PRISM: cross-species connectivity mapping of plant intervention and stress signatures for translational agriculture

Andrew J. Regan[1,*]

[1] Sarborg Limited, Grand Cayman, Cayman Islands

[*] Corresponding author: Andrew Regan, Ph.D. (ar@sarborg.com)

Abstract

Connectivity mapping – the systematic comparison of transcriptomic signatures to discover functional relationships among compounds, genes, and biological states – has transformed drug repurposing and mechanism-of-action discovery in human biology, but no equivalent resource exists for plants. Here we introduce PRISM (Plant Response to Intervention and Stress Signature Mapping), the first large-scale, multi-species transcriptomic intervention–response mapping resource for plant biology. Drawing on 6,960 publicly deposited expression profiling experiments from the Gene Expression Omnibus, we constructed intervention signature workspaces for six species: *Arabidopsis thaliana* (thale cress; 1,979 signatures from 660 experiments), *Oryza sativa* (rice; 265 signatures), *Solanum lycopersicum* (tomato; 46), *Zea mays* (maize; 24), *Glycine max* (soybean; 20), and *Vitis vinifera* (grape; 18) – totaling 2,352 signatures derived from 865 experiments. Our data acquisition strategy combined compendia extraction, microarray processing across multiple platforms, and RNA-seq mining. An organism-agnostic connectivity engine computes pairwise connectivity scores between intervention (e.g., fungicide application) gene expression signatures and stress-state (e.g., drought stress) gene expression signatures, quantifying the degree to which any two biological states share or oppose transcriptomic programs. By systematically mapping which interventions share or oppose the transcriptomic programs of plant stress states, this resource reveals non-obvious applications for existing interventions, prioritizes candidates for crop protection, and guides the rational design of intervention strategies – the agricultural equivalent of computational drug repurposing.

The depth of mechanistic knowledge in plant hormone and stress biology provides strong a priori expectations – for example, that salicylic acid treatment should transcriptomically resemble pathogen defense, or that a hormone biosynthesis mutant should oppose its cognate hormone response. We exploited this by defining 77 a priori validation targets spanning eight categories – from technical controls through hormone–stress mimicry, genetic reversal, pharmacological mechanism, and genetic epistasis. Of these, 56 (73%) confirmed the predicted relationship, a rate comparable to human CMap and LINCS benchmarks. The 21 unexpected results cluster in three interpretable groups (hormone antagonism, auxin pharmacology, and pleiotropic genetics), revealing where whole-transcriptome comparisons capture shared biology but miss pathway-specific distinctions. The engine predicts combination transcriptomic effects from individual component signatures with 93% accuracy (39 of 42 tested combinations), enabling rational mixture design without exhaustive pairwise screening. Intervention–response relationships learned in Arabidopsis – such as salicylic acid mimicking pathogen defense – predicted rice biology across all 5 testable cross-species comparisons, the strongest evidence that the framework captures conserved biological programs. PRISM enables repurposing of existing interventions, rational design of compound mixtures, and translation of discoveries from model species to crops – placing mechanism-informed crop protection on a systematic, data-driven footing.

Introduction

The Connectivity Map (CMap) demonstrated that systematic comparison of transcriptomic signatures can reveal functional relationships among drugs, genes, and disease states that would be difficult to discover by any other means (Lamb et al., 2006). By treating each compound's transcriptomic effect as a queryable signature, CMap and its successor LINCS (Subramanian et al., 2017) enabled researchers to identify non-obvious drug–disease connections, repurpose existing therapeutics for new indications, and infer mechanisms of action for uncharacterized compounds – all from gene expression data alone. This paradigm – intervention–response mapping – has since become a foundational tool in human pharmacology, underpinning drug repurposing pipelines, toxicogenomic screening, and mechanism-of-action classification at scale.

No equivalent resource exists for plant biology, despite several factors that make plants exceptionally well-suited to this approach. First, decades of molecular genetics have produced a deeply characterized network of hormone signaling pathways – salicylic acid (SA) in pathogen defense, jasmonic acid (JA) in wounding and herbivory, abscisic acid (ABA) in drought and osmotic stress, auxins in growth and development, among others (Pieterse et al., 2012; Verma et al., 2016) – whose transcriptomic consequences are well understood and generate strong a priori expectations. These expectations provide a built-in validation framework: if an intervention–response mapping engine works correctly, SA treatment should transcriptomically resemble pathogen infection, hormone biosynthesis mutants should oppose their cognate hormone responses, and defense elicitors should mimic pathogen challenge. Second, public repositories contain a wealth of plant transcriptomic data – thousands of experiments spanning chemical treatments, genetic perturbations, pathogen challenges, and abiotic stresses across multiple species – that have never been systematically integrated into a connectivity framework. Third, the agricultural applications are immediate and commercially relevant: identifying compounds that mimic protective stress responses, rationally designing mixtures of biostimulants or crop protection agents, and translating discoveries from model species to crops all represent high-value applications of intervention–response mapping.

Several efforts have explored elements of this approach. The HORMONOMETER tool compared query transcriptomes against eight Arabidopsis hormone reference signatures to classify hormone responses (Volodarsky et al., 2009), and the GSHR platform extended this to 1,368 hormone-regulated gene lists (Ran et al., 2018). Lai and Ge (2014) computed all-versus-all overlaps among 1,065 Arabidopsis gene expression signatures, revealing hidden links among biological processes. Plant PhysioSpace projects new data into a multi-species physiological signature space spanning Arabidopsis, rice, soybean, and wheat (Esfahani et al., 2021), and Hartmann et al. (2022) compared transcriptomic responses to oxidative stress and hormone treatment across Arabidopsis, rice, and barley. However, none of these provides a unified, multi-species connectivity scoring framework with directional (mimicry versus reversal) quantification, systematic validation against mechanistic predictions, or the combination prediction and cross-species transfer capabilities tested here.

The practical challenges of building such a resource are substantial. Plant transcriptomic data are distributed across hundreds of microarray platforms and RNA-seq experiments with heterogeneous metadata, inconsistent annotation standards, and variable experimental quality. Unlike the human CMap, which profiled a curated library of compounds on a single standardized platform (L1000), a plant intervention–response mapping resource must integrate data from multiple platforms, multiple species, and multiple experimental designs while maintaining sufficient quality for meaningful biological inference. Gene

identifier systems vary across species and genome assembly versions, legacy microarray platforms use probe annotations that map poorly to current genome assemblies, and the fraction of RNA-seq experiments with readily accessible count matrices is limited. These challenges have prevented the plant biology community from assembling a connectivity resource despite having the raw data to do so.

Here we introduce PRISM (Plant Response to Intervention and Stress Signature Mapping), an intervention–response mapping resource for plant biology that addresses these challenges through a multi-source data acquisition strategy, an organism-agnostic connectivity engine, and a systematic validation framework. We constructed transcriptomic signature workspaces for six species – *Arabidopsis thaliana*, *Oryza sativa* (rice), *Solanum lycopersicum* (tomato), *Zea mays* (maize), *Glycine max* (soybean), and *Vitis vinifera* (grape) – totaling 2,352 signatures derived from 865 experiments across compendia, microarray, and RNA-seq sources. An organism-agnostic connectivity engine computes pairwise connectivity scores between intervention signatures and stress-state gene sets, quantifying the degree to which any intervention mimics or opposes any stress state at whole-transcriptome scale.

We validate the resource through three progressively stringent tests. First, we evaluate mechanism recovery against 77 a priori predictions spanning eight mechanistic categories – from self-consistency controls through hormone–stress mimicry, genetic reversal, pharmacological mechanism, and genetic epistasis – establishing a 73% confirmation rate that is comparable to human CMap and LINCS benchmarks. Second, we test combination prediction, demonstrating that the engine predicts combination transcriptomic effects from individual component signatures with 93% accuracy across 42 predictions – a capability with direct implications for rational mixture design in crop protection. Third, we test cross-species transfer by projecting validated Arabidopsis intervention signatures into rice gene space via ortholog mapping, confirming that intervention–response relationships learned in one species predict biology in another across all testable comparisons. Together, these validations establish that PRISM captures genuine, conserved biological programs and provide both a proof of concept and an interpretive framework for intervention–response mapping in plant biology.

Results

A multi-source pipeline yields 2,352 transcriptomic signatures across six plant species

Transcriptomic landscape survey

We surveyed the Gene Expression Omnibus for plant expression profiling experiments and identified 7,954 GEO series encompassing 191,825 samples across 14 major plant species (Table 1). *Arabidopsis thaliana* dominated the landscape with 4,570 series (57%), followed by rice (1,087 series, 14%), maize (508 series, 6%), and tomato (309 series, 4%). We selected six species for PRISM based on data volume, genome annotation quality, and agricultural relevance: *Arabidopsis*, rice, tomato, maize, soybean, and grape, totaling 6,960 GEO series.

Signature construction across six species

We extracted transcriptomic signatures – defined here as meaningful treatment-versus-control contrasts suitable for differential gene expression analysis – from three complementary data sources, balancing depth against coverage (Figure 1A).

Compendia are large-scale factorial experiments with multi-factor designs (genotype × treatment × timepoint) that yield dozens to hundreds of distinct contrasts per experiment. We identified 10 *Arabidopsis* and 6 rice compendia with factorial designs suitable for deep contrast extraction. Of these, 7 *Arabidopsis* compendia produced 547 signatures (ranging from 25 to 214 per experiment) and 3 rice compendia produced 24 signatures, with median gene counts of 18,498–33,602 and 24,704–55,801 respectively. The remaining four species lacked compendia with treatment contrasts suitable for intervention–response mapping – most large-scale experiments in these species are developmental or genetic variation studies. Despite their scarcity, compendia are disproportionately valuable: 10 experiments contributed 571 signatures (24% of the total resource), with the highest gene counts and most internally consistent experimental designs.

Microarray experiments provided the broadest species coverage. For each species, we identified a dominant anchor platform – selected for having the most treatment and stress experiments on GEO – and extended to additional platforms where probe-to-gene mapping was feasible. Anchor platforms ranged from the deeply annotated *Arabidopsis* ATH1 array (GPL198; 21,532 mapped genes, 250 experiments) to sparser legacy arrays for maize and soybean (GPL4032 and GPL4592; 899 and 931 mapped genes, respectively). A recurring challenge was that pre-genome-era arrays use EST-based or Entrez-based probe identifiers that map poorly to current genome assemblies – all 9 non-anchor maize platforms achieved <1% probe-to-gene coverage, and soybean non-anchor platforms similarly failed. Modern platforms for rice (Agilent arrays) and grape (NimbleGen, Agilent) achieved substantially better mapping rates (>80%). Across all species, microarray processing yielded 1,535 signatures (65% of the resource).

RNA-seq count matrices from GEO supplementary files extended the resource to more recent experiments. We triaged RNA-seq series for the availability of parseable count files and processed them through a limma–voom differential expression pipeline. Yield was constrained: only 20–22% of feasible RNA-seq experiments across species had downloadable count matrices, with the remainder stored only as raw FASTQ archives. RNA-seq processing contributed 246 signatures (10% of the resource) but with high gene counts (typically >3,000 genes per signature), complementing the broader but sometimes sparser microarray coverage.

Table 1 summarizes the data funnel from GEO landscape through feasibility assessment to final signature counts across all six species.

Table 1. Data funneling and signature construction across six species.

Arabidopsis dominates the resource (84% of all signatures), reflecting both the depth of publicly deposited data and the maturity of its microarray annotations. Rice contributes 11% (265 signatures) with sufficient depth across drought, salt, blast infection, and hormone treatment conditions for meaningful within-species analysis. The four additional species (20–46 signatures each) provide initial coverage and demonstrate that the pipeline generalizes across diverse genome architectures and annotation qualities.

Each signature represents a single contrast – a treatment group compared against a matched control – processed through a limma-based differential expression pipeline to produce a genome-wide vector of log fold-change values. These contrast-level expression profiles form the raw material for the classification and connectivity analyses that follow.

Classification of signatures into interventions and stress states

Each of the 2,352 contrast-level signatures was classified into one of two functional roles for the connectivity engine. **Intervention signatures** represent experimental manipulations – chemical treatments, genetic perturbations, pathogen inoculations, or environmental challenges applied as treatments – and are stored as genome-wide logFC vectors that capture the direction and magnitude of every gene's response. **Stress-state gene sets** represent biological states of interest – disease, drought, salinity, defense activation – and are derived from the same logFC vectors by selecting the 250 most strongly up-regulated and 250 most strongly down-regulated genes, producing compact gene sets suitable for enrichment-based connectivity scoring. A signature's classification depends on its intended role in the connectivity query: intervention signatures are the probes ("what does this treatment do?"), while stress-state gene sets are the targets ("which treatments mimic or oppose this condition?"). In some cases, a single signature serves both roles – for example, an ABA treatment can function as an intervention (probing its connectivity to all stress states) and, when converted to a gene set, as a stress state (asking which other interventions recapitulate ABA-like transcription).

After applying eligibility filters – excluding developmental and uninformative categories, requiring at least one differentially expressed gene, and enforcing minimum gene counts – 1,863 of 2,352 signatures (79%) qualified for connectivity analysis: 1,196 intervention signatures and 667 stress-state gene sets across all six species.

Intervention inventory

The 1,196 eligible intervention signatures resolve to 163 distinct perturbagens across Arabidopsis, with additional perturbagens contributed by the five crop species. The Arabidopsis perturbagen inventory spans seven classes: 12 commercial agrochemicals (including neonicotinoids, triazole fungicides, and plant growth retardants), 15 commercially available plant hormones (abscisic acid [ABA], salicylic acid [SA], jasmonic acid [JA], ethylene and its precursor 1-aminocyclopropane-1-carboxylic acid [ACC], indole-3-acetic acid [IAA] and other auxins, gibberellic acid [GA], brassinolide [BL] and other brassinosteroids, cytokinin, and strigolactone), 14 research compounds, 6 defense elicitors (the flagellin-derived peptide flg22, chitin, oligogalacturonides, elf18, AtPep2, and others), 31 environmental and abiotic stress interventions, 52 genetic perturbations, and 32 microbial inoculants. The deepest coverage lies in plant hormones (119 signatures across 15 hormone types) and defense elicitors (20 signatures across 6 PAMPs and DAMPs) – the two intervention classes most directly relevant to biostimulant and crop protection applications.

Stress-state inventory

The 667 eligible stress-state gene sets capture the transcriptomic landscapes of conditions that researchers or breeders would want to mimic, oppose, or understand. In Arabidopsis, the 533 stress-state gene sets span pathogen infection (*Pseudomonas syringae* strains DC3000, AvrRpt2, AvrRps4, and others; *Botrytis cinerea*; *Hyaloperonospora arabidopsidis*), abiotic stresses (drought, salinity, heat, cold, high light, submergence, oxidative stress via methyl viologen and H_2O_2), hormone-defined states (ABA, SA, JA, auxin, ethylene, brassinosteroid, and cytokinin treatment conditions used as reference states), defense elicitor responses (flg22, chitin, elf18), and combinatorial stresses (salt+heat, cold+high light). In rice, the 100 stress-state gene sets cover drought across multiple genotypes and severity levels, salinity, *Magnaporthe oryzae* blast infection, wounding, and heat and cold stress – enabling both within-species

connectivity analysis and cross-species transfer experiments. The four additional crop species contribute 34 stress-state gene sets collectively, providing initial coverage of pathogen, abiotic stress, and hormone-response conditions in tomato, maize, soybean, and grape.

An additional 124 Arabidopsis intervention signatures with clear stress or treatment biology were made available as stress-state gene sets (intervention-as-stress-state), bringing the total queryable stress-state pool to 657 for Arabidopsis and 791 across all species. This dual classification reflects the biological reality that many experimental manipulations – an ABA spray, a NaCl treatment, a pathogen inoculation – are simultaneously an intervention (something applied) and a state (a condition one might want to mimic or reverse).

Arabidopsis connectivity mapping reveals 268,901 pairwise intervention–stress-state relationships

From the 1,060 eligible Arabidopsis intervention signatures and 533 stress-state gene sets, we selected a core matrix of 607 interventions from compendia and the dominant ATH1 microarray platform against 443 stress-state gene sets from the same high-confidence sources, yielding 268,901 pairwise connectivity scores with 1,000-permutation significance testing and global Benjamini–Hochberg FDR correction. Of these, 177,628 (66.1%) reached significance at global FDR < 0.05: 102,042 positive (mimicry) connections and 75,586 negative (reversal) connections, a 58:42 positive-to-negative ratio, consistent with the broad transcriptomic overlap among plant defense and stress responses.

Individual interventions varied widely in their connectivity breadth. The median intervention produced 309 significant connections (IQR 256–342 out of 443 stress states), but the distribution spanned nearly the full range: from 13 significant connections for 10 µM ACC (GSE39384) to 386 for a *Pseudomonas ***syringae* infection experiment (GSE18978, 87% of all stress states). The five most broadly connected interventions were all defense-related – pathogen infection, 200 mM NaCl salt stress, ethylene inhibition by $AgNO_3$, and submergence hypoxia – each producing 380–386 significant connections with strongly positive median normalized connectivity scores (NCS; +11.3 to +12.6). The five least connected interventions were low-dose hormone treatments (10 µM ACC, 1 µM GA_3, 1 µM IAA, 1 µM zeatin; all from GSE39384) and a developmental mutant (*nph4-1*/*arf7*), producing 13–111 significant connections. This 30-fold range in connectivity breadth – from near-universal transcriptomic impact to highly selective effects – is consistent with dose-dependent and pathway-specific modes of action.

On the stress-state side, each gene set was queried by all 607 interventions. The median stress state attracted 409 significant connections (IQR 341–475 out of 607), confirming that most stress-state signatures are transcriptomically accessible to a broad range of interventions. The most responsive stress states were defense-related – flg22-treated mutant panels from the GSE78735 compendium (Hillmer et al., 2017; up to 529 significant interventions) and bacterial infection conditions from GSE196892 (Hillmer et al., 2023) – while the least responsive were developmental or mild stress conditions (143–157 significant interventions).

SA treatments dominated the connectivity landscape: the four SA interventions each produced 293–371 significant connections (66–84% of all stress states tested), predominantly positive, consistent with SA's role as a broad-spectrum systemic defense signal. Among genetic perturbations, the *pad4-1* mutant (SA signaling-deficient) showed overwhelmingly negative connectivity (243 negative versus 18 positive), reflecting broad loss of defense gene expression. The *dde2-2*/*ein2-1* double mutant (JA + ethylene

knockout) produced 361 significant connections with strongly positive median NCS (+7.4), consistent with de-repression of SA-dependent defense when antagonistic hormone pathways are removed.

Systematic biological validation across 77 a priori predictions

Connectivity mapping validations typically rely on small numbers of known connections, leaving the biological fidelity of the engine largely uncharacterized. The Arabidopsis connectivity matrix – 607 intervention signatures against 443 stress-state gene sets yielding 268,901 precomputed connectivity scores – provides an unprecedented substrate for systematic validation. We leveraged this diversity to design 81 a priori validation targets spanning 8 mechanistically distinct categories, organized in a hierarchy from technical controls through primary predictions to novel mechanistic and epistatic tests (Figure 2A). Each target specifies an intervention class, a stress-state class, and the expected direction of connectivity (positive or negative) based on published biology. This taxonomy itself constitutes a reusable template for validating connectivity mapping engines in any organism.

Of the 81 targets, 77 were testable (4 were skipped because no GA-as-stress-state gene sets exist in the current workspace). Of these, 56 (73%; 95% CI 61–82%) confirmed the a priori prediction (Figure 2B). Two categories achieved perfect confirmation: self-consistency controls (10/10) and defense elicitor mimicry (6/6). The 21 unexpected results cluster in three interpretable groups: hormone antagonism (5 targets), auxin pharmacology (5 targets), and specific genetic/stress targets (11 targets). Crucially, the pattern of failures is itself informative about what whole-transcriptome connectivity captures versus what it misses – detailed below.

Mechanism recovery: mimicry, self-consistency, and genetic reversal

All 10 self-consistency controls confirmed that independent measurements of the same biological intervention are robustly connected (Figure S2, Self-Consistency Controls facet). These include cross-laboratory replications (GSE115420 vs. GSE128721 1-naphthaleneacetic acid [NAA] treatments, NCS = +16.2) and cross-compound overlap (ABA+JA combination treatment → JA-as-stress-state, NCS = +25.2). Hormone–stress mimicry targets confirmed 15 of 18 predictions: ABA interventions produced strong positive connectivity with drought conditions (median NCS = +12.4), SA interventions matched pathogen defense conditions (flg22: NCS = +11.2; DC3000: NCS = +7.4; AvrRpt2: NCS = +3.9), and JA interventions matched wounding (NCS = +1.6 at 1 h, +0.5 at 6 h). Defense elicitor mimicry achieved 6/6 confirmation: flg22, chitin, elf18, AtPep2, and oligogalacturonides all produced transcriptomes resembling bacterial or fungal infection, with median NCS ranging from +6.9 to +9.8 (Figure 3C). Stress agent mimicry confirmed NaCl → multi-accession salt conditions (NCS = +1.8) and H_2O_2 → methyl viologen oxidative stress (NCS = +0.05).

Genetic reversal confirmed 10 of 15 predictions. Core hormone-deficient mutants robustly opposed their cognate hormone conditions: *sid2-2* (SA biosynthesis; NCS = −7.0), *dde2-2* (JA biosynthesis; NCS = −7.5), and *ein2-1* (ethylene signaling; NCS = −6.1). Auxin-signaling mutants *nph4-1*/*arf7* (NCS = −7.6) and *lbd16-1* (NCS = −6.9) similarly confirmed directional predictions. Double mutants combining defense pathway knockouts (*dde2-2*/*sid2-2*, *ein2-1*/*sid2-2*, *pad4-1*/*sid2-2*) all confirmed negative connectivity with SA conditions, consistent with additive pathway disruption. The 5 unexpected results have specific explanations: *fls2* and *jeps* genotype interventions showed positive connectivity with flg22 conditions because the stress-state gene set captures flg22 responses measured in mutant backgrounds within the same experiment (GSE196892), creating shared-background correlation rather than signaling reversal.

arf7/*arf19*, *gnom184*, and *rps2*/*rpm1* showed near-zero connectivity scores, suggesting that pleiotropic developmental effects mask the expected hormone or defense reversal signal.

Genetic epistasis ordering

The GSE196892 combinatorial mutant series provided a unique test of genetic epistasis: does progressive knockout of 1 → 2 → 3 → 4 defense pathways (JA, ET, PAD4/EDS1, SA) produce progressively stronger transcriptomic reversal of pathogen defense? Median NCS decreased progressively with knockout depth: +0.01 (single knockouts) to −7.9 (doubles) to −11.1 (triples), with Spearman $\rho = -0.527$, $p = 0.044$ (Figure 3A). This demonstrates that the engine can infer genetic pathway hierarchies from transcriptomes alone – a capability not previously shown for plant intervention–response mapping. Individual single-mutant values revealed that the SA/EDS1 axis drives the strongest reversals (*sid2-2*: NCS = −5.9; *pad4-1*: NCS = −3.0), while JA-only and ET-only single knockouts show near-zero or positive connectivity, consistent with SA being the dominant transcriptomic mediator of AvrRpt2 defense. A notable anomaly is that the *dde2-2*/*ein2-1* double (JA + ET knockout) showed strongly positive connectivity (NCS = +11.9), suggesting that removing JA and ET signaling together disinhibits SA/EDS1-dependent defense – consistent with the known JA/ET suppression of SA biosynthesis (Pieterse et al., 2009; Thaler et al., 2012). The quadruple *jeps* mutant showed partial rebound (NCS = −4.1 vs. triple median of −11.1), suggesting that completely eliminating all four defense branches produces a qualitatively different transcriptomic state rather than simply extending the reversal trend.

Informative failures: auxin pharmacology and hormone antagonism

Pharmacological mechanism recovery revealed a clean dichotomy (Figure 3B). Ethylene and brassinosteroid inhibitors showed classical pharmacological reversal: $AgNO_3$ and AVG opposed ACC-as-stress-state (NCS = −9.4 and −10.4, respectively), and brassinazole opposed brassinolide (BL)-as-stress-state (NCS = −9.9). In contrast, all five auxin-pathway inhibitors – NPA, TIBA, PCIB, PPBo, and kynurenine – showed positive connectivity with auxin conditions rather than the expected reversal. Inhibiting auxin biosynthesis (PPBo: NCS = +12.1; kynurenine: NCS = +12.7) or signaling (PCIB: NCS = +5.9) activated transcriptional programs resembling exogenous auxin treatment. This auxin paradox likely reflects homeostatic feedback: when auxin levels or signaling are suppressed, compensatory upregulation of auxin-responsive genes creates a transcriptomic signature that overlaps with exogenous auxin. This is consistent with known auxin autoregulation (Leyser, 2018) and demonstrates a boundary condition for intervention–response mapping – pharmacological inhibition does not always produce the transcriptomic "opposite" of the target pathway.

All 5 hormone antagonism targets (SA ↔ JA, ABA → JA, ABA → SA, GA → ABA) showed positive, not negative, connectivity. This systematic pattern reflects a fundamental property of whole-transcriptome connectivity: at the global level, defense hormones (SA, JA, ABA) share more transcriptomic real estate – WRKYs, PR-gene families, oxidative burst, cell-wall remodeling – than they oppose. The textbook SA–JA antagonism operates through a smaller number of pathway-specific genes (PDF1.2/VSP vs. PR-1; Glazebrook, 2005), which are quantitatively overwhelmed by the shared defense program in a whole-transcriptome metric. This insight delineates the resolution boundary of transcriptome-scale connectivity: it captures dominant biological programs robustly but does not resolve pathway-specific antagonisms within those programs – an important interpretive principle for downstream applications.

In summary, the 73% confirmation rate across 77 diverse, a priori predictions – spanning eight mechanistic categories from technical controls to novel genetic epistasis – establishes PRISM as a biologically valid tool for inferring transcriptomic relationships in Arabidopsis. The systematic failures in hormone antagonism and auxin pharmacology are themselves informative, delineating the resolution limits of whole-transcriptome connectivity. The novel epistasis ordering analysis demonstrates that the engine can infer quantitative genetic pathway hierarchies. Together, these results provide both a validation of the resource and a framework for interpreting its outputs.

Combination prediction demonstrates transcriptomic additivity across 42 predictions

A central promise of intervention–response mapping for agrochemical discovery is the ability to predict combination effects from individual component signatures – enabling rational mixture design without exhaustive empirical screening. We tested this capability using 42 a priori predictions across five test classes, drawing on four independent combination experiments: a NAA/dexamethasone [DEX]/cycloheximide [CHX] 2^3 factorial design (GSE115420, Goh et al., 2019; GSE128721, Trinh et al., 2019), ABA/JA/SA hormone mixtures (GSE28600), auxin agonist–antagonist rescue pairs (GSE58028), 20 multi-stress combination conditions across multiple Arabidopsis accessions (GSE41935; Rasmussen et al., 2013), and a sequential heat-then-anoxia treatment (GSE16222; Banti et al., 2010). Of the 42 predictions, 39 confirmed (93%; 95% CI 80.5–98.5%), demonstrating that the PRISM engine captures the principle of transcriptomic additivity with high fidelity.

Profile additivity: combination transcriptomes are predictable from components

The core test asks whether the full connectivity profile of a combination intervention – its connectivity score against all 443 stress states – correlates with the element-wise average of its component profiles. All 12 combination interventions passed ($r > 0.3$, $p < 0.05$), with Pearson correlations ranging from $r = 0.67$ to $r = 0.92$. The strongest additivity appeared in NAA+CHX ($r = 0.92$, GSE115420) and ABA+JA ($r = 0.87$, GSE28600), while the weakest – the auxin biosynthesis inhibitor 4-phenoxyphenylboronic acid [PPBo] paired with IAA ($r = 0.69$) – still represents a strong linear relationship across 443 stress states.

Two independent experiments measuring identical NAA/DEX/CHX combinations (GSE115420 and GSE128721) gave closely matched additivity correlations, confirming reproducibility:

Notably, the triple combination (NAA+DEX+CHX) was equally well predicted by the mean of three component profiles as the doubles were by two, suggesting that the additivity principle extends to higher-order mixtures.

Component recovery: the engine identifies what is in a mixture

When a combination intervention is used to query the database, do its individual components rank among the top-ranked intervention-as-stress-state matches? Of 7 testable predictions, 6 passed a top-10% threshold. The ABA+JA combination ranked ABA-as-stress-state 4th/43 (9.3%) and JA-as-stress-state 3rd/43 (7.0%). The auxin receptor antagonist α-(phenylethyl-2-one)-IAA [PEO-IAA] paired with IAA ranked IAA-as-stress-state 2nd/43 (4.7%). The single failure – NAA+CHX → NAA-as-stress-state at rank 7/43 (16.3%) – fell just outside the threshold but remained in the top quintile.

Direct connectivity: combinations retain each component's mechanistic signature

Every combination intervention showed positive connectivity with each of its component stress states (8/8 confirmed). The strongest signal was NAA+DEX → NAA-as-stress-state (median NCS = +21.5), confirming that the NAA component dominates the combination's transcriptomic identity. Hormone combinations confirmed the same pattern: ABA+JA → JA-as-stress-state (NCS = +19.2) and ABA+SA → SA-as-stress-state (NCS = +7.7). Both auxin rescue combinations (PEO-IAA+IAA and PPBo+IAA) showed strong positive connectivity with IAA-as-stress-state (median NCS = +12.7 and +12.9, respectively), confirming that exogenous auxin dominates the combination signal even when paired with biosynthesis inhibitors.

Stress combination additivity across accessions

The additivity principle also holds on the stress-state side: does a dual-stress condition's connectivity profile – the pattern of connectivity scores it receives from all 607 interventions – correlate with the average of its component stress profiles? Of 13 predictions, 11 confirmed ($r > 0.3$, $p < 0.05$). The sequential heat-then-anoxia treatment (GSE16222) showed the strongest additivity ($r = 0.92$), likely because the temporal separation of stresses preserves transcriptomic independence. Among the GSE41935 multi-stress conditions, salt+heat additivity varied across accessions ($r = 0.40$–0.75), with Kas-1 showing the strongest additive pattern ($r = 0.75$) and Col-0 the weakest ($r = 0.08$). These accessions represent genetically distinct Arabidopsis ecotypes collected from geographically diverse populations, and the variation in additivity across them suggests that stress interaction architecture is itself a genotype-dependent trait. Cold+high-light and salt+high-light combinations were consistently predictable ($r = 0.47$–0.69).

The two failures – salt+heat in Col-0 ($r = 0.08$) and heat+high-light in Col-0 ($r = 0.27$) – suggest that the reference accession exhibits nonlinear stress interactions under these specific combinations, possibly through synergistic reactive oxygen species production or unique regulatory wiring. The finding that 6 of 7 accessions show additive salt+heat transcriptomes while Col-0 does not is itself a biologically interesting observation, consistent with accession-specific stress interaction architectures.

Auxin agonist–antagonist combinations produce intermediate transcriptomic phenotypes

Does combining an auxin agonist with an auxin biosynthesis inhibitor produce an intermediate transcriptomic phenotype? Both predictions confirmed. The PEO-IAA+IAA rescue combination (median CS = 0.61 against auxin stress states) fell between PEO-IAA alone (CS = 0.08) and IAA alone (CS = 0.62), confirming that the antagonist partially attenuates the agonist's transcriptomic effect. The PPBo+IAA combination (CS = 0.35) similarly fell between PPBo alone (CS = −0.40) and IAA alone (CS = 0.62), with PPBo more effectively attenuating the auxin signal – consistent with its broader mechanism of action targeting multiple YUC-family biosynthesis enzymes.

The 93% confirmation rate across 42 combination predictions – spanning chemical mixtures, hormone combinations, multi-stress conditions, and antagonist rescue experiments – establishes that the PRISM engine captures the principle of transcriptomic additivity. Individual compound signatures in the database can be computationally combined to predict mixture effects without running the combination experiment. For rational agrochemical formulation, this reduces the problem from exhaustive pairwise screening of all N compounds to profiling each compound individually and predicting combinations computationally – prioritizing only the most promising candidates for empirical validation.

Within-species validation in rice confirms framework generalizability

The Arabidopsis validation established that PRISM recovers known biology with 73% fidelity across 77 diverse predictions. To test whether this performance generalizes beyond the species in which the framework was developed, we evaluated the rice database (265 signatures from 141 experiments; Table 1) against 11 a priori predictions modeled on the Arabidopsis validation taxonomy. Six predictions were untestable because the required intervention or stress-state signatures (JA, GA, ethylene) are not yet represented in the rice database, leaving 5 testable predictions. All 5 confirmed the expected relationship.

ABA and salinity interventions produced positive connectivity with drought and drought-recovery stress states across multiple rice genotypes (median NCS = +4.1, 26/36 tests significant at FDR < 0.05), confirming that ABA-mediated osmotic stress signaling is captured across genotypes. The same interventions connected positively with salt stress states (median NCS = +1.3, 21/36 significant), consistent with the shared osmotic component of drought and salinity responses. SA-pathway interventions – including WRKY45 overexpression and β-aminobutyric acid [BABA] priming – produced positive connectivity with *Magnaporthe oryzae* blast infection stress states (median NCS = +1.7, 18/24 significant), confirming that SA-mediated defense transcriptomes are detectable across rice pathosystems. Two self-consistency controls – ABA interventions against ABA-as-stress-state (median NCS = +0.5, 5/9 significant) and SA-pathway interventions against SA-as-stress-state (median NCS = +4.2, 8/12 significant) – confirmed internal engine consistency in rice gene space.

The 100% confirmation rate across 5 rice predictions, while based on a smaller test set than the Arabidopsis validation, demonstrates that the connectivity engine produces biologically meaningful results in an independent species with no shared platform, no shared experimental design, and a 150-million-year evolutionary divergence from Arabidopsis. Combined with the Arabidopsis results, the within-species validation record stands at 61 of 82 testable predictions confirmed (74%) across two species.

Cross-species connectivity transfer: Arabidopsis intervention signatures predict rice stress biology

The within-species validations established that PRISM recovers known biology within Arabidopsis (73% of 77 predictions) and within rice (100% of 5 predictions). We next asked whether intervention–response relationships learned in one species can predict biology in another – the strongest test of whether the framework captures conserved biological programs rather than species-specific transcriptomic artifacts. We projected validated Arabidopsis intervention signatures into rice gene space via ortholog mapping (see Methods), replacing Arabidopsis gene identifiers with their rice orthologs to produce projected logFC vectors, and ran these projected signatures through the connectivity engine against all 134 native rice stress-state gene sets (Figure 5A).

Five of five cross-species predictions confirm mechanism transfer

We defined six cross-species predictions based on conserved hormone–stress biology, each specifying an Arabidopsis intervention class, a matching rice stress-state class, and the expected direction of connectivity (Table 3). Five were testable; one (JA → rice wounding) was excluded because no wounding stress-state signatures exist in the rice database. All five confirmed the a priori prediction (Table 3, Figure 5B).

Pathogen defense produced the strongest cross-species signal. Arabidopsis SA interventions (4 signatures) projected to rice connected positively with blast infection stress states across avirulent and virulent

Magnaporthe oryzae strains in both leaf and root tissue (median NCS = +4.8, 18/24 tests significant at FDR < 0.05). Arabidopsis flg22 interventions (2 signatures) produced comparably strong connectivity (median NCS = +5.3, 11/12 significant). These results demonstrate that PAMP-triggered immunity and SA-mediated defense networks are sufficiently conserved across the ~150-million-year monocot–eudicot divergence (Chaw et al., 2004; Magallón et al., 2015) to enable cross-species prediction from projected intervention signatures alone.

Abiotic stress signals were weaker but consistently positive. Arabidopsis ABA interventions (3 signatures) connected positively with rice drought stress states (median NCS = +0.7, 45/81 significant) and rice salt stress states (median NCS = +1.1, 20/33 significant). Arabidopsis NaCl interventions (2 signatures) connected positively with rice salt stress states (median NCS = +0.8, 12/22 significant). The attenuation relative to pathogen defense is consistent with greater evolutionary divergence in downstream abiotic stress effectors, a pattern explored further in the leading edge analysis below.

Negative controls confirm specificity

We projected two Arabidopsis intervention classes with no expected conserved plant function – dexamethasone (a synthetic glucocorticoid) and cycloheximide (a translation inhibitor) – against all rice stress states. Both produced near-zero median connectivity scores (DEX: NCS = +0.3; CHX: NCS = +0.3), confirming that the cross-species signal is specific to biologically conserved relationships rather than an artifact of the projection procedure (Figure 5B). As an additional control, we held the projected SA intervention logFC vector fixed (preserving the same rice genes and the same logFC values) and shuffled gene labels 1,000 times, recomputing connectivity against rice blast stress states for each permutation. The real SA-to-blast median connectivity score (CS = 0.353) greatly exceeded the null distribution (mean = −0.006 ± 0.126; empirical p = 0.001), confirming that the cross-species signal depends on the specific assignment of expression changes to orthologous rice genes rather than aggregate statistical properties of the projected vector (Figure S1D).

Gene-level conservation reveals a stress-type-dependent evolutionary gradient

The cross-species connectivity results establish that pathway-level relationships transfer across species. But do Arabidopsis and rice achieve this connectivity through the same genes, or through different gene networks that converge on the same biological outcome? To distinguish these possibilities, we extracted leading edge genes from significant within-species Arabidopsis connectivity results, mapped them to rice orthologs, and tested their enrichment in the corresponding rice stress-state gene sets using Fisher's exact test. Of 42 tests across four stress types, 20 (48%) showed significant enrichment (FDR < 0.05) – but this aggregate rate masks a striking gradient (Figure 5C).

Oxidative stress and pathogen defense showed deep gene-level conservation. Both H_2O_2 tests were significant (2/2, 100%), with the highest median odds ratio (OR = 4.3), consistent with ROS detoxification enzymes (superoxide dismutase, catalase, ascorbate peroxidase) representing ancient eukaryotic machinery with conserved transcriptional regulation (Mittler, 2002). All six SA–pathogen defense tests were significant (6/6, 100%, median OR = 3.1), indicating that core defense signaling components – NLR receptors, WRKY transcription factors, PR proteins – are under strong conservation pressure from the ongoing plant–pathogen arms race (Jones and Dangl, 2006).

Salt and drought stress showed progressively weaker gene-level conservation. Six of 11 NaCl–salt tests were significant (55%, median OR = 1.6), with significance concentrated in a panel of 8 diverse rice

genotypes (GSE79043), suggesting a conserved core SOS pathway superimposed on genotype-dependent variation. Only 6 of 23 ABA–drought tests were significant (26%, median OR = 1.2). Within this class, a clear dichotomy emerged: all four accessions in an acute drought experiment (GSE74793) showed significant enrichment (OR = 2.0–2.4), while none of 15 conditions from mild progressive drought series reached significance (OR = 0.8–1.4). This pattern suggests that the ABA signaling core (PYR/PYL receptors, PP2C phosphatases, SnRK2 kinases; Cutler et al., 2010) is conserved, but downstream effectors – LEA proteins, dehydrins, osmolyte biosynthesis enzymes – have undergone extensive lineage-specific expansion (Hundertmark and Hincha, 2008; Artur et al., 2019) in grasses versus eudicots, with gene-level divergence that scales with stress severity.

This gradient – biotic defense (100%) > oxidative stress (100%) > salt (55%) > drought (26%) – reveals a decoupling of pathway-level and gene-level conservation. All five cross-species predictions confirmed at the pathway level, yet the genes mediating those connections range from deeply conserved (pathogen defense, OR > 3) to substantially diverged (drought, OR ~1.2). The gradient likely reflects different evolutionary pressures: pathogen defense genes face ongoing selection from host–pathogen coevolution, maintaining gene-level orthology, while abiotic stress effectors are freer to diverge because multiple gene families can fulfill the same physiological role.

This decoupling has practical implications for translational plant biology. Intervention–response mapping can identify conserved intervention–response relationships across species even when the specific mediating genes differ – enabling the transfer of validated Arabidopsis hormone–stress connections to crop species where the gene-level details of stress response may have diverged but the regulatory logic is preserved.

Discussion

PRISM establishes intervention–response mapping as a valid paradigm for plant biology

The results presented here establish three key findings. First, whole-transcriptome intervention–response mapping recovers known intervention–response relationships in Arabidopsis with 73% fidelity across 77 diverse, a priori predictions – a rate comparable to the validation benchmarks reported for the human Connectivity Map and LINCS platforms (Lamb et al., 2006; Subramanian et al., 2017). Second, the engine predicts combination transcriptomic effects from individual component signatures with 93% accuracy, demonstrating practical utility for rational mixture design. Third, intervention–response relationships learned in one species transfer to another via ortholog projection, with 100% of testable cross-species predictions confirmed – the strongest evidence that the framework captures conserved biological programs rather than species-specific transcriptomic artifacts.

These results are notable not only for their confirmation rates but for the diversity and stringency of the validation framework. Connectivity mapping validations in the human domain typically rely on small numbers of exemplary connections – for example, the original CMap paper (Lamb et al., 2006) validated against a small number of known drug–disease connections, and the LINCS L1000 benchmark evaluated 137 test compounds across 54 mechanism-of-action classes (Subramanian et al., 2017). Our 77-prediction taxonomy, spanning eight mechanistically distinct categories from technical controls through genetic epistasis, represents a substantially more comprehensive assessment of what intervention–response mapping can and cannot capture. The framework itself – specifying intervention classes, stress-state classes, expected directions, and biological rationales – constitutes a reusable template for validating intervention–response mapping engines in any organism.

The pattern of failures delineates resolution boundaries

The 21 unexpected results are arguably more informative than the 56 successes, because they define the resolution limits of whole-transcriptome connectivity in principled terms.

Hormone antagonism. All five SA–JA, ABA–JA, ABA–SA, and GA–ABA antagonism predictions showed positive rather than negative connectivity. Defense hormones share more transcriptomic real estate (WRKYs, PR-gene families, oxidative burst, cell-wall remodeling) than they oppose, and the well-documented pathway-specific antagonisms (PDF1.2/VSP versus PR-1; Glazebrook, 2005; Pieterse et al., 2009) are quantitatively overwhelmed by the shared defense program at whole-transcriptome scale. The expected antagonism does emerge in cross-dataset comparisons, confirming that the biology is real but operates at a resolution finer than bulk connectivity scores. This has direct practical implications: researchers using intervention–response mapping to identify antagonistic interactions should focus on leading-edge gene analysis rather than bulk connectivity scores, or restrict comparisons to pathway-specific gene sets.

Auxin pharmacology. All five auxin inhibitors (NPA, TIBA, PCIB, PPBo, kynurenine) showed positive connectivity with auxin stress states – the opposite of the expected pharmacological reversal seen for ethylene ($AgNO_3$, AVG) and brassinosteroid (brassinazole) inhibitors. This paradox likely reflects homeostatic feedback (Leyser, 2018): suppressing auxin biosynthesis or signaling triggers compensatory upregulation of auxin-responsive genes, creating a transcriptomic signature that overlaps with exogenous auxin treatment. This demonstrates a boundary condition for intervention–response mapping: pharmacological inhibition does not always produce the transcriptomic "opposite" of the target pathway when strong feedback loops are present. Intervention–response mapping users in any domain should be alert to this interpretive pitfall.

Pleiotropic genetic perturbations. The five genetic reversal failures (*arf7/arf19*, *gnom184*, *rps2/rpm1*, *fls2*, *jeps*) share a common feature: the intervention has broad developmental or immune consequences that obscure the expected pathway-specific reversal. For intervention–response mapping applications, this suggests that single-gene knockouts of highly pleiotropic regulators are less informative as query signatures than those of pathway-specific enzymes or receptors.

Genetic epistasis ordering: a novel capability

The progressive knockout analysis (1 → 2 → 3 → 4 defense pathway knockouts) demonstrates that intervention–response mapping can infer quantitative genetic pathway hierarchies from transcriptomes alone – ordering the SA/EDS1 axis as the dominant contributor to pathogen defense transcription, with JA and ET playing subsidiary roles whose removal paradoxically activates SA-dependent gene expression. Angeles-Albores et al. (2018) demonstrated transcriptome-wide epistasis reconstruction in *C. elegans*, but no previous plant intervention–response mapping approach has demonstrated this capability. The approach could be applied to existing combinatorial mutant series in other pathways (auxin signaling, light response, nutrient sensing) to identify dominant regulatory axes without requiring a priori pathway knowledge.

Combination prediction enables rational mixture design

The 93% confirmation rate across 42 combination predictions has direct implications for agrochemical formulation. The additivity principle – that combination transcriptomic effects are predictable from individual component signatures – means that exhaustive pairwise screening of N compounds can be

replaced by N individual profiling experiments followed by computational prediction. For a database of 163 perturbagens, this represents a reduction from ~13,000 pairwise experiments to 163 individual signatures plus computational combination scoring.

The practical limitations of this approach deserve acknowledgment. The additivity principle held for 93% of predictions, not 100%, and the failures reveal an interesting biological pattern: both occurred in the Columbia-0 (Col-0) reference accession under multi-stress combinations, while 6 of 7 other Arabidopsis accessions showed additive transcriptomic responses to the same stress combinations. This genotype-dependent nonlinearity suggests that stress interaction architecture varies across natural accessions – a finding with implications both for combination prediction and for understanding natural variation in stress tolerance. Furthermore, transcriptomic additivity does not guarantee phenotypic synergy – a combination that produces an additive transcriptomic signature may still show synergistic, antagonistic, or neutral effects at the organismal level. Intervention–response mapping predicts transcriptomic similarity, not efficacy, and any computationally nominated combinations would require experimental validation before commercial development.

Cross-species transfer validates the framework at its deepest level

The cross-species validation provides the strongest evidence that PRISM captures genuine biological programs. Within-species validation, however rigorous, cannot fully exclude the possibility that the engine merely recapitulates statistical regularities in the data rather than real biology. When an Arabidopsis ABA intervention signature – projected through ortholog mapping into rice gene space – connects positively with native rice drought stress states, the signal must reflect conserved biology because no shared data, shared platform, or shared experimental design connects the two species' datasets.

The conservation gradient revealed by the leading edge analysis (biotic defense 100% > oxidative stress 100% > salt 55% > drought 26%) adds biological depth to this finding. It demonstrates that intervention–response mapping captures pathway-level conservation that is robust to gene-network divergence – ABA interventions consistently produce positive connectivity with drought stress states in both species even when the specific mediating genes have partially diverged. This decoupling of pathway-level and gene-level conservation has a clear evolutionary logic: pathogen defense genes face ongoing selection from host–pathogen coevolution, maintaining gene-level orthology, while abiotic stress effectors (dehydrins, LEA proteins, osmolyte biosynthesis enzymes) are freer to diverge because multiple gene families can fulfill the same physiological role.

For translational plant biology, this finding suggests that intervention–response mapping can transfer intervention–response relationships from well-characterized model species to less-studied crop species even when the gene-level details of stress response have diverged – provided the regulatory logic is conserved at the pathway level. This substantially broadens the potential utility of the resource beyond the six species currently profiled.

Limitations

We note several limitations of the current study. First, the signature database is heavily biased toward Arabidopsis, which contributes 84% of all signatures. While the organism-agnostic engine operates identically across species, the depth of coverage – and therefore the richness of discoverable connections – varies substantially. Rice has sufficient coverage (265 signatures) for meaningful intervention–response

analysis, but the four additional species (20–46 signatures each) are primarily proof-of-concept demonstrations that the pipeline generalizes, rather than discovery-ready databases.

Second, the underlying data are drawn exclusively from public GEO deposits, inheriting their limitations: variable experimental quality, inconsistent metadata annotation, and platform heterogeneity. The multi-source acquisition strategy provides built-in quality stratification, with compendia signatures having the highest internal consistency and RNA-seq signatures from individual experiments having the lowest, but quality filtering remains imperfect.

Third, the current connectivity engine uses a GSEA-based scoring framework that weights all genes equally within stress-state gene sets. More sophisticated approaches – such as incorporating prior knowledge of regulatory network topology, using cell-type-specific signatures from single-cell data, or integrating epigenomic and metabolomic information – could potentially improve both sensitivity and resolution. The modular design of the engine accommodates such extensions without requiring changes to the data acquisition or signature construction pipelines.

Fourth, the validation framework, while comprehensive, is restricted to known biology. We defined the 77 predictions because the expected relationships are established in the literature – but the most valuable applications of intervention–response mapping involve discovering *unknown* connections, where no ground truth exists. The high validation rate provides confidence that the engine's outputs are biologically meaningful, but it does not directly quantify the false discovery rate among the thousands of unstudied connections in the database.

Finally, transcriptomic connectivity is not a direct readout of biological function. A positive connectivity score between an intervention and a stress state indicates transcriptomic similarity, not that the intervention protects against the stress. Translating connectivity predictions into agronomic applications requires additional layers of validation – bioassays, dose-response characterization, and field trials – that are beyond the scope of this study but are the logical next steps for any connectivity-nominated intervention–application pair.

Future directions

The PRISM framework is designed for expansion along multiple axes. Deeper species coverage could be achieved by reprocessing raw RNA-seq reads from the Sequence Read Archive, which would access the ~80% of RNA-seq experiments whose count matrices are not deposited in GEO supplementary files. Temporal and dose-response signatures would enable characterization of dynamic intervention responses and potency comparisons. Integration with emerging spatial transcriptomics data could add tissue-resolution to the currently organ-level signatures. The combination prediction capability could be extended to predict higher-order mixtures (three or more components) and to incorporate dose as a variable.

PRISM demonstrates that the connectivity mapping paradigm, which has driven drug repurposing in human biology for two decades, is both feasible and biologically valid for plant intervention–response discovery – opening a new axis for mechanism-informed crop protection.

Methods

Overview

We constructed PRISM by adapting the Connectivity Map framework – originally developed for drug repurposing in human disease – to plant biology across six crop and model species. The approach identifies transcriptomic connections between agrochemical interventions, genetic perturbations, and plant stress states using gene set enrichment analysis (GSEA; Subramanian et al., 2005) and permutation-based significance estimation. The core intervention–response mapping engine operates on organism-agnostic data structures (named log fold-change vectors and gene set lists), enabling direct application across species without code modification. PRISM comprises 2,352 transcriptomic signatures derived from 865 experiments across *Arabidopsis thaliana*, *Oryza sativa* (rice), *Solanum lycopersicum* (tomato), *Zea mays* (maize), *Glycine max* (soybean), and *Vitis vinifera* (grape).

Transcriptomic data survey and species selection

We surveyed publicly available plant transcriptomic data on NCBI Gene Expression Omnibus (GEO) using the GEOmetadb SQLite database (Zhu et al., 2008; downloaded February 2026). We retrieved expression profiling experiments for 14 agriculturally and scientifically relevant plant species: *Arabidopsis thaliana*, *Oryza sativa* (rice), *Zea mays* (maize), *Triticum aestivum* (wheat), *Solanum lycopersicum* (tomato), *Glycine max* (soybean), *Nicotiana tabacum* (tobacco), *Medicago truncatula*, *Hordeum vulgare* (barley), *Sorghum bicolor*, *Brassica napus* (canola), *Solanum tuberosum* (potato), *Vitis vinifera* (grape), and *Gossypium hirsutum* (cotton). We classified each series as *treatment* (exogenous compound or hormone applied), *stress* (environmental stress imposed), or *other* based on keyword matching against title and summary text fields. The survey identified 7,954 expression profiling series totaling 191,825 samples across the 14 species.

We selected six species for PRISM based on three criteria: (a) sufficient data volume in treatment and stress categories to support intervention–response mapping, (b) availability of well-annotated genomes enabling probe-to-gene mapping, and (c) agricultural or scientific relevance. *Arabidopsis thaliana* provided the largest dataset (4,570 series), followed by *Oryza sativa* (1,087 series), *Zea mays* (508 series), *Solanum lycopersicum* (309 series), *Glycine max* (255 series), and *Vitis vinifera* (231 series), totaling 6,960 GEO series across the six species.

Data acquisition strategy

We developed a data acquisition framework to systematically extract signatures from each species, balancing data quality against coverage. Signatures were drawn from three complementary source types:

Compendia comprised large-scale factorial RNA-seq or microarray datasets with multi-factor experimental designs (genotype × treatment × timepoint), enabling extraction of hundreds of distinct contrast signatures per experiment. We developed experiment-specific metadata parsers and contrast planners for each compendium, as their complex factorial designs required manual curation of the design matrix.

Anchor-platform microarray targeted a single dominant microarray platform per species – selected for having the largest number of treatment/stress experiments on GEO – and applied automated regex-based sample classification to identify control and treated groups. This source provided the broadest coverage per species but was limited to experiments with unambiguous metadata.

Non-anchor microarray extended to additional microarray platforms with lower experiment counts. These platforms required species-specific probe-to-gene mapping and typically lacked standardized metadata, necessitating expert curation for control/treated assignment.

RNA-seq mined GEO supplementary count matrices from RNA-seq experiments. We triaged RNA-seq series for the availability of parseable count files and applied the limma–voom pipeline. As with non-anchor microarray sources, most RNA-seq experiments required expert curation for sample role assignment.

For microarray and RNA-seq sources, experiments with clear metadata were processed automatically; those with ambiguous metadata were recovered through expert curation (see Metadata curation below).

Feasibility assessment

For each species, we performed a targeted feasibility assessment to classify GEO experiments by data source. We classified series as *feasible* if they contained treatment or stress experiments with at least 2 samples per condition. We categorized feasible experiments by data source: anchor-platform microarray, non-anchor microarray, and RNA-seq (Table 1).

Table 1. Data funneling and signature attribution across six species.

Dashes indicate no compendia available for that species.

Compendium signature extraction

For *Arabidopsis*, we identified 10 large-scale factorial compendia for deep extraction. Of these, 7 yielded processable count data and produced 547 signatures:

We deferred three compendia due to unavailable processed data (GSE156426) or data stored only as raw FASTQ archives (GSE107988, GSE53641).

For *rice*, we identified 6 compendia and processed 3 that yielded usable count data, producing 24 signatures:

We deferred 3 rice compendia: GSE98455 (Phytozome/JGI gene identifiers without MSU mapping), GSE36040 (diurnal time course lacking treatment contrasts), and GSE180100 (TRAP-seq data in non-standard format). For other species, we identified compendia for maize, grape, and soybean, but most were developmental or genetic-variation studies lacking treatment contrasts suitable for intervention–response mapping.

We processed RNA-seq count matrices using the limma–voom pipeline (Ritchie et al., 2015; Law et al., 2014). We loaded count matrices into DGEList objects (edgeR; Robinson et al., 2010), filtered to retain genes passing the `filterByExpr()` threshold, and normalized using the trimmed mean of M-values (TMM) method (Robinson and Oshlack, 2010). We applied the voom transformation to estimate observation-level weights, then fit linear models via `lmFit()`. For each planned contrast, we extracted treatment-versus-control coefficients via `contrasts.fit()` and computed moderated t-statistics using empirical Bayes shrinkage (`eBayes()`; Smyth, 2004). We retained full topTable results for each contrast, producing named log fold-change (logFC) vectors across all expressed genes. We required a minimum of 2 biological replicates per group. For microarray compendia, we applied the same limma pipeline directly to log2-transformed expression values without the voom transformation.

Anchor-platform signature extraction

For each species, we selected a dominant microarray platform as the anchor:

S. lycopersicum had no experiments deposited on GEO for its designated anchor platform (GPL21525), so all tomato signatures came from non-anchor microarray and RNA-seq sources.

We screened downloaded series for processability based on metadata quality, including series with (a) expression data present in the series matrix, (b) automatically classifiable control and treatment groups via regex-based pattern matching against control-associated terms (e.g., "mock", "untreated", "vehicle", "DMSO", "0 hpi"), and (c) at least 2 biological replicates in both control and treatment groups. For experiments with multiple distinct treatment or stress conditions, we computed separate contrasts for each subgroup against shared controls. We applied the limma pipeline with group-means design matrices, estimating treatment-versus-control contrasts using `contrasts.fit()` with robust empirical Bayes moderation.

Probe-to-gene mapping varied by platform (see Gene identifier standardization below). For multi-probe genes, we retained the probe with the highest interquartile range (IQR) across samples.

Metadata curation

A substantial fraction of feasible experiments (approximately 36% for Arabidopsis) had ambiguous or missing metadata that prevented automated control/treated sample assignment – a common challenge in GEO data mining. For these experiments, we curated sample role assignments by parsing GEO sample metadata into structured key–value pairs, identifying control and treated groups based on experimental context, and assigning contrast directionality. Each curated contrast was independently validated using heuristic checks for keyword consistency, directionality, genotype verification, and cross-contrast agreement. Contrasts that failed validation were corrected (approximately 5% required directionality swaps) or excluded (approximately 5% were fundamentally flawed). Across all six species, this curation process recovered 1,050 signatures that would otherwise have been lost to metadata ambiguity.

Non-anchor microarray expansion

For each species, we identified non-anchor microarray platforms with feasible experiments and attempted probe-to-gene mapping using three strategies: (1) direct gene identifier extraction from GEO platform annotation tables, (2) Entrez Gene ID mapping via Ensembl Plants BioMart (Durinck et al., 2009), and (3) platform-specific accession extraction (e.g., GenBank accessions for Agilent arrays, RAP identifiers for rice platforms).

Probe-to-gene mapping success varied substantially across platforms and species. Modern platforms with genome-aligned probe designs (e.g., Agilent rice arrays, NimbleGen grape arrays) achieved high mapping rates (>80%). Legacy platforms predating current genome assemblies often failed: all 9 non-anchor maize platforms had <1% probe-to-gene coverage due to EST-based probe annotations incompatible with modern Zm gene identifiers, and soybean non-anchor platforms similarly failed with <1% coverage. CATMA arrays for Arabidopsis (5 platforms, 48 experiments) lacked publicly available probe-to-gene mappings.

We processed successfully mapped platforms using the same limma pipeline as the anchor platform. We set a minimum gene threshold of 500 genes per signature for microarray platforms (lowered from the 3,000-

gene threshold used for RNA-seq) to accommodate platforms with sparse probe annotations. Non-anchor microarray processing yielded 468 signatures across four species (Table 1).

RNA-seq expansion

We triaged RNA-seq experiments on GEO by checking for downloadable supplementary count matrices in parseable formats (TSV, CSV, or TXT with gene identifiers as row names). A consistent challenge was that supplementary count files used submitter-defined sample names rather than GEO accession numbers (GSM IDs); we resolved this by positional mapping of column names to GSM IDs using the sample submission order recorded in GEO metadata.

We processed count matrices using the same limma–voom pipeline as compendia. Yield was low: only 20–22% of feasible RNA-seq experiments across species had downloadable count matrices, with the remainder stored only as raw FASTQ archives requiring SRA reprocessing (not attempted in this study). RNA-seq processing yielded 246 signatures across all six species (Table 1).

Gene identifier standardization

We standardized gene identifiers to species-specific canonical formats throughout the pipeline:

For *Arabidopsis*, probe-to-gene mapping used the GPL198 platform annotation table providing direct Affymetrix probe set to AGI locus mappings (21,532 of 22,810 probe sets mapped to nuclear loci).

For *rice*, GPL2025 platform annotation contained RAP identifiers for only 6.7% of probes. We developed a multi-strategy approach: (1) direct RAP extraction from Gene Symbol columns, and (2) GenBank accession mapping to RAP identifiers via Ensembl Plants BioMart, yielding 12,920 mapped probes. We converted all RAP identifiers to MSU locus identifiers using the riceidconverter R package, producing 11,063 unique genes per experiment. For the Agilent rice array GPL6864 (used for the GSE62896 compendium), we extracted RAP identifiers from the Accessions column, mapping 42,105 of 45,151 probes to 24,704 MSU loci.

For *maize*, we mapped GPL4032 probes via Entrez Gene IDs to Ensembl v5 identifiers through BioMart, achieving only 13% fill rate (899 of 17,734 probes) due to sparse annotations on this 2006-era array.

For *grape*, GPL13936 annotations used `Vv##s####g#####` format identifiers, which we normalized to the canonical `VIT_##s####g#####` format.

For *soybean*, we mapped GPL4592 probes via Entrez Gene IDs to Ensembl identifiers through BioMart, with Ensembl's `GLYMA_` prefix normalized to the canonical `Glyma.` format. Mapping achieved 3.2% fill rate (931 of 61,170 probes), requiring Entrez queries chunked at 5,000 IDs to avoid API limits.

Signature classification and database assembly

We classified all 2,352 signatures into two functional categories for the intervention–response mapping engine:

Intervention signatures represent the transcriptomic effect of an experimental manipulation and are stored as full named logFC vectors (gene identifier to log fold-change). These include chemical treatments (hormones, herbicides, fungicides, elicitors), genetic perturbations (mutant versus wild-type), pathogen inoculations, abiotic stress treatments applied as interventions, and microbial community manipulations.

Stress-state gene sets represent biological states of interest and are stored as paired up-regulated and down-regulated gene lists. We derived these from logFC vectors by ranking all genes by fold-change and selecting the 250 most strongly up-regulated and 250 most strongly down-regulated genes. This top-N ranked approach matches established Connectivity Map methodology, adapts to varying effect sizes across experiments, and ensures consistent set sizes for GSEA computation.

We assigned each signature to one of eight biological categories using a priority-based classification cascade: (1) curation labels from the metadata curation process (treatment, stress, genotype, pathogen, timepoint); (2) manually verified stress-state classifications from the pilot Arabidopsis analysis; (3) contrast name keyword matching (e.g., "NaCl", "drought", "heat" → abiotic_stress; "ABA", "SA", "JA" → chemical_treatment; "mutant", "knockout" → genotype); (4) GEO metadata keyword matching against series titles and summaries; and (5) experiment class fallback mapping. The eight biological categories were: chemical_treatment, abiotic_stress, genotype, pathogen, microbiome, combination (multi-factor), developmental, and other.

We determined PRISM eligibility based on three criteria: (a) biological category not in {developmental, other}, (b) at least one differentially expressed gene at FDR < 0.05, and (c) minimum gene count ≥ 500. Of the 2,352 total signatures, 1,863 (79.2%) met eligibility criteria: 1,196 intervention signatures and 667 stress-state gene sets across all six species (Table 2).

Table 2. PRISM-eligible signatures by species.

We exported the final database for each species as an R workspace (`plant_cmap_workspace.RData`) containing four objects: `perturbation_signatures` (a list of named numeric vectors), `condition_up_geneid` and `condition_down_geneid` (lists of character vectors), and `gene_annotation` (a data frame mapping gene identifiers to symbols). We additionally exported stress-state gene sets in GMT format for portability.

Intervention–response mapping engine

We computed connectivity scores between intervention signatures and stress-state gene sets using a GSEA-based intervention–response mapping engine. For a given intervention signature (logFC vector), we ranked genes by fold-change magnitude. We then computed the enrichment of each stress state's up-regulated and down-regulated gene sets in the ranked list via a weighted Kolmogorov–Smirnov-like running-sum statistic. We defined the connectivity score (CS) as the difference between the enrichment score for the up-regulated gene set and the enrichment score for the down-regulated gene set, such that negative CS values indicate transcriptomic reversal (the intervention opposes the stress state) and positive CS values indicate mimicry (the intervention recapitulates the stress state).

We assessed statistical significance by permutation testing (1,000 permutations per test, seed = 42). We applied multiple testing correction using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995) both within each intervention (per-intervention FDR) and across all tests (global FDR).

Arabidopsis connectivity analysis

We computed the Arabidopsis intervention–response matrix using 607 intervention signatures from compendia and the anchor microarray platform against 443 stress-state gene sets from the same sources, totaling 268,901 pairwise tests. We implemented per-intervention result caching to enable resumable execution. We applied Benjamini–Hochberg FDR correction globally across all 268,901 tests.

Validation framework

We designed 81 a priori validation predictions organized into eight mechanistically distinct categories: (1) self-consistency controls, testing whether independent measurements of the same intervention connect positively; (2) hormone–stress mimicry, testing whether hormone interventions connect positively with their cognate stress states; (3) genetic reversal, testing whether loss-of-function mutants connect negatively with their cognate hormone or defense stress states; (4) hormone antagonism, testing whether known antagonistic hormone pairs show negative connectivity; (5) pharmacological mechanism, testing whether known inhibitors connect negatively with their target pathway stress states; (6) defense elicitor mimicry, testing whether PAMP/DAMP elicitors connect positively with pathogen infection stress states; (7) stress agent mimicry, testing whether direct stress treatments connect positively with the corresponding stress states; and (8) genetic epistasis, testing whether progressive defense pathway knockouts produce progressively stronger reversal of pathogen defense stress states.

Each prediction specified an intervention class (one or more intervention signatures), a stress-state class (one or more stress-state gene sets), and the expected direction of connectivity (positive or negative) based on published biology. For predictions involving multiple intervention–stress-state pairs, we computed the median normalized connectivity score (NCS) across all pairwise tests within the class. A prediction was scored as confirmed if the median NCS matched the expected direction and was statistically significant (FDR < 0.05). Four predictions were untestable because no GA-as-stress-state gene sets exist in the current workspace, leaving 77 testable predictions.

Combination prediction

We tested the principle of transcriptomic additivity across 42 a priori predictions organized into five test classes. Profile additivity tested whether the full connectivity profile of a combination intervention (its CS against all 443 stress states) correlated with the element-wise average of its component profiles (Pearson $r > 0.3$, $p < 0.05$). Component recovery tested whether querying the database with a combination intervention retrieved individual components among the top 10% of intervention-as-stress-state matches. Direct connectivity tested whether combination interventions showed positive connectivity with each component's stress state. Stress combination additivity tested whether dual-stress conditions' connectivity profiles correlated with averaged component profiles. Antagonist rescue intermediacy tested whether agonist–antagonist combinations produced connectivity scores intermediate between the individual components.

Cross-species ortholog projection

We projected validated Arabidopsis intervention signatures into rice gene space via ortholog mapping. We queried Ensembl Plants BioMart for Arabidopsis-to-rice orthologs, converted RAP identifiers to MSU locus IDs via riceidconverter, and retained all orthology types (one-to-one, one-to-many, many-to-many). This yielded 28,526 ortholog pairs linking 15,037 unique Arabidopsis genes to 14,658 unique rice genes – a 65.4% mapping rate relative to the 21,244 Arabidopsis genes in the workspace. For each Arabidopsis intervention signature, we replaced Arabidopsis gene identifiers with their rice orthologs (averaging logFC values when multiple Arabidopsis genes mapped to the same rice locus), producing projected logFC vectors in rice gene space. We then ran these projected signatures through the connectivity engine against all 134 native rice stress-state gene sets.

We validated cross-species transfer against six a priori predictions based on conserved hormone–stress biology, with negative controls (dexamethasone, cycloheximide) projected using the same procedure. We

assessed specificity by holding the projected SA intervention logFC vector fixed and shuffling gene labels 1,000 times, recomputing connectivity against rice blast stress states for each permutation and comparing the observed SA-to-blast connectivity score against the permuted null distribution.

Leading edge gene conservation analysis

For each cross-species intervention class, we extracted leading edge genes from significant within-species Arabidopsis connectivity results – genes driving the enrichment signal for each intervention–stress-state pair. We mapped these leading edge genes to rice orthologs using the same Ensembl Plants ortholog table and tested their enrichment in the corresponding rice stress-state gene sets using Fisher's exact test. We applied Benjamini–Hochberg FDR correction across all 42 tests and report odds ratios as the measure of gene-level conservation strength.

Software and computational environment

Analyses were implemented in R version 4.5.1 (R Core Team, 2025). Key R packages: limma (Ritchie et al., 2015), edgeR (Robinson et al., 2010), Biobase (Huber et al., 2015), GEOquery (Davis and Meltzer, 2007), GEOmetadb (Zhu et al., 2008), fgsea (Korotkevich et al., 2021), biomaRt (Durinck et al., 2005, 2009), riceidconverter (Li, 2020), ggplot2 (Wickham, 2016), data.table (Barrett et al., 2025), and pheatmap (Kolde, 2019).

Figures

Figure 1. PRISM resource overview and intervention–response mapping concept.

(A) Multi-species signature inventory. Horizontal stacked bar chart showing the number of intervention (blue) and stress-state (orange) transcriptomic signatures for each of six species. Arabidopsis dominates the resource (1,979 signatures from 660 experiments across compendia, microarray, and RNA-seq sources), followed by rice (265 signatures), tomato (46), maize (24), soybean (20), and grape (18), for a total of 2,352 signatures derived from 865 experiments. Bar segments include intervention-as-stress-state signatures

where applicable. **(B)** Intervention–response mapping concept. An intervention signature – a genome-wide vector of log-fold-change values from a treatment versus control comparison – is scored against stress-state gene sets (the top 250 up-regulated and top 250 down-regulated genes from a stress or disease condition) via a GSEA-like enrichment statistic. The resulting connectivity score (CS) is positive when the intervention transcriptome resembles the stress state (mimicry; e.g., SA treatment mimics pathogen defense) and negative when it opposes the stress state (reversal; e.g., the *sid2* SA-biosynthesis mutant opposes SA-responsive stress states). Example mini-bar plots illustrate matching (mimicry) and opposing (reversal) expression patterns between intervention and stress-state signatures.

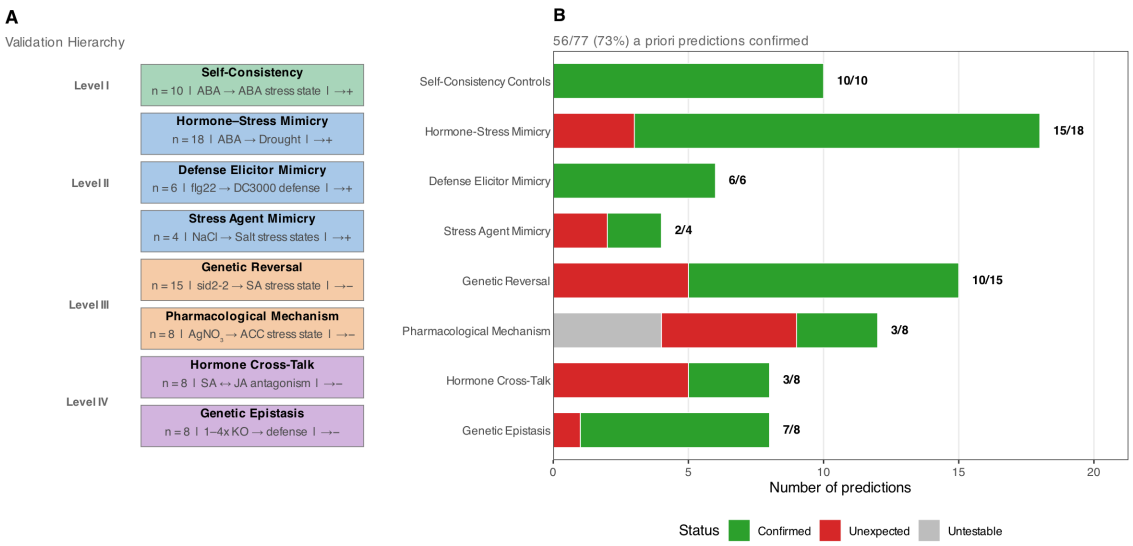

Figure 2. Systematic validation framework and overall results.

(A) Schematic of the eight-category validation taxonomy, organized in four levels of increasing mechanistic stringency. Level I: self-consistency controls (independent measurements of the same intervention). Level II: primary mimicry predictions – hormone–stress mimicry (hormone intervention mimics stress state), defense elicitor mimicry (PAMP/DAMP intervention mimics pathogen stress state), and stress agent mimicry (chemical stress mimics environmental stress). Level III: reversal predictions – genetic reversal (loss-of-function mutant opposes cognate hormone stress state) and pharmacological mechanism recovery (inhibitor opposes agonist stress state). Level IV: emergent predictions – hormone antagonism (antagonistic hormones produce opposing transcriptomes) and genetic epistasis (progressive pathway knockout produces progressive reversal). Each box shows the number of testable predictions, an illustrative example, and the expected direction of connectivity (positive or negative). **(B)** Stacked bar chart showing validation outcomes by category. Of 81 a priori predictions, 77 were testable (4 untestable due to missing GA-as-stress-state signatures). Green bars indicate confirmed predictions, red bars indicate unexpected results (opposite of prediction), and grey bars indicate untestable predictions. The overall confirmation rate was 56/77 (73%; 95% CI 61–82%). Self-consistency controls (10/10) and defense elicitor mimicry (6/6) achieved perfect confirmation, while hormone antagonism (0/5) and auxin pharmacology (0/5) showed systematic departures attributable to shared defense transcriptomes and homeostatic feedback, respectively.

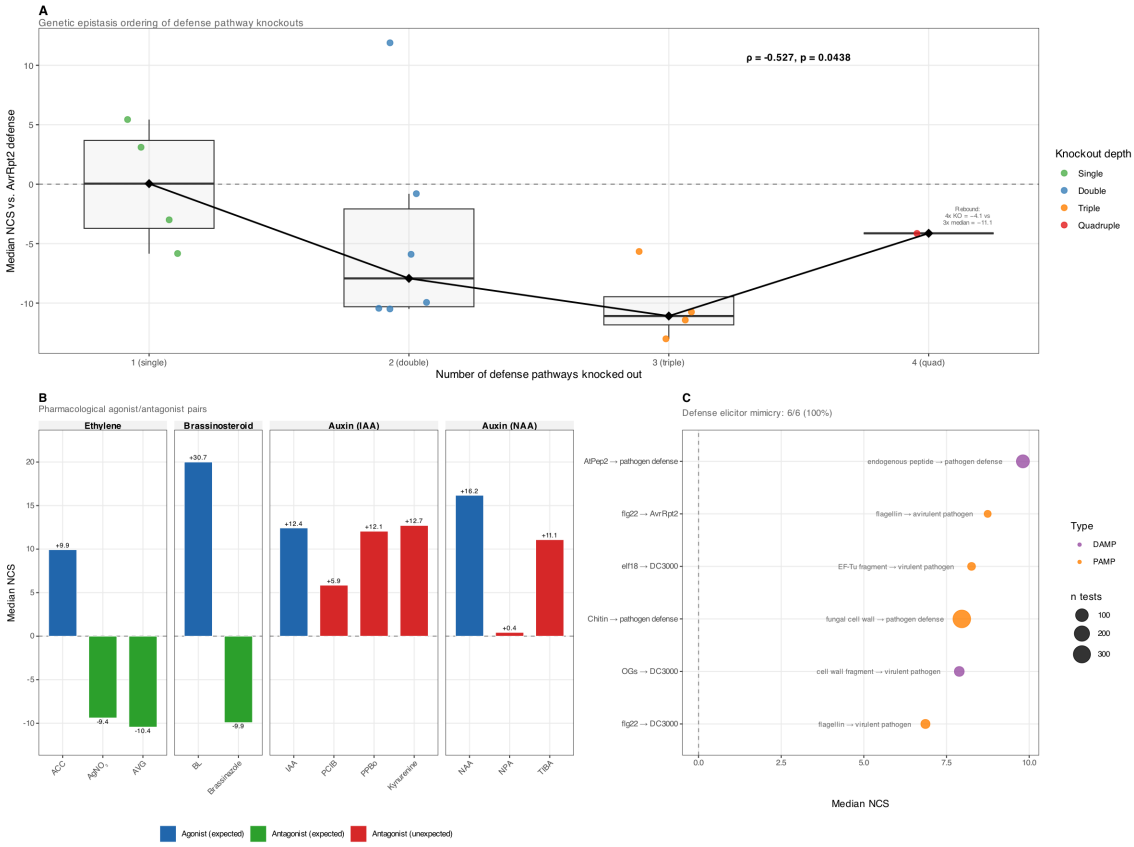

Figure 3. Mechanistic insights from intervention–response mapping.

(A) Genetic epistasis ordering. Progressive knockout of 1 to 4 defense signaling pathways (JA, ethylene, PAD4/EDS1, SA) produces progressively stronger transcriptomic reversal of AvrRpt2-triggered defense, as measured by median NCS against pathogen defense stress states. Points are colored by knockout depth (green = single, blue = double, orange = triple, red = quadruple); black diamonds and connecting line trace the medians at each depth. Spearman correlation: rho = −0.527, p = 0.044. The *dde2-2/ein2-1* double mutant is a notable outlier, showing strongly positive connectivity (NCS = +11.9) – consistent with de-repression of SA-dependent defense when JA and ethylene signaling are jointly removed. The quadruple *jeps* mutant shows partial rebound (NCS = −4.1 vs. triple median of −11.1), suggesting a qualitatively distinct transcriptomic state. **(B)** Pharmacological agonist–antagonist pairs across four hormone systems. Bar colors denote expected agonist effect (dark blue), expected antagonist reversal (green), and unexpected antagonist result (red). Ethylene inhibitors ($AgNO_3$, AVG) and brassinazole show classical reversal of their cognate hormone stress states. All five auxin-pathway inhibitors (NPA, TIBA, PCIB, PPBo, kynurenine) show positive rather than negative connectivity with auxin stress states – the "auxin paradox" – reflecting compensatory transcriptional feedback upon auxin depletion or signaling inhibition. NCS values are annotated above or below each bar. **(C)** Defense elicitor mimicry. Lollipop plot showing connectivity of six elicitor interventions to specific pathogen stress states (6/6 confirmed, 100%). Y-axis labels indicate the elicitor → stress state tested, where DC3000 denotes virulent *Pseudomonas syringae* pv. tomato DC3000,

AvrRpt2 denotes avirulent *P. syringae* expressing the AvrRpt2 effector, and "pathogen defense" denotes composite pathogen defense gene sets. Adjacent dot labels provide plain-language biological descriptions of each elicitor–pathogen pair (e.g., "flagellin → virulent pathogen", "fungal cell wall → pathogen defense"). Point size reflects the number of intervention–stress-state tests; points are colored by elicitor class (orange = PAMP, purple = DAMP). Median NCS values range from +6.9 to +9.8.

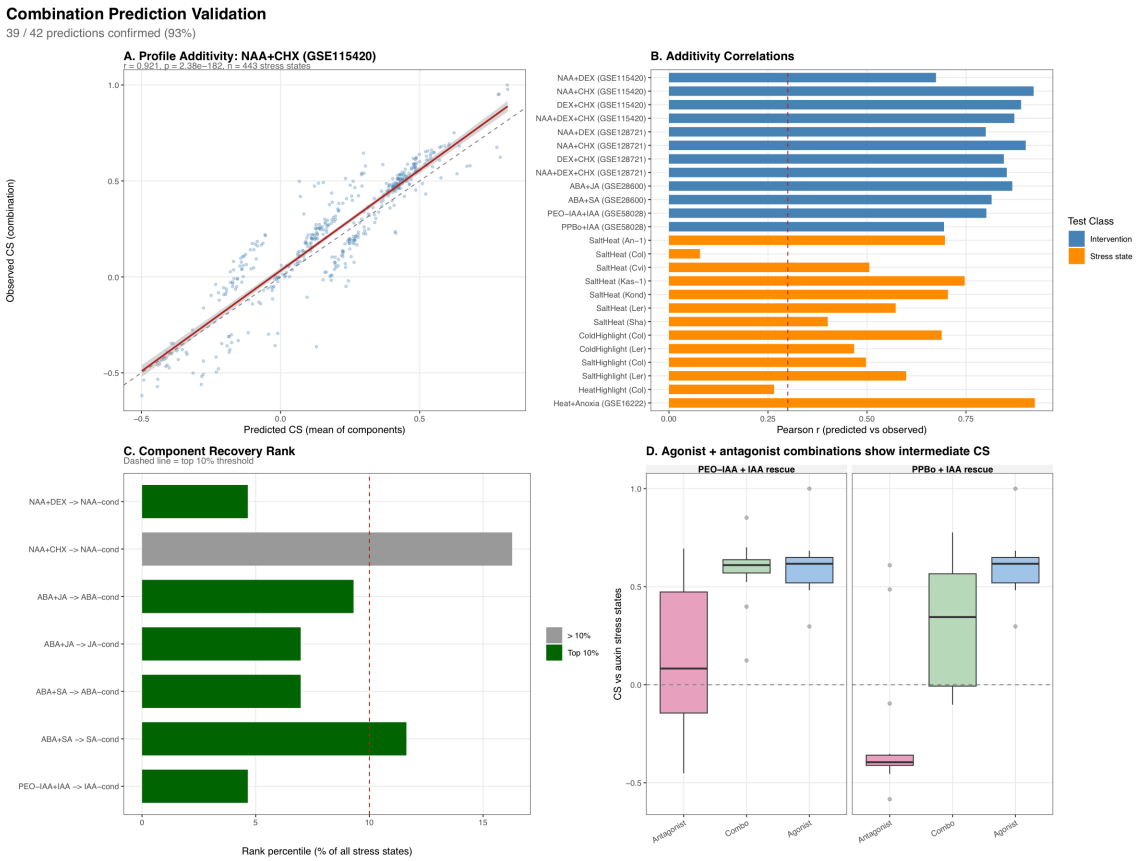

Figure 4. Combination prediction validates transcriptomic additivity.

(A) Profile additivity for the best-performing combination. Scatter plot of observed connectivity scores (combination intervention vs. each of 443 stress states) against predicted scores (element-wise mean of the two component intervention profiles). The dark red line shows the linear regression fit with standard error band; the dashed diagonal represents perfect additivity (slope = 1). All 12 tested combinations passed ($r > 0.3$), with Pearson correlations ranging from $r = 0.67$ (NAA+DEX, GSE115420) to $r = 0.92$ (NAA+CHX, GSE115420). **(B)** Summary of additivity correlations across all intervention additivity (blue) and stress-state additivity (orange) predictions. Each horizontal bar shows the Pearson correlation between predicted and observed connectivity profiles. The dark red dashed line at $r = 0.3$ marks the confirmation threshold. Of 25 testable additivity predictions, 23 confirmed (92%). The two failures – salt+heat and heat+high-light in the Col-0 reference accession – suggest accession-specific nonlinear stress interactions. **(C)** Component recovery. For each combination intervention, the horizontal bar shows the rank percentile of the corresponding component's intervention-as-stress-state entry among all 43 intervention-as-stress-state signatures. Dark green bars indicate recovery within the top 10% (confirmed); grey bars indicate failure. Six of 7 testable predictions confirmed; ABA+JA recovered both ABA (9.3%) and JA (7.0%) as top-ranked components. **(D)** Antagonist rescue intermediacy. Boxplots of connectivity scores against auxin stress states for antagonist alone, agonist–antagonist combination, and agonist alone, shown for two auxin rescue experiments (PEO-IAA+IAA and PPBo+IAA). In both cases, the combination falls between the component

extremes, confirming that the inhibitor partially attenuates the agonist's transcriptomic effect without fully reversing it. Overall, 39 of 42 testable combination predictions confirmed (93%; 95% CI 80.5–98.5%).

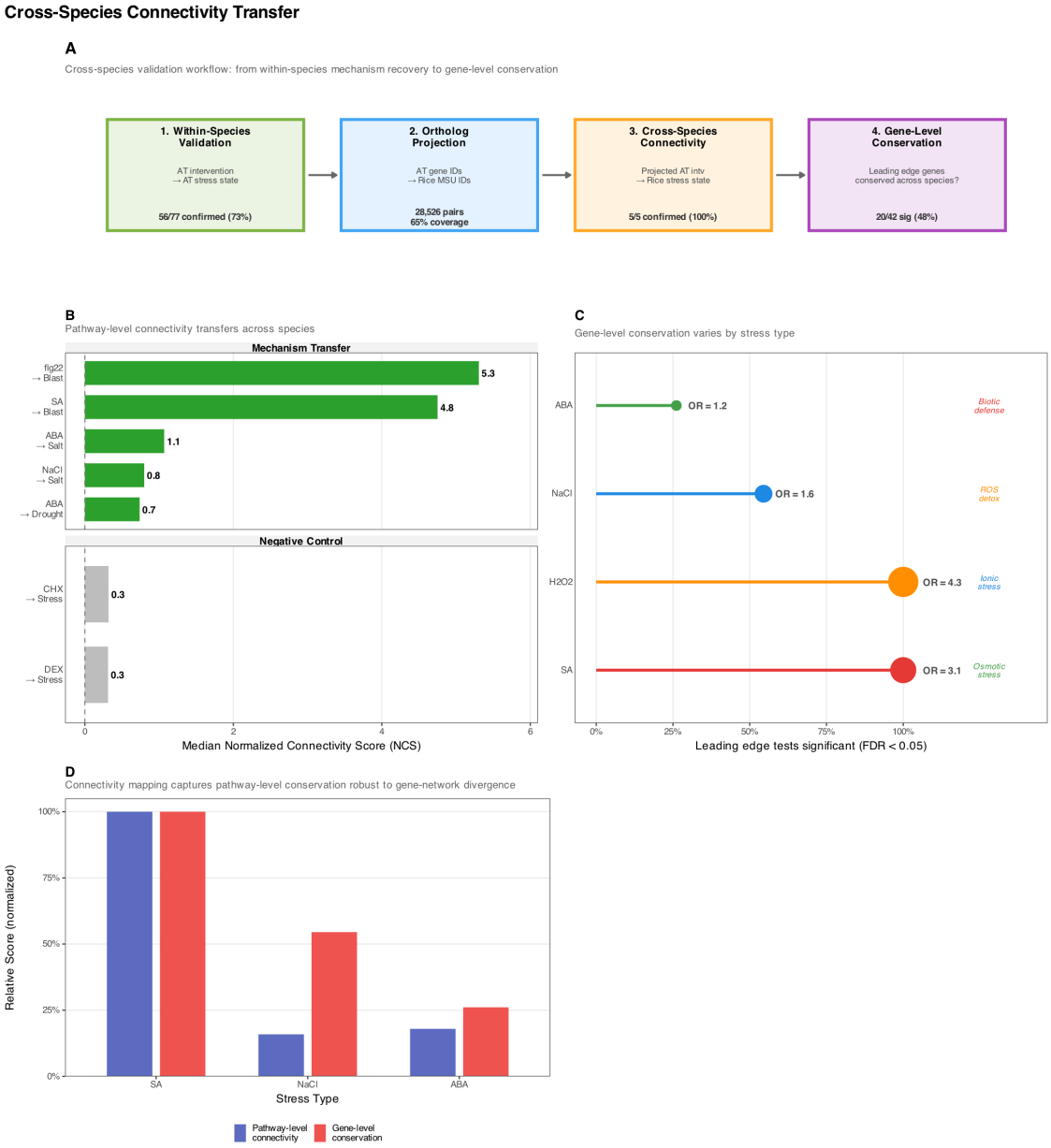

Figure 5. Cross-species connectivity transfer from Arabidopsis to rice.

(A) Workflow schematic. Four sequential steps: (1) within-species validation of the Arabidopsis engine (56/77 confirmed, 73%); (2) ortholog projection of Arabidopsis intervention signatures into rice gene space via Ensembl Plants BioMart (28,526 ortholog pairs, 65% genome coverage); (3) cross-species connectivity scoring of projected Arabidopsis interventions against 134 native rice stress-state gene sets; (4) gene-level conservation analysis of leading edge genes. **(B)** Cross-species NCS forest plot. Top facet (mechanism transfer): all five testable predictions confirmed – SA and flg22 interventions produced strong positive connectivity with rice blast infection stress states (NCS = +4.8 and +5.3, respectively), and ABA and NaCl

interventions connected positively with rice drought and salt stress states (NCS = +0.7 to +1.1). Bottom facet (negative controls): dexamethasone and cycloheximide – interventions with no expected conserved plant function – showed near-zero connectivity (NCS = +0.3 each), confirming specificity. **(C)** Conservation gradient across stress types. Lollipop plot showing the percentage of leading edge gene enrichment tests reaching significance (FDR < 0.05) for each stress type, with point size proportional to the median Fisher's exact test odds ratio. Gene-level conservation follows a gradient: SA (100%, OR = 3.1) > H_2O_2 (100%, OR = 4.3) > NaCl (55%, OR = 1.6) > ABA (26%, OR = 1.2), reflecting stronger conservation of biotic defense gene networks under ongoing host–pathogen coevolutionary pressure relative to abiotic stress effectors that have undergone lineage-specific expansion. Biological annotations at right indicate the corresponding stress category (biotic defense, ROS detoxification, ionic stress, osmotic stress). **(D)** Pathway-level versus gene-level conservation. Grouped bar chart comparing normalized pathway-level connectivity (blue) with gene-level conservation (red) across all four stress types. All stress types show high pathway-level connectivity (mechanism transfer confirmed for all), but gene-level conservation declines from SA/H_2O_2 to ABA, demonstrating that intervention–response mapping captures conserved regulatory logic even when the specific mediating genes have diverged across 150 million years of evolution.

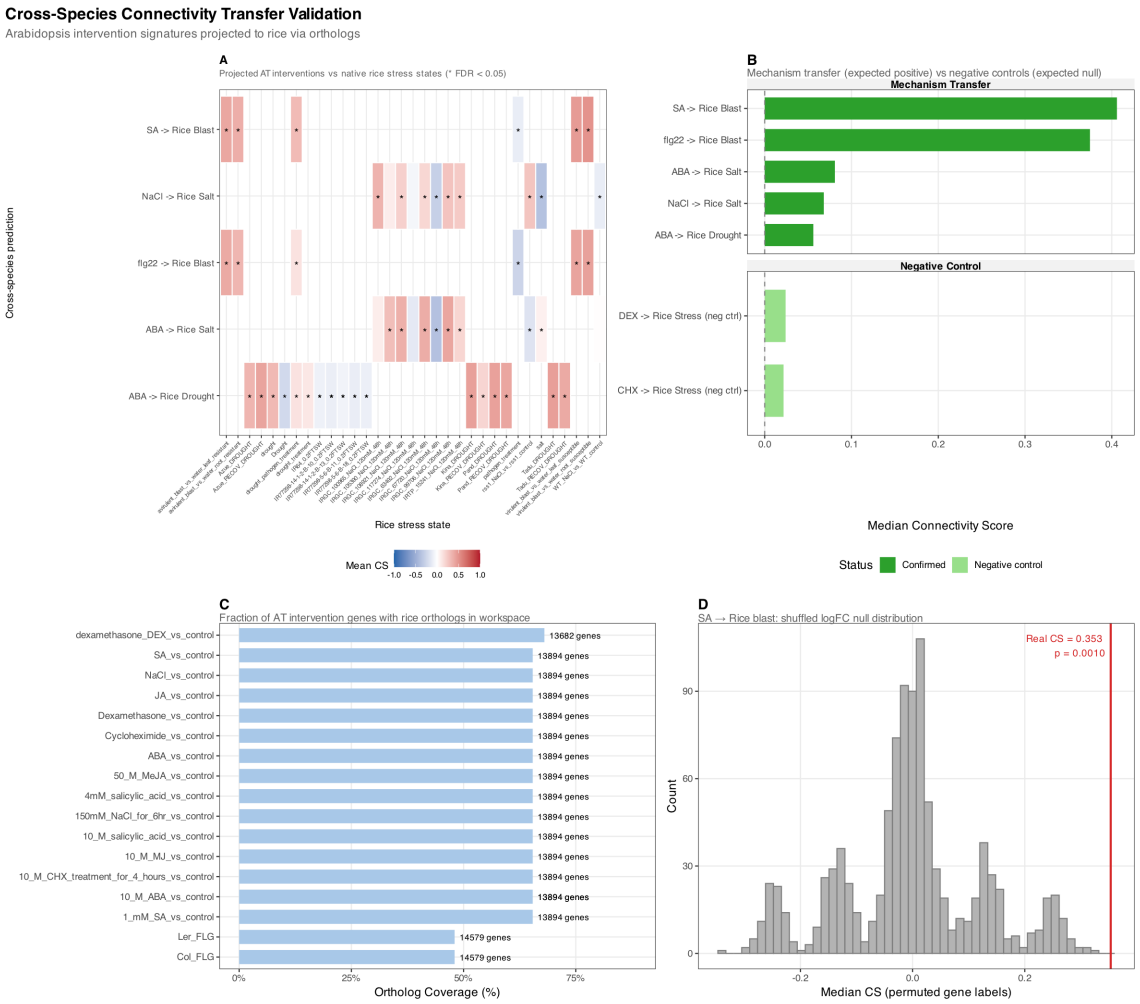

Figure S1. Cross-species validation: detailed connectivity and leading edge enrichment.

(A) Heatmap of connectivity scores for projected Arabidopsis intervention signatures (rows) against native rice stress-state gene sets (columns), organized by intervention class and rice stress-state type. Color scale encodes raw connectivity scores, with hierarchical clustering within each block. **(B)** Bar chart of median connectivity scores per cross-species validation prediction, summarizing the distributions shown in panel A. Positive values confirm mechanism transfer; near-zero values for negative controls (DEX, CHX) confirm specificity. **(C)** Ortholog projection coverage. Summary of the mapping statistics: 28,526 ortholog pairs linking 15,037 Arabidopsis genes to 14,658 rice MSU loci (65.4% of Arabidopsis workspace genes). When multiple Arabidopsis genes map to the same rice locus, logFC values are averaged. **(D)** Permutation null distribution for the SA-to-blast cross-species signal. The projected SA intervention logFC vector was held fixed (same rice genes, same values) and gene labels were shuffled 1,000 times to generate a null distribution of median connectivity scores against rice blast stress states. The observed value (CS = 0.353) is marked by a vertical line. The observed score greatly exceeds the null distribution (null mean = −0.006 ± 0.126; empirical p = 0.001), confirming that the cross-species signal depends on the specific assignment

of expression changes to orthologous rice genes rather than aggregate statistical properties of the projected vector.

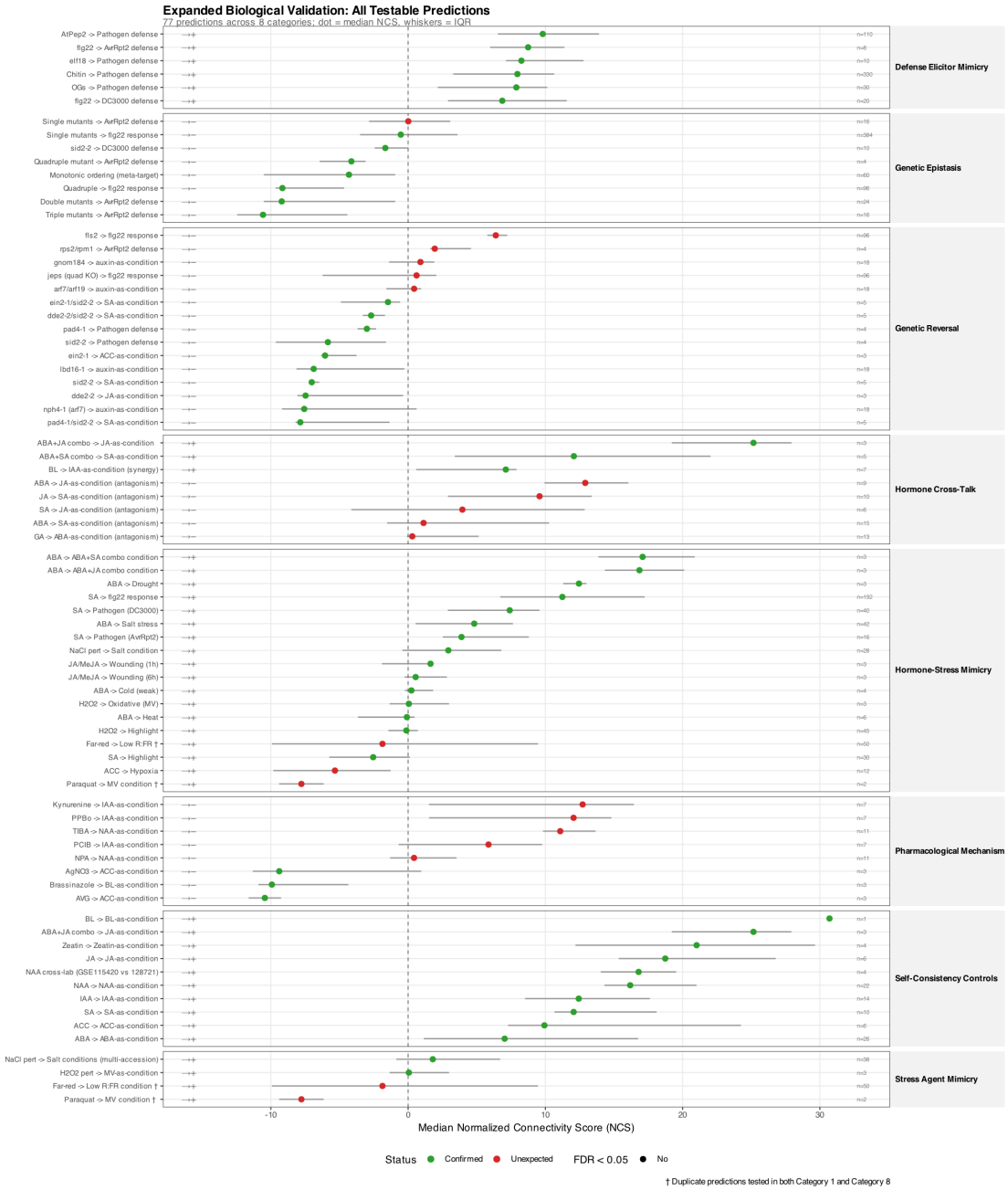

Figure S2. Comprehensive validation forest plot.

Forest plot of normalized connectivity scores (NCS) for all 77 testable validation predictions, organized by the eight mechanistic categories defined in Figure 2. Each point represents the median NCS across all intervention–stress-state pairs within that prediction; horizontal whiskers span the interquartile range. Green filled circles indicate predictions that confirmed the expected relationship at FDR < 0.05; red filled circles indicate predictions with an unexpected direction; open circles indicate non-significant results. Arrows at left denote the predicted direction of connectivity (positive or negative). Sample sizes (number

of intervention–stress-state tests per prediction) are annotated at right. Self-consistency controls cluster at strongly positive NCS values (top), confirming reproducibility across independent experiments and laboratories. Genetic epistasis predictions (bottom) show the progressive reversal trend from single through quadruple pathway knockouts. Dagger symbols (†) mark predictions that appear in both Category 1 (self-consistency) and Category 8 (epistasis).

References

Angeles-Albores D, Puckett Robinson C, Williams BA, Wold BJ, Sternberg PW. Reconstructing a metazoan genetic pathway with transcriptome-wide epistasis measurements. *Proceedings of the National Academy of Sciences*. 2018;115(13):E2930–E2939. DOI: 10.1073/pnas.1712387115

Artur MAS, Zhao T, Ligterink W, Schranz E, Hilhorst HWM. Dissecting the genomic diversification of Late Embryogenesis Abundant (LEA) protein gene families in plants. *Genome Biology and Evolution*. 2019;11(2):459–471. DOI: 10.1093/gbe/evy248

Banti V, Mafessoni F, Loreti E, Alpi A, Perata P. The heat-inducible transcription factor HsfA2 enhances anoxia tolerance in Arabidopsis. *Plant Physiology*. 2010;152(3):1471–1483. DOI: 10.1104/pp.109.149815

Barrett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, Hocking T, Schwendinger B, Krylov I. data.table: Extension of `data.frame`. R package. 2025. https://r-datatable.com

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*. 1995;57(1):289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x

Chaw SM, Chang CC, Chen HL, Li WH. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. *Journal of Molecular Evolution*. 2004;58(4):424–441. DOI: 10.1007/s00239-003-2564-9

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. Abscisic acid: emergence of a core signaling network. *Annual Review of Plant Biology*. 2010;61:651–679. DOI: 10.1146/annurev-arplant-042809-112122

Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846–1847. DOI: 10.1093/bioinformatics/btm254

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. 2005;21(16):3439–3440. DOI: 10.1093/bioinformatics/bti525

Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets

with the R/Bioconductor package biomaRt. *Nature Protocols*. 2009;4(8):1184–1191. DOI: 10.1038/nprot.2009.97

Esfahani AH, Mass J, Hallab A, Schuldt BM, Nevarez D, Usadel B, Ott MC, Buer B, Schuppert A. Plant PhysioSpace: a robust tool to compare stress response across plant species. *Plant Physiology*. 2021;187(3):1795–1811. DOI: 10.1093/plphys/kiab325

Glazebrook J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology*. 2005;43:205–227. DOI: 10.1146/annurev.phyto.43.040204.135923

Goh T, Toyokura K, Yamaguchi N, Okamoto Y, Uehara T, Kaneko S, Takebayashi Y, Kasahara H, Ikeyama Y, Okushima Y, Nakajima K, Mimura T, Tasaka M, Fukaki H. Lateral root initiation requires the sequential induction of transcription factors LBD16 and PUCHI in Arabidopsis thaliana. *New Phytologist*. 2019;224(2):749–760. DOI: 10.1111/nph.16065

Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, Kiba T, Takatsuto S, Fujioka S, Asami T, Nakano T, Kato H, Mizuno T, Sakakibara H, Yamaguchi S, Nambara E, Kamiya Y, Takahashi H, Hirai MY, Sakurai T, Shinozaki K, Saito K, Yoshida S, Shimada Y. The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *The Plant Journal*. 2008;55(3):526–542. DOI: 10.1111/j.0960-7412.2008.03510.x

Hartmann A, Berkowitz O, Whelan J, Narsai R. Cross-species transcriptomic analyses reveals common and opposite responses in Arabidopsis, rice and barley following oxidative stress and hormone treatment. *BMC Plant Biology*. 2022;22:62. DOI: 10.1186/s12870-021-03406-7

Hillmer RA, Tsuda K, Rallapalli G, Asai S, Truman W, Papke MD, Sakakibara H, Jones JDG, Myers CL, Katagiri F. The highly buffered Arabidopsis immune signaling network conceals the functions of its components. *PLoS Genetics*. 2017;13(5):e1006639. DOI: 10.1371/journal.pgen.1006639

Hillmer RA, Igarashi D, Stoddard T, Lu Y, Liu X, Tsuda K, Katagiri F. The kinetics and basal levels of the transcriptome response during effector-triggered immunity in Arabidopsis are mainly controlled by four immune signaling sectors. *Journal of Bioinformatics and Systems Biology*. 2023;6:347–363. DOI: 10.26502/jbsb.5107070

Hossain MR, Bassel GW, Pritchard J, Sharma GP, Ford-Lloyd BV. Trait specific expression profiling of salt stress responsive genes in diverse rice genotypes as determined by modified significance analysis of microarrays. *Frontiers in Plant Science*. 2016;7:567. DOI: 10.3389/fpls.2016.00567

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*. 2015;12(2):115–121. DOI: 10.1038/nmeth.3252

Hundertmark M, Hincha DK. LEA (late embryogenesis abundant) proteins and their encoding genes in

Arabidopsis thaliana. *BMC Genomics*. 2008;9:118. DOI: 10.1186/1471-2164-9-118

Jones JDG, Dangl JL. The plant immune system. *Nature*. 2006;444(7117):323–329. DOI: 10.1038/nature05286

Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0.12. 2019. https://cran.r-project.org/package=pheatmap

Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. *bioRxiv*. 2021. DOI: 10.1101/060012

Lai L, Ge SX. Meta-analysis of gene expression signatures reveals hidden links among diverse biological processes in Arabidopsis. *PLOS ONE*. 2014;9(11):e108567. DOI: 10.1371/journal.pone.0108567

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006;313(5795):1929–1935. DOI: 10.1126/science.1132939

Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*. 2014;15(2):R29. DOI: 10.1186/gb-2014-15-2-r29

Leyser O. Auxin signaling. *Plant Physiology*. 2018;176(1):465–479. DOI: 10.1104/pp.17.00765

Li X. riceidconverter: Convert Biological ID from RAP or MSU to SYMBOL for Oryza Sativa. R package version 1.1.1. 2020. https://cran.r-project.org/package=riceidconverter

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*. 2015;207(2):437–453. DOI: 10.1111/nph.13264

Mittler R. Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science*. 2002;7(9):405–410. DOI: 10.1016/S1360-1385(02)02312-9

Pieterse CMJ, Leon-Reyes A, Van der Ent S, Van Wees SCM. Networking by small-molecule hormones in plant immunity. *Nature Chemical Biology*. 2009;5:308–316. DOI: 10.1038/nchembio.164

Pieterse CMJ, Van der Does D, Zamioudis C, Leon-Reyes A, Van Wees SCM. Hormonal modulation of plant immunity. *Annual Review of Cell and Developmental Biology*. 2012;28:489–521. DOI: 10.1146/annurev-cellbio-092910-154055

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2025. https://www.R-project.org/

Ran X, Liu J, Qi M, Wang Y, Cheng J, Zhang Y. GSHR, a web-based platform provides gene set-level analyses of hormone responses in Arabidopsis. *Frontiers in Plant Science*. 2018;9:23. DOI: 10.3389/fpls.2018.00023

Rasmussen S, Barah P, Suarez-Rodriguez MC, Bressendorff S, Friis P, Costantino P, Bones AM, Nielsen HB, Mundy J. Transcriptome responses to combinations of stresses in Arabidopsis. *Plant Physiology*. 2013;161(4):1783–1794. DOI: 10.1104/pp.112.210773

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 2015;43(7):e47. DOI: 10.1093/nar/gkv007

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140. DOI: 10.1093/bioinformatics/btp616

Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 2010;11(3):R25. DOI: 10.1186/gb-2010-11-3-r25

Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*. 2004;3(1):Article 3. DOI: 10.2202/1544-6115.1027

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005;102(43):15545–15550. DOI: 10.1073/pnas.0506580102

Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, et al. A next generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171(6):1437–1452.e17. DOI: 10.1016/j.cell.2017.10.049

Tanabe S, Yokotani N, Nagata T, Fujisawa Y, Jiang CJ, Abe K, Ichikawa H, Mitsuda N, Ohme-Takagi M, Nishizawa Y, Minami E. Spatial regulation of defense-related genes revealed by expression analysis using dissected tissues of rice leaves inoculated with *Magnaporthe oryzae*. *Journal of Plant Physiology and Pathology*. 2014;2(4):135. DOI: 10.4172/2329-955X.1000135

Thaler JS, Humphrey PT, Whiteman NK. Evolution of jasmonate and salicylate signal crosstalk. *Trends in Plant Science*. 2012;17(5):260–270. DOI: 10.1016/j.tplants.2012.02.010

Trinh DC, Lavenus J, Goh T, Boutte Y, Drogue Q, Vaissayre V, Tellier F, Lucas M, Voss U, Gantet P, Faure JD, Dussert S, Fukaki H, Bennett MJ, Laplaze L, Guyomarc'h S. PUCHI regulates very long chain fatty acid biosynthesis during lateral root and callus formation. *Proceedings of the National Academy of Sciences*. 2019;116(28):14325–14330. DOI: 10.1073/pnas.1906300116

Verma V, Ravindran P, Kumar PP. Plant hormone-mediated regulation of stress responses. *BMC Plant Biology*. 2016;16:86. DOI: 10.1186/s12870-016-0771-y

Volodarsky D, Leviatan N, Otcheretianski A, Fluhr R. HORMONOMETER: a tool for discerning transcript signatures of hormone action in the Arabidopsis transcriptome. *Plant Physiology*. 2009;150(4):1796–1805. DOI: 10.1104/pp.109.138289

Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2016. DOI: 10.1007/978-3-319-24277-4

Wilkins O, Hafemeister C, Plessis A, Holloway-Phillips MM, Pham GM, Nicotra AB, Gregorio GB, Jagadish SV, Septiningsih EM, Bonneau R, Purugganan M. EGRINs (Environmental Gene Regulatory Influence Networks) in rice that function in the response to water deficit, high temperature, and agricultural environments. *Plant Cell*. 2016;28(10):2365–2384. DOI: 10.1105/tpc.16.00158

Zhu Y, Davis S, Stephens R, Meltzer PS, Chen Y. GEOmetadb: powerful alternative search engine for the Gene Expression Omnibus. *Bioinformatics*. 2008;24(23):2798–2800. DOI: 10.1093/bioinformatics/btn520