.
11. M. Mascher, H. Gundlach, A. Himmelbach, S. Beier, S.O. Twardziok, T. Wicker, et al., “A chromosome conformation capture ordered sequence of the barley genome,” Nature, vol. 544, pp. 427-433, April 2017.
12. M. Mousavi‐Derazmahalleh, P.E. Bayer, J.K. Hane, B. Valliyodan, H.T. Nguyen, M.N. Nelson, et al., “Adapting legume crops to climate change using genomic approaches,” Plant Cell Environ, vol. 42, pp. 6-19, January 2019.
13. S.P. Chand, S. Debnath, M. Rahimi, M.S. Ashraf, P. Bhatt, S.A. Rahin, “Contextualization of Trait Nexus and Gene Action for Quantitative and Qualitative Characteristics in Indian Mustard,” J. Food Qual, vol. 2022, pp. 1-24, May 2022.
14. M.W. Blair, M.M. Torres, M.C. Giraldo, and F. Pedraza, “Development and diversity of Andean-derived, gene-based microsatellites for common bean (*Phaseolus vulgaris* L.),” BMC Plant Biol, vol. 9, pp. 1-4, December 2009.
15. A. Heesacker, V.K. Kishore, W. Gao, S. Tang, J.M. Kolkman, A. Gingle, et al., “SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility,” Theor. Appl. Genet, vol. 117, pp. 1021-1029, November 2008.
16. R.V. Kantety, M. La Rota, D.E. Matthews, and M.E. Sorrells, “Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat,” Plant Mol. Biol, vol. 48, pp. 501-510, March 2002.
17. S. Kaur, N.O. Cogan, G. Ye, R.C. Baillie, M.L. Hand, A.E. Ling, et al., “Genetic map construction and QTL mapping of resistance to blackleg (*Leptosphaeria maculans*) disease in Australian canola (*Brassica napus* L.) cultivars,” Theor. Appl. Genet, vol. 120, pp. 71-83, December 2009.
18. R. Kota, R.K. Varshney, T. Thiel, K.J. Dehmer, and A. Graner, “Generation and comparison of EST‐derived SSRs and SNPs in barley (*Hordeum vulgare* L.),” Hereditas, vol. 135, pp. 145-151, December 2001.
19. R. Liu, H. Zhang, P. Zhao, Z. Zhang, W. Liang, Z. Tian, et al., “Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data,” Plant Mol. Biol. Rep, vol. 30, pp. 297-308, April 2012.
20. J.I. Said, Z. Lin, X. Zhang, M. Song, and J. Zhang, “A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton,” BMC Genom, vol. 14, pp. 1-22, December 2013.
21. Z.M. Qi, Q. Wu, X. Han, Y.N. Sun, X.Y. Du, C.Y. Liu, et al., “Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes,” Euphytica, vol. 179, pp. 499-514, June 2011.
22. E. Hanocq, A. Laperche, O. Jaminon, A.L. Lainé, J. Le Gouis, “Most significant genome regions involved in the control of earliness traits in bread wheat, as revealed by QTL meta-analysis,” Theor. Appl. Genet, vol. 114, pp. 569-584, February 2007.
23. K. Gacek, P.E. Bayer, I. Bartkowiak-Broda, L. Szala, J. Bocianowski, D. Edwards, et al., “Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*,” Front. Plant Sci, vol. 7, pp. 2062, January 2017.
24. X. Huang, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, et al., “Genome-wide association studies of 14 agronomic traits in rice landraces,” Nat. Genet, vol. 42, pp. 961-967, November 2010.
25. H. Sonah, L. O'Donoughue, E. Cober, I. Rajcan, and F. Belzile, “Identification of loci governing eight agronomic traits using a GBS‐GWAS approach and validation by QTL mapping in soya bean,” Plant Biotechnol. J, vol. 13, pp. 211-221, February 2015.
26. F. Tian, P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, et al., “Genome-wide association study of leaf architecture in the maize nested association mapping population,” Nat. Genet, vol. 43, pp. 159-162, February 2011.
27. J.E. Rutkoski, J.A. Poland, R.P. Singh, J. Huerta‐Espino, S. Bhavani, H. Barbier, et al., “Genomic selection for quantitative adult plant stem rust resistance in wheat,” Plant Genome, vol. 7, pp. plantgenome2014-02, November 2014.
28. D.S. Johnson, A. Mortazavi, R.M. Myers, and B. Wold, “Genome-wide mapping of in vivo protein-DNA interactions,” Sci, vol. 316, pp. 1497-1502, June 2007.
29. J.D. Buenrostro, B. Wu, H.Y. Chang, and W.J. Greenleaf, “ATAC‐seq: a method for assaying chromatin accessibility genome‐wide,” Curr. Protoc. Mol. Biol, vol. 109, pp. 21-29, January 2015.
30. F.M. Pauler, S.H. Stricker, K.E. Warczok, and D.P. Barlow, “Long-range DNase I hypersensitivity mapping reveals the imprinted Igf2r and Air promoters share cis-regulatory elements,” Genome Research, vol. 15, pp. 1379-1387, October 2005.
31. S. Rombauts, K. Florquin, M. Lescot, K. Marchal, P. Rouzé, and Y. Van de Peer, “Computational approaches to identify promoters and cis-regulatory elements in plant genomes,” Plant Physiol, vol. 132, pp. 1162-1176, July 2003.
32. J.V. de Velde, K.S. Heyndrickx, and K. Vandepoele, “Inference of transcriptional networks in *Arabidopsis* through conserved noncoding sequence analysis, Plant Cell, vol. 26, pp. 2729-2745, July 2014.
33. T. Rabilloud, “When 2 D is not enough, go for an extra dimension,” Proteomics, vol. 13, pp. 2065-2068, July 2013.
34. G. Mastrobuoni, S. Irgang, M. Pietzke, H.E. Aßmus, M. Wenzel, W.X. Schulze, et al., “Proteome dynamics and early salt stress response of the photosynthetic organism *Chlamydomonas reinhardtii*,” BMC Genom, vol. 13, pp. 1-3, December 2012.
35. G.K. Sim, F.C. Kafatos, C.W. Jones, M.D. Koehler, A. Efstratiadis, and T. Maniatis, “Use of a cDNA library for studies on evolution and developmental expression of the chorion multigene families,” Cell, vol. 18, pp. 1303-1316, December 1979.
36. R. Lowe, N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, “Transcriptomics technologies,” PLoS Comput. Biol, vol. 13, pp. e1005457, May 2017.
37. V.E. Velculescu, L. Zhang, B. Vogelstein, and K.W. Kinzler, “Serial analysis of gene expression,” Sci, vol. 270, pp. 484-487, October 1995.
38. R. Ramautar, G.W. Somsen, and G.J. de Jong, “CE‐MS in metabolomics,” Electrophoresis, vol. 30, pp. 276-291, January 2009.
39. M.Y. Hirai, M. Yano, D.B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, et al., “Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*,” PNAS, vol. 101, pp. 10205-10210, July 2004.
40. M. Lakshmanan, S.H. Lim, B. Mohanty, J.K. Kim, S.H. Ha, and D.Y. Lee, “Unraveling the light-specific metabolic and regulatory signatures of rice through combined in silico modeling and multiomics analysis,” Plant Physiol, vol. 169, pp. 3002-3020, December 2015.
41. L.T. Pazhamala, P. Chaturvedi, P. Bajaj, S. Srikanth, A. Ghatak, A. et al., “Multiomics approach unravels fertility transition in a pigeonpea line for a two‐line hybrid system,” Plant Genome, vol. 13, pp. e20028, July 2020.
42. A. Singh, G. Vaidya, V. Jagota, D.A. Darko, R.K. Agarwal, S, Debnath, et al., “Recent Advancement in Postharvest Loss Mitigation and Quality Management of Fruits and Vegetables Using Machine Learning Frameworks,” J. Food Qual, vol. 2022, pp. 1-9, June 2022.
43. A.L. Samuel, “Some studies in machine learning using the game of checkers,” IBM J. Res. Dev, vol. 44, pp. 206-226, January 2000.
44. S. Ghosal, D. Blystone, A.K. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar, “An explainable deep machine vision framework for plant stress phenotyping,” PNAS, vol. 115, pp. 4613-4618, May 2018.
45. H. Wang, E. Cimen, N. Singh, and E. Buckler, “Deep learning for plant genomics and crop improvement,” Curr. Opin. Plant Biol, vol. 24, pp. 34-41, April 2020.
46. F. Crick, “Central dogma of molecular biology,” Nature, vol. 227, pp. 561-563, August 1970.
47. A.K. Singh, I.R. Khan, S. Khan, K. Pant, S. Debnath, and S. Miah, “Multichannel CNN Model for Biomedical Entity Reorganization,” Biomed Res. Int, vol. 2022, pp. 1-11.
48. D. Rodríguez-Leal, Z.H. Lemmon, J. Man, M.E. Bartlett, and Z.B. Lippman, “Engineering quantitative trait variation for crop improvement by genome editing,” Cell, vol. 171, pp. 470-480, October 2017.