Plant Physiology®

*Plant Physiology*, 2025, **199**, kiaf618
**https://doi.org/10.1093/plphys/kiaf618**
Advance access publication 18 December 2025
**Letter**

# Modeling omics integration with HIVE identifies response signatures to multifactorial stress in plants

Giulia Calia,[1,2,3] Sophia Marguerit,[1] Ana Paula Zotta Mota,[1] Manon Vidal,[1] Hannes Schuler,[2,3] Ana Cristina Miranda Brasileiro,[4] Patricia Messenberg Guimaraes,[4] Silvia Bottini[1,*]

[1]INRAE, Université Côte d'Azur, CNRS, Institut Sophia Agrobiotech, Sophia-Antipolis 06903 France
[2]Competence Centre for Plant Health, Free University of Bolzano, Bolzano 39100, Italy
[3]Faculty of Agricultural, Environmental and Food Science, Free University of Bolzano, Bolzano 39100, Italy
[4]Embrapa, Department of Genetic Resources and Biotechnology, Brasilia, DF 70770-91, Brazil
*Author for correspondence: silvia.bottini@inrae.fr
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is: Silvia Bottini (silvia.bottini@inrae.fr).

Dear Editor,

Plants as sessile organisms are exposed to diverse stresses which can act singly, sequentially or in a multifactorial manner, and their combination can have dramatic consequences for plant survival even when the effect of each stress applied individually is negligible (Pandey et al. 2015, 2017; Peyraud et al. 2017; Zandalinas et al. 2021; Dutt et al. 2022; Zandalinas and Mittler 2022). The flourishing of omics techniques has led to the possibility of studying complex biological systems, through systematic analysis of its content at the molecular level (Großkinsky et al. 2018). However, because of the experimental complexity of studying the response of one organism to multiple stressors simultaneously, usually experiments are conducted considering one stress factor at a time. An alternative consists of performing *in silico* integration of data on single-stress response. Currently used methods to integrate unpaired experiments are based on meta-analysis or condition-specific differential expression analysis followed by selection of commonly regulated genes. Multiple challenges need to be addressed (Subramanian et al. 2020; Mohammadi-Shemirani et al. 2023); on top of all, the notorious batch effects may hinder a joint analysis (Leek et al. 2010; Goh et al. 2017). Although these approaches yield valuable results, they mainly identify specific signatures in response to one stress and lack those modulated differently in each condition.

To address these challenges, we developed HIVE (Horizontal Integration analysis using Variational AutoEncoders), a method to jointly analyze multiple transcriptomics data from different experiments (i.e. unpaired). The current implementation is based on the use of a VAE, a generative model that learns low-dimensional representations of the observed data, using a variational Bayes methodology in an unsupervised framework. By coupling a random forest regression model and the SHAP explainer, HIVE selects relevant genes for the studied phenotype. The application of a nested stratified cross-validation technique allowed us not only to treat datasets with unbalanced classes but also to overcome the small sample size challenges (Fig. 1, Supplementary Note S1, Supplementary Figs. S1 to S5, and Supplementary Tables S1 to S3).

To illustrate the functionalities of HIVE, by re-using publicly available transcriptomics data from Expression Atlas or PlaD, we constituted 7 multi-stress datasets of transcriptomic data from either RNA-sequencing or microarray for 5 different plants, namely maize (*Zea mays*), rice (*Oryza sativa*), wheat (*Triticum aestivum*), grapevine (*Vitis vinifera*), and Arabidopsis (*Arabidopsis thaliana*) (Supplementary Note S2, Supplementary Table S4). We focus on the results obtained from the grapevine dataset as a showcase to show the suitability of HIVE to handle non-model plants. The results on the other datasets are reported in Supplementary Notes S3 to S6 to show the generality of findings and the robustness of our method.

To evaluate the batch effect reduction, we performed a logistic regression approach on features extracted with either PCA, t-SNE, UMAP, or the HIVE VAE. By considering each latent feature separately, HIVE achieved better results in batch effect reduction compared to the other 2 methods for all datasets (Supplementary Note S3, Supplementary Figs. S6 to S9, and Supplementary Tables S5 to S7). To evaluate the global batch effect reduction, we implemented a k-means clustering analysis of samples by considering their transcriptomics profiles in the raw data and upon application of either PCA, t-SNE, UMAP or the HIVE VAE, and the "reduced latent space" obtained by removing the 10% of the latent features with higher association to the batch effect. We evaluated the distribution of batches in the clusters obtained from the 5 setups by using common metrics, including the Silhouette index, Jaccard score, and the Shannon entropy. We expect that if the batch effect is reduced, the samples do not cluster according to the batch. Strikingly, as we can observe in Fig. 1A both the Silhouette score and the Jaccard index on the HIVE original and "reduced" latent space show the lowest values and the highest entropy, independent of the number of clusters in the chosen range, compared to the correspondent scores obtained on raw data, PCA, UMAP, or t-SNE.

Received May 26, 2025. Accepted November 5, 2025.

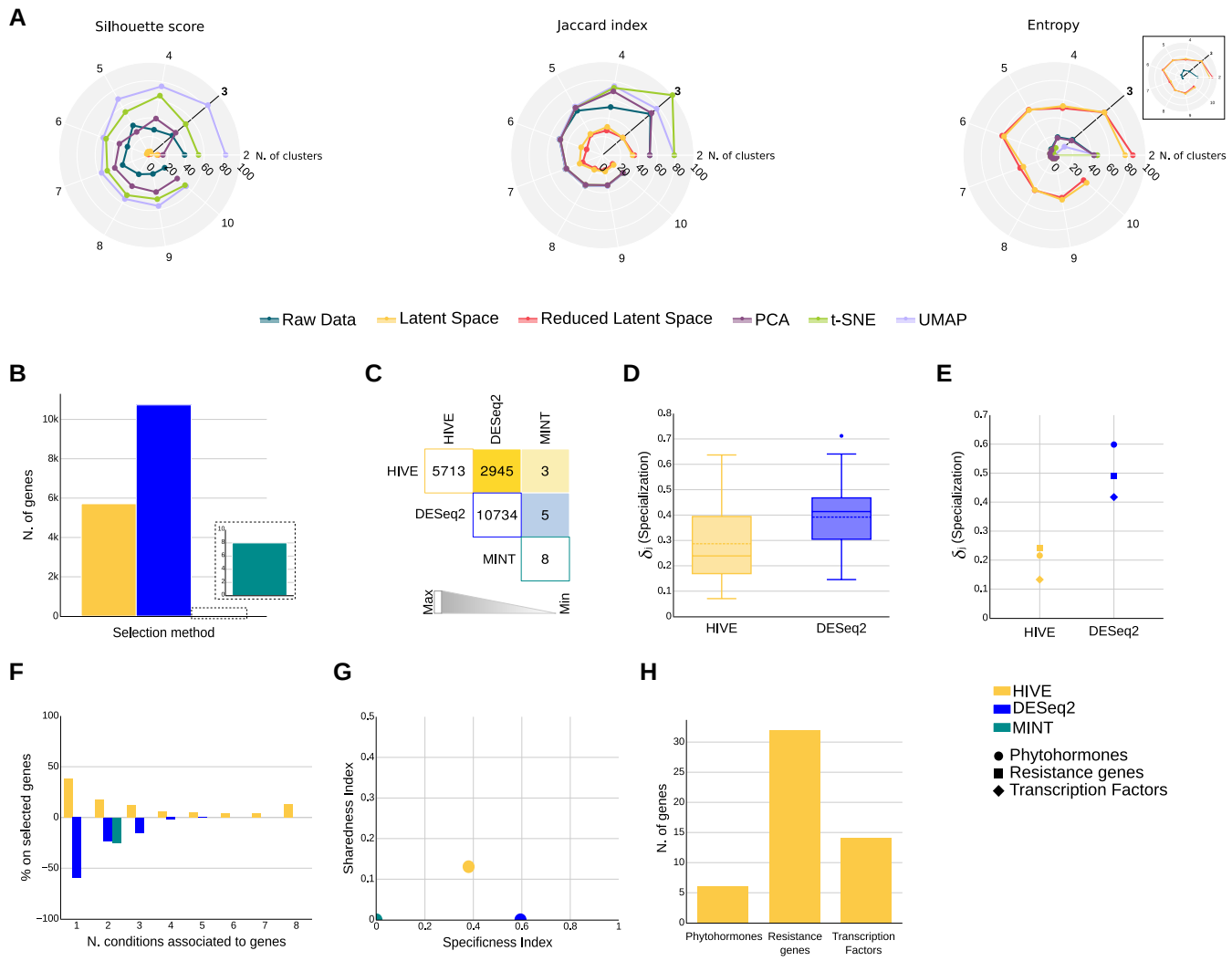


**Figure 1.** Benchmark on batch effect reduction and HIVE, DESeq2 and MINT comparison for gene selection, GO term enrichment, and gene association to condition for grapevine RNA-seq dataset. **A)** Radar plots summarizing the agreement between K-means clustering with increased number of clusters (*k*) and batch belongings with 3 metrics, Silhouette score, Jaccard index and entropy. The inset in the plot related to the entropy highlights the results obtained on the raw data, since they are not visible in the main plot because of superposition with the other tracks. The number of original batches is highlighted in bolded font on the plots. Colors accordingly to the legend at the bottom of the radar plots. See Supplementary Note S3 for further details on the calculation of "reduced" latent space. **B)** Number (N.) of selected genes per benchmarked method. Because of the different scale, the inset shows the number of genes found by MINT. **C)** Pair-wise agreement in gene selection of the benchmarked methods. **D)** Distribution of specialization index for all categories in MERCATOR4 from annotations of either HIVE or DESeq2. **E)** Specialization index of the 3 key categories for plant response to stress found by genes only in HIVE list and only in DESeq2 list. The shape of points corresponds to the categories as indicated in the legend at the bottom right of the figure. **F)** The % of selected genes associated to them by each method for each number of associated conditions. **G)** Relationship between the Sharedness Index and the Specificness Index of each method. **H)** Number (N.) of core genes, namely genes found deregulated in all conditions, belonging to the 3 categories with pivotal roles in plant defense mechanisms. (See Supplementary Note S4 for more details on the indices calculation). Colors for panels from B to H are indicated at the bottom right. For each boxplot the solid line represents the median while the dashed line, the mean of the distribution; box limits are the first (lower limit, Q1) and third (upper limit, Q3) quartiles; whiskers represent upper and lower fences; dots symbolize outliers.

We performed extensive benchmark against other state-of-the art tools, namely DESeq2 (Love et al. 2014) and MINT, from the MixOmics R package (Rohart et al. 2017a, 2017b) to test the ability of HIVE to extract biologically relevant genes. In Fig. 1B, we observe that DESeq2 finds by far the largest number of genes deregulated, while MINT the lowest. Pair-wise comparison of genes selected by the 3 methods shows very low agreement (Fig. 1C). With a similar aim, we performed Mercator annotation to study the biological processes in which selected genes by each method are involved. To quantify the gene expression variation across conditions in those processes, we calculated the specialization index from information theory (Martínez and Reyes-Valdés 2008). In Fig. 1D, we observe that pathways from DESeq2 have higher specialization values compared with HIVE selection, meaning that genes are expressed specifically in one or very few conditions. Similar results were obtained by considering only the lists of genes involved in important processes to regulate plant defenses against stresses, namely phytohormone related, transcription factors, and resistance genes only found by HIVE or by DESeq2 (Fig. 1E). An important aspect of integrative analysis is the association of genes to conditions. To quantitatively measure the difference among methods, we defined 2 novel indexes: the sharedness and the specificness, which quantify the ratio of common or specific signature compared to the total number of selected

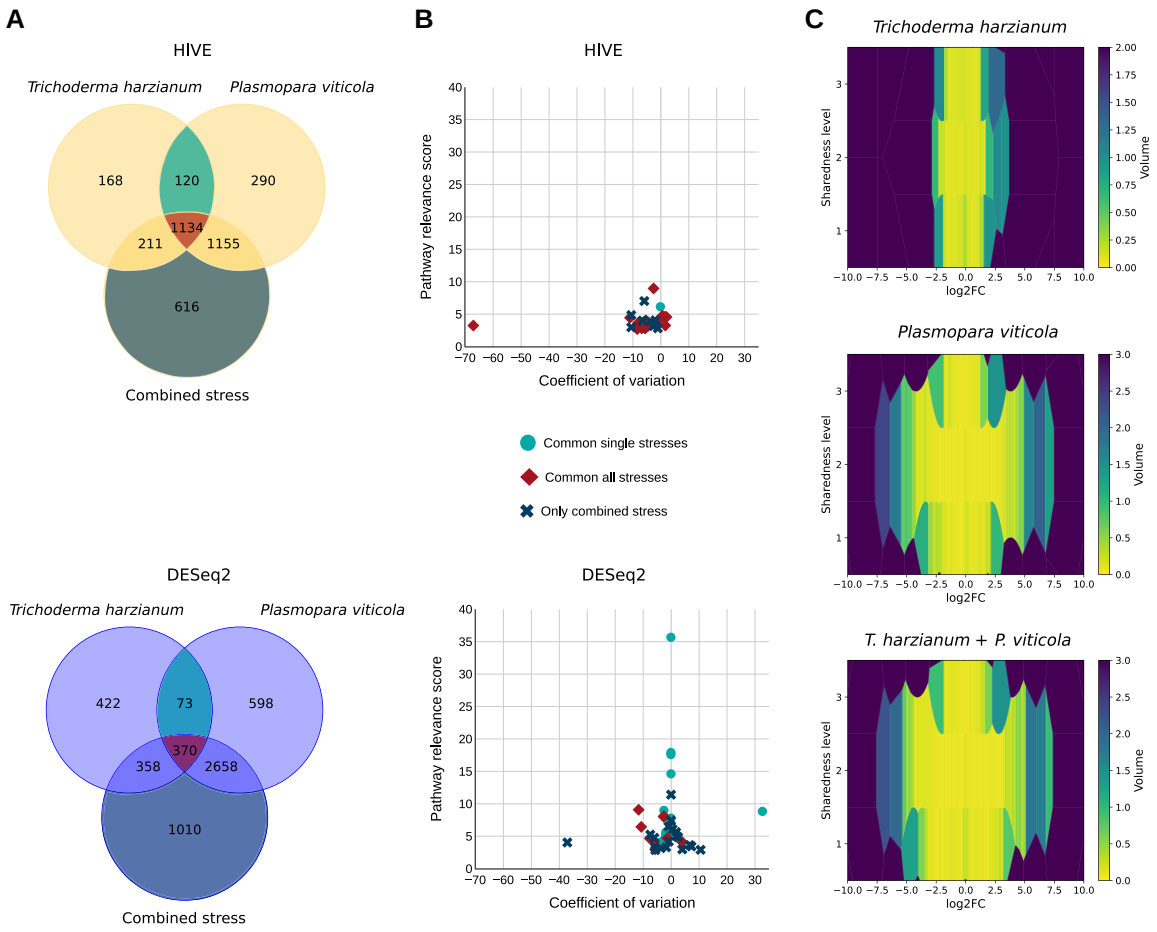


**Figure 2.** Comparison of HIVE and DESeq2 for multi-stress associated genes and Voronoi maps for sharedness gene expression in grapevine RNA-seq dataset. **A)** Venn diagrams representing overlaps between single-stress genes and the respective combination of stresses for grapevine dataset. **B)** Number (N.) of genes from the intersection among the 3 experiments obtained by the analysis with HIVE or DESeq2 annotated as belonging to key categories for plant response to stresses. **C)** Voronoi maps of gene expression associated to the single stresses (first 2 maps) and associated to the combined stress (last map). Genes were mapped accordingly to their level of sharedness-gene expression value on a x-y plot, each gene were then enclosed in a Voronoi cell, the volume of each cell is represented by a color scale, accordingly to each plot.

signatures, respectively. HIVE identifies a balanced ratio of common and specific signatures, compared to the other methods obtaining the highest sharedness for all datasets; therefore, the highest number of genes deregulated in all conditions (Fig. 1, F and G). To inspect whether HIVE can identify novel signatures not found by the other 2 tools, we defined the "core genes" as the genes found deregulated in all conditions. Then we inspected the core genes identified by the different methods and quantified the number of genes involved in important functions to better understand plant response to multiple stresses. As reported in Fig. 1H, not only did DESeq2 and MINT not identify any core genes, but HIVE core genes are composed of genes in the 3 categories: resistance genes, genes related to phytohormones, and transcription factors. Finally, we investigated the level of gene regulation found by each method as specifically deregulated in one condition or shared by all (Supplementary Note S4, Supplementary Fig. S10, and Supplementary Tables S8 to S11). From the cumulative distributions, we can observe that the genes selected by HIVE as specific to each condition show the most extreme values of modulation when compared to the background, either up or down regulation, with the most significant *P*-value compared to the other methods *versus* background. This result highlights that the genes found by HIVE as responsive specifically to one condition have an overall level of regulation that is stronger than the genes selected by other methods for the same category (Supplementary Note S4, Supplementary Fig. S11, and Supplementary Table S12).

To test the ability of HIVE to identify signatures common to multiple stresses responses when only single-stress experiments are available, we dispose, of an experiment in which the grapevine is subjected to 2 stresses contemporaneously (*Trichoderma harzianum* and/or *Plasmopara viticola*) and also to only one of the two at a time. By comparing the number of genes found in the 3 conditions, we showed that HIVE found only a few genes associated only with the combined stress while a high number of genes in common between the combined stress and both the single stresses were retrieved, compared to the selection by DESeq2 in which a higher percentage of genes were found deregulated only in the combined stress and not in the 2 single-stress experiments (Fig. 2A). To show the validity of those genes in common between the 2 single stresses and validated by the combined stress experiment, we quantified the number of genes involved in key processes for plant defense against stresses. In Fig. 2B, we show that, especially regarding resistance genes, HIVE identifies more of those key genes than DESeq2 in the common selection. Similar findings were reported for the other 2 datasets as shown in Supplementary Note S5 (Supplementary Fig. S12 and Supplementary Tables S13 to S15). This represents a proof-of-concept that HIVE can be used to reliably extract multi-stress signatures from experiments

performed on single stresses. To further inspect the relationship between the number of conditions in which genes were found deregulated (sharedness) and their level of expression, we used the Voronoi maps. We observe that the overall effect of *T. harzianum* on plant response is very mild, opposed to the effect of *P. viticola*, and the combined stress shows a very similar pattern as the *P. viticola* alone (Fig. 2C). Overall, we found that the patterns of sharedness-expression from HIVE Voronoi on the single stresses are comparable to the combined, therefore strengthening the ability of HIVE to perform *in silico* multi-stress integration from experiments conducted on single stresses (Supplementary Note S6, Supplementary Fig. S13 and S14, and Supplementary Table S16).

In conclusion, we showed that integrative analysis with HIVE can highlight novel signatures and biological insights that cannot be found by other analysis methods. HIVE is a valuable tool that can be applied to any phytopathosystem and will provide improvements in our understanding of the mechanisms set up by the plant to respond to multiple stresses.

## Author contributions

G.C.: methodology, software, validation, investigation, writing original draft, visualization; S.M.: methodology, software, formal analysis; A.P.Z.M.: resources, data curation, writing review & editing, M.V.: software; H.S.: supervision, writing review & editing; A.C.M.B.: investigation, resources; P.M.G.: investigation, resources, writing review & editing; S.B.: conceptualization, methodology, validation, writing original draft, supervision, project administration.

## Supplementary data

The following materials are available in the online version of this article.

**Supplementary Note S1.** HIVE framework overview.

**Supplementary Note S2.** Datasets collection, preprocessing and integration.

**Supplementary Note S3.** Metrics for benchmarking batch effect correction.

**Supplementary Note S4.** Metrics for benchmarking methods performances towards gene selection.

**Supplementary Note S5.** Combined stress analysis.

**Supplementary Note S6.** Voronoi diagrams.

**Supplementary Figure S1.** HIVE framework.

**Supplementary Figure S2.** Workflow focus.

**Supplementary Figure S3.** 50 VAE performances.

**Supplementary Figure S4.** $\mu$ and $\log(\sigma^2)$ distributions.

**Supplementary Figure S5.** Explanation of HIVE genes selection and association to conditions.

**Supplementary Figure S6.** Intra- and extra-batch correlation of samples.

**Supplementary Figure S7.** F1-score distributions latent space compared with PCA and UMAP.

**Supplementary Figure S8.** Clustering performances for the benchmarked RNA sequencing and microarray data.

**Supplementary Figure S9.** Max values of Jaccard Index for Microarray and RNA-seq datasets.

**Supplementary Figure S10.** Benchmark of HIVE, DESeq2, and MINT regarding gene selection, Specialization index, and gene association to condition for *Arabidopsis thaliana* and *Zea mays*.

**Supplementary Figure S11.** Cumulative distribution of log2FC for genes associated to different conditions from each benchmarked tool.

**Supplementary Figure S12.** Comparison of multi-stress associated genes' selection from HIVE and DESeq2.

**Supplementary Figure S13.** Voronoi maps representing gene expression in different sharedness levels across conditions.

**Supplementary Figure S14.** Volumes distributions per ranges of genes log2FC for Voronoi regions at each sharedness level.

**Supplementary Table S1.** Total loss and learning error of 50 VAE.

**Supplementary Table S2.** KL-divergence of 50 VAE.

**Supplementary Table S3.** Mean and standard deviation of $\mu$ and $\log(\sigma 2)$.

**Supplementary Table S4.** Extended information on analyzed integrated datasets.

**Supplementary Table S5.** F1-scores of Logistic Regressions.

**Supplementary Table S6.** K-means clustering performances.

**Supplementary Table S7.** Pairwise comparison of max Jaccard score in K-means clustering.

**Supplementary Table S8.** Gene association to condition for all datasets and all benchmarked tools.

**Supplementary Table S9.** MERCATOR4 annotated categories for each tool gene-list.

**Supplementary Table S10.** Overlap of tools gene-lists with PRGdb.

**Supplementary Table S11.** MERCATOR4 annotated categories and three key categories for core genes.

**Supplementary Table S12.** Kolmogorov-Smirnov test *P*-values for each pair of log2FC cumulative distribution.

**Supplementary Table S13.** Comparison of selected genes involved in 2 single stresses, in combined stress or an overlap of the three by both HIVE and DESeq2.

**Supplementary Table S14.** MERCATOR4 annotated categories and three key categories for genes common to all experiments.

**Supplementary Table S15.** GO-terms used to calculate the pathway relevance score.

**Supplementary Table S16.** Voronoi regions volume distribution comparisons.

## Funding

This work was supported by the French government, through the UCA JEDI Investments in the Future project managed by the French National Research Agency (ANR) under reference number ANR-15-IDEX-01. H.S. was funded by a joint project of the Province of Bozen-Bolzano and the Austrian Science Fund (FWF).

*Conflict of interest statement*. None declared.

## Data availability

Data, scripts, and HIVE source code are available in the GitHub repository: https://github.com/Plant-Net/HIVE.git (https://doi.org/10.5281/zenodo.14944705).

## References

Dutt A, Andrivon D, Le May C. Multi-infections, competitive interactions, and pathogen coexistence. *Plant Pathol*. 2022:71(1):5–22. https://doi.org/10.1111/ppa.13469

Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. *Trends Biotechnol*. 2017:35(6):498–507. https://doi.org/10.1016/j.tibtech.2017.02.012

Großkinsky DK, Syaifullah SJ, Roitsch T. Integration of multi-omics techniques and physiological phenotyping within a holistic

phenomics approach to study senescence in model and crop plants. *J Exp Bot*. 2018:69(4):825–844. https://doi.org/10.1093/jxb/erx333

Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010:11(10):733–739. https://doi.org/10.1038/nrg2825

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014:15(12):550. https://doi.org/10.1186/s13059-014-0550-8

Martínez O, Reyes-Valdés MH. Defining diversity, specialization, and gene specificity in transcriptomes through information theory. *Proc Natl Acad Sci USA*. 2008:105(28):9709–9714. https://doi.org/10.1073/pnas.0803479105

Mohammadi-Shemirani P, Sood T, Paré G. From 'omics to multi-omics technologies: the discovery of novel causal mediators. *Curr Atheroscler Rep*. 2023:25(2):55–65. https://doi.org/10.1007/s11883-022-01078-8

Pandey P, Irulappan V, Bagavathiannan MV, Senthil-Kumar M. Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Front Plant Sci*. 2017:8:537. https://doi.org/10.3389/fpls.2017.00537

Pandey P, Ramegowda V, Senthil-Kumar M. Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. *Front Plant Sci*. 2015:6:723. https://doi.org/10.3389/fpls.2015.00723

Peyraud R, Dubiella U, Barbacci A, Genin S, Raffaele S, Roby D. Advances on plant-pathogen interactions from molecular toward systems biology perspectives. *Plant J*. 2017:90(4):720–737. https://doi.org/10.1111/tpj.13429

Rohart F, Eslami A, Matigian N, Bougeard S, Lê Cao K-A. MINT: a multivariate integrative method to identify reproducible molecular signatures across independent experiments and platforms. *BMC Bioinformatics*. 2017a:18(1):128. https://doi.org/10.1186/s12859-017-1553-8

Rohart F, Gautier B, Singh A, Cao K-AL. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017b:13(11):1005752. https://doi.org/10.1371/journal.pcbi.1005752

Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights*. 2020:14:1177932219899051. https://doi.org/10.1177/1177932219899051

Zandalinas SI, Mittler R. Plant responses to multifactorial stress combination. *New Phytol*. 2022:234(4):1161–1167. https://doi.org/10.1111/nph.18087

Zandalinas SI, Sengupta S, Fritschi FB, Azad RK, Nechushtai R, Mittler R. The impact of multifactorial stress combination on plant growth and survival. *New Phytol*. 2021:230(3):1034–1048. https://doi.org/10.1111/nph.17232