

INTEGRATING MULTI-OMICS, GENOME EDITING AND ARTIFICIAL INTELLIGENCE FOR PRECISION BREEDING: ADVANCES IN MOLECULAR GENETICS AND CROP IMPROVEMENT

Dr. K S Usharani¹, Dr. Mala C. Patil², Raghul S³, Dr. Nayan Deep⁴, Dr. V. Karpagam⁵, Dr. Polagani Nagarjuna⁶ and Nidhi Sharma⁷

¹Assistant Professor (Plant Breeding and Genetics), Kumaraguru Institute of Agriculture, Sathy-Athani-Bhavani Main Road, Nachimuthupuram, Erode District, Tamil Nadu, India.

²Assistant Professor of Computer Science, University of Horticultural Sciences, Bagalkot

³Assistant Professor, Department of Agricultural Engineering, Bannari Amman Institute of Technology, Erode, Tamil Nadu, India

⁴Ph.D. Research Scholar, Department of Vegetable Science, Dr. Y.S. Parmar University of Horticulture and Forestry, Himachal Pradesh, 173230

⁵Assistant Professor (Plant Breeding and Genetics), MIT College of Agriculture and Technology, Vellalapatti, Musiri, 621211

⁶Assistant Professor, Agronomy, College of Agriculture, Koneru Lakshmaiah Education Foundation, 522502

⁷Assistant Professor, Department of Horticulture, Lovely Professional University, 144411

*Corresponding author: Dr. K S Usharani, E-mail: usharaniagri@gmail.com

ABSTRACT

Precision breeding is increasingly shaped by the ability to connect molecular measurements with decisions made in breeding populations and fields. Multi-omics can help explain how inherited variation is expressed through regulatory activity, proteins and metabolites; genome editing can test selected hypotheses and create particular alleles; and artificial intelligence (AI) can support prediction, phenotyping and the prioritization of experiments. These contributions are complementary, but they are not interchangeable. Molecular associations may reflect consequences rather than causes, an edit that produces the expected molecular change may not improve crop performance and a model that performs well on a familiar dataset may fail in another population or environment. This review critically examines how multi-omics, genome editing and AI can be combined for crop improvement, drawing on research in tomato, maize, rice, wheat, soybean and other plant systems. The evidence is strongest when candidate discovery is followed by functional testing and replicated agronomic evaluation. We outline a closed-loop approach in which molecular data generate hypotheses, AI helps rank candidates and select informative tests, editing or crossing evaluates causality and field results are used to update subsequent decisions. Priorities include better causal designs, representative training populations, independent validation, reproducible data workflows, realistic economic assessment and wider access to enabling technologies. The value of precision breeding should ultimately be judged by reliable genetic gain under target production conditions, not by the amount of data collected, the apparent sophistication of a model or editing efficiency alone.

KEYWORDS: crop improvement; multi-omics; CRISPR; genomic selection; artificial intelligence; precision breeding

1. INTRODUCTION

Breeders are being asked to improve yield, quality and resilience at the same time. Heat, drought, salinity and changing pest and pathogen pressures make this difficult, while the long process of crossing, testing and releasing varieties slows the assembly of useful alleles. Molecular genetics has already shortened parts of this process through genetic mapping, marker-assisted selection and genomic selection. Sequencing, transcriptomics, metabolomics, advanced phenotyping, genome editing and AI now provide further ways to identify candidates and evaluate them. None removes the need for field selection. Their value lies in making the route from inherited variation to crop performance more explicit and, where possible, faster to test.

Because the term precision breeding is used in different ways, this review applies it to a decision process in which molecular and phenotypic evidence guides choices about alleles, genotypes, crosses and target environments. Multi-omics means the coordinated analysis of two or more molecular layers, such as the genome, transcriptome, epigenome, proteome and metabolome; simply collecting several datasets does not constitute integration. Genome editing refers to targeted changes in DNA sequence or gene regulation. AI includes statistical learning and machine-learning methods used to make predictions, extract phenotypes or prioritize decisions. It is treated here as decision support, not as an autonomous substitute for breeders or biological judgement.

Each technology answers a different question. Multi-omics can link variation to molecular states and suggest mechanisms. Genome editing can test whether a candidate change contributes to a phenotype and, in some cases, install that change in breeding material. AI can scale prediction and measurement across many lines, traits and environments. Overstating any one of these roles weakens the evidence. An association among omics layers does

not establish causation; an edited plant with a changed molecular phenotype is not necessarily a better crop; and high internal prediction accuracy does not guarantee transfer to new germplasm or growing conditions. A convincing claim therefore needs to move from discovery through functional testing to replicated agronomic performance.

Research in tomato shows that selection for visible fruit traits can also reshape metabolite profiles, sometimes through linkage with neighbouring genes rather than direct selection for those biochemical changes (Schauer *et al.*, 2006; Riedelsheimer *et al.*, 2012; Zhu *et al.*, 2018). Pan-genomic and structural-variation studies have likewise revealed that a single reference genome may omit useful alleles or poorly represent the variation present in breeding material. At the same time, genome editing has been used to test specific disease-resistance, composition and domestication-related targets, while AI methods have expanded the scale of phenotyping and prediction. The practical question is not simply whether these technologies can be combined, but which combinations deliver repeatable improvement at a realistic cost.

This review brings together primary research on multi-omics, genome editing, AI-enabled phenotyping and genomic prediction, with particular attention to causal inference, transferability and field validation. The six tables compare evidence layers and crop examples, summarize editing and AI applications, set out an integrated breeding workflow and identify minimum standards for evaluating claims. The central argument is that precision breeding works best as a cycle of hypothesis, test and revision: molecular and computational predictions should lead to experiments whose outcomes matter to crop performance.

2. Materials and methods: scope and evidence selection

This is a critical narrative review of original, peer-reviewed research on crop multi-omics and pan-genomics, CRISPR-based editing, AI-assisted phenotyping and genomic prediction and the integration of these approaches in breeding decisions. Foundational methods papers are retained where they are needed to explain a platform or its limitations; empirical crop studies are used preferentially when discussing biological effects or breeding outcomes. The review is not a registered systematic review or meta-analysis and it does not claim to include every relevant study. Its aim is to assess what kinds of conclusions the existing evidence can support and where important uncertainties remain.

Studies were considered in terms of the inference they support: identifying an association, explaining a molecular process, testing a candidate through perturbation, predicting performance within a sampled population, transferring predictions to independent material, or demonstrating value in field trials. These are distinct levels of evidence. Multi-omics was treated as integrated when data layers were analysed together or explicitly connected to a common trait. AI studies were assessed in light of their baselines, data splits and validation populations. Editing studies were considered more informative when the target, resulting allele, molecular consequence and phenotype were described. Results from model plants were used to clarify mechanisms, but were not assumed to demonstrate performance in elite crop backgrounds.

3. Multi-omics as a bridge from genotype to phenotype

3.1. What integration contributes beyond parallel profiling

The omics layers describe different parts of the biological process. Genomics captures inherited sequence variation; transcriptomics measures RNA abundance and regulation; epigenomics describes features such as DNA methylation and chromatin state; proteomics measures proteins and some of their modifications; and metabolomics records downstream biochemical states. Phenomics adds observations of plant structure, development, physiology and performance over time. The layers differ in stability, measurement error and proximity to the trait. Their integration can help assemble a plausible chain from variant to expression, enzyme activity, metabolism and phenotype. That chain is useful only if sample identities, tissues, developmental stages and experimental conditions are sufficiently matched to make comparisons meaningful.

Crop studies offer several examples of what integrated measurements can reveal. In tomato, analysis of genomes, transcripts and metabolites across diverse accessions showed that fruit-size selection and other breeding decisions had broad effects on fruit chemistry; some metabolite changes were linked to neighbouring genes rather than being direct targets of selection (Zhu *et al.*, 2018). In hybrid maize, genomic and metabolite profiles captured partly distinct signals for complex hybrid performance, supporting their complementary use in prediction (Riedelsheimer *et al.*, 2012b). Studies of cereal grain metabolism connected natural biochemical variation with genome-wide associations and agronomic traits, generating candidates for further investigation (Wen *et al.*, 2014; Chen *et al.*, 2016). Earlier work in rice, Arabidopsis and potato similarly linked metabolite variation to genetic architecture and tuber quality (Keurentjes *et al.*, 2006; McNally *et al.*, 2009; Carreño-Quintero *et al.*, 2012; Matsuda *et al.*, 2015). Together, these studies show that biochemical data can add useful information to genetic analysis. They do not show that metabolomics can replace field evaluation or that every measured metabolite is a useful selection target.

Integration is not a single statistical operation. In early integration, normalized features from several layers are combined before modelling; this is straightforward but can be dominated by the most complete or highest-dimensional dataset. Intermediate approaches learn shared components, factors or networks across layers. Late integration combines predictions made separately from each data type and can be useful when the layers contain different samples. Causal approaches use prior biological knowledge or experimental perturbations to test directed relationships. The most suitable strategy depends on the question and the completeness of the data. Adding more layers is not inherently better if the extra variables mainly duplicate existing information or introduce batch effects.

One of the hardest problems is distinguishing a molecular cause from a molecular consequence. Stress-responsive transcripts, for example, may record damage after it begins. A metabolite may be associated with tolerance because a tolerant plant has maintained photosynthesis, not because the metabolite caused tolerance. Time-series sampling, tissue-specific measurements, segregating populations, allele-specific analysis and targeted perturbation can help separate these alternatives. Wang *et al.* (2024) illustrate the promise of combining genomic, transcriptomic and methylomic information to predict plant traits and prioritize genes. Such predictions remain hypotheses until the proposed genes are tested across genetic backgrounds and, where relevant, under field conditions.

Cell-resolved multi-omics may help explain traits whose control differs among tissues or cell types. Linking chromatin accessibility with expression, for example, can reveal regulatory candidates that would be obscured in whole-organ measurements. Yet detailed molecular atlases are expensive, sample preparation can bias which cells are recovered and controlled sampling may not represent responses in a changing field environment. Their greatest value may be targeted rather than universal: use fine-resolution data to resolve a regulatory mechanism or explain a robust locus, then identify a simpler marker, assay or edit that can be used in routine breeding.

3.2. Pan-genomes, structural variation and the limits of a single reference

A reference genome is essential for many analyses, but it represents one accession and can therefore make some forms of variation difficult to detect. Pan-genomes and multiple high-quality assemblies broaden the comparison to include gene presence-absence differences, copy-number changes, transposable-element insertions and larger rearrangements, any of which may affect gene dosage or regulation. In tomato, the pan-genome helped link rare variation in the *TomLoxC* promoter to fruit flavour (Gao *et al.*, 2019). Long-read analysis also showed that structural variants can affect genes and regulatory regions in ways that SNP-only analyses may miss (Alonge *et al.*, 2020). Pan-genomics consequently expands the set of alleles and mechanisms that can be examined; it is more than a larger version of a single linear reference.

Work in other crops supports the same point. The annotated wheat reference genome improved resources for comparative analysis and breeding, while the sorghum reference added an improved assembly, gene annotation and a transcriptome atlas (International Wheat Genome Sequencing Consortium, 2018; McCormick *et al.*, 2018). Studies in Brassica and soybean have documented gene-content variation beyond a single linear reference (Golicz *et al.*, 2016; Torkamaneh *et al.*, 2021). In soybean, comparisons across cultivated and wild accessions further exposed gene-content diversity relevant to breeding resources (Liu *et al.*, 2020). Large accession studies in *Arabidopsis thaliana* illustrate the broader value of sampling diverse genotypes rather than relying on a single reference background (The 1001 Genomes Consortium, 2016). These advances matter because an apparently missing gene or uncommon marker may reflect how the reference was assembled, rather than a true absence across the crop. Long-read sequencing and graph-based references can reduce this bias, although cost and computational demands still limit routine use.

A larger variant catalogue also increases the risk of false leads. With more candidate variants available, it becomes easier to construct plausible stories after the fact, especially when population structure and linkage disequilibrium are not adequately controlled. A structural variant may cause a trait, tag another causal variant or simply be common in the population where the trait occurs. Before editing a candidate, researchers should test whether the association remains after accounting for relatedness and whether the variant adds information beyond ancestry and nearby markers. Mixed-model association approaches help control for relatedness, but functional evidence is still needed to establish causality (Yu *et al.*, 2006). The useful connection is therefore between pan-genomic discovery, multi-omic interpretation and experimental testing—not merely the accumulation of variants in a database.

Table 1. Molecular and phenotypic data layers: breeding value, limitations and integration needs.

Data layer	Primary signal	Useful breeding inference	Main failure mode	Best integration practice	Selected primary evidence
Genome / pan-genome	SNPs, indels, structural and presence-absence variants	Discover alleles, define haplotypes and support genomic prediction	Reference bias, population structure and linkage disequilibrium	Sample diverse references, use pan-genome or graph methods and validate independently	Gao <i>et al.</i> (2019); Golicz <i>et al.</i> (2016); Liu <i>et al.</i> (2020); Torkamaneh <i>et al.</i> (2021)
Epigenome	DNA methylation, chromatin accessibility, histone state	Characterize regulatory states and environment-responsive expression	Strong dependence on tissue and stage; direction of causality may be unclear	Match tissues and time points, then test key regulatory hypotheses experimentally	Wang <i>et al.</i> (2024)

Data layer	Primary signal	Useful breeding inference	Main failure mode	Best integration practice	Selected primary evidence
Transcriptome	RNA abundance, splicing, allele-specific expression	Prioritize expressed candidates and regulatory networks	Expression may be a response to stress or development, not its cause	Repeat across environments and connect expression QTLs with trait loci	Zhu <i>et al.</i> (2018); Wang <i>et al.</i> (2024)
Proteome / phosphoproteome	Protein abundance and modification	Determine whether RNA changes translate into active protein machinery	Incomplete coverage and strong tissue or temporal variation	Use targeted assays for prioritized pathways and pair them with functional evidence	Integrated with functional assays rather than used alone
Metabolome	Metabolite abundance and biochemical profiles	Investigate quality traits, stress biomarkers and pathway state	Environmental sensitivity, uncertain annotation and batch effects	Use standards, replicated trials and metabolic QTL analysis	Schauer <i>et al.</i> (2006); Riedelsheimer <i>et al.</i> (2012a); Wen <i>et al.</i> (2014)
Phenomics / enviromics	Image, sensor, weather, soil and management variables	Track developmental responses and model genotype-by-environment effects	Sensor drift, confounding and limited transfer between environments	Align genotypes, environmental records and ground-truth measurements	Ghosal <i>et al.</i> (2018); Shahhosseini <i>et al.</i> (2021)

4. Genome editing: from candidate genes to causal tests and designed alleles

4.1. Editing modalities are not interchangeable

CRISPR nucleases, base editors and prime editors produce different kinds of changes and should be chosen accordingly. Nucleases create a double-strand break that is repaired by the cell, often generating insertions or deletions useful for gene disruption, regulatory deletions or multiplex knockout experiments. Repair is not always predictable, however and unwanted alleles or larger structural changes can occur. Base editors use a programmable DNA-binding component and a deaminase to convert selected bases without relying on a conventional double-strand break. Prime editors combine a nickase and reverse transcriptase with a prime-editing guide RNA to specify substitutions and small insertions or deletions. Each method has its own target constraints, by-products and delivery requirements (Komor *et al.*, 2016; Gaudelli *et al.*, 2017; Anzalone *et al.*, 2019). The choice should follow from the intended allele and the evidence needed, rather than from a general preference for the newest editing system. The underlying RNA-guided cleavage mechanism was established before these tools were adapted to plants (Jinek *et al.*, 2012).

Early studies established that CRISPR-mediated editing could be achieved in rice, wheat, tomato, tobacco and other plant systems (Feng *et al.*, 2013; Miao *et al.*, 2013; Li *et al.*, 2013; Nekrasov *et al.*, 2013; Shan *et al.*, 2013). Subsequent work moved from demonstrating edits to testing useful crop traits. In wheat, editing MLO homoeologues produced heritable resistance to powdery mildew, a clear example of why redundant gene copies must be considered in polyploid crops (Wang *et al.*, 2014). In rice, changes to promoter elements in susceptibility genes offered a way to alter disease response without eliminating all gene function (Oliva *et al.*, 2019). Editing the rice CDP-diacylglycerol synthase gene produced a resistance allele that was also examined in small-scale field trials, although its broader breeding value depends on maintaining yield and growth (Sha *et al.*, 2023). Editing of *CsLOB1* in citrus provided another example of targeting a host susceptibility gene to reduce disease (Jia *et al.*, 2017). Earlier rice work showed that homozygous targeted edits could be recovered within one generation under suitable experimental conditions (Zhang *et al.*, 2014). The common lesson is that editing can test practical hypotheses, but each trait still requires a careful assessment of trade-offs.

Base editing has broadened the range of alleles that can be made without supplying a donor DNA template. Cytosine base editing was demonstrated in rice, wheat and maize and later editors expanded the available conversion types or editing windows (Zong *et al.*, 2017, 2018; Li *et al.*, 2018). Prime editing has also been adapted to plants, including rice and wheat, to introduce specified substitutions and small insertions or deletions, although success varies sharply by target and regeneration context (Lin *et al.*, 2020). These approaches could be useful for recreating naturally occurring alleles, changing a single amino acid or modifying a regulatory motif. Their promise should not be confused with routine efficiency. Bystander changes, sequence context, editor design and the ability to regenerate plants all affect outcomes. Optimized prime editors and paired pegRNAs have improved results at

selected targets, but crop- and locus-specific testing remains essential (Jiang *et al.*, 2020; Lin *et al.*, 2021; Xu *et al.*, 2022). Work in rice and tomato likewise showed that editor configuration affects the position and frequency of cytidine conversions (Shimatani *et al.*, 2017).

A successful molecular edit is only the first step toward a useful variety. High editing rates in protoplasts do not guarantee fertile regenerated plants and the presence of an intended edit does not rule out mosaicism or other sequence changes. Nor does a changed molecular trait demonstrate yield stability, quality or adaptation under farm conditions. A practical pipeline should document delivery and regeneration efficiency, zygosity, unwanted alleles, off-target assessment, segregation of editing constructs and performance across environments. When a favourable allele is already available in germplasm, conventional crossing or marker-assisted introgression may be simpler and less costly than editing. Delivery of preassembled CRISPR/Cas9 ribonucleoproteins has produced transgene-free edited tomato and potato plants, but the same limits of regeneration and genotype dependence remain (Veillet *et al.*, 2019).

The case for integration is strongest when each technology addresses a different uncertainty. Multi-omics may help identify a candidate gene and the regulatory context in which it acts. AI can prioritize candidates by expected effect, uncertainty, possible pleiotropy and relevance across environments. Editing then tests whether the proposed change produces the expected molecular and phenotypic result. The process can also run in the other direction: results from edited lines can improve mechanistic models and help train future predictors. Without perturbation, AI can mistake correlation for mechanism; without diverse genetic and environmental testing, an edit can appear promising only in the background in which it was first evaluated.

4.2. Editing beyond knockout: dosage, regulation and domestication

Many useful crop traits depend on the amount, timing and location of gene expression rather than on whether a gene is simply present or absent. Knocking out a regulatory gene may improve one trait yet compromise flowering, fertility, architecture or stress response. Editing promoter or other cis-regulatory sequences can sometimes tune expression while preserving essential gene function. Allele-specific editing may also recreate a natural variant without transferring a larger donor segment and its linked alleles. These strategies are especially relevant to quantitative traits, but they require a better understanding of regulatory architecture and more careful phenotyping than a simple knockout experiment. Programmable transcriptional regulation and CRISPR-Cas12a systems expand the available interventions, although the resulting expression changes still need to be shown to produce stable, heritable phenotypes (Piatek *et al.*, 2015; Tang *et al.*, 2017).

Experiments in wild tomato and other underutilized crops have shown that several domestication-related loci can be edited together to introduce traits such as compact growth, changes in flowering and fruit-size characteristics (Zsögön *et al.*, 2018; Lemmon *et al.*, 2018). These studies are useful proof of concept, but they also expose the distance between modifying a few conspicuous traits and producing a competitive crop variety. Adaptation, yield stability, seed quality, harvestability and farmer preference reflect many interacting genes and environments. Multiplex editing can create promising starting material; it does not remove the need to combine those alleles with adapted germplasm and test their interactions. Other examples illustrate the specificity of editing outcomes: disrupting *GmFT2a* delayed flowering in soybean, while editing the *FAD2* gene family altered soybean oil quality (Cai *et al.*, 2018; Haun *et al.*, 2014).

Editing can turn an association-based multi-omics hypothesis into a testable causal model. If a targeted change produces the predicted effects on expression, protein or metabolite abundance and the final phenotype, the evidence for that mechanism becomes stronger. If the molecular change occurs but the desired agronomic outcome does not, that is not a useless result. It may reveal compensation by other pathways, dependence on genetic background, or a mistaken assumption about how the molecular trait affects performance. Recording such outcomes is important. A dataset containing only successful edits would give an overly optimistic picture of how often candidate changes translate into breeding value and would bias AI models trained on those data.

Table 2. Crop studies connecting molecular evidence with agronomic traits.

Crop / system	Integrated evidence	Principal contribution	Breeding relevance	Critical limitation	Supporting references
Tomato fruit quality	Genome, transcriptome and metabolome across diverse genotypes	Breeding selection altered fruit metabolites, including through linked effects	Can inform selection for flavour and composition	Separate direct selection from linkage and genetic-background effects	Schauer <i>et al.</i> (2006); Zhu <i>et al.</i> (2018)
Hybrid maize	Genome-wide markers and metabolites	Marker and metabolite profiles captured partly distinct signals for complex hybrid traits	Metabolites may add information to genomic prediction for selected traits	Routine use depends on sampling design, heritability and assay cost	Riedelsheimer <i>et al.</i> (2012b)

Crop / system	Integrated evidence	Principal contribution	Breeding relevance	Critical limitation	Supporting references
Cereal grain metabolism	Metabolome-wide and phenome-wide association mapping	Connected biochemical variation with genetic loci and agronomic traits	Helps prioritize pathways and candidate genes for further testing	Association alone does not prove the biochemical mechanism	Wen <i>et al.</i> (2014); Chen <i>et al.</i> (2016)
Tomato pan-genome	Diverse accessions, sequence variation and trait genetics	Identified rare promoter variation and gene-content diversity linked to fruit flavour	Expands the allelic search beyond a single linear reference	Variant frequency and linkage can complicate transfer into elite lines	Gao <i>et al.</i> (2019)
Soybean pan-genome	Multiple cultivated genomes	Characterized conserved gene content alongside variation among cultivated accessions	Supports haplotype discovery and more representative selection resources	Presence in a pan-genome does not itself demonstrate a trait effect	Torkamaneh <i>et al.</i> (2021)
Arabidopsis trait prediction	Genomic, transcriptomic and methylomic profiles	Combined molecular layers to predict traits and prioritize genes	Provides a methodological example for multilayer prediction	Crop transfer requires new training data and independent validation	Wang <i>et al.</i> (2024)

Table 3. Genome-editing approaches and their suitability for precision-breeding objectives.

Modality	Molecular action	Best-fit use	Key benefits	Common failure modes / checks	Supporting primary studies
CRISPR nuclease knockout	Double-strand break followed by DNA repair	Test gene function, disrupt susceptibility genes or reduce gene dosage	Guide-dependent outcomes, large deletions, mosaicism and polyploid complexity; sequence the target locus	Indels, large deletions, mosaicism, homoeologue complexity; sequence target locus fully	Jiang <i>et al.</i> (2013); Li <i>et al.</i> (2013); Miao <i>et al.</i> (2013); Wang <i>et al.</i> (2014); Xing <i>et al.</i> (2014)
Promoter / cis-regulatory editing	Deletion or sequence alteration in regulatory DNA	Tune expression or remove pathogen-responsive regulatory motifs	Regulatory effects depend on context; several alleles may need testing	Regulatory logic may be context-specific; multiple promoter alleles may need testing	Oliva <i>et al.</i> (2019)
Cytosine base editing	C-to-T / complementary G-to-A conversions in an editing window	Recreate natural variants or introduce selected amino-acid substitutions	Bystander changes, editing-window limits and sequence-context effects	Bystander edits, window constraints, editing-context dependence	Komor <i>et al.</i> (2016); Zong <i>et al.</i> (2017, 2018); Li <i>et al.</i> (2018)
Adenine base editing	A-to-G / complementary T-to-C conversion	Install point mutations not accessible with cytosine editors	Target sequence and editing window restrict available sites	Editing window and sequence context constrain targets	Gaudelli <i>et al.</i> (2017)

Modality	Molecular action	Best-fit use	Key benefits	Common failure modes / checks	Supporting primary studies
Prime editing	Programmed replacement, small insertion or deletion via pegRNA and reverse transcriptase	Introduce specified substitutions or short insertions and deletions	Efficiency can be low and target-dependent; recover and sequence independent events	Efficiency can be low and strongly target-dependent; regenerate and sequence independent events	Anzalone <i>et al.</i> (2019); Lin <i>et al.</i> (2020, 2021); Xu <i>et al.</i> (2022); Jiang <i>et al.</i> (2020)
CRISPR-Cas12a / related editors	Alternative RNA-guided nuclease systems	Address alternative target contexts, AT-rich sites or multiplex designs	Efficiency and repair outcomes require crop- and locus-specific testing	Crop- and target-specific efficiency and repair outcomes require validation	Zetsche <i>et al.</i> (2015); Tang <i>et al.</i> (2017)
Multiplex editing / de novo domestication	Simultaneous edits at several genes or regulatory regions	Combine multiple validated alleles or modify domestication-related traits	Epistasis, pleiotropy and background effects become more difficult to predict as edits accumulate	Epistasis, pleiotropy and background interactions increase rapidly with edit number	Zsögön <i>et al.</i> (2018); Lemmon <i>et al.</i> (2018)

5. Artificial intelligence in precision breeding

5.1. Genomic prediction: useful nonlinear models, but no automatic advantage

Genomic selection estimates breeding values from genome-wide marker data and allows candidates to be ranked before every trait has been measured in mature plants or across many environments. Meuwissen *et al.* (2001) established the core principle of predicting total genetic value from dense marker information; later work adapted and evaluated the approach in maize, wheat, rice and other crops (Crossa *et al.*, 2010; Heslot *et al.*, 2012; Poland *et al.*, 2012; Rutkoski *et al.*, 2012; Spindel *et al.*, 2015). Linear mixed models remain important baselines because they are efficient, comparatively easy to interpret and often competitive when additive effects explain much of the variation. Kernel and relationship-matrix methods can capture additional structure, while models that account for dominance or population-specific effects may be useful for particular hybrid-breeding settings (de los Campos *et al.*, 2010; Gianola and van Kaam, 2008; Endelman, 2011; Endelman and Jannink, 2012; Technow *et al.*, 2012). Machine-learning methods offer ways to represent nonlinear marker effects, epistasis and interactions between genotype and environment. Deep neural networks, random forests, boosting and kernel methods can identify patterns that simple additive models miss, but their flexibility also makes them capable of fitting noise or structure that will not persist in new material. Existing comparisons do not support a general claim that deep learning is always better for genomic prediction. Relative performance depends on population size, trait architecture, marker density, training design and validation strategy (Montesinos-López *et al.*, 2018; Azodi *et al.*, 2020). Apparent improvements can be misleading when close relatives occur in both training and test sets or observations from the same environment leak across validation folds.

Validation should reflect how the model will actually be used. Random cross-validation estimates prediction within a population that resembles the training sample and may be appropriate for selecting among related lines in the same breeding program. Leaving families or future generations out provides a harder test of whether predictions extend to new genetic material. When deployment involves unfamiliar locations or seasons, leave-environment-out validation is also needed. Changes in allele frequency, linkage disequilibrium or ancestry can erode accuracy even when internal validation looks strong. Results should be compared with a transparent baseline and reported with uncertainty and calibration measures; ultimately, subsequent breeding cycles should establish whether the model improves realized genetic gain.

Multi-omics can improve prediction when a molecular layer contains information that markers alone do not capture. But such layers may be costly and unstable: transcript abundance changes with tissue, stage and stress, while metabolite profiles respond to recent environmental conditions as well as genotype. Their contribution should therefore be measured in a genuinely independent test set, then weighed against the cost of collecting them. DNNGP applies deep learning to genomic prediction with multi-omics features; its reported performance should be interpreted in the context of the size and composition of the datasets on which it was evaluated (Wang *et al.*, 2023). Crop-specific validation is needed before such models become routine breeding tools.

Average prediction error is not always the outcome breeders care about most. A small reduction in error may leave crossing decisions unchanged, whereas better identification of the strongest candidates can materially affect the next breeding cycle. A model can also perform well on average while overlooking rare genotypes that carry valuable traits. Evaluation should therefore include selection success near the top of the ranking, realized gain, genetic diversity retained, stability across environments and cost per useful candidate—not only one accuracy

statistic. The relevant question is whether the model changes a breeding decision for the better.

5.2. Computer vision, phenomics and environmental data

Computer vision can make phenotyping more frequent and consistent by measuring traits such as leaf area, canopy architecture, biomass, disease symptoms, flowering and senescence. Deep-learning methods have shown promise for plant segmentation and trait estimation when training and test images are collected under comparable conditions (Pound *et al.*, 2017; Ubbens and Stavness, 2017; Ubbens *et al.*, 2018). Image-based models have also been related to physiologically meaningful stress responses, although a saliency map or feature-importance score does not establish a biological mechanism (Ghosal *et al.*, 2018). The practical benefit is the ability to turn labour-intensive observations into repeatable time-series measurements that can be analysed alongside genomic and environmental information.

Image models often fail when the conditions change. Lighting, camera angle, plant spacing, growth stage, soil background, cultivar architecture and management practices can all shift the data distribution. A model trained on isolated greenhouse plants may perform poorly in a field canopy; nutrient stress may resemble a disease symptom; or a biomass model may learn background colour that happens to correlate with yield in the training site. Reliable deployment requires representative data, explicit metadata, ground-truth measurements, external field tests and monitoring for drift. Where confidence is low or the image falls outside the model's training conditions, the system should flag uncertainty rather than produce a forced label.

Weather, soil and management data can help interpret why genotypes perform differently across environments. In the US Corn Belt, combining machine learning with crop modelling improved yield prediction, illustrating the value of pairing data-driven and process-based approaches (Millet *et al.*, 2019; Shahhosseini *et al.*, 2021). Yet models built from historical farm records can confuse management differences with genetic effects and location variables may act as shortcuts for other conditions. The breeding problem is more demanding than predicting yield in familiar fields: it is to estimate relative genetic performance in target environments, including environments that are poorly represented in the training data. This requires replicated genotypes across sites, reliable management records and explicit treatment of genotype-by-environment interaction.

For breeding, AI is often most defensible when used to measure, flag or prioritize rather than make unreviewable decisions. It can segment leaves, estimate disease severity, identify unusual growth trajectories or select plants for more detailed biochemical assays. Allowing an opaque model to discard germplasm or prescribe a gene edit without further evidence carries greater risk. Outputs should be connected to traits that breeders understand, checked across relevant subgroups and reported with uncertainty. Where sample sizes are modest, a simpler statistical model may be more robust and easier to update than a deep network.

Table 4. AI applications in breeding: decisions supported and minimum validation standards.

AI application	Inputs	Decision supported	Primary risk	Minimum validation	Supporting research
Genomic prediction	SNPs, haplotypes, pedigrees, omics	Rank lines and predict breeding values	Leakage, ancestry effects and non-transferable linkage disequilibrium	Use family-, generation- and environment-aware tests; compare with a genomic best linear unbiased prediction baseline	Meuwissen <i>et al.</i> (2001); Crossa <i>et al.</i> (2010); Montesinos-López <i>et al.</i> (2018)
Multi-omics prediction	Genome, transcripts, methylation, metabolites	Predict traits and prioritize informative molecular features	Batch effects, missing data and features acting as environmental proxies	Test independent germplasm, conduct ablation analysis and weigh gains against cost	Wang <i>et al.</i> (2023, 2024)
Computer vision phenotyping	RGB, multispectral, hyperspectral, thermal images	Measure architecture, disease, stress responses and maturity	Domain shift, data leakage and background confounding	Test independent field seasons against calibrated manual measurements	Pound <i>et al.</i> (2017); Ubbens and Stavness (2017); Ghosal <i>et al.</i> (2018)
Yield prediction	Genotype, weather, soils, management, remote sensing	Estimate performance and help prioritize trials	Confounding of location with genotype and weak climate extrapolation	Hold out locations or years and report calibrated uncertainty	Khaki and Wang (2019); Shahhosseini <i>et al.</i> (2021)

AI application	Inputs	Decision supported	Primary risk	Minimum validation	Supporting research
Candidate-gene prioritization	GWAS, QTL, regulatory networks, expression and editing records	Prioritize candidate genes and alleles for experiments	Circular annotations and bias toward well-studied genes	Validate prospectively and report negative experiments	Gao <i>et al.</i> (2019); Zhu <i>et al.</i> (2018)
Cross / parent optimization	Predicted values, kinship, diversity, allele frequencies	Choose parents and crossing combinations	Short-term gain may reduce diversity or robustness	Simulate candidate crosses, then test while monitoring diversity and inbreeding	Jarquín <i>et al.</i> (2014); Spindel <i>et al.</i> (2015)
Field-trial allocation	Uncertainty, target environments and trial cost	Select sites, seasons and genotypes for phenotyping	Trial allocation may miss rare stresses or farmer environments	Compare prospectively with standard trial allocation	Pound <i>et al.</i> (2017); Shahhosseini <i>et al.</i> (2021)

6. Integrating the technologies as a closed-loop breeding process

Integration should start with a breeding objective that can be measured in the field. Examples include maintaining grain yield under heat stress, improving disease resistance without delaying maturity, or preserving fruit quality under high temperature. The objective should specify the target population of environments, management conditions, market or quality requirements and acceptable trade-offs. Neither omics nor AI can rescue a poorly defined goal; they may simply optimize a convenient but incomplete proxy. Farmers' needs, seed systems and local management must therefore be part of the objective from the outset.

The first step is to establish what genetic diversity and evidence are already available. Pan-genomes, marker data, pedigrees, existing QTLs and historical trial records should be brought together with reliable information on location, season, soil, management and measurement methods. Relatedness and population structure should be examined before fitting prediction models. If the available panel does not represent the germplasm or environments for which decisions will be made, the priority is to improve sampling and records rather than compensate with a more complex algorithm.

The next step is to use multi-omics to identify and rank possible genes, regulatory elements, pathways or haplotypes. The most persuasive candidates are supported by more than one independent line of evidence—for example, a mapped locus, an expression change, a metabolite connection and a plausible physiological mechanism. Evidence should be graded rather than treated as all-or-nothing. Co-localization is a useful starting point; fine mapping and allele-specific evidence can narrow the explanation; and near-isogenic material, segregation or targeted perturbation can strengthen the causal case. Competing hypotheses and negative results should remain in the record so that prioritization can be updated honestly.

AI can then help decide which candidate to test and how to use limited experimental resources. A practical ranking should consider expected effect, probability of success, pleiotropic risk, editability, regeneration burden, uncertainty, transferability across backgrounds and cost. In some projects, the best recommendation will be a parent or cross rather than an edited allele. Image-based models may instead identify the stress trajectories or plants worth sampling for detailed molecular analysis. Active-learning approaches can help select informative genotypes and environments, but decision rules should also protect genetic diversity and account for groups that are poorly represented in existing data.

The next stage is an experiment designed to distinguish cause from correlation. For editing, that usually means multiple independent events, suitable non-edited controls and null segregants where possible. In polyploid crops, relevant homoeologues and unintended dosage effects need explicit assessment. Target-locus sequencing is essential and long-read or structural assays may be warranted where a double-strand break could produce larger rearrangements. In crossing experiments, haplotypes and recombination should be tracked so the target allele can be separated from unfavourable linked variation. The experimental design and criteria for success should be set before the results are known.

Controlled experiments can explain physiology, but replicated trials across locations and seasons determine whether a candidate is useful to breeders. Field evaluation should measure the target trait together with likely penalties and correlated traits, such as maturity, plant height, lodging, fertility, grain or fruit quality, disease spectrum and stability. AI predictions should be tested prospectively on material not used to fit the model wherever possible. These results then feed back into the next cycle, updating genomic predictions, environmental-response models and candidate rankings.

Before a successful line is deployed, the program still needs to address genetic stability, seed quality, performance under farm management, intellectual property and the applicable regulatory framework. The status of genome-edited products differs among jurisdictions and may depend on the final product as well as the development

process. Regulatory assumptions should therefore be checked for the specific crop and location, not generalized internationally. An edit becomes a breeding achievement only when the resulting product can be multiplied reliably and offers a benefit that growers and users can realize.

Table 5. A closed-loop workflow for integrating omics, AI, genome editing and field validation.

Stage	Key question	Data / technology	Decision gate	Evidence required to advance	Failure response
1. Define objective	What useful phenotype is required and under which conditions?	Target population of environments, trait definition and market or farmer requirements	Operational phenotype and constraints agreed	Predefined primary and secondary traits; clear target-environment map	Refine the objective before collecting expensive molecular data
2. Map diversity	Which alleles and genetic backgrounds are represented?	Pan-genome, genotyping, pedigree and historical trial records	Reference and training populations fit the intended use	Coverage of relevant ancestry and environments; complete metadata	Add germplasm and repair sampling gaps
3. Prioritize mechanisms	Which genes or pathways plausibly explain the variation?	Multi-omics, QTL/GWAS, regulatory networks and metabolite profiles	Ranked hypotheses with explicit confidence levels	Independent evidence and predictions that can be tested	Retain competing explanations and avoid premature certainty
4. Predict and design	Which lines, crosses or edits are most likely to be useful?	Genomic prediction, AI, optimization and uncertainty estimates	Candidate lines, crosses or edits selected for testing	External validation, comparison with a baseline and diversity constraints	Simplify the model or collect informative data if predictions do not transfer
5. Perturb and validate	Does the candidate change cause the predicted biological effect?	Genome editing, crosses, segregants and complementation	Causal hypothesis supported or revised	Independent events, complete genotyping and appropriate controls	Revise the proposed mechanism and retain negative evidence
6. Field test	Does the change deliver agronomic gain in target environments?	Replicated multi-location trials, sensors, omics and multi-environment models	Advance, redesign or stop	Yield, quality or stability benefit with acceptable trade-offs	Do not equate a molecular effect with breeding success
7. Deploy and learn	Can the product be deployed reliably, lawfully and usefully?	Seed multiplication, regulatory review, extension and farmer feedback	Release or continue testing	Current jurisdiction-specific review, seed quality and user-relevant benefit	Adapt the deployment pathway and update the evidence and prediction models

Table 6. Minimum evidence required for claims about omics-, editing- and AI-enabled breeding.

Claim being evaluated	Evidence required	Common overclaim	Recommended validation	Reporting minimum	Supporting references
Omics identifies a trait mechanism	Independent genetic evidence plus expression, protein or metabolite results consistent with a biological model	Treating co-variation or pathway enrichment as proof of causality	Fine mapping, allele-specific evidence, matched tissues/time points and targeted perturbation	Report sample size, batches, missingness, normalization and uncertainty	Chen <i>et al.</i> (2016); Wen <i>et al.</i> (2014); Wang <i>et al.</i> (2024)

Claim being evaluated	Evidence required	Common overclaim	Recommended validation	Reporting minimum	Supporting references
Candidate gene is causal	Allelic evidence, functional perturbation and rescue or complementation where feasible	Choosing a candidate solely because it lies near a GWAS peak or scores highly in a model	Independent edited events, segregating material, appropriate controls and molecular confirmation	Distinguish association, functional support and causal evidence	Gao <i>et al.</i> (2019); Alonge <i>et al.</i> (2020); Sha <i>et al.</i> (2023)
An edit is precise and heritable	Sequence-confirmed intended allele, assessment of by-products and transmission to progeny	Reporting editing efficiency without describing unwanted alleles or regeneration bias	Independent events, amplicon/long-read checks where warranted, off-target assessment and progeny testing	Specify editing system, target, generation, allele spectrum and assay limit	Li <i>et al.</i> (2018); Zong <i>et al.</i> (2018); Lin <i>et al.</i> (2021); Xu <i>et al.</i> (2022)
AI improves selection	Improved prediction against a strong baseline on data excluded from model fitting	Treating random cross-validation as evidence of deployable breeding gain	Family- or population-aware splits, leave-environment/year-out tests and external validation	Report baseline, splits, calibration, uncertainty and data leakage checks	Azodi <i>et al.</i> (2020); Sandhu <i>et al.</i> (2021); Wang <i>et al.</i> (2023)
A candidate improves a crop	Replicated agronomic performance with quality, stability and penalty traits recorded	Equating a molecular phenotype or single-site result with commercial value	Multi-location, multi-season trials and target-population-of-environments analysis	Report effect size, environment interaction, negative results and practical costs	Riedelsheimer <i>et al.</i> (2012b); Spindel <i>et al.</i> (2015); Sha <i>et al.</i> (2023)
A model generalizes across germplasm	Performance across distinct populations, backgrounds and environmental ranges	Assuming a model trained on elite material will transfer to unrelated landraces or future climates	Prospective test set, subgroup performance, recalibration and drift monitoring	Document ancestry, relatedness, genotype-by-environment effects and scope of use	Huang <i>et al.</i> (2010); Huang <i>et al.</i> (2012); Wang <i>et al.</i> (2023, 2024)
Integration is operationally worthwhile	Incremental benefit from added data and technology exceeds costs and throughput constraints	Assuming the most data-intensive workflow is necessarily the most useful	Ablation studies, cost-per-decision analysis and independent comparison with simpler workflows	Report cost, time, failure rate, reproducibility and deployment limits	Riedelsheimer <i>et al.</i> (2012b); Khaki and Wang (2019); Shahhosseini <i>et al.</i> (2021)

7. Persistent barriers and critical research priorities

7.1. Causality, confounding and reproducibility

Multi-omics studies often use relatively small and expensive datasets, whereas prediction models need populations that represent the diversity and environments in which they will be applied. If molecular assays are concentrated in a few elite lines or one environment, a model may learn those conditions rather than the biology of the trait. Sample identities must be traceable across assays; batch effects, missing data and exclusion criteria should be documented; and quality thresholds should be defined before analysis. The same applies to phenotyping, where instrument drift and inconsistent field protocols can produce apparent genetic differences. Data dictionaries, versioned code, clear model documentation and lawful access to underlying data are basic requirements for reproducibility.

A model may assign high importance to a gene because it marks ancestry or developmental stage, not because it regulates the trait. Feature-importance estimates also depend on the model and training set and correlated variables can divide or exchange apparent importance. Population-aware controls, independent cohorts, negative controls where appropriate and targeted perturbation can help test whether a candidate has a biological role. Interpretability is useful, but it should not be confused with causality: an explanation of how a model made a prediction is not, by itself, an explanation of how a plant trait is produced.

The literature is also vulnerable to publication bias. Successful edits, impressive prediction scores and neat mechanistic stories are easier to publish than failed experiments or results that contradict the hypothesis. The published record may therefore overstate the expected success rate of an integrated workflow. Standard benchmarking panels, independent challenge datasets, prospective reporting of editing outcomes and publication of null results would make comparisons more reliable. Failed edits and poor predictions are valuable data: they define the limits of a method and help subsequent models avoid repeating the same errors.

7.2. Scale, delivery and equitable access

The cost of technologies differs widely. Short-read genotyping and some cloud-computing workflows may be affordable to programs for which long-read pan-genomics, extensive metabolomics, proteomics or high-throughput transformation remain prohibitive. The sensible objective is not to measure everything, but to identify the smallest combination of measurements that changes a decision in a useful way. Targeted assays, marker panels or low-cost spectral proxies may retain much of the predictive value of more elaborate data. Breeding programs should compare the cost of a workflow with realized genetic gain, not judge it by cost per sample or model accuracy alone.

Transformation and regeneration are still major constraints for many crops and elite genotypes. An edit demonstrated in a model cultivar may be difficult to recover in the germplasm that breeders actually need. Delivery systems, transient reagents and regeneration protocols that work across genotypes are therefore as important as improvements in editor design. The best balance between editing, genomic selection and conventional crossing will differ among hybrid crops, long-generation species and vegetatively propagated crops. An effective program needs a plan tailored to the biology and breeding system of each crop rather than a fixed package of technologies. Who is represented in the data matters for both model performance and the distribution of benefits. Training datasets may overrepresent large public programs, widely grown cultivars and well-funded production environments. Models developed on those data may be less reliable for minor crops, landraces or low-input farmer fields. Data governance should address consent, benefit sharing, access rights and the treatment of locally important germplasm. Open data can strengthen reproducibility, but sensitive breeding information and genetic resources should be shared under transparent, appropriate governance rather than released without conditions.

7.3. Regulatory, biosafety and stewardship considerations

An edited plant may contain a change that is indistinguishable at the DNA level from a naturally occurring variant, but regulatory treatment of genome-edited crops differs across jurisdictions and continues to change. It is therefore not accurate to describe all edited crops as either unregulated or subject to the same rules as conventional transgenic organisms. Developers should retain clear records of the editing method, final genotype, presence or absence of introduced DNA, off-target evaluation and intended use. Requirements should be checked with the relevant authority before field testing or commercialization. Public engagement and product stewardship also matter when changes affect food composition, gene flow, weediness or culturally important varieties.

AI-supported selection raises its own stewardship questions. Repeatedly choosing the same lines because a model predicts high short-term yield can narrow the genetic base and weaken future adaptability. Selection objectives should account for genetic gain alongside diversity, resilience and performance stability. Data security, model reproducibility, uncertainty, human review and accountability become particularly important when automated rankings determine which costly experiments are conducted or which lines are discarded. AI should make breeding decisions more transparent and better informed, not remove responsibility from the people making them.

8. DISCUSSION: WHAT CONSTITUTES CONVINCING INTEGRATION?

The evidence for integration is convincing only when three conditions are met. First, each data layer should add information that is not already available from a simpler or cheaper source. An omics feature that mainly tracks ancestry, or an image feature that duplicates a routine field score, may have little practical value. Second, the biological explanation should survive an appropriate test. Depending on the question, this may involve editing, complementation, allele replacement, segregation or another functional experiment. Third, the effect should persist in relevant breeding material under realistic management and across target environments. Technically elaborate studies that do not make an explicit decision or validate its consequences are less informative than studies that test a smaller number of clear hypotheses rigorously.

A useful integrated program need not apply every technology to every trait. For a disease-resistance target with a major-effect susceptibility gene, genetic analysis and editing may be decisive, with little need for comprehensive metabolomics. For a complex quality trait, pan-genomic variation and metabolite profiles may help narrow candidates, while editing becomes useful only once the causal allele is sufficiently resolved. For yield under variable environments, genomic prediction and environmental data may offer the greatest immediate benefit; profiling every line by transcriptomics could be too costly to justify. The combination should be chosen according to the genetic architecture, available resources and decision that needs to be improved.

It is also important to separate discovery value from selection value. A molecular measurement may explain a pathway without improving prediction, while a low-cost sensor feature may improve selection without revealing the underlying mechanism. Both can be worthwhile, but they support different claims. Discovery should be judged by the strength of the causal explanation and whether it identifies an actionable target; prediction by performance on independent material; editing by the intended molecular change and its biological effect; and breeding by genetic gain, stability and value to growers or end users.

Integration should be treated as a process that is revisited as evidence accumulates. Genotype-phenotype relationships depend on genetic background, developmental stage, environment and management. An edit may interact with other alleles; a prediction model may lose accuracy as the breeding population changes; and a resistance model may no longer apply when pathogen populations shift. Programs therefore need to monitor performance, recalibrate models when justified and set explicit criteria for retraining. Environmental variation is part of the breeding problem, not merely noise to be removed from the data.

Future studies would be easier to compare if they consistently reported the germplasm and environments represented; the provenance and quality control of each data layer; the design of training, validation and test sets; effect sizes and uncertainty rather than significance alone; independent validation of high-priority causal claims; the full spectrum of editing outcomes; and multi-environment agronomic performance with economic context. Such reporting would help distinguish real progress from improvements in model accuracy, editing efficiency or data volume that do not translate into better crops.

9. CONCLUSIONS

Multi-omics, genome editing and AI can strengthen precision breeding because they address different stages of the path from inherited variation to crop performance. Multi-omics helps connect genotype with molecular state and identify plausible mechanisms. Editing tests whether selected changes contribute to a phenotype and can create targeted alleles. AI supports prediction and measurement across large populations and variable environments. Their integration can make crop improvement more efficient and more explanatory, but it cannot replace experimental testing or field selection. The strongest evidence comes from studies that connect molecular discovery with causal validation and a durable agronomic benefit.

The next step is not to maximize the number of technologies in a breeding program, but to build workflows with clear evidence thresholds and practical decision gates. Data must represent the germplasm and environments where a variety is intended to perform; models must be challenged outside their training data; edits must be characterized and tested across generations; and success must be measured in yield, quality, resilience, stability and outcomes relevant to farmers. Costs, regulatory requirements and negative findings should be reported alongside successful examples. Precision breeding will fulfil its promise when molecular resolution, predictive tools and causal experiments work together in a repeatable breeding process and the strength of each claim remains proportionate to the evidence.

ACKNOWLEDGEMENT

The authors sincerely acknowledge the support and guidance received from their respective institutions and departments during the preparation of this article. The authors are grateful to all the faculty members, scientists, research scholars and colleagues who provided valuable suggestions, encouragement and assistance. The authors also acknowledge the institutional facilities and resources that contributed to the successful completion of this work.

REFERENCES

1. Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., Levy, Y., Harel, T.H., Shalev-Schlosser, G., Amsellem, Z., Razifard, H., Caicedo, A.L., Tieman, D.M., Klee, H., Kirsche, M., Aganezov, S., Ranallo-Benavidez, T.R., Lemmon, Z.H., Kim, J., Robitaille, G., Kramer, M., Goodwin, S., McCombie, W.R., Hutton, S., Van Eck, J., Gillis, J., Eshed, Y., Sedlazeck, F.J., van der Knaap, E., Schatz, M.C. and Lippman, Z.B. (2020). Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, 182, 145-161.e23. <https://doi.org/10.1016/j.cell.2020.05.021>
2. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 576, 149-157. <https://doi.org/10.1038/s41586-019-1711-4>
3. Azodi, C.B., Tang, J. and Shiu, S.-H. (2020). Benchmarking parametric and machine learning models for genomic prediction of complex traits. *G3: Genes, Genomes, Genetics*, 10, 3691-3702. <https://doi.org/10.1534/g3.120.401490>
4. Cai, Y., Chen, L., Liu, X., Guo, C., Sun, S., Wu, C., Jiang, B., Han, T. and Hou, W. (2018). CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soybean. *Plant Biotechnology Journal*, 16, 1761-1771. <https://doi.org/10.1111/pbi.12758>
5. Carreño-Quintero, N., Acharjee, A., Maliepaard, C., Bachem, C.W.B., Mumm, R., Bouwmeester, H., Visser, R.G.F. and Keurentjes, J.J.B. (2012). Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. *The Plant Cell*, 24, 239-251. <https://doi.org/10.1105/tpc.111.092221>

6. Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., Wang, S., Shi, L., Zhou, B., Li, Z., Peng, X., Yang, C., Qu, L., Liu, X. and Luo, J. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nature Communications*, 7, 12767. <https://doi.org/10.1038/ncomms12767>
7. Crossa, J., de los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J.L., Makumbi, D., Singh, R.P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M. and Braun, H.-J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186, 713-724. <https://doi.org/10.1534/genetics.110.118521>
8. de los Campos, G., Gianola, D., Rosa, G.J.M., Weigel, K.A. and Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research*, 92, 295-308. <https://doi.org/10.1017/S0016672310000285>
9. Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, 4, 250-255. <https://doi.org/10.3835/plantgenome2011.08.0024>
10. Endelman, J.B. and Jannink, J.-L. (2012). Improved shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genetics*, 2, 1405-1413. <https://doi.org/10.1534/g3.112.004259>
11. Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y. and Zhu, J.-K. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, 23, 1229-1232. <https://doi.org/10.1038/cr.2013.114>
12. Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D.M., Burzynski-Chang, E.A., Fish, T.L., Stromberg, K.A., Sacks, G.L., Thannhauser, T.W., Foolad, M.R., Diez, M.J., Blanca, J., Cañizares, J., Xu, Y., van der Knaap, E., Huang, S., Klee, H.J., Giovannoni, J.J. and Fei, Z. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics*, 51, 1044-1051. <https://doi.org/10.1038/s41588-019-0410-2>
13. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I. and Liu, D.R. (2017). Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature*, 551, 464-471. <https://doi.org/10.1038/nature24644>
14. Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A. and Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 4613-4618. <https://doi.org/10.1073/pnas.1716999115>
15. Gianola, D. and van Kaam, J.B.C.H.M. (2008). Reproducing kernel Hilbert spaces regression methods for genomic-assisted prediction of quantitative traits. *Genetics*, 178, 2289-2303. <https://doi.org/10.1534/genetics.107.084285>
16. Golicz, A.A., Bayer, P.E., Barker, G.C., Edger, P.P., Kim, H., Martinez, P.A., Chan, C.K.K., Severn-Ellis, A., McCombie, W.R., Parkin, I.A.P., Pires, J.C., Sharpe, A.G., Tang, H., Town, C.D., Batley, J. and Edwards, D. (2016). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications*, 7, 13390. <https://doi.org/10.1038/ncomms13390>
17. Haun, W., Coffman, A., Clasen, B.M., Demorest, Z.L., Lowy, A., Ray, E., Retterath, A., Stoner, M., Showalter, M., Head, G., Spangler, R. and Pohl, B. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnology Journal*, 12, 934-940. <https://doi.org/10.1111/pbi.12201>
18. Heslot, N., Yang, H.-P., Sorrells, M.E. and Jannink, J.-L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Science*, 52, 146-160. <https://doi.org/10.2135/cropsci2011.09.0497>
19. Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E.S., Qian, Q., Zhang, Q.-F., Li, J. and Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42, 961-967. <https://doi.org/10.1038/ng.695>
20. Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., Fan, D., Lu, Y., Weng, Q., Liu, K., Zhou, T., Jing, Y., Si, L., Dong, G., Huang, T., Lu, T., Feng, Q., Qian, Q., Li, J. and Han, B. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics*, 44, 32-39. <https://doi.org/10.1038/ng.1018>
21. International Wheat Genome Sequencing Consortium (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361, eaar7191. <https://doi.org/10.1126/science.aar7191>
22. Jarquín, D., Crossa, J., Lacaze, X., Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Huston, C. and McCouch, S.R. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, 127, 595-607. <https://doi.org/10.1007/s00122-013-2243-1>
23. Jia, H., Zhang, Y., Orbović, V., Xu, J., White, F.F., Jones, J.B. and Wang, N. (2017). Genome editing of the disease susceptibility gene *CsLOB1* in citrus confers resistance to citrus canker. *Plant Biotechnology Journal*, 15, 817-823. <https://doi.org/10.1111/pbi.12677>
24. Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. and Weeks, D.P. (2013). Demonstration of CRISPR/Cas9/sgrRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Research*, 41, e188. <https://doi.org/10.1093/nar/gkt780>
25. Jiang, Y.-Y., Chai, Y.-P., Lu, M.-H., Han, X.-L., Lin, Q., Zhang, Y., Zhang, Q., Zhou, Y., Wang, X.-C. and Chen, Q.-J. (2020). Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in

maize. *Genome Biology*, 21, 257. <https://doi.org/10.1186/s13059-020-02170-1>

26. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821. <https://doi.org/10.1126/science.1225829>
27. Keurentjes, J.J.B., Fu, J., de Vos, C.H.R., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H.W., Jansen, R.C., Vreugdenhil, D. and Koornneef, M. (2006). The genetics of plant metabolism. *Nature Genetics*, 38, 842-849. <https://doi.org/10.1038/ng1815>
28. Khaki, S. and Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10, 621. <https://doi.org/10.3389/fpls.2019.00621>
29. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533, 420-424. <https://doi.org/10.1038/nature17946>
30. Lemmon, Z.H., Reem, N.T., Dalrymple, J., Soyk, S., Swartwood, K.E., Rodriguez-Leal, D., Van Eck, J. and Lippman, Z.B. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. *Nature Plants*, 4, 766-770. <https://doi.org/10.1038/s41477-018-0249-7>
31. Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M. and Sheen, J. (2013). Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and tobacco using CRISPR/Cas9. *Nature Biotechnology*, 31, 688-691. <https://doi.org/10.1038/nbt.2654>
32. Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., Zhang, R. and Gao, C. (2018). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biology*, 19, 59. <https://doi.org/10.1186/s13059-018-1443-z>
33. Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., Wang, Y., Anzalone, A.V., Raguram, A., Doman, J.L., Liu, D.R. and Gao, C. (2020). Prime genome editing in rice and wheat. *Nature Biotechnology*, 38, 582-585. <https://doi.org/10.1038/s41587-020-0455-x>
34. Lin, Q., Jin, S., Zong, Y., Yu, H., Zhu, Z., Liu, G., Kou, L., Wang, Y., Qiu, J.-L., Li, J., Anzalone, A.V., Raguram, A., Doman, J.L., Liu, D.R. and Gao, C. (2021). High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nature Biotechnology*, 39, 923-927. <https://doi.org/10.1038/s41587-021-00868-w>
35. Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.-A., Zhang, H., Liu, Z., Shi, M., Huang, X., Li, Y., Ma, M., Zhang, M., Wang, Z., Zhu, B., Han, B., Liang, C. and Tian, Z. (2020). Pan-genome of wild and cultivated soybeans. *Cell*, 182, 162-176.e13. <https://doi.org/10.1016/j.cell.2020.05.023>
36. Lu, Y. and Zhu, J.-K. (2017). Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Molecular Plant*, 10, 523-525. <https://doi.org/10.1016/j.molp.2016.11.013>
37. Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J.-I., Ebana, K., Yano, M. and Saito, K. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *The Plant Journal*, 81, 13-23. <https://doi.org/10.1111/tpj.12681>
38. McCormick, R.F., Truong, S.K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., Kennedy, M., Amirebrahimi, M., Weers, B.D., McKinley, B.A., Mattison, A.J., Morishige, D.T., Grimwood, J., Schmutz, J. and Mullet, J.E. (2018). The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas and signatures of genome organization. *The Plant Journal*, 93, 338-354. <https://doi.org/10.1111/tpj.13781>
39. McNally, K.L., Bruskiewich, R., Mackill, D., Buell, C.R., Leach, J.E. and Leung, H. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 12273-12278. <https://doi.org/10.1073/pnas.0900992106>
40. Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819-1829. <https://doi.org/10.1093/genetics/157.4.1819>
41. Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H. and Qu, L.-J. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Research*, 23, 1233-1236. <https://doi.org/10.1038/cr.2013.123>
42. Millet, E.J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., Charcosset, A., Welcker, C., van Eeuwijk, F. and Tardieu, F. (2019). Genomic prediction of maize yield across environments using genomic, environmental and management data. *The Plant Genome*, 12, 180074. <https://doi.org/10.3835/plantgenome2018.09.0074>
43. Montesinos-López, O.A., Montesinos-López, A., Crossa, J., Toledo, F.H., Pérez-Hernández, O., Eskridge, K.M. and Campos, G. (2018). Multi-trait, multi-environment genomic prediction of plant breeding values using deep learning and Bayesian methods. *G3: Genes, Genomes, Genetics*, 8, 3829-3840. <https://doi.org/10.1534/g3.118.200585>
44. Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D.G. and Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using the RNA-guided Cas9 endonuclease. *Nature Biotechnology*, 31, 691-693. <https://doi.org/10.1038/nbt.2655>
45. Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D. and Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Scientific Reports*, 7, 482. <https://doi.org/10.1038/s41598-017-00578-x>
46. Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J.C., Perez-Quintero, A., Li, T., Eom, J.-S., Li, C., Nguyen, H., Liu, B., Auguy, F., Sciallano, C., Luu, V.T., Dossa, G.S., Cunnac, S., Schmidt, S.M., Slamet-Loedin,

- I.H., White, F.F., Yang, B. and Ronald, P.C. (2019). Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nature Biotechnology*, 37, 1344-1350. <https://doi.org/10.1038/s41587-019-0267-z>
47. Piatek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., Aouida, M. and Mahfouz, M.M. (2015). RNA-guided transcriptional regulation in planta via CRISPR/Cas9. *Nature Plants*, 1, 15014. <https://doi.org/10.1038/nplants.2015.14>
48. Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M. and Jannink, J.-L. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome*, 5, 103-113. <https://doi.org/10.3835/plantgenome2012.06.0006>
49. Pound, M.P., Atkinson, J.A., Townsend, A.J., Wilson, M.H., Griffiths, M., Jackson, A.S., Bulat, A., Tzimiropoulos, G., Wells, D.M., Murchie, E.H., Pridmore, T.P. and French, A.P. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*, 6, gix083. <https://doi.org/10.1093/gigascience/gix083>
50. Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., Altmann, T., Stitt, M., Willmitzer, L. and Melchinger, A.E. (2012a). Genome-wide association mapping of leaf metabolic profiles for dissecting the genetic architecture of metabolic traits in maize. *Genetics*, 192, 991-1005. <https://doi.org/10.1534/genetics.112.143602>
51. Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L. and Melchinger, A.E. (2012b). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics*, 44, 217-220. <https://doi.org/10.1038/ng.1033>
52. Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., Jannink, J.-L. and Sorrells, M. (2012). Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *The Plant Genome*, 5, 51-61. <https://doi.org/10.3835/plantgenome2012.02.0001>
53. Sandhu, K., Lozada, D.N., Zhang, Z., Pumphrey, M.O. and Carter, A.H. (2021). Multitrait machine- and deep-learning models for genomic selection using spectral information in a wheat breeding program. *The Plant Genome*, 14, e20119. <https://doi.org/10.1002/tpg2.20119>
54. Schauer, N., Semel, Y., Balbo, I., Steinfath, M., Repsilber, D., Selbig, J., Tohge, T., Fernie, A.R., Zamir, D. and Willmitzer, L. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnology*, 24, 447-454. <https://doi.org/10.1038/nbt1194>
55. Sha, G., Sun, P., Kong, X., Han, X., Sun, Q., Fouillen, L., Zhao, J., Li, Y., Yang, L., Wang, Y., Gong, Q., Zhou, Y., Zhou, W., Jain, R., Gao, J., Huang, R., Chen, X., Zheng, L., Zhang, W., Qin, Z., Zhou, Q., Zeng, Q., Xie, K., Xu, J., Chiu, T.-Y., Guo, L., Mortimer, J.C., Boutté, Y., Li, Q., Kang, Z., Ronald, P.C. and Li, G. (2023). Genome editing of a rice CDP-DAG synthase confers multipathogen resistance. *Nature*, 618, 1017-1023. <https://doi.org/10.1038/s41586-023-06205-2>
56. Shahhosseini, M., Hu, G., Huber, I. and Archontoulis, S.V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, 11, 1606. <https://doi.org/10.1038/s41598-020-80820-1>
57. Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L. and Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology*, 31, 686-688. <https://doi.org/10.1038/nbt.2650>
58. Shimatani, Z., Fujikura, U., Ishii, H., Matsui, Y., Suzuki, M., Ueke, Y., Taoka, K.-I., Terada, R., Nishida, K., Kondo, A., Nishimasu, H., Nureki, O. and Toki, S. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nature Biotechnology*, 35, 441-443. <https://doi.org/10.1038/nbt.3833>
59. Spindel, J.E., Begum, H., Akdemir, D., Virk, P., Collard, B.C.Y., Redoña, E.D., Atlin, G.N., Jannink, J.-L. and McCouch, S.R. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genetics*, 11, e1004982. <https://doi.org/10.1371/journal.pgen.1004982>
60. Tang, X., Lowder, L.G., Zhang, T., Malzahn, A.A., Zheng, X., Voytas, D.F., Zhong, Y., Qi, Y. and Zhang, Y. (2017). A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nature Plants*, 3, 17018. <https://doi.org/10.1038/nplants.2017.18>
61. Technow, F., Riedelsheimer, C., Schrag, T.A. and Melchinger, A.E. (2012). Genomic prediction of hybrid performance in maize with models incorporating dominance and population-specific marker effects. *Theoretical and Applied Genetics*, 125, 1181-1194. <https://doi.org/10.1007/s00122-012-1905-x>
62. The 1001 Genomes Consortium (2016). 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, 166, 481-491.e15. <https://doi.org/10.1016/j.cell.2016.05.063>
63. Torkamaneh, D., Lemay, M.-A. and Belzile, F. (2021). The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. *Plant Biotechnology Journal*, 19, 1852-1862. <https://doi.org/10.1111/pbi.13600>
64. Ubbens, J.R. and Stavness, I. (2017). Deep Plant Phenomics: a deep learning platform for complex plant phenotyping tasks. *Frontiers in Plant Science*, 8, 1190. <https://doi.org/10.3389/fpls.2017.01190>
65. Ubbens, J.R., Cieslak, M., Prusinkiewicz, P., Parkin, I. and Stavness, I. (2018). The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods*, 14, 6. <https://doi.org/10.1186/s13007-018-0276-8>

66. Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M.-P., Guyon-Debast, A., Chauvin, J.-E., Nogu  , F. and Mazier, M. (2019). Transgene-free genome editing in tomato and potato plants using *Agrobacterium*-mediated delivery of CRISPR/Cas9 ribonucleoprotein complexes. *Plant Biotechnology Journal*, 17, 1176-1178. <https://doi.org/10.1111/pbi.13081>
67. Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.-L. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, 32, 947-951. <https://doi.org/10.1038/nbt.2969>
68. Wang, K., Abid, M.A., Rasheed, A., Crossa, J., Hearne, S. and Li, H. (2023). DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Molecular Plant*, 16, 279-293. <https://doi.org/10.1016/j.molp.2022.11.004>
69. Wang, P., Lehti-Shiu, M.D., Lotreck, S., Segura Ab  , K., Krysan, P.J. and Shiu, S.-H. (2024). Prediction of plant complex traits via integration of multi-omics data. *Nature Communications*, 15, 6856. <https://doi.org/10.1038/s41467-024-50701-6>
70. Wei, J., Wang, A., Li, R., Qu, H., Wang, J., Zhang, X. and Jia, Z. (2018). Metabolome-wide association studies for agronomic traits of rice. *Heredity*, 120, 342-355. <https://doi.org/10.1038/s41437-017-0032-3>
71. Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J. and Yan, J. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications*, 5, 3438. <https://doi.org/10.1038/ncomms4438>
72. Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., Wang, X.-C. and Chen, Q.-J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology*, 14, 327. <https://doi.org/10.1186/s12870-014-0327-y>
73. Xu, W., Yang, Y., Yang, B., Krueger, C.J., Xiao, Q., Zhao, S., Zhang, L., Kang, G., Wang, F., Yi, H., Ren, W., Li, L., He, X., Zhang, C., Zhang, B., Zhao, J. and Yang, J. (2022). A design optimized prime editor with expanded scope and capability in plants. *Nature Plants*, 8, 45-52. <https://doi.org/10.1038/s41477-021-01043-4>
74. Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S. and Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38, 203-208. <https://doi.org/10.1038/ng1702>
75. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., Koonin, E.V. and Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163, 759-771. <https://doi.org/10.1016/j.cell.2015.09.038>
76. Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Zhang, H., Xu, N. and Zhu, J.-K. (2014). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnology Journal*, 12, 797-807. <https://doi.org/10.1111/pbi.12200>
77. Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., Lin, T., Qin, M., Peng, M., Yang, C., Cao, X., Han, X., Wang, X., van der Knaap, E., Zhang, Z., Cui, X., Klee, H., Fernie, A.R., Luo, J. and Huang, S. (2018). Rewiring of the fruit metabolome in tomato breeding. *Cell*, 172, 249-261.e12. <https://doi.org/10.1016/j.cell.2017.12.019>
78. Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.-L., Wang, D. and Gao, C. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nature Biotechnology*, 35, 438-440. <https://doi.org/10.1038/nbt.3811>
79. Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., Qiu, J.-L. and Gao, C. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nature Biotechnology*, 36, 950-953. <https://doi.org/10.1038/nbt.4261>
80. Zs  g  n, A.,   erm  k, T., Naves, E.R., Notini, M.M., Edel, K.H., Weinl, S., Freschi, L., Voytas, D.F., Kudla, J. and Peres, L.E.P. (2018). De novo domestication of wild tomato using genome editing. *Nature Biotechnology*, 36, 1211-1216. <https://doi.org/10.1038/nbt.4272>