Article

# Machine-learning meta-analysis reveals ethylene as a central component of the molecular core in abiotic stress responses in Arabidopsis

Raul Sanchez-Munoz [1,6], Thomas Depaepe [1], Marketa Samalova [2,3], Jan Hejatko [3,4], Isiah Zaplana [5] ✉ & Dominique Van Der Straeten [1] ✉

Understanding how plants adapt their physiology to overcome severe and often multifactorial stress conditions in nature is vital in light of the climate crisis. This remains a challenge given the complex nature of the underlying molecular mechanisms. To provide a comprehensive picture of stress-mitigation mechanisms, an exhaustive analysis of publicly available stress-related transcriptomic data has been conducted. We combine a meta-analysis with an unsupervised machine-learning algorithm to identify a core of stress-related genes active at 1-6 h and 12-24 h of exposure in *Arabidopsis thaliana* shoots and roots. To ensure robustness and biological significance of the output, often lacking in meta-analyses, a triple validation is incorporated. We present a 'stress gene core': a set of key genes involved in plant tolerance to ten adverse environmental conditions and ethylene-precursor supplementation rather than individual conditions. Notably, ethylene plays a key regulatory role in this core, influencing gene expression and acting as a critical factor in stress tolerance. Additionally, the analysis provides insights into previously uncharacterized genes, key genes within large families, and gene expression dynamics, which are used to create biologically validated databases that can guide further abiotic stress research. These findings establish a strong framework for advancing multi-stress-resilient crops, paving the way for sustainable agriculture in the face of climate challenges.

Environmental factors such as light, temperature, and water availability are key signals steering plant growth and development. The current climate crisis generated by global warming steadily increases the number of geographical regions suffering from extreme environmental conditions[1]. The latter create situations in which plants, even if their survival is not severely compromised, grow under suboptimal conditions, which ultimately hamper their growth and reduce crop yield[2]. Considering that the vast majority of the world's

[1]Laboratory of Functional Plant Biology, Department of Biology, Faculty of Sciences, Ghent University, Gent B-9000, Belgium. [2]Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czech Republic. [3]CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic. [4]National Centre for Biotechnological Research, Faculty of Science, Masaryk University, Brno, Czech Republic. [5]Institute of Industrial and Control Engineering (IOC), Universitat Politècnica de Catalunya - BarcelonaTech (UPC), Barcelona 08028, Spain. [6]Present address: Department of Agri-Food Engineering and Biotechnology (DEAB), Universitat Politècnica de Catalunya – BarcelonaTech (UPC), Castelldefels 08860 Barcelona, Spain. ✉e-mail: isiah.zaplana@upc.edu; Dominique.VanDerStraeten@UGent.be

arable-land area is exposed to biotic- and abiotic-stress conditions—with estimations of up to 30% and 82% of worldwide crop productivity decay, respectively—and that climate change hits the planet faster than anticipated (IPCC report, 2023), we face an imminent global-agronomic crisis[3,4]. Nevertheless, plants have evolved sophisticated mechanisms to cope with adverse environmental conditions.

Several efforts have focused on unraveling stress-mitigation pathways with the aim of engineering strategies to improve plant stress tolerance, though with limited success[5]. This could be caused by the diversity in stress responses, also across taxa, as well as the complexity of each stress type and its associated signaling pathway. In addition, environmental conditions usually impose multifactorial stimuli rather than a juxtaposition of individual stresses, which plants perceive as a combination of different inputs[6]. Thus, plants require a strict and coordinated communication between different tissues and organs to optimally respond to unfavorable conditions[7]. Systemic responses between distant tissues or organs have been demonstrated by the application of localized stress stimuli, confirming the existence of efficient communication routes within plants[8].

Besides these systemic and coordinated responses, different tissues also perform specific tasks, related to their primary physiological functions[9]. Therefore, inter-organ communication and tissue-specific responses are crucial for optimal stress mitigation. In addition, responses to stress also depend on its duration. At early stages, stress signaling is focused on rapid biochemical, biophysical, molecular and physiological alterations that avoid and/or reduce irreversible cellular damage[10]. As stress persists, the acclimation process starts, modulating growth and development to ensure survival and, concomitantly regulating the mechanisms initiated during the early stages[11–13]. The gaseous hormone ethylene is one of the main players in environmental adaptation[14]. Transcriptional analyses have detected several groups of important stress-related transcription factors that are directly influenced by ethylene, such as AP2/ERF (APETALA 2/ETHYLENE RESPONSIVE FACTOR), NAC (NAM (NO APICAL MERISTEM), ATAF1/2 (ACTIVATING FACTOR 1/2), CUC2 (CUP-SHAPED COTYLEDON 2) and WRKYs, among others[15]. Nevertheless, many aspects of ethylene-mediated stress adaptation are still poorly understood, often clouded by the intricate crosstalk with other regulatory pathways. Additionally, while ethylene has traditionally been associated with specific stress responses, its overarching role as a central regulator across multiple stress conditions remains largely unexplored.

Stress responses are generally defined as highly specific and sometimes potentially antagonistic[16], however several clues hint at the existence of a shared molecular core acting as a signaling hub to coordinate multiple stress stimuli[17]. This is further supported by the finding that stress priming—the exposure to mild stress conditions in order to develop subsequent stress tolerance—can induce cross-tolerance[18]. Heat and cold cross-tolerance is a common and well-known example due to their physiological similarities[19], but this relationship has also been identified between heat and cadmium[20], or cold, salt, and drought[21]. Together, this points towards the existence of a 'stress gene core', responsible for the coordination of specific responses towards physiological adaptation upon compound stresses, though the exact players remain to be determined.

Several factors impede the identification of such a stress core, predominantly related to the enormous complexity of the network controlling plant stress tolerance and the multifactorial nature of stress. A direct consequence of this complexity is the fact that the modification of specific DNA elements with the aim to enhance stress tolerance can generate unpredictable and unwanted consequences[22]. Therefore, a deeper understanding of factors shared between stress signaling routes, both in spatial and temporal contexts, is vital. Moreover, the large quantity of potential genes involved in stress responses infers that the experimental efforts needed to understand all the players involved are overwhelming. Hence, it is clear that robust in silico methods are crucial to gain insights into the systems' complexity and guide experimental confirmation.

The extensive study of transcriptional changes by means of RNA sequencing methods provides a rich and diverse library of data. Nevertheless, single transcriptomic analyses can produce contradictory conclusions, driven by experimental differences such as the type of treatment, the severity of stress stimuli, the time range of treatment, tissue and plant age, and/or sample size[23], or lead to biased, misleading results. The combination of multiple transcriptomic data with meta-analysis approaches has been proposed as a method to bypass these limitations[24], providing solid input for the definition of a gene core. However, the design and implementation of meta-analyses are not trivial, since they require the combination of powerful statistical methods without losing biological significance[25].

For that reason, here we performed a meta-analysis combining all publicly available stress-related *Arabidopsis thaliana* transcriptomic data, after careful and individual consideration of the suitability of each transcriptome in order to ensure biological significance of the analysis. Our aim was to identify an abiotic-stress gene core, given the impact of abiotic stress on crop yield[4] and exacerbation of stress by global climate change. Such complex datasets require, in addition to a reliable meta-analysis method, a potent data-mining tool to extract information. The potential of machine-learning techniques for data analysis has been extensively demonstrated, as well as the limitations and flaws that need to be taken into consideration for its proper usage and interpretation[26].

The high-dimensional dataset derived from the combination of all single transcriptomes requires an efficient machine-learning method able to cope with such vast data. Support vector machine (SVM) is easy to implement for classification of complex multi-dimensional datasets[27]. In particular, an unsupervised version of the standard SVM, called SVM Clustering, was selected for this work, as it preserves all the key properties of a standard SVM while, concurrently, avoiding the limitations of purely supervised methods, e.g. overfitting[28–30]. To properly control the reproducibility and robustness of our methodology, and to increase the biological significance of the results, often lacking in meta-analyses, we have designed a triple validation, including experimental validation of some key stress-related genes detected in our analysis.

This work presents the first machine learning-driven meta-analysis of abiotic-stress-related plant genes including all publicly available Arabidopsis datasets. The final output is a list of genes forming the plant 'abiotic-stress gene core'. Rather than being stress-specific—as when derived from single transcriptomic analyses—these genes represent potential hubs of general stress responses in plants. Overall, this methodology, along with the derived datasets, represents a data-driven launchpad for informed crop-engineering efforts toward realizing sustainable agriculture.

## Results
### Construction of differentially-expressed-gene libraries and hierarchical clustering
After screening and filtering all available abiotic-stress-related transcriptomic datasets ($n = 945$) from the Gene-Expression Omnibus (GEO) database, 500 individual transcriptomes (from 23 selected datasets) were analyzed for differentially expressed genes (DEGs) under 10 different stress conditions: cold, complete submergence, drought, heat, high light, osmotic, salt, partial submergence, UV (UV-B), wounding, and exogenous treatment with the ethylene precursor 1-aminocyclopropane-1-carboxylate (ACC) (Supplementary Data file 1). An overview of the complete analysis is presented in Fig. S1. Based on the kinetics of individual transcriptomic analyses, the lists of DEGs were combined in early (from 1 to 6 h) and late responses to stress conditons (from 12 to 24 h) (Supplementary Data file 2; see the "Methods" section). Prior to further analyses, a first validation step was
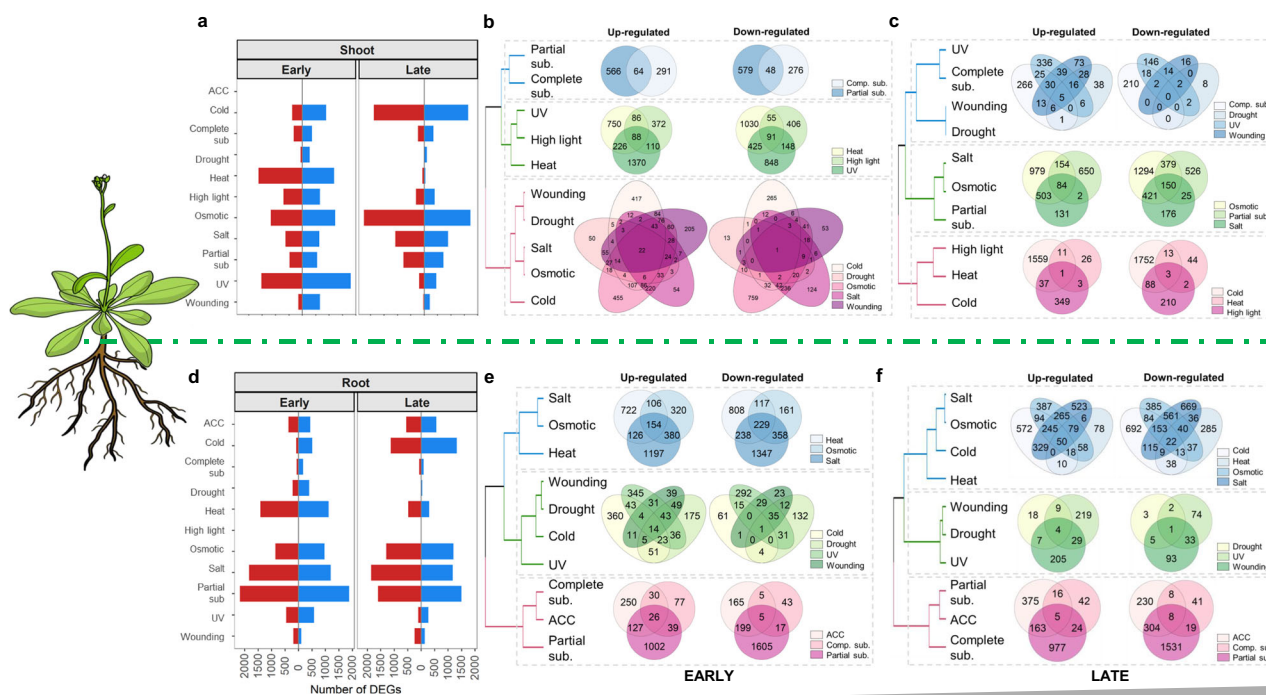
**Fig. 1 | Distribution of differentially expressed genes per stress, tissue, and time point. a** Number of DEGs detected in each of the studied conditions in shoot tissue for early and late time points. Up-regulated DEGs are colored in blue, down-regulated in red. **b** and **c** HCA of the different stress conditions by DEG modules in shoots for early **b** and late **c** time points. **d** Number of DEGs detected in each of the studied conditions in root tissue for early and late timepoints. Up-regulated DEGs are colored in blue, down-regulated in red. **e** and **f** HCA of the different stress conditions by DEG modules in roots for early **e** and late **f** time points. Branch length is based on cluster distance.

performed. To confirm the suitability and biological relevance of each DEG library, six marker genes were selected per stress condition based on their empirically determined expression. Subsequently, the expression of these markers was assessed for each stress, taking into account temporal and spatial specificity (Supplementary Data file 3). All passed the first biological validation test with at least five out of six markers being differentially expressed, demonstrating both the suitability and the accuracy of the DEG libraries.

The number of DEGs obtained in each stress, tissue, and timepoint combination provided a first insight into the regulation of abiotic-stress responses (Fig. 1a, d). In general, the responses were balanced considering up- and down-regulation and number of DEGs between tissues and timepoints. A hierarchical clustering analysis (HCA) grouped certain stresses by DEG modules, suggesting physiological resemblances between them (e.g. salt and osmotic stress in roots and shoots; partial and complete submergence in roots; wounding and drought in shoots; and early exposure to UV and high light in shoots; Fig. 1b, c, e, and f).

Analysis of stress-related transcriptional responses found clear spatial and temporal differences in gene expression. In general, root responses were more stable over time, with minimal changes in clustering between early and late responses. For example, the same stress clusters were identified during early and late stress responses in roots, suggesting consistent underlying mechanisms (Figs. 1, S2 and S3). In contrast, shoot responses exhibited more dynamic shifts between early and late stress conditions. Notably, stresses like osmotic, cold, and UV elicited stronger transcriptional responses in shoots, with a higher number of DEGs detected during the early exposure. However, late responses in shoots showed a shift, with fewer up-regulated genes and a higher presence of transcriptional repression as down-regulated DEGs, highlighting the temporal complexity of shoot stress responses (Figs. 1, S2 and S3). This supports the existence of a tissue-specific mechanism to respond to stress conditions, especially when

maintained over time (for more details, see Supplementary Information file).

## Support vector machine (SVM) clustering classifies genes into stress cores

To identify a set of central actors participating in stress signaling, i.e. as part of a stress gene core, we computed meta-$p$-values for all studied genes (see the "Methods" section). We took into account their transcriptional changes in all surveyed stress conditions in four different datasets: roots early, roots late, shoots early and shoots late, and used SVM Clustering—an unsupervised version of standard SVM. First, a frequency-based pre-classification was performed. The genes appearing as DEGs in at least five of the studied conditions were assigned to the positive class (class 1), while the rest was assigned to the negative class (class 0). Subsequently, the meta-$p$-value dataset containing the information for all 10 stress conditions for the complete set of around 12,000 genes, was re-classified using SVM Clustering. This analysis classifies genes depending on their distribution in the 10-dimensional space taking into account the distribution of meta-$p$-values (reflecting how statistically significant the expression changes are under each condition).

SVM Clustering categorized the vast majority of genes as not relevant (class 0; approximately 99%), coinciding with the frequency-based pre-classification (Fig. 2a). Around 5–30% of the genes pre-classified as relevant (32, 6, 82 and 14 genes for the early-root, late-root, early-shoot and late-shoot responses, respectively) were refuted after SVM Clustering and deemed not relevant (1 → 0). In contrast, a few genes pre-classified as not relevant(0) were included in the final SVM gene core (0 → 1) (2, 4, 7 and 0 genes from the early-root, late-root, early-shoot and late-shoot responses, respectively). The number of genes pre-classified as relevant but considered irrelevant by SVM Clustering highlights the discriminatory power of this classification approach, refining the pre-classification established on the
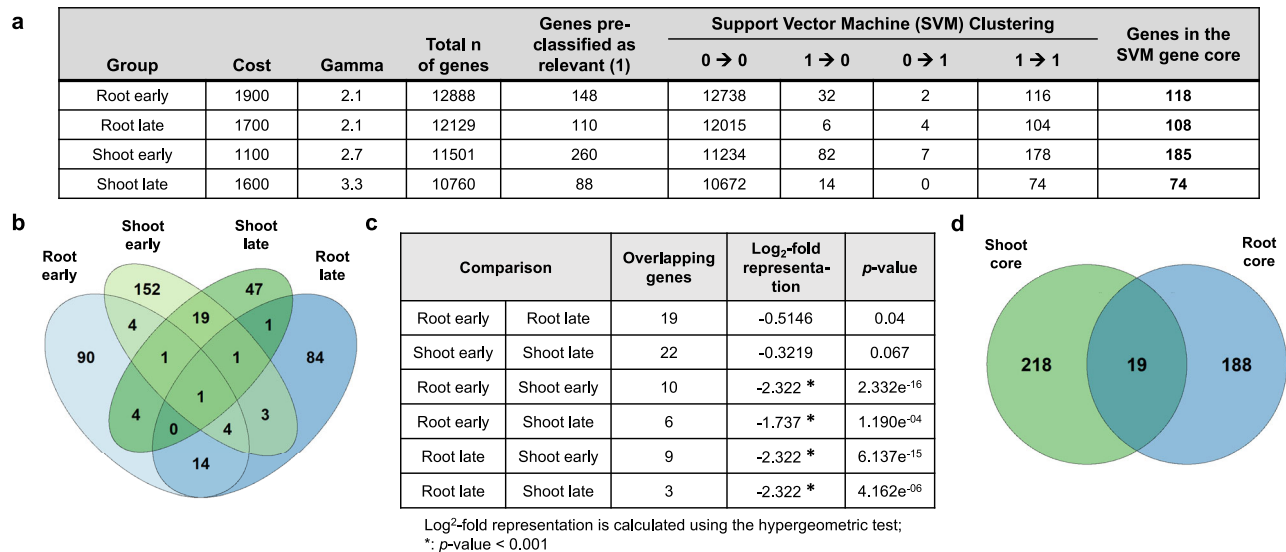
**a**

| Group | Cost | Gamma | Total n of genes | Genes pre-classified as relevant (1) | Support Vector Machine (SVM) Clustering | | | | Genes in the SVM gene core |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 0 → 0 | 1 → 0 | 0 → 1 | 1 → 1 | |
| Root early | 1900 | 2.1 | 12888 | 148 | 12738 | 32 | 2 | 116 | **118** |
| Root late | 1700 | 2.1 | 12129 | 110 | 12015 | 6 | 4 | 104 | **108** |
| Shoot early | 1100 | 2.7 | 11501 | 260 | 11234 | 82 | 7 | 178 | **185** |
| Shoot late | 1600 | 3.3 | 10760 | 88 | 10672 | 14 | 0 | 74 | **74** |

**b**



**c**

| Comparison | | Overlapping genes | Log₂-fold representation | p-value |
|---|---|---|---|---|
| Root early | Root late | 19 | -0.5146 | 0.04 |
| Shoot early | Shoot late | 22 | -0.3219 | 0.067 |
| Root early | Shoot early | 10 | -2.322 * | 2.332e⁻¹⁶ |
| Root early | Shoot late | 6 | -1.737 * | 1.190e⁻⁰⁴ |
| Root late | Shoot early | 9 | -2.322 * | 6.137e⁻¹⁵ |
| Root late | Shoot late | 3 | -2.322 * | 4.162e⁻⁰⁶ |

Log²-fold representation is calculated using the hypergemetric test;
*: p-value < 0.001

**d**



**Fig. 2 | Support vector machine (SVM) Clustering for the determination of the stress gene cores. a** SVM-specific hyperparameters (cost and gamma) used for the SVM Clustering of each tissue and timepoint dataset. The number of genes forming each SVM gene core is highlighted in bold at the right. A pre-classification was performed first, whereby relevant genes appearing as a DEG in at least five of the stress conditions studied were assigned to the positive class (class 1). The results of SVM Clustering are defined as 0 → 0 (genes marked as not relevant and stated as such by the algorithm), 1 → 0 (genes marked as relevant but not stated as such by the algorithm), 0 → 1 (genes not marked as relevant but stated as such by the algorithm) and 1 → 1 (genes marked as relevant and stated as such by the algorithm). **b** Venn diagram representing the overlap between different SVM gene cores. **c** Overlap between the SVM gene cores. The value, positive (over-representation) or negative (under-representation), and the statistical significance of the overlaps are indicated. A hypergeometric test was used to detect the statistical significance of the overlaps, considering a p-value < 0.001 as significant due to the reduced number of genes in the comparisons. No adjustments were applied for statistical comparissons. **d** Venn diagram representing the overlap between the root and shoot gene cores.

distribution of the complete set of meta-p-values. Based on this, four sets of core genes, coined SVM gene cores, with significant transcriptional alterations in all the studied conditions, were identified: 118 genes for early responses in roots, 108 for late responses in roots, 185 for early responses in shoots and 74 for late responses in shoots (Fig. 2a).

The comparison between tissues (root versus shoot in all timepoint combinations; Fig. 2b, c), as opposed to comparisons between timepoints within a tissue (root early versus root late; shoot early versus shoot late), revealed a significant under-representation of overlapping genes. Therefore, considering gene-response composition, tissue specificity is stronger than temporal specificity, supporting the results obtained by the qualitative DEG classification (Figs. 1 and S3). For that reason, and due to the statistically non-significant differences between timepoints (Fig. 2c), the genes belonging to the SVM gene cores per tissue were combined, resulting in a final dataset of 207 genes forming the root gene core and 237 genes forming the shoot gene core after removal of duplicates (Fig. 2d). Despite the tissue specificity of the SVM gene cores, 19 genes are shared between the root and shoot cores, which encompass fundamental proteins with tissue-independent functions. These predominantly cover genes involved in cell-wall maintenance and membrane integrity (*EXPANSIN A1* (*EXPA1*), *LIPID TRANSFER PROTEIN 2* (*LTP2*), dehydrins, such as *COLD-REGULATED 47* (*COR47*) and *LOW TEMPERATURE-INDUCED 30* (*LTI30*), and *BLUE COPPER BINDING PROTEIN* (*BCB*))[31–33] in addition to some uncharacterized genes (e.g. *AT1G19380* and *AT5G19875*). The complete list of genes that form the different SVM gene cores as well as their overlap is in Supplementary Data file 4.

We studied the composition of the gene cores in terms of annotated biological functions (Gene Ontology (GO) enrichment analysis) and gene families. The most significant function enriched in the shoot core was the response to water (GO:0009415) and response to hypoxia appeared in roots (GO:0071456) (Figs. S2 and S4). This reflects that stressed shoots prioritize maintenance of water homeostasis, while roots mostly try to maintain normoxia. In addition, amino-acid

transporters were enriched in the shoot core, while EXPANSINs, related to cell-wall remodeling, appeared on the forefront in the root core (for a complete GO and gene family analysis, see Supplementary Information file).

## Protein networks related to the SVM gene cores

To further elucidate the functionality of shoot and root cores, we constructed a protein network representing both physical and functional interactions. Subsequently, a k-means clustering method was applied to obtain protein clusters based on known interactions in order to provide further evidence of their biological roles (Material and Methods).

For shoots, four clusters were identified (Fig. 3a). The blue and red clusters contained the largest number of proteins (28%). However, given its higher degree of connectivity, the blue cluster was considered to be a key cluster within the shoot core (Fig. 3a, b). To support the biological relevance of the different clusters, we performed a biological-processes GO enrichment for each cluster individually (Fig. 3c–f). As expected, the biological responses of the blue cluster largely overlapped with those of the overall shoot core (Fig. S4), with 'response to water deprivation' (GO:0009414) the most significant GO term (Fig. 3f). Three WRKY transcription factors appeared to act as central nodes in the interaction network, strongly interacting among themselves (WRKY33, WRKY46 and WRKY18) (Fig. 3b). In addition, MAP KINASE KINASE 9 (MKK9), a Mitogen Activated Protein Kinase (MAPK) protein, directly interacts with WRKY33, the central protein in the interactome, suggesting an important regulatory role for MKK9 as well. Furthermore, the interaction network also highlighted other known stress genes as part of the stress signaling core, including the mitochondrial *ALTERNATIVE OXIDASE 1A* (*AOX1a*), as well as 31 unannotated genes, hence uncovering their function (Supplementary Data 5).

The red cluster contained proteins mainly related to maintenance of cell-wall integrity (GO:0042546, GO:0010411, GO:0006949) (Fig. 3c). The green cluster was marked by 'alpha-amino acid metabolic process'
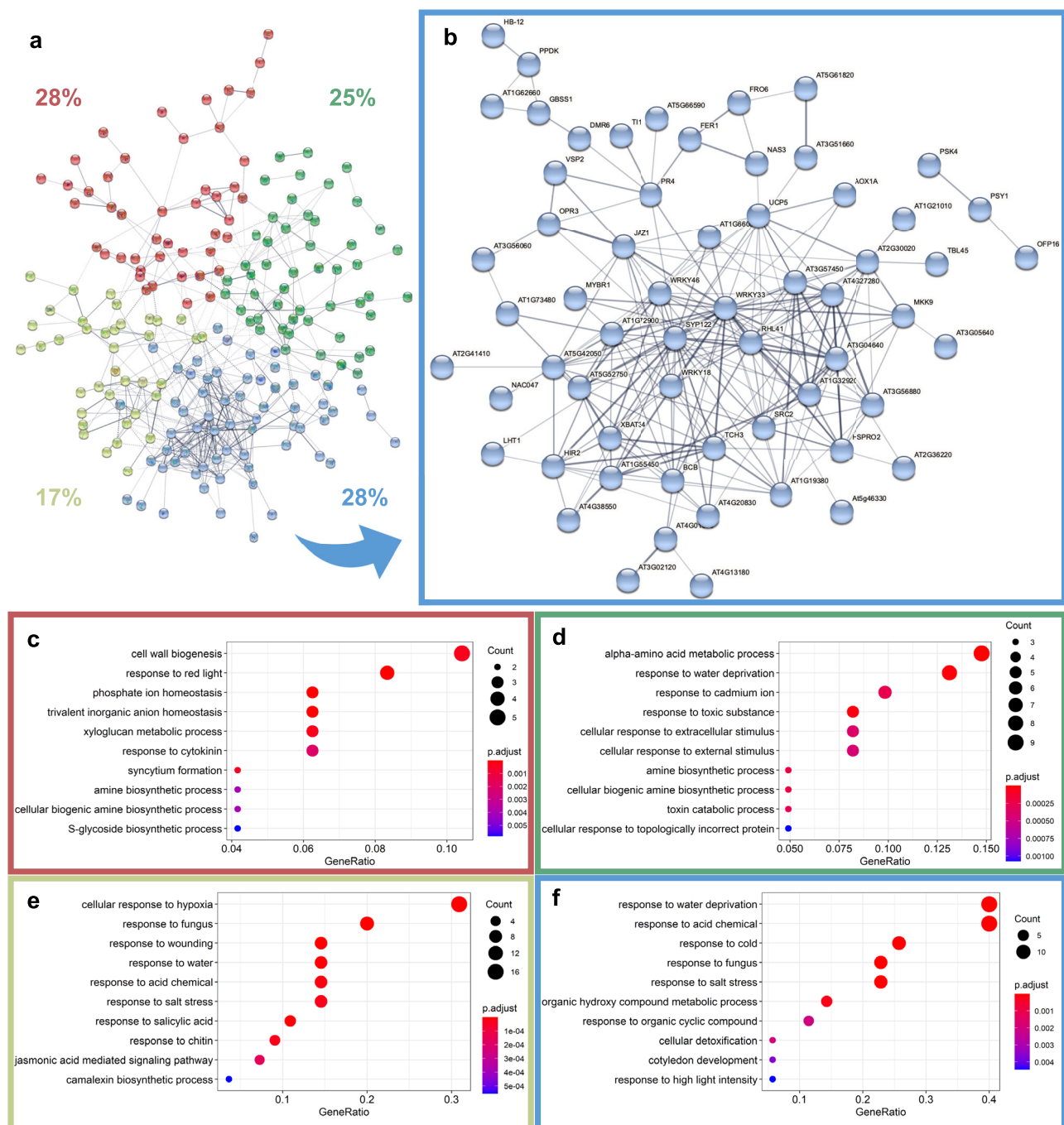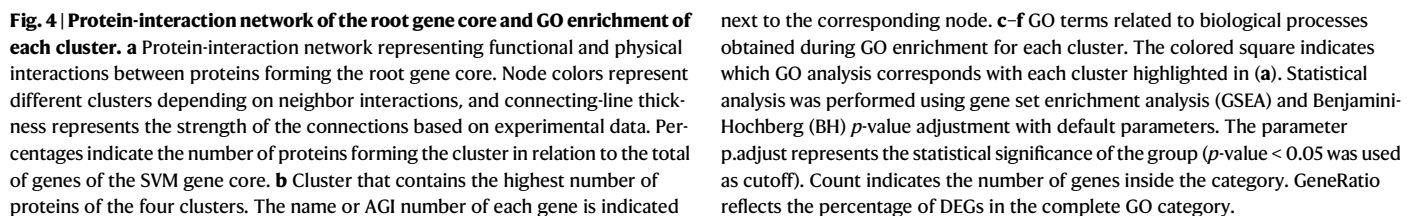
**Fig. 3 | Protein-interaction network based on the shoot gene core and GO enrichment of each cluster. a** Protein-interaction network representing functional and physical interactions between proteins encoded by the shoot gene core. Node colors represent different clusters depending on neighbor interactions; connecting-line thickness represents the strength of the connections based on experimental data. Percentages indicate the number of proteins forming the cluster in relation to the total number of genes of the SVM gene core. **b** Cluster that contains the highest number of proteins of the four clusters. The name or AGI number of each gene is indicated next to the corresponding node. **c–f** GO terms related to biological processes obtained during the GO enrichment of each cluster. The colored square indicates which GO analysis corresponds with each cluster highlighted in (**a**). Statistical analysis was performed using gene set enrichment analysis (GSEA) and Benjamini–Hochberg (BH) $p$-value adjustment with default parameters. The parameter p.adjust represents the statistical significance of the group ($p$-value < 0.05 was used as cutoff). Count indicates the number of genes inside the category. GeneRatio reflects the percentage of DEGs in the complete GO category.

(GO:1901605) and 'response to water deprivation' (GO:0009414) as main GO terms, reflecting its role in both metabolism and responses to water availability (Fig. 3d). Lastly, the smallest cluster (yellow) contained proteins involved in hypoxia responses (GO:0071456), together with proteins related to other stress responses (biotic and wounding stress) (Fig. 3e). In conclusion, it is evident that shoot-stress signaling mainly

mitigates alterations in water status and conserves cellular water homeostasis. In addition, MKK9 and WRKY transcription factors, specifically WRKY33, appear to be pivotal in the regulation of these responses.

A similar topology was obtained for the root-core interaction network (Fig. 4a). Of the four clusters, two contained the maximum

**Fig. 4 | Protein-interaction network of the root gene core and GO enrichment of each cluster. a** Protein-interaction network representing functional and physical interactions between proteins forming the root gene core. Node colors represent different clusters depending on neighbor interactions, and connecting-line thickness represents the strength of the connections based on experimental data. Percentages indicate the number of proteins forming the cluster in relation to the total of genes of the SVM gene core. **b** Cluster that contains the highest number of proteins of the four clusters. The name or AGI number of each gene is indicated next to the corresponding node. **c–f** GO terms related to biological processes obtained during GO enrichment for each cluster. The colored square indicates which GO analysis corresponds with each cluster highlighted in (**a**). Statistical analysis was performed using gene set enrichment analysis (GSEA) and Benjamini-Hochberg (BH) *p*-value adjustment with default parameters. The parameter p.adjust represents the statistical significance of the group (*p*-value < 0.05 was used as cutoff). Count indicates the number of genes inside the category. GeneRatio reflects the percentage of DEGs in the complete GO category.

number of proteins (representing 28% of the total number in the core) and, one of them (colored in blue), exhibited the highest number of connections within the network (Fig. 4b). As expected, the blue cluster showed the GO category that characterized the root core (GO:0071456: 'cellular response to hypoxia'). The second-most relevant GO term ('secondary metabolic process', GO:0019748) covered genes involved in lignan biosynthesis, such as *BCB*, and phenylpropanoid biosynthesis (*KISS ME DEADLY 1* and *4*; *KMD1/4*). In addition, defense responses seemed to play an important role in this blue cluster (GO:0031347), as

well as responses to external stimuli (GO:0009605) and to ethylene (GO:0009723), indicating a relevant role in environmental interactions (Fig. 4f). The MAPK protein MPK11 was found at a central position in the interaction network, possibly coordinating the activity of the remaining members of the blue cluster. Interestingly, the ethylene receptor ETHYLENE RESPONSE 2 (ETR2) and the downstream transcription factor ETHYLENE RESPONSE FACTOR 2 (ERF2), which is induced by ethylene[34], were also present in this cluster, corroborating a pivotal role of this hormone in the generic stress response, at least in roots.

The green cluster (Fig. 4d) was characterized by GO terms related to oxidative-stress responses (GO:0006979), cellular metabolism of amino acids (GO:0009063) and vitamins (GO:0009110; GO:0006766), and transport of inorganic compounds (GO:0006829), revealing a potential role for the maintenance of shoot central metabolism and physiology in root responses. Finally, both red and yellow clusters showed a reduced number of proteins and interaction levels compared to the previous ones (Fig. 4c, e). The yellow cluster showed GO terms involved mainly in cell-wall homeostasis and modification (GO:0009826; GO:0016049; GO:0009828; GO:0006949; GO:0010025) while the red one included GO terms water responses (GO:0009415), and hypoxia (GO:0071456), among others.

Overall, we conclude that shoot stress responses are mostly related to the maintenance of water potential and homeostasis and, secondary, to the maintenance of normoxia levels; while in roots, the opposite trend is observed. In addition, growth regulation and metabolism as well as cell-wall homeostasis are important aspects of core stress signaling in both tissues. The complete list of the genes in the SVM gene cores classified in the four clusters (blue, yellow, red and green) is found in Supplementary Data file 5.

### Role of ethylene in the SVM gene cores

Because of its key role in a multitude of stresses, and given the presence of ETR2 and ERF2 in the root core, as well as MKK9–known to play a pivotal role in the activation of MPK6 under ethylene signaling[35]–in the shoot core, the ethylene responsiveness of the genes within the SVM gene cores was investigated. To define a robust list of such genes, we combined the publicly available data of an ETHYLENE INSENSITIVE 3 (EIN3) ChIP-seq analysis[15] with a set of DEGs under early (4 h, GSE14247) and late (24 h, GSE83573) ethylene treatment, forming an ethylene-responsiveness database (Supplementary Data file 6). More than 50% of the genes in the SVM gene core for shoots and roots were ethylene responsive (Fig. 5a), underpinning the relevance of ethylene in both gene cores. Remarkably, the number of ethylene-responsive genes increased to 77% and 62.7% in the blue cluster of shoots and roots cores, respectively, further substantiating the central role of ethylene in core stress signaling.

The subgroup of genes in the blue clusters detected as ethylene-related genes were used to construct a protein-interaction network (hereafter defined as ethylene-related clusters). In the case of the shoots, ethylene-related proteins showed the same interconnected pattern as in the complete network (Figs. 3b and 5c). Moreover, WRKY33 still appeared as a central node in shoot stress signaling, together with MKK9. Notably, MKK9, along with MPK3 and MPK6, has been directly linked to both ethylene biosynthesis and signaling[35,36]. In addition, LYSINE HISTIDINE TRANSPORTER 1 (LHT1), an amino-acid importer responsible for the transport of the ethylene precursor 1-aminocyclopropane-1-carboxylate (ACC[37]); was also part of the ethylene-related shoot cluster, as well as the mitochondrial AOX1a, which connects the regulation of respiration to stress signaling in an ethylene-dependent manner[38].

The central role of WRKY33 in the shoot gene core highlights its potential in regulating stress responses. To further explore this connection, we performed an in silico study of the presence of WRKY33 binding sites, identified by the binding motif TTGACY, which was empirically determined through ChIP-seq analysis[39] (Supplementary Date file 7). Given the putative central role of WRKY33 in the gene core, it is unsurprising that 70% and 75% of genes forming the blue clusters in shoots and roots, respectively, contained WRKY33 binding motifs (Fig. 5b). Furthermore, when comparing the genes presenting this motif with the ethylene-responsive genes calculated previously, more than 54% of genes overlap between these two conditions, underscoring the close relation between ethylene signaling and the potential function of WRKY33. This connection is further supported by the analysis genes related to both ethylene biosynthesis and signaling

(Fig. 5f). More than 50% of the biosynthetic genes, including SAMS4, several ACSs (ACS2, ACS5–8, and ACS11), and all ACOs (ACO1–5), as well as key genes involved in ethylene signaling, such as the receptors ETR2 and ERS2, CTR1, EIN3–EIL1, and EBF2, along with one of the primary ethylene transcription factors ERF1, are also targeted by WRKY33. This further reinforces the relationship between WRKY33 and ethylene responses.

In roots, the number of genes constituting the interaction network was reduced (Figs. 4b and 5d). Consequently, the number of interactions was also decreased. Nevertheless, the ETR2–ERF2 module again appeared at its center, coordinating other nodes of the network. In addition, the core network also contained AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE (ARGOS), which is part of a negative feedback mechanism to attenuate ethylene responses, further highlighting the importance of coordinated ethylene signaling[40]. Several transcription factors that have been experimentally linked to specific stresses or processes, including RAP2.6L/ERF113 (wounding), NAC047 (partial submergence), and NAC6 (leaf senescence), also appeared in this cluster, suggesting a more general function for all.

Lastly, to further demonstrate the role of ethylene in the regulation of both SVM gene cores, we studied the transcriptional responses of key genes from the stress gene cores under stress conditions in wild-type Col-0 and in the ethylene-insensitive ein2-5 mutant (Fig. 5e). We observed that most of the studied genes exhibited clear down-regulation in the ein2-5 mutant compared to Col-0 under control conditions, empirically confirming that ethylene signaling is required for the transcriptional activation of these genes, particularly at early time points (1 h). The exception was WRKY33, which remained partially unaltered, especially under cold and heat conditions, indicating functioning upstream ethylene signaling during the regulation of stress responses.

The high degree of interconnection within the blue clusters of both SVM gene cores, coupled with the presence of numerous genes related to ethylene responses, underscores the central role of ethylene in regulating stress responses in both shoots and roots. This is further supported by the transcriptional data from the ethylene-insensitive ein2-5 mutant, which shows a general down-regulation of genes in both gene cores, particularly at early time points, confirming the requirement of ethylene signaling for the activation of these stress-related genes.

### A central role for EXPANSINS, AP2/ERFs, WRKYs, and MAPKs in the SVM gene cores

Certain gene families were identified as crucial players in the SVM gene cores, such as EXPs, AP2/ERFs, WRKYs and MAPKs. In addition, novel gene families were also identified by the SVM Clustering algorithm, including USPs. To provide a complete and detailed map of the function of these gene families in stress responses, we investigated the transcriptional alterations of all their family members under all conditions in our meta-analysis (Fig. 6; Supplementary Data file 8). In addition, as a second validation, experimentally validated data about specific members of the selected families corroborated our results (Supplementary Data file 9). Select families are covered in the next section; with further details are presented in the Supplementary Information file.

### EXPANSIN (EXP) superfamily

EXPs were detected as the main enriched gene family in the root core (Fig. S4). They enable cell expansion and increase cell-wall flexibility[41,42]. The EXP superfamily is divided into four groups: EXPA, EXPB, expansins-like A (EXPLA) and EXPLB. In our DEG database, both EXPA and EXPB subfamilies are predominantly down-regulated in several stress conditions, with tissue-specific patterns (Fig. 6a; Supplementary Data file 8). Three members of EXPA (EXPA1, EXPA8 and EXPA15) and one EXPB (EXPB3) were part of SVM stress cores. While

**a**

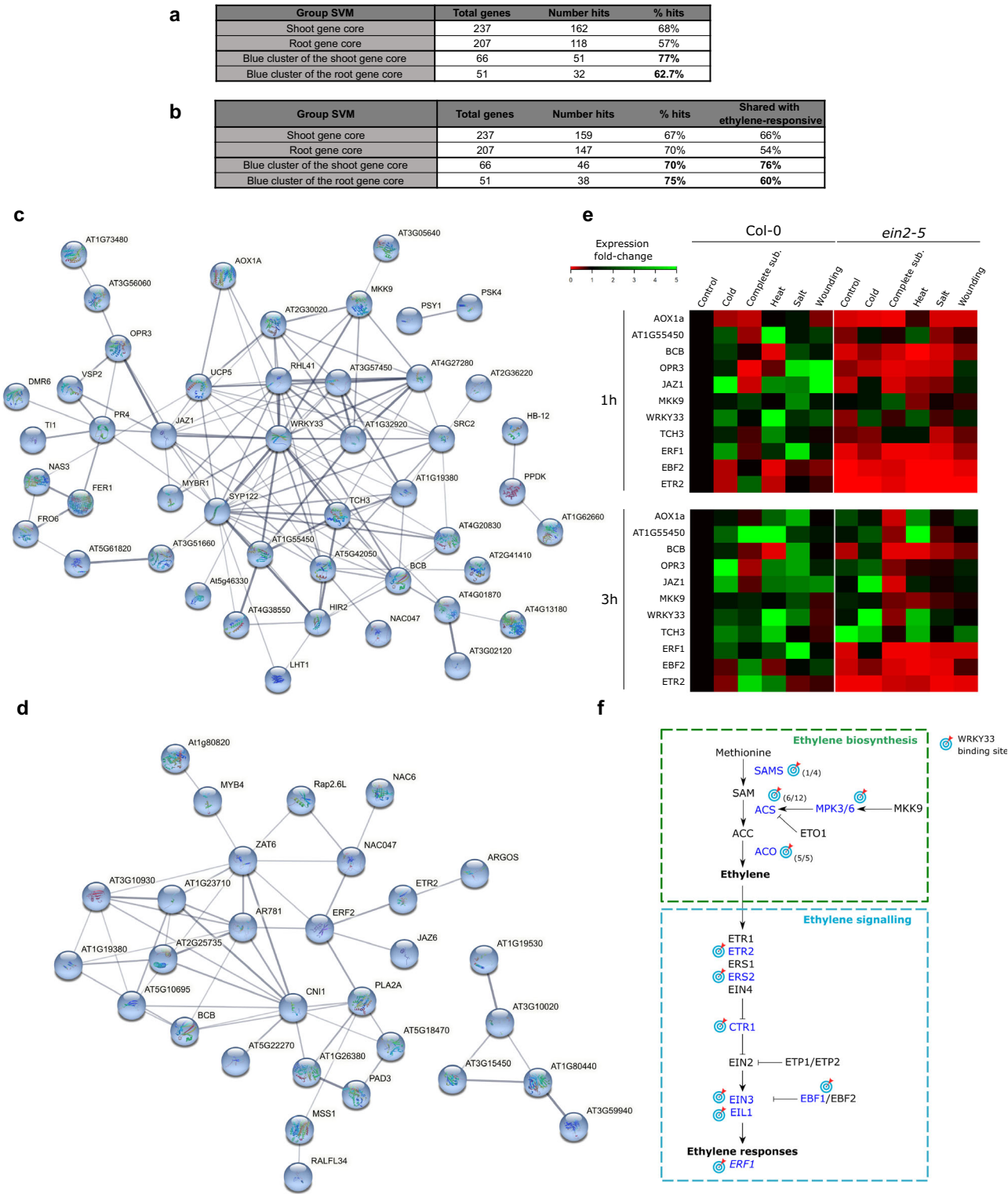| Group SVM | Total genes | Number hits | % hits |
|---|---|---|---|
| Shoot gene core | 237 | 162 | 68% |
| Root gene core | 207 | 118 | 57% |
| Blue cluster of the shoot gene core | 66 | 51 | **77%** |
| Blue cluster of the root gene core | 51 | 32 | **62.7%** |

**b**

| Group SVM | Total genes | Number hits | % hits | Shared with ethylene-responsive |
|---|---|---|---|---|
| Shoot gene core | 237 | 159 | 67% | 66% |
| Root gene core | 207 | 147 | 70% | 54% |
| Blue cluster of the shoot gene core | 66 | 46 | **70%** | **76%** |
| Blue cluster of the root gene core | 51 | 38 | **75%** | **60%** |



**Fig. 5 | Ethylene-related genes extracted from the shoot and root gene cores. a** Number and percentage of genes of each gene core present in the ethylene-responsiveness database. Number hits refers to the number of genes from the core present in the ethylene-responsiveness database (Supplementary Data file 6), while the percentage indicates the number of hits compared to the total number of genes present in the complete core. **b** Number and percentage of genes of each gene core with putative WRKY33 binding sites. Number hits refers to the number of genes from the core that present WRKY33 binding motifs, while the percentage indicates the number of hits compared to the total number of genes present in the complete core. The last column shows the overlap between the presence of WRKY33 binding sites and ethylene responsiveness. **c, d** STRING interaction network of the ethylene-response genes from the blue cluster of the shoot **c** and root **d** gene cores. The connection between different nodes indicates protein–protein association. Thickness of connecting lines represents the strength of the connections based on experimental data. **e** Transcriptional responses of key genes of the SVM gene cores under stress conditions in WT (Col-0) and the ethylene-insensitive mutant *ein2-5* after 1 and 3 h of stress exposure. Fold-change expression is calculated relative to the expression levels of Col-0 control samples at both 1 and 3 h. The color scale indicates downregulation (0–1; red), no change (1; black), and upregulation (1–5; green). Source data are provided as a Source Data file. **f** Map of predicted WRKY33 binding-site motifs in ethylene-related genes.
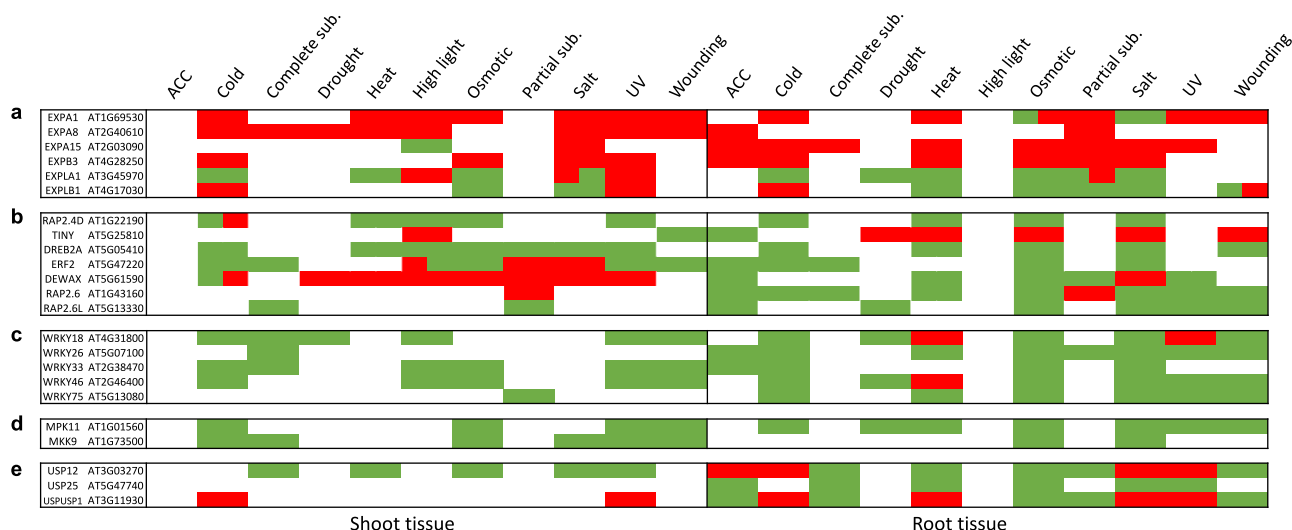
**Fig. 6 | Summary of the presence of different key gene families of the abiotic stress core in the different conditions studied.** Shown are members of the expansin (*EXP*) (**a**), ethylene-response factors (*ERF*) (**b**), *WRKYs* (**c**), mitogen-activated protein kinases (*MAPK*) (**d**) and universal stress proteins (*USP*) (**e**) families across the conditions analyzed. Green and red represent up- or down-regulation in each condition. Bicolored cells indicate that the behavior of the gene is variable depending on early or late responses (for complete information on all members of the family, see Supplementary File 7).

transcriptional alterations of *EXPA8*, *EXPA15* and *EXPB3* were observed for roots under certain conditions, *EXPA1* showed transcriptional alterations in all tissues and timepoints in 7 out of 10 studied stresses. These findings highlight the importance of the tissue specificity of *EXPs* in stress signaling and of *EXPA1* as a main stress regulator. Interestingly, though most *EXPs* were down-regulated in response to stress, the *EXPLA* and *EXPLB* groups (with two genes present in the SVM gene cores, *EXPLA1* and *EXPLB1*) were up-regulated under several stress conditions, suggesting a potential positive role for these subfamilies in stress responses.

## Ethylene response factor (ERF) family

The ERF family of transcription factors is considered to be crucial in both growth and defense responses[43]. ERFs are part of the AP2/ERF superfamily, comprising 147 member divided in three sub-families: 18 *AP2s*, 122 *ERFs* and six *RELATED TO ABSCISIC ACID INSENSITIVE 3/VIVIPAROUS 1* (*RAVs*), as well as a not-yet-classified gene (AT4G13040)[44]. Whereas few to no transcriptional alterations were found for the *AP2* family, the *RAV* and *ERF* subfamilies revealed to be highly affected by stress, often with very distinct expression patterns (Fig. 6b; Supplementary Data file 8).

Almost all *ERF* subgroups displayed substantial transcriptional changes under the studied stress conditions (Supplementary Data file 8). Seven *ERFs* were classified as members of the SVM gene core, six in the root core (*ERF2*, *TINY* (*ERFO40*), *DREB2A* (*ERFO45*), *DEWAX* (*ERF107*), *RAP2.6* (*ERF108*), and *RAP2.6L* (*ERF113*)), and only one (*RAP2.4D* (*ERFO58*)) in the shoot core (Fig. 6b). Some of these *ERFs* showed tissue specificity. For instance, *RAP2.6/ERF108* and *RAP2.6L/ERF113* showed a similar transcriptional pattern in root tissue, yet distinct patterns in shoots. While *RAP2.6/ERF108* was down-regulated under the early exposure to partial submergence, *RAP2.6L/ERF113* was up-regulated under both types of submergence. In contrast, *ERF2* and *DREB2A/ERFO45* were broadly expressed in both roots and shoots in most conditions. Since the direct relationship between the members of the *ERF* family and ethylene is not always clear, we investigated their responsiveness to ACC as well as to ethylene (Supplementary Data files 6 and 8). *ERF2*, *RAP2.6/ERF108* and *RAP2.6L/ERF113* were up-regulated by ACC, as well as present in the ethylene-responsiveness dataset. The other *ERFs* were either only up-regulated by ACC (*TINY/ERFO40* and *DEWAX/ERF107*), only ethylene responsive (*DREB2A/ERFO45*), or not detected under ACC, nor ethylene treatment (*RAP2.4D/ERFO58*).

## Mitogen-activated protein kinase (MAPK) superfamily

MAPKs are important signaling proteins in many intracellular responses to developmental, physiological and/or environmental stimuli[45]. MAPK signaling cascades are typically characterized by a sequence of phosphorylation and activation events along three levels comprising members of mitogen-activated protein kinase kinase kinases (MAPKKK, MKKK or MEKK), mitogen-activated protein kinase kinases (MAPKK or MKK), and mitogen-activated protein kinases (MAPK or MPK). In *A. thaliana* 69 MAPKKKs, 10 MAPKKs and 20 MAPKs have been described[45].

Four out of the ten members of the *MAPKK* group were unresponsive to any of the conditions studied (Supplementary Data file 8). However, of the stress-responsive *MAPKKs*, *MKK9* stood out, being up-regulated by six different conditions, mainly in shoot tissue (Fig. 6d). High salinity and osmotic stress elevated *MKK9* transcription in roots. Not surprisingly, MKK9 appeared as part of the shoot core (Fig. 3b), indicative for its central regulatory role in abiotic stress responses.

Out of 20 *MAPKs*, only six did not show a transcriptional effect under any of the stress conditions. Conversely, *MPK11*, *MPK3*, *MPK5* and *MPK19* stood out given transcriptional alteration under seven, six, five and four different stress conditions, respectively (Supplementary Data file 8). However, only *MPK11* was retained by SVM Clustering as a stress core gene (Fig. 6d). Cold, osmotic and UV stresses up-regulated *MPK11* expression in all tissues, while salt induction was root-specific. Wounding (shoots and roots), drought (roots) and heat (roots) induced *MPK11* transcription predominantly at early timepoints, implying its relevance specifically during the initial stages of the stress response.

## Universal stress protein (USP) superfamily

USPs are proteins involved in, as their name suggests, a broad range of metabolic processes related to stress, such as nutrient starvation, heat shock and oxidative stress[46]. Nevertheless, their specific roles and molecular mechanisms remain largely unknown. In *A. thaliana*, 41 genes encode for USP proteins, cataloged according to domain organization. From our analysis, the *USP* family and the single gene belonging to the double USP domain group (*USPUSP*) appear to be involved in all stress conditions (Supplementary Data file 8).

*USP12*, *USP25*, and *USPUSP1* were detected as part of the root gene core. *USP12*, characterized as a gene involved in ROS modulation in anoxia conditions, is up-regulated in submergence conditions, but also in heat and osmotic conditions in both tissues (Fig. 6e; Supplementary Data file 9). *USP25* and *USPUSP1* appear to be tissue-specific players in the general stress response, being only expressed in roots. It is evident that the function of USPs deserves more scrutiny, given their highly specific expression patterns as well as the central role of certain family members in general stress signaling.

### Experimental validation: the role of *EXPAs, WRKY33, MKK9,* and *LHT1* in general stress responses

As part of a third validation supporting the role of members of the gene cores as central stress regulators, we first studied the transcriptional alterations of three members of the *EXPA* family (*EXPA1*, *EXPA10*, and *EXPA14*), represented in the root core, and *EXPA1* as a part of both gene cores, using transgenic translational-reporter lines (*pEXPA1::EXPA1–mCherry* (Fig. 7a), *pEXPA10::EXPA10–mCherry* (Fig. 7b) and *pEXPA14::EXPA14–mCherry* (Fig. 7c)[41]. We exposed the three lines to cold, salt, and osmotic stress for 1 h. The expression patterns in control conditions corresponded with the patterns described by Samalova et al. (2023)[41]. Short-term exposure to stress saw changes in mCherry intensity levels, indicating a change in EXPA abundance in the studied tissue. Specifically, salt treatment dramatically increased the levels of EXPA1–mCherry and EXPA10–mCherry, while decreasing the level of EXPA14–mCherry (Fig. 7d), corroborating the expression changes after the equivalent treatment obtained in our meta-analysis (Fig. 7e). Osmotic treatment modestly increased the levels of EXPA1–mCherry and EXPA10–mCherry, without affecting the level of EXPA14–mCherry. Upon cold treatment, none of the levels of the studied EXPAs differed from those of the control samples, again confirming the expression data. In conclusion, data from the translational reporter lines matched with the transcriptional data derived from the meta-analysis, experimentally validating the robustness of our analysis, and positioning *EXPAs* as key regulators of multiple stress responses as evidenced by their relevance in both shoot and root gene cores.

Secondly, we investigated the function of both WRKY33 and MKK9, putative key regulators in the shoot core (Fig. 3b). We assessed stress tolerance by comparing alterations in rosette growth between the loss-of-function mutants *wrky33-2* and *mkk9-1* and the wild-type Col-0 (Fig. 8a–d). To consider a representative set of stress conditions, we selected five conditions studied in our analysis, covering each of the clusters obtained by HCA: complete submergence, cold, heat, salt, and wounding (Fig. 1).

The *wrky33-2* mutant exhibited notable phenotypic differences when exposed to various stress conditions (Fig. 8a and b). Under cold, heat, and wounding conditions, *wrky33-2* was hypersensitive, evidenced by reductions in both rosette area (Fig. 8a and b) and relative biomass compared to control conditions (Fig. S5a). In contrast, under complete submergence and salt stress, *wrky33-2* mutants responded differently. While Col-0 plants experienced a marked decrease in both rosette size and relative biomass, *wrky33-2* mutants displayed either non-significant changes or even improvements in these parameters. On the other hand, *mkk9-1* mutants behaved differently from the wild type in all tested conditions and displayed an increased resistance to the vast majority of stress conditions. *mkk9-1* mutants had increased tolerance under cold, heat, and wounding stress, with rosettes statistically significantly larger than those of treated Col-0 plants under the same conditions (Fig. 8c and d). However, under salt stress, *mkk9-1* exhibited hypersensitivity, showing a more pronounced reduction in rosette size compared to Col-0. Lastly, under complete submergence, *mkk9-1* rosettes showed enhanced growth, being larger than both treated and untreated Col-0 plants. These observations were further supported by the relative biomass measurements, which followed the same trends as rosette size (Fig. S5b).

Lastly, we analyzed the response of the *lht1-5* loss-of-function mutant to the different stress treatments (Fig. 8e and f). Both its presence as a member of the shoot core, as well as its function in amino acid and ACC transport, suggest that LHT1 could act as a vital component of general stress signaling. Indeed, plants that lack functional LHT1 are hypersensitive to cold, complete submergence, heat, and salt stress. In contrast, *lht1-5* plants were less affected than Col-0 plants upon wounding. Similar to the above-mentioned results, these findings were corroborated by corresponding changes in the relative biomass compared to the control condition measure (Fig. S5c).

To gain deeper insights into the roles of WRKY33 and MKK9 in the response to the tested stress conditions and the connection with ethylene, its production was analyzed under the same treatments and compared to wild-type (Col-0) (Fig. S6). Ethylene emanation increased in response to all tested abiotic stresses except cold stress, wherein a decrease was observed compared to control conditions. In WRKY33-deficient plants, ethylene levels were statistically significantly higher under control conditions and in response to heat, wounding, and complete-submergence stress, whereas slight reductions were observed under cold and salt stress. Similarly, *mkk9-1* mutants displayed hypersensitivity to heat, complete submergence, and cold stress, as reflected by altered ethylene levels, although these did not reach the higher levels produced by *wrky33-2*. These findings highlight the regulatory influence of WRKY33 and MKK9 on ethylene production and underscore their critical roles in stress signaling mechanisms, intertwined with the ethylene pathway.

Further investigation of the transcriptional alterations of key genes from the SVM gene cores (*WRKY33*, *MKK9*, *BCB*, *AOX1a*, *AT1G55450*, *JAZ1*, *OPR3*, and *TCH3*) and ethylene-related genes (*ETR2*, *ERF1*, and *EBF2*) was conducted by real-time quantitative PCR analysis under the same set of stresses (Fig. S7). In the *wrky33-2* mutant, a marked down-regulation of most assayed genes was seen under stress conditions, wherein the mutant exhibited hypersensitivity, particularly cold and wounding. However, in complete submergence, whereby *wrky33-2* mutants performed better in terms of rosette size and relative biomass, some genes, such as *BCB*, *ERF1*, and *ETR2*, were up-regulated compared to Col-0. Conversely, transcriptional changes in the *mkk9-1* mutant showed the opposite trend. Under stress conditions when the mutant displayed resistance (cold, complete submergence, heat, and wounding), a substantial number of genes were up-regulated relative to Col-0. In contrast, under salt stress, wherein the *mkk9-1* mutant exhibited hypersensitivity, most assayed genes were statistically significantly down-regulated. These results further support contrasting regulatory roles for WRKY33 and MKK9 in stress-specific transcriptional responses.

The results of these analyses emphasize the overarching role of ethylene in regulating plant stress responses, as well as the pivotal contributions of WRKY33 and MKK9 in modulating the transcriptional activity of the stress gene core across various conditions. These findings not only validate the biological significance of the SVM gene core but also confirm its robustness, reproducibility, and suitability as a foundational framework for understanding general stress regulatory networks in plants.

## Discussion

### Machine learning as a tool for rapid identification of the stress gene core: strengths and limitations

Machine learning approaches represent a powerful tool for data analysis. Yet, the design, scientific question, and biological significance need to be carefully addressed to avoid incorrect interpretations or lack of reproducibility of the generated output[26]. In this study, we aimed to identify an abiotic stress gene core, a critical step toward a deeper understanding of the genetic basis of stress responses in crop plants under the increasing, multifactorial pressures due to climate change. This overarching objective guided the development and
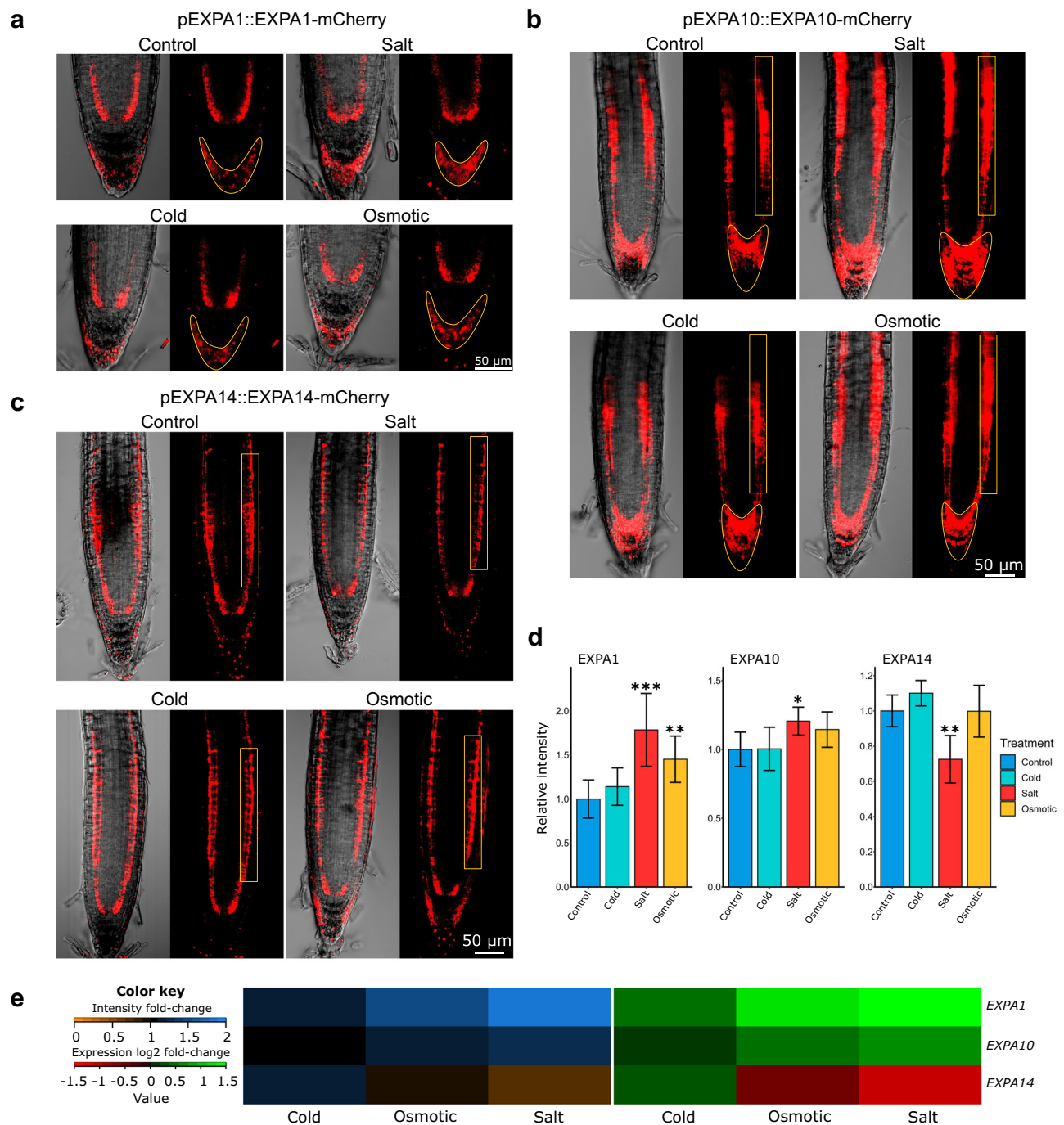
**Fig. 7 | Experimental validation of the transcriptional alterations of *EXPAs*. a–c** mCherry fluorescence intensity for *EXPA_pro:EXPA-mCherry* reporter lines for EXPA1 (**a**), EXPA10 (**b**), and EXPA14 (**c**) in the lateral root cap for EXPA1, in the root tip and a region of the cortex for EXPA10 and in a region of the cortex for EXPA14. The measured region of interest (ROI) is framed with a yellow box. Scale = 50 μm. **d** Fluorescence-intensity changes in relation to control conditions (*n* for EXPA1 = 34 (control), 17 (cold), 15 (salt), 20 (osmotic); for EXPA10 = 38 (control), 20 (cold), 20 (salt), 19 (osmotic); for EXPA14 = 28 (control), 13 (cold), 16 (salt), 12 (osmotic)). Data are presented as mean values ± SD. A *p*-value of *p* < 0.05 was assumed as statistically significant (*<0.05; **<0.01; ***<0.001). **e** Heatmap showing intensity fold-changes represented as relative marker-gene intensity compared to control conditions (ranging from 0 (orange) to 2 (blue)) and expression fold changes based on transcriptomic analyses (represented as the log2-fold change compared to control conditions, ranging from –1.5 (red) to 1.5 (green)) for each stress condition. Data used for constructing the heatmaps are provided in the Data Source File. Shapiro's test and Levene's test were used to assess normality and variance of each dataset. To compare the intensity of the different treatments with the control samples, Student's *T*-test and Wilcoxon rank sum exact test were used for parametric and non-parametric testing, respectively. Complete statistical analysis is available in Supplementary Data file 15.

application of our pipeline, encompassing a solid methodological foundation to generate reproducible results aligned with biological research questions relevant to the societal context.

To ensure the robustness of our analysis, we first applied multiple quality metrics, including the sum-of-squared error (SSE) and clustering entropy (Supplementary Data file 10). Additionally, iterative re-classifications showed high stability, with over 95% of genes remaining unchanged after the first iteration, further supporting the reliability of our results (Supplementary Data file 11). To assess the discriminatory power of the analysis, we conducted
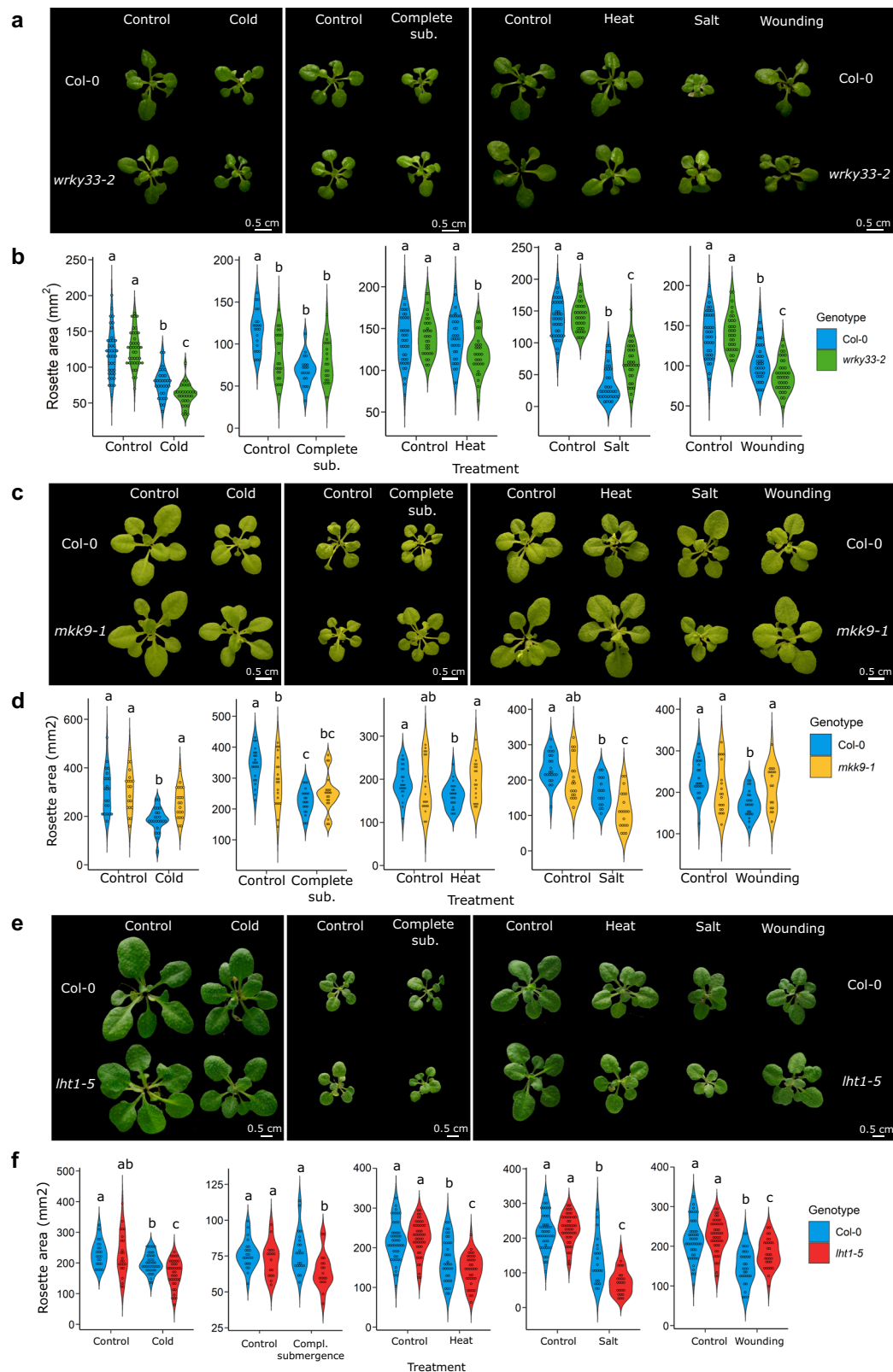
a supplementary analysis by reducing the sets of genes under analysis (Supplementary Data file 12). To test the biological significance of the obtained output, we included a triple validation. Firstly, we used experimentally determined stress markers to assess the quality of the generated DEG libraries (Supplementary Data file 3). Secondly, expression patterns of specific members of gene families with well-characterized transcriptional behavior

under different stress conditions were used as an additional level of validation (Supplementary Data file 9). Finally, to validate the genes forming the proposed SVM gene core, we empirically analyzed the effect of stress exposure on alterations in transcription (Fig. 7a–d) and the function of three key genes in physiologically representative stress conditions, one of which was previously not known to be linked to abiotic stress conditions

**Fig. 8 | Experimental validation of the role of *WRKY33*, *MKK9*, and *LHT1* as central hubs in stress responses.** Rosette area of Col-0 and *wrky33-2* (**a** and **b**), *mkk9-1* (**c** and **d**) *lht1-5* (**e** and **f**) plantlets. Rosette area was measured after 4 days of cold (4 °C), complete submergence, heat (42 °C), salt stress (100 mM NaCl), and wounding, and after 5 days in recovery conditions for *wrky33-2* plants and 10 days for *mkk9-1* and *lht1-5* (*n* > 20 plants per sample; specific *n* values are specified in Supplementary Data file 15). Scale = 0.5 cm. Lighting conditions differed between the sets, and exposure of panel (**e**) was increased equally to 1.5 to improve the quality of the images. Source data are provided as a Source Data file. A *p*-value <

0.05 was assumed as statistically significant and different letters represent statistical differences between samples (specific *p*-values are specified in Supplementary Data file 15). Shapiro's test and Levene's test were used to assess normality and variance of each dataset. To compare rosette areas, a two-way ANOVA followed by post-hoc Tukey's test were used as parametric tests, while a Kruskal–Wallis rank sum test followed by Dunn's Multiple Comparison test were used as non-parametric tests. Complete statistical analysis is available in Supplementary Data file 15.

(Fig. 8a–f). Altogether, this provides a solid basis to put confidence in the previously uncharacterized genes that form part of the stress gene core, including *USPs*, offering strong evidence of their putative biological function. In addition, it strengthens the validity of our methodology to study complex processes. The proof of concept presented in this study could be extended to, for example, the determination of the central players in biotic-stress responses, as well as to more deeply understand the differences between responses induced by necrotrophic, biotrophic, and hemibiotrophic pathogens.

Building on our validation framework, this study also offers unique contributions in the landscape of machine-learning-based meta-analyses of plant stress responses (Supplementary Data file 13). Unlike prior studies that typically target gene sets specific to individual stress conditions, we focused on identifying genes that participate in responses shared across all abiotic stresses. Our approach, using an unsupervised machine-learning algorithm, avoids the need for pre-existing training data. This design minimizes potential biases toward specific stress types, providing a more objective view of stress-responsive gene networks while also avoiding one of the main limitations of supervised learning, namely data leakage. Data leakage occurs when information from the test set is incorporated into the training process, leading to circular reasoning and overfitting[26]. In our approach, this issue is inherently avoided, as no training step is required.

Another key difference is that our method directly identifies shared genes without relying on preliminary differential-expression analyses, which helps further reduce biases in gene selection. We also included a broad range of conditions, covering ten abiotic stresses and ACC treatment, representing a broader range of conditions than similar studies, such as Shaik and Ramakrishna (2013)[47] in *O. sativa* (seven conditions) and Ma et al. (2014)[48] in Arabidopsis (six conditions). By analyzing 500 transcriptomes related to stress, we work with a larger dataset than previous meta-analyses, enhancing the robustness of our identified gene core. This approach, therefore, provides a comprehensive and innovative framework for discovering genes that are key to abiotic stress responses across diverse conditions.

### Clustering patterns of stress responses across tissues and timepoints

The HCA revealed interesting trends, indicating shared responses between stress conditions but varying dynamics depending on tissue type and exposure time (Fig. 1). On the one hand, in roots, ACC treatment clusters with both partial and complete submergence at early and late timepoints, likely due to the high levels of ACC that accumulate in submerged plants[49]. This connection is further supported by the ethylene-response gene *ETR2*, which is shared across these three conditions at both timepoints. In shoots, this similarity holds at early timepoints, while at later timepoints, partial submergence clusters with salt and osmotic stress, whereas complete submergence with UV, wounding, and drought (Figs. 1 and S3). These changes are also mirrored in the GO analysis within each cluster (Fig. S2). In the early stages, partial and complete submergence are enriched for terms related to water responses. However, at later stages, complete submergence shows enrichment for hypoxia

responses, often associated with oxidative stress responses typical also of UV stress[50], while partial submergence clusters with osmotic and salt stress, wherein dehydration responses play a crucial role[51].

On the other hand, during the early responses in shoots, drought, wounding, salt, osmotic, and cold stress cluster together. Although these stresses have distinct characteristics, their molecular mechanisms to mitigate water stress likely overlap. For instance, both drought and cold responses involve the expression of molecular chaperones like *HEAT SHOCK PROTEINS* (*HSPs*) and *LATE EMBRYOGENESIS ABUNDANT* (*LEA*) genes[52]. Representative members of these families, such as *LEA14*, *HSP70*, and *HSP90*.1, are highlighted as shared genes among all these conditions. Additionally, jasmonic-acid-related genes, including *JAZ1* and *MYC2*, are found in the same intersection, linking drought and wounding responses with the rest of the cluster[53], with these two stresses consistently co-clustering in all studied groups (Figs. 1, S2 and 3). However, under prolonged stress, the role of water stress seems to shift, and while drought and wounding continue to cluster together, they now group with complete submergence and UV. This change may be explained by the accumulation of reactive oxygen species (ROS) triggered by wounding and drought conditions, as oxidative stress becomes more prominent[54–56].

Finally, other relevant clusters include temperature responses (heat and cold) together with osmotic stresses (salt and osmotic stress) mainly in root tissue (Figs. 1 and S3). It is not surprising to see heat stress grouping with high salt and osmotic pressures, as elevated temperatures increase water evaporation, which raises osmotic pressure in root tissues[56]). However, the inclusion of cold stress in this cluster is less intuitive. In early time points, cold clusters with drought and wounding, but over time, it aligns with heat, salt, and osmotic stress. Prolonged cold conditions are known to trigger responses that help maintain cytosolic osmolarity, preventing ice-crystal formation[57]. These responses involve the expression of osmoprotectants like galactinol, which is associated with both drought and cold responses[58]. *GALACTINOL SYNTHASE 4* appears in the intersection of this cluster, along with dehydrin family members such as *COR47*, reflecting the shared osmoprotective mechanisms among these stress conditions.

### To stress stimuli and beyond: the physiological role of SVM gene cores in stress responses

Most environmental changes are first sensed by displacement of the cell wall–membrane interface[31], and plant genomes have evolved mechanisms to monitor and ensure membrane integrity and cell-wall rigidity[59,60]. Representatives of these mechanisms were found in the SVM gene cores. *EXPANSINs*, a gene family extensively studied for its implication in cell-wall loosening and cell growth[41], was the main family enriched in the root core (Fig. S4) and, moreover, *EXPA1* was included in the 19 genes shared between root and shoot cores (Fig. 6a). This supports the contention that cell-wall loosening and remodeling, apart from its role in normal growth, is also crucial for the adaptation to environmental stresses, especially in roots. Typically, stresses that lead to ROS production and loss of water alter the expression of EXPs[42]. However, the precise mechanism of action of EXPs in stress mitigation remains unclear. To provide empirical validation of our meta-analysis as a third validation layer, we compared the effects of short-term stress

exposure on EXPA1, EXPA10, and EXPA14 accumulation with the transcriptional changes detected in our analysis revealing robust parallelism between the two datasets (Fig. 7a–e). *EXPA* and *EXPB* are well-characterized groups and were down-regulated in the majority of stress responses (Fig. 6a, Supplementary Data file 8), supporting the importance of cell wall and membrane rigidity in stress responses. Among *EXPAs* and *EXPBs*, *EXPA1* stands out as a pivotal gene in stress responses, while the other members seem to have a more specific role. The other two *EXP* groups, and especially *EXPLA1* and *EXPLB1* showed an opposite trend, being mainly up-regulated. Since the functions of both subgroups have not been elucidated to date and our data suggest that they could play opposite roles compared to *EXPA* and *EXPB*, it will be worthwhile to further characterize their roles.

Alterations at the level of the cell membrane often serve as initiators of stress signaling, with the activation of membrane-anchored $Ca^{2+}$ channels as one of the most relevant stress-response inducers[61]. Calcium influx and signaling are implicated in drought, cold, salt, osmotic, hypoxia, and flooding responses, highlighting their importance in most stresses[62,63]. Nevertheless, little is known about the specific genes controlling this signaling network[62]. One of the gene families in the shoot core corresponded to ion exchangers, specifically cation/$Ca^{2+}$ exchangers (Fig. S4). On the one hand, *CALCIUM EXCHANGER 1* (*CCX1*), is up-regulated in UV, complete submergence, wounding, osmotic, and salt stresses during the early exposure to stress in shoots. On the other hand, its paralog *CCX2* is present in the root gene core, up-regulated in drought, cold, heat, osmotic, and salt stresses, and also during early exposure. Both are linked to ROS accumulation, while a *CCX2* loss-of-function mutant is hypersensitive to salt stress[64,65]. The presence of *CCX1* and *CCX2* in the shoot and root cores, respectively, highlights their potential involvement during the early stress responses. In addition, they could be excellent targets for the study of $Ca^{2+}$ channels in the systemic communication between root and shoot.

In addition to $Ca^{2+}$ influxes, interpretation of $Ca^{2+}$ waves by Calmodulins (CaM) and calmodulin-like (CML) proteins is required for proper stress signaling and inter-organ communication[66]. *TCH2/CML24* is vital for heavy-metal tolerance in *A. thaliana* owing to its interaction with WRKY46 (present in the shoot gene core; Fig. 3a)[67]. Another CML, TCH3/CML12, plays a central role in the interaction network of the shoot core, interacting with the central protein WRKY33. Moreover, the fact that these genes have been related to other stresses apart from the ones included in our analysis (such as heavy-metal tolerance), endorses the general nature of this core in stress signaling. On the other hand, it can verify genes that play a central role in the interaction network (*TCH3*–*WRKY33*) as key players in general stress-response coordination.

Following the activation of $Ca^{2+}$ waves, MAPK signaling cascades are highlighted as one of the main coordinators, facilitating downstream signaling processes[45]. Nevertheless, the study of MAPKs is hindered by the complexity of MAPK signaling cascades and their regulation[68]. Certain MAPKs arose in our analysis, such as MKK9, whereas others, such as MPK6—shown to be part of the senescence-related module MKK9–MPK6 but mostly controlled by post-translational regulation[68]—did not appear. Our study revealed an interesting and particular role of MKK9 in the shoot interaction network, binding to the central WRKY33 and TCH3 (Fig. 3b). Functional analysis of a loss-of-function *mkk9-1* mutant supported its role in the coordination of stress responses (Fig. 8c and d). In addition, our results corroborated previous observations, whereby the *mkk9-1* mutant exhibited hypersensitivity to high salinity[36]. This validation not only supports the potential role of the SVM gene cores in stress responses but also provides strong evidence for the role of MKK9 as a part of a hub in stress responses, giving additional insights into its biological function.

In many stress-signaling cascades, hormones are activated after initial stress sensing and signal relay. Ethylene is such a key stress

hormone, with a negative effect on cold tolerance[69]. Conversely, ethylene positively influences the survival rate and tolerance to flooding conditions, mediated by ERF-VIIs[70], and it also improves salt tolerance through *ERF1* induction[71]. Here, we provide evidence for broad transcriptional alterations of *ERFs* in multiple stress conditions. Especially *ERF2*, up-regulated in five and six conditions in roots and shoots respectively, was identified as a central player in the root core interaction network (Fig. 6b). In addition, the ERF2–ETR2 module demonstrated that one of the main root core functions is related to ethylene responses, supporting the pivotal role of the hormone in root stress responses (Fig. 4f).

The amino-acid transporter LHT1 was also part of the shoot stress core (Fig. 3b). LHT1 was previously shown to transport ACC in *Arabidopsis*, and *lht1-5* mutants display an early-senescence phenotype[37]. Here, we show that loss of *LHT1* leads to an altered tolerance to all of the tested stresses (Fig. 8e and f). These results indicate that LHT1 plays a prime role in abiotic stress responses in addition to its previously reported function during pathogen infection[72]. Though LHT1 clearly appears to act as an important node that simultaneously regulates cellular ACC availability—and thus ethylene—as well as levels of other (non)-proteinogenic amino acids, more work on its precise mode of action is needed. Besides this direct link, using our ethylene-responsiveness database we were able to detect the relevance of ethylene in the regulation of more than 50% of the genes in both gene cores (Fig. 5a). In addition, some of the abovementioned key genes (*CCX1*, *TCH3*, *WRKY33* and *MKK9*) were also related to ethylene responses, further supporting the pivotal role of ethylene in expression of these central core genes (Fig. S8).

Ethylene is a primary hormone mediating stress responses across various conditions, and amino-acid transporters also play an essential, complementary role. Such transporters, which are enriched in the shoot SVM gene core, appear to be crucial for contributing to the proper induction of stress responses across tissues (Fig. S4). Amino-acid transporters are key links between abiotic and biotic stress tolerance, with LHT1 playing a critical role[72]. Specifically, LHT1-mediated increases in amino-acid levels, such as L-proline (L-Pro) are pivotal in biotic-stress responses[72,73]. Therefore, it is not surprising that these channels play a significant role in the stress gene cores. Notably, LHT1, involved in the transport of L-Pro as well as ACC[37], appears in the shoot SVM gene core within the blue cluster. This supports the potential role of LHT1 as a stress response facilitator through amino-acid transport in leaves, underscoring the broader importance of amino-acid metabolism in stress responses. This is further corroborated by the observed stress hypersensitivity in the *lht1-5* loss-of-function mutant (Fig. 8e and f) and serves as a link between amino-acid metabolism and ethylene regulation.

Our findings position ethylene as a key regulator of plant responses to abiotic stress, acting as a previously unappreciated overarching regulator that influences multiple stress conditions rather than being specific to individual stresses. Analysis of the ethylene-insensitive mutant *ein2-5* revealed that ethylene signaling is required for the transcriptional activation of most assayed genes, with the exception of *WRKY33* (Fig. 5e). Given that many ethylene-related genes contain WRKY33 binding motifs, WRKY33 appears to be a key component in the regulation of general stress responses by ethylene, as supported by the known interaction between WRKY33 and ERF1[74]. Additionally, ethylene emanation was altered under all tested stresses, with increased levels in response to heat, wounding, complete submergence, and salt stress, and decreased levels under cold stress (Fig. S6). Mutant analyses revealed that WRKY33-deficient plants had elevated ethylene levels, correlating with a hypersensitive phenotype under most stresses. Similarly, *mkk9-1* mutants showed altered ethylene production, particularly under cold and submergence stress. These results accentuate the dynamic regulation of ethylene by WRKY33 and MKK9, with its levels of fine-tuning stress

tolerance. While further investigation is needed, our work underscores ethylene as a central component in the core of abiotic stress response genes, with the potential for developing stress-resilient crops. This knowledge serves as a stepping stone toward a more rational design of stress-tolerant crops by defining the intricate gene networks involved. For instance, although the upregulation or knockdown of the central hub WRKY33 may result in undesirable effects due to increased defense responses, a CRISPR-based targeted mutagenesis strategy focusing on optimizing its interactions with other key elements in the network—such as ERF1, EIN3, or MKK9—could provide a promising approach to enhance stress resistance while minimizing negative impacts on growth.

In conclusion, we demonstrate the suitability of an unsupervised machine-learning technique—SVM Clustering —to gain insights into complex biological processes. The methodology is robust and provides a comprehensive view of plant responses to a particular set of growth conditions. The stress cores cover a number of genes previously linked to specific stresses, corroborating the solidity of the cores and emphasizing the power of the SVM Clustering approach to identify genes involved in general stress signaling. This serves as a stepping stone for studies on the impact of global climate change on plants. Moreover, the approach enables high-confidence discovery of central players in the processes of interest, with either another unrelated or no function previously assigned to such genes. Hence, SVM Clustering is a robust mining tool to rapidly gain a holistic view of a gene network at the center of a set of responses. Furthermore, we demonstrate the vital role of ethylene signaling in the core stress-signaling network, particularly highlighting the critical regulatory functions of WRKY33 and MKK9 in modulating ethylene production and stress tolerance. Our findings underscore ethylene as a central signaling hub that integrates diverse environmental cues, while its dynamic regulation is tightly linked with WRKY33 and MKK9 activity. This integrated network plays a key role in plant stress responses, demonstrating the complexity of stress signaling and the importance of ethylene in mediating adaptive responses across various abiotic stresses. These results provide valuable insights into the broader regulatory framework that governs plant stress tolerance. Lastly, different databases were generated that unify valuable information regarding plant stress-responsiveness. Those datasets present comprehensive expression patterns of complete stress-related gene families (*EXP*, *WRKY*, *AP2/ERF*, and *MAPK*) as well as the previously poorly characterized *USPs* and the *EXP* subgroups *EXPLA* and *EXPLB*, providing new insights into their biological roles. These will nurture future functional analyses of as-yet uncharacterized genes and relevant members within large gene families, unchartered territory that hampers a full understanding of complex stress responses in plants.

While corroborating the role of a number of genes known to be stress-related, the gene cores present strong candidates for the engineering of plant tolerance to a wide range of adverse conditions, surpassing the limitations of single stress-related empirical studies including single transcriptomic analyses. We demonstrated that ethylene plays a crucial regulatory role within the cores, further underscoring its significance in multi-stress tolerance. Therefore, this core has the potential to serve as a foundation for exploring the development of multi-stress-resistant crop varieties. Secondly, the analysis offers high-confidence information on the temporal and spatial expression of stress genes and regulatory gene families. Thirdly, clear insights are gained into the presently unknown functionality of relevant genes, as well as into key members within large multigene families. Both our approach and the obtained results represent an important step forward in the field of plant systems biology, offering a powerful methodology to identify biologically relevant core genes, supporting more robust engineering strategies for the future development of stress-resistant plants.

## Methods

### Database selection

*Arabidopsis thaliana* transcriptomes were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/; Accessed in June 2021). The keywords "Arabidopsis" and "stress" provided a list of 945 series of data after filtering for organisms ("*Arabidopsis thaliana*"; filter aimed to exclude analyses using *A. thaliana* genes in other species) comprising both microarrays and RNA-seq platforms. The specific words for retrieving stress-related transcriptomes from the GEO database were "osmotic", "salinity", "drought", "oxidative", "heat", "cold", "hypoxia", "submergence", "light", "UV", "wounding", and "cadmium". In addition, the keywords "ACC/1-aminocyclopropane carboxylic acid" as the direct precursor of the plant hormone ethylene, and "ethylene" were used. For each stress type, the datasets corresponding to specific time points were selected, taking the following criteria into account: (1) at least duplicates per treatment, (2) availability of raw data, (3) tissue specificity of raw data is known (root or shoot), and (4) the age of analyzed plants was between 1 and 4 weeks, reducing the bias of the developmental stage.

The final list was composed of 23 data series comprising 500 single transcriptomes (Supplementary Data file 1). The composition of the final 500 individual transcriptomes was: 56 of which for ACC (the direct precursor of ethylene), 6 for cadmium, 40 for cold, 22 for complete submergence, 40 for drought, 56 for heat, 8 for hypoxia, 34 for excess of light, 82 for osmotic stress, 32 for partial submergence, 44 for high salinity, 40 for UV, and 40 for wounding. The sets were divided according to temporal and spatial determinants, with different time points divided into early (1, 3, and 6 h) and late responses (12 and 24 h), and tissue types (root and shoot tissues) (Supplementary Data file 3). The selection of time points is related to the variability of differentially expressed genes (DEG) peaks in literature, the maximum usually being between 1 and 3 h for early responses but displaced in some examples up to 6 h[75,76]. In this way, the inclusion of a 6-h timepoint in the category of early responses encompasses most early transcriptomic changes. Only the datasets containing information for root early and late and/or shoot early and late responses were considered for subsequent analysis, thus excluding cadmium and hypoxia. All datasets used are specified in the Supplementary Data file 1.

### Data pre-processing and normalization

Data pre-processing and normalization were performed using the robust multi-array average (RMA) method, considered the best method to increase comparability between different platforms[77]. Data from one-color microarrays were imported from the raw CEL files using the *affy* R package and then normalized. Normalization consisted of a background correction using a convolutional model, a quantile normalization, and a gene expression summarization (median-polish). For two-color microarray data, GPR files were imported with the *limma* R package, and their background was corrected by a normal-exponential convolutional model (method equivalent to the convolutional model in the RMA method). Then, data were normalized using quantile normalization. Finally, gene expression was obtained by summarization of the normalized data. In the case of RNA-seq data, both TXT and/or TSV files were imported using the *edgeR* package of R. Counts were transformed to log-CPM (counts-per-million) units, and unexpressed tags were filtered by the *edgeR's* filterByExpr function, which reduces the bias in large library sizes. Filtered log-CPM data were normalized with the method timed mean of *M*-values (TMM) to scale library sizes. Then, *limma's* voom function was used to perform the quantile normalization of the RNA-seq data (which is also equivalent to the RMA method). For gene annotation, the R package *AnnotationDbi* was employed.

### Quality control and data filtering

Unexpressed tags were detected and filtered by the MAS 5.0 method (tags with an alpha value > 0.06 were discarded, as default parameter

defined by Affimetrix) and by using the *edgeR* package, for microarray and RNA-seq respectively, and redundant tags were expressed as the average of the redundant probes. All expression data were expressed as log2/log2-CPM to increase comparability. A quality check of the pre-processed data was performed to remove low-quality datasets from the analysis. The quality was assessed by computing MA-plots to verify individual array quality, log2 intensities boxplots, and density-estimate plots to evaluate homogeneity between arrays, and heatmaps of the between-array distances for the in-between array quality check[78]. Datasets showing low quality or presented as outliers in some of the quality tests were removed for the downstream steps. This quality analysis was performed using the R package *arrayQualityMetrics*.

### Detection of differentially expressed genes (DEGs)

DEGs in each individual dataset (per stress, time-point, and tissue) were detected by using the R package *limma* and computing a linear model using an empirical Bayes method, which has been demonstrated to provide high-precision analyses of transcriptomic data in both microarray and RNA-seq analyses moderating the standard error of the log-fold changes between each probe set[77]. Genes with a log-fold change $(LFC) > |1|$ (equivalent to a fold change $> |2|$) and a false-discovery rate (FDR; obtained by the Benjamini and Hochberg corrected $p$-value[79]) $< 0.05$ were assigned as DEGs. This strict threshold was defined in order to provide a DEG list containing only the most relevant up- or down-regulated genes.

### Hierarchical clustering analysis (HCA) of the different stress modules

Hierarchical clustering analysis (HCA) was selected as an unsupervised machine-learning method to detect similarities and differences between the DEGs modules of each stress, taking into account time and tissue. DEG modules were created considering up- and down-regulated DEGs in each stress condition for early (1, 3, and 6 h) and late (12 and 24 h) time points for both root and shoot tissues. In case of discordance between some of the points (when a gene was up-regulated in one timepoint but down-regulated in another, or vice versa) the earliest time point for the early module and the latest time point for the late module were selected as representatives. The computation of the clusters was performed by an agglomerative HCA method with Gower's distance as a dissimilarity matrix[80]. Complete-linkage was used as an agglomeration method for the hclust function of the stats R package, allowing us to find the maximal differences between clusters, and suggested when small and specific clusters are expected. To assess the differences between the four computed hierarchical clusters (root-early, root-late, shoot-early, and shoot-late), pairs of dendrograms were compared. Baker's gamma index was calculated as the numerical value for the similarities between different dendrograms (with 0 being totally different dendrograms and 1 exactly the same). Significance was assessed with the $p$-value after computing 1000 random permutations of the compared dendrograms, as described in ref. 81, and assuming a $p$-value of 0.05 as statistically significant. For the computation of the HCA and the plotting of the dendrograms, the CRAN R packages dplyr, cluster, ggplot2, ggdendro, and dendextend were used.

### Meta-$p$-value computation and support vector machine (SVM) Clustering-based classification

To overcome the variability of sample-specific transcriptomes, meta-$p$-values for each stress, tissue, and timepoint were computed by the generalized weighted Fisher's method with sample-sizes correction[82]. The selection of this method is based on the meta-analysis decision scheme proposed in ref. 24, considering the source data (different platforms) and the heterogeneity of the dataset. The Benjamini and Hochberg corrected $p$-values obtained from the individual DEG detection analysis were used as input[83]. As output, a complex dataset

composed of 11 meta-$p$-values (one per stress condition and ACC treatment), per tissue, and timepoint for each gene is given.

Due to the multidimensionality of the dataset, we used SVM Clustering as an unsupervised classification method for the definition of the 'stress gene cores'. Particularly, it performs clustering of the input data in two different clusters with an algorithm that is typically used for binary classification[28]. SVM Clustering takes advantage of all the machinery provided by standard SVM to create a partition in a dataset, i.e., compute the decision boundaries between a set of user-pre-defined clusters. To do that, it requires pre-classification of the data, to later apply the SVM Clustering algorithm to re-classify the data while accounting for its multidimensional representation (the distribution of meta-$p$-values in all studied conditions). Based on the premise that genes belonging to the putative gene core should appear as DEGs in a significant number of stress conditions, we pre-classify the data assigning to each gene a binary label (1, present as DEG in five or more conditions, or 0 otherwise), thus obtaining the 'pre-defined clusters'. This is the input for the SVM Clustering algorithm that re-classifies, i.e., computes novel and more accurate decision boundaries (that in turn defines a new set of clusters). In addition, since SVM Clustering is an unsupervised machine learning method, it significantly reduces the potential bias inherent to supervised learning methods, as it does not depend on previous knowledge of the data. We selected the radial basis method (RBF) as an SVM kernel function, highly powerful for non-linear high-dimensional datasets and allowing for a more accurate classification[84]. The hyperparameters cost ($c$) and gamma ($\gamma$), crucial for robust and trustable results, together with the chosen kernel method, were optimized individually for each tissue by time set (root-early, root-late, shoot-early, and shoot-late). To select the optimal $c$ and $\gamma$ values, an iterative search from 1 to 5000 for $c$ and 0.1–9.9 for $\gamma$ was used. The selection of optimal values was based on the stabilization of the results, meaning the results remained consistent across subsequent value increments (known as the grid-search strategy)[85]. The output consisted of four stress gene cores, denoted as SVM gene cores. *e1071* R package (version 1.7–9) was used for computing all SVM calculations.

To assess the robustness of the analysis, several standard clustering metrics, specifically for SVM clustering, were calculated[28]. First, the sum-of-squared error (SSE) and clustering entropy were computed as quality metrics for each SVM clustering analysis (root early, root late, shoot early, and shoot late). Additionally, the SVM-clustering algorithm involves iterative re-classifications of the entire dataset until the resulting clusters stabilize, ensuring consistency. The algorithm includes three stop criteria: (1) the quality metrics for the clusters are satisfied, (2) further iterations no longer produce changes in the clusters, and (3) the predefined maximum number of iterations is reached[28]. Alongside the quality metrics, several iterations of clustering were performed to assess the stability of the classification method for each dataset. The SSE for the gene core clusters were between 0.45 and 0.68, indicating compactness, as lower SSE values (below 1) suggest that the genes in the core are spatially close in the multi-dimensional space. This compactness is expected for stress-related genes, as their meta-$p$-values are typically lower across conditions compared to non-significant genes. Additionally, clustering entropies were between 0.06 and 0.008, with values close to zero, indicating minimal disorder and highly ordered clustering (see Supplementary Data file 10). Supporting these quality metrics, the iterative analysis showed that after the first iteration, no significant changes were observed in the gene core, with over 95% of genes remaining unchanged in subsequent iterations. For example, the early-root set showed 97% stability, while both the late-root and late-shoot sets exhibited 100% stability after the first iteration. This significant stability (assuming a significance level of $\alpha = 0.05$) confirms that the clusters obtained are stable and reliable, further supporting the robustness of our analysis (Supplementary Data file 11).

To assess the discriminatory power of the analysis, we conducted a supplementary analysis by reducing the sets of genes under analysis. Since the vast majority of the genes are tagged as non-relevant(0), we performed the analysis with a reduced number of genes to identify possible biases towards the non-relevant set. To achieve this, we removed from the dataset all genes that were present as DEGs in fewer than four conditions, retaining only those appearing in four or more conditions. We then performed a pre-classification by binarizing the data: genes identified as DEGs in four conditions were labeled as non-relevant(0), and genes appearing as DEGs in five or more conditions were labeled as relevant[1]. The reduction in the gene pool was as follows: from 12,888 to 403 in the roots early dataset, from 12,129 to 327 in roots late, from 11,501 to 639 in the shoots early dataset, and from 10,760 to 275 in shoots late (Supplementary Data file 12).

## Gene-core analysis

The comparisons between the SVM gene cores were performed by computing the overlap between different datasets and were visualized using Venn Diagrams (*VennDiagram* R package was used; DOI: 10.32614/CRAN.package.VennDiagram). The hypergeometric test was used to assess the statistical significance and over- or under-representation of the different overlapping groups of genes as described in ref. 86, using the formula:

$$C(d, x) \cdot C(n - d, n - x)/C(N, n), \qquad (1)$$

where $C(A, B)$ denotes the number of combinations of $A$ elements in groups of $B$ elements, $x$ is the number of overlapping genes between the two groups, $n$ and $d$ are the number of genes in groups 1 and 2, respectively; and $N$ is the total of genes in the comparison.

For GO enrichment, *A. thaliana* genomic information was retrieved using the *org.At.tair.db* R package and analyzed by the *topGO*, *goProfiles* and *clusterProfiler* R packages. GO terms showing a *p*-value < 0.05 were considered representative, and GO redundancy was minimized using the *simplify* function of the *clusterProfiler* R package.

GenFAM (Gene Families)[87] and the search tool of recurring instances of neighboring genes (STRING; accessed in August 2022; version 11.1)[88], were used for gene-family enrichment and the interaction-network construction, respectively. For SVM gene-cores clustering, the *k*-means method from STRING was used.

## Detection of gene responsiveness to ethylene

To detect the relation between the SVM gene cores and ethylene responses, experimentally validated EIN3 interactions[15] and the transcriptomic response to ethylene were used (GEO accession numbers: GSE14247, GSE83573) to create an ethylene-responsiveness database (Supplementary Data file 6). The subset of ethylene-responsive genes within the SVM gene cores was used as input in STRING (version 11.1)[88], and their interaction network was obtained.

## Identification of WRKY33 binding sites

To predict the presence of the WRKY33 binding motif, data from a WRKY33 ChIP-seq analysis[39] were retrieved and used for an in silico analysis. The analysis aimed to identify the presence of the empirically determined WRKY33 binding motif (TTGACY) in the promoter regions of the genes in our stress gene core. For this purpose, 1000 bp of sequence upstream of the translational start codon was retrieved from the TAIR database. Find individual motif occurrences (FIMO[89]) was employed to scan for statistically significant occurrences of the WRKY33 binding motif. A *p*-value threshold of <0.01 was applied to select statistically significant occurrences of the motif within the analyzed sequences.

## Real-time quantitative PCR

For all stress treatments (except heat), tissue samples from 10-day-old plants were collected at either 1 or 3 h after treatment initiation. For heat treatments, plants were subjected to a 40-min heat exposure at 42 °C, followed by recovery at room temperature. Tissue was collected 20 and 140 min after the start of the recovery period. Approximately 100 mg of whole plant tissue was homogenized using a Retsch mill, and RNA was extracted using the GeneJET Plant RNA Purification Kit (Thermo Fisher, Belgium). Genomic DNA was removed from the total RNA using DNase I (Thermo Fisher). RNA quality and quantity were assessed using an NP80 NanoPhotometer (Implen, Germany). Subsequently, cDNA synthesis was performed using the Bio-Rad iScript cDNA Synthesis Kit with 1 µg RNA. Target genes and their primers for real-time quantitative PCR analysis are listed in Supplementary Data file 14.

Reactions were carried out using qPCRBIO SyGreen Mix with Fluorescein (PCR Biosystems, UK), with a final primer concentration of 400 nM. qPCR was conducted on a CFX Opus 384 Real-time PCR System, using the following thermal cycling conditions: initial denaturation at 95 °C for 2 min, followed by 40 cycles of 95 °C for 5 s (denaturation) and annealing at a variable temperature for 20 s (see Supplementary Data file 14 for annealing temperatures). Gene expression was quantified using the ΔΔCq method, and reference-gene stability was assessed using Bio-Rad Maestro software. Expression values were normalized to the reference genes *ACTIN 2* (*ACT2*; AT3G18780), *UBIQUITIN 10* (*UBQ10*; AT4G05320), and *PROTEIN PHOSPHATASE 2A SUBUNIT A2* (*PP2A*; AT3G25800), based on three technical replicates and three biological replicates per treatment.

## Plant material and analysis of stress effects

EXPA abundance was assayed using transgenic *pEXPA:EXPA−mCherry* translational reporter fusions for *EXPA1*, *EXPA10*, and *EXPA14*[41]. Seeds were surface-sterilized according to[90] and subsequently plated on half-strength Murashige and Skoog medium containing 1% w/v sucrose and 0.8% w/v agar (hereafter MS 1/2). After 3 days of stratification at 4 °C, plates were transferred to a tissue-culture room and grown in a 16/8-h photoperiod (70 µmol photons m$^{-2}$ s$^{-1}$) for 4 days at 21 °C, placed in a vertical position. Subsequently, the plants were exposed to a short-term (1 h) stress treatment. For such treatment, 4-day-old plantlets grown in vertical conditions were transferred to specific conditions. For salt and osmotic treatments, plantlets were transferred to treatment plates containing 100 mM NaCl (salt) or 150 mM mannitol (osmotic). For cold, plantlets were transferred to MS 1/2 medium and immediately placed at 4 °C. Control plantlets were transferred to MS 1/2. After transfer to the respective treatments, plantlets were incubated for 1 h in the same growing conditions (16/8-h photoperiod (70 µmol m$^{-2}$ s$^{-1}$) at 21 °C; except cold treatment at 4 °C). Confocal laser-scanning microscopy images were obtained with an inverted Nikon TiE-C2 confocal microscope. Roots were imaged with a ×20 CFI Plan Apochromat VC objective lens (NA 0.75, dry). Images (1024 × 1024; ×1.5 scanner zoom) were collected by exciting mCherry with a solid-state 561-nm laser and emission was collected from 571 to 700 nm. The same settings were kept to compare fluorescence intensities between treatments of specific transgenic lines. Quantification of the mCherry signal was performed in ImageJ. Regions of interest (ROI) were defined for each line focusing on the expression patterns as described in ref. 41. Signal intensity was quantified as gray value intensity normalized by the ROI size in µm². Shapiro's test and Levene's test were used to assess the normality and variance of each dataset. To compare the intensity of the different treatments with the control samples, the Student's *T*-test and Wilcoxon rank sum exact test were used for parametric and non-parametric testing, respectively. Complete statistical analysis is available in Supplementary Data file 15.

For the analyses of *lht1-5*, *mkk9-1*, and *wrky33-2* (GABI_324B11, 39) mutants, *A. thaliana* ecotype Columbia 0 (Col-0) was used as

wild-type control. The *mkk9-1* and *lht1-5* mutants (Col-0 background) were obtained from the NASC Arabidopsis stock center (SALK_017378 and SALK_115555C, respectively). Seeds were surface-sterilized, plated on MS 1/2 medium, stratified, and transferred to tissue culture in the same conditions as mentioned above. Hereafter, plants were exposed to their respective long-term stress treatments. For such treatments, 10-day-old plantlets were transferred to specific conditions. For salt and osmotic stresses, plantlets were transferred to treatment plates containing 100 mM NaCl or 150 mM mannitol, respectively. For cold stress, plantlets were transferred to 12-well MS 1/2 plates and grown at 4 °C for 4 days in a 16/8-h photoperiod (a specific control with a similar light source was included). For heat treatment, plantlets were subjected to 42 °C for 40 min and then transferred to the tissue culture room (same conditions). For wounding, plants were wounded in the two main leaves immediately after transfer and subsequently wounded in the new leaves for 4 consecutive days. In the case of complete submergence, seeds were directly sown in 12-well MS 1/2 plates, to facilitate root anchorage and to avoid that plants float. After 10 days of growing in the tissue chamber (same conditions), the wells were filled with distilled water to completely cover the plantlets and they were maintained for 4 days in dark conditions (a dark control was included). After the stress treatments, plantlets were transferred to recovery plates (MS 1/2 medium). After 5 days (for *wrky33-2*) or 10 days (for *lht1-5* and *mkk9-1*) of recovery, plants were imaged (CANON EOS 550D camera (Canon, Japan)) and the rosette area was analyzed using the ImageJ plug-in (National Institutes of Health) *Rosette tracker*[91]. Rosette areas were plotted as violin plots using the *ggplot2* R package. Complete statistical analysis is available in Supplementary Data file 15.

For biomass analyses, the same stress treatments were performed and biomass was measured after 10 days in recovery plates for *mkk9-1* and *lht1-5* samples, or after 5 days in the case of *wrky33-2* plantlets ($n > 20$ plants per sample). Data were represented relative to control conditions to minimize bias in sample handling.

### Ethylene emanation analysis

To measure ethylene emanation in response to stress conditions, two-week-old Col-0, *mkk9-1*, or *wrky33-2* plants were transferred to 10 mL chromatography vials (Chromacol, VWR, Leuven, Belgium) containing MS 1/2 medium supplemented with the appropriate chemical stress inducers, or were placed directly into the designated stress conditions (see previous section for more details). Each vial contained one plant to prevent ethylene production resulting from overcrowding. The vials were sealed with rubber septa and snap-caps (Chromacol), and ethylene accumulation was allowed to proceed for 24 h following the initiation of the treatment. Ethylene concentration was measured using a laser-based photoacoustic detector (ETD-300, Sensor Sense, The Netherlands). Average ethylene production was normalized both to time (per hour) and to plant biomass, based on data from five to eight biological replicates.

### Statistical analysis

Shapiro's test and Levene's test were used to assess the normality and variance of each dataset. To compare the intensity of the different treatments with the control samples, the Student's *T*-test and Wilcoxon rank sum exact test were used for parametric and non-parametric testing, respectively. A *p*-value of <0.05 was assumed as statistically significant (*<0.05; **<0.01; ***<0.001). To compare rosette areas, a two-way ANOVA followed by post-hoc Tukey's test was used as parametric test, while a Kruskal–Wallis rank sum test followed by Dunn's Multiple Comparison test was used as non-parametric test. A *p*-value of <0.05 was assumed as statistically significant. For statistical comparison of biomass measures, a one-way ANOVA with Brown–Forsythe and Welch ANOVA correction for heteroscedastic followed by post-hoc Dunnett T3 tests (*p*-value < 0.05) with correction for multiple

pairwise comparisons. For non-parametric analysis, the Kruskal–Wallis test followed by post-hoc Dunnett T3 tests (*p*-value < 0.05) with correction for multiple pairwise comparisons was used. Statistically significant differences between means are indicated with *$p$-value < 0.05, **$p$-value < 0.005, and ***$p$-value < 0.001. The statistical comparisons of ethylene production were performed by a one-way ANOVA with Brown–Forsythe and Welch ANOVA correction for heteroscedastic data (*p*-value < 0.05; $5 < n < 8$) followed by post-hoc Dunnett T3 tests (*p*-value < 0.05) with correction for multiple pairwise comparisons. Statistically significant differences between means are indicated with *$p$-value < 0.05, **$p$-value < 0.005, and ***$p$-value < 0.001. Complete statistical analysis for all the assays is available in Supplementary Data file 15.

## Data availability
All data supporting the findings of this study are available within the paper and its Supplementary Information file. Source data are provided with this paper as Source Data File. All datasets used to construct the meta-analysis are deposited in the Gene Expression Omnibus (GEO) database, and the accession numbers are specified in Supplementary Data file 1. To construct the ethylene responsiveness dataset, the transcriptomes GSE14247 and GSE83573, available in the GEO database, were used. All software used is cited in the corresponding sections of the manuscript. The R packages used in this study include *affy, AnnotationDbi, affyPLM, arrayQualityMetrics, oligo, limma, edgeR, metaPro, e1071* (version 1.7-9), *VennDiagram, topGO, goProfiles, clusterProfiler*, and *ggplot2* (available at https://cran.r-project.org/). Image analysis was performed using ImageJ (https://imagej.net/downloads). GenFAM and STRING were accessed online. qPCR analysis was conducted using Bio-Rad Maestro software (Bio-Rad). Statistical analyses were performed using R. Source data are provided with this paper.

## References

1. Food and Agriculture Organization of the United Nations. *The Impact of Disasters and Crises on Agriculture and Food Security: 2021* (Food & Agriculture Organization, 2021).
2. Pandey, P., Irulappan, V., Bagavathiannan, M. V. & 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.* **8**, 537 (2017).
3. Oshunsanya, S. O., Nwosu, N. J. & Li, Y. Abiotic stress in agricultural crops under climatic conditions. In *Sustainable Agriculture, Forest and Environmental Management* (eds Jhariya, M., Banerjee, A., Meena, R. & Yadav, D.) 71–100 (Springer, Singapore, 2019).
4. Savary, S. et al. The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* **3**, 430–439 (2019).
5. Zhang, H., Zhao, Y. & Zhu, J.-K. Thriving under stress: how plants balance growth and the stress response. *Dev. Cell* **55**, 529–543 (2020).
6. Zandalinas, S. I. & Mittler, R. Plant responses to multifactorial stress combination. *N. Phytol.* **234**, 1161–1167 (2022).
7. Kuromori, T., Fujita, M., Takahashi, F., Yamaguchi-Shinozaki, K. & Shinozaki, K. Inter-tissue and inter-organ signaling in drought stress response and phenotyping of drought tolerance. *Plant J.* **109**, 342–358 (2022).
8. Li, H., Testerink, C. & Zhang, Y. How roots and shoots communicate through stressful times. *Trends Plant Sci.* **26**, 940–952 (2021).
9. Singh, A. et al. Tissue specific and abiotic stress regulated transcription of histidine kinases in plants is also influenced by diurnal rhythm. *Front. Plant Sci.* **6**, 711 (2015).
10. Choudhury, F. K., Devireddy, A. R., Azad, R. K., Shulaev, V. & Mittler, R. Rapid accumulation of glutathione during light stress in Arabidopsis. *Plant Cell Physiol.* **59**, 1817–1826 (2018).
11. Moore, M., Vogel, M. O. & Dietz, K. J. The acclimation response to high light is initiated within seconds as indicated by upregulation of

*AP2/ERF* transcription factor network in *Arabidopsis thaliana*. *Plant Signal Behav*. **9**, 976479 (2014).

12. Kollist, H. et al. Rapid responses to abiotic stress: priming the landscape for the signal transduction network. *Trends Plant Sci*. **24**, 25–37 (2019).

13. Depaepe, T. et al. At the crossroads of survival and death: the reactive oxygen species–ethylene–sugar triad and the unfolded protein response. *Trends Plant Sci*. **26**, 338–351 (2021).

14. Depaepe, T. & Van Der Straeten, D. Tools of the ethylene trade: a chemical kit to influence ethylene responses in plants and its use in agriculture. *Small Methods* **4**, 1900267 (2020).

15. Chang, K. N. et al. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in *Arabidopsis*. *Elife* **2**, e00675 (2013).

16. Anderson, J. P. et al. Antagonistic interaction between abscisic acid and jasmonate–ethylene signaling pathways modulates defense gene expression and disease resistance in *Arabidopsis*. *Plant Cell* **16**, 3460–3479 (2004).

17. Van den Broeck, L. et al. From network to phenotype: the dynamic wiring of an *Arabidopsis* transcriptional network induced by osmotic stress. *Mol. Syst. Biol*. **13**, 961 (2017).

18. Hossain, M. A. et al. Heat or cold priming-induced cross-tolerance to abiotic stresses in plants: key regulators and possible mechanisms. *Protoplasma* **255**, 399–412 (2018).

19. Zhang, X., Shen, L., Li, F., Meng, D. & Sheng, J. *Arginase* induction by heat treatment contributes to amelioration of chilling injury and activation of antioxidant enzymes in tomato fruit. *Postharvest Biol. Technol*. **79**, 1–8 (2013).

20. Chou, T.-S., Chao, Y.-Y. & Kao, C. H. Involvement of hydrogen peroxide in heat shock- and cadmium-induced expression of ascorbate peroxidase and glutathione reductase in leaves of rice seedlings. *J. Plant Physiol*. **169**, 478–486 (2012).

21. Hossain, M. A., Mostofa, M. G. & Fujita, M. Cross protection by cold-shock to salinity and drought stress-induced oxidative stress in mustard (*Brassica campestris* L.) seedlings. *Mol. Plant Breed*. **4**, 50–70 (2013).

22. Atkinson, N. J. & Urwin, P. E. The interaction of plant biotic and abiotic stresses: from genes to the field. *J. Exp. Bot*. **63**, 3523–3543 (2012).

23. Panahi, B., Frahadian, M., Dums, J. T. & Hejazi, M. A. Integration of cross species RNA-seq meta-analysis and machine-learning models identifies the most important salt stress-responsive pathways in microalga. *Front. Genet*. **10**, 752 (2019).

24. Toro-Domínguez, D. et al. A survey of gene expression meta-analysis: methods and applications. *Brief. Bioinform*. **22**, 1694–1705 (2021).

25. Meta-analysis in basic biology. *Nat. Methods* **13**, 959–959 https://doi.org/10.1038/nmeth.4102 (2016).

26. Gibney, E. Could machine learning fuel a reproducibility crisis in science? *Nature* **608**, 250–251 (2022).

27. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn* **20**, 273–297 (1995).

28. Winters-Hilt, S. & Merat, S. SVM clustering. *BMC Bioinform*. **8**(Suppl. 7), S18 (2007).

29. Krishnaveni, N. & Radha, V. Performance evaluation of clustering-based classification algorithms for detection of online spam reviews. In *Data Intelligence and Cognitive Informatics. Algorithms for Intelligent Systems* 255–266 (eds Jeena Jacob, I., Kolandapalayam Shanmugam, S., Piramuthu, S. & Falkowski-Gilski, P.) (Springer, Singapore, 2021).

30. Yan, J. & Wang, X. Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology. *Plant J*. **111**, 1527–1538 (2022).

31. Codjoe, J. M., Miller, K. & Haswell, E. S. Plant cell mechanobiology: greater than the sum of its parts. *Plant Cell* **34**, 129–145 (2022).

32. Tenhaken, R. Cell wall remodeling under abiotic stress. *Front. Plant Sci*. **5**, 771 (2014).

33. Ghosh, D. & Xu, J. Abiotic stress responses in plant roots: a proteomics perspective. *Front. Plant Sci*. **5**, 6 (2014).

34. Huang, P.-Y., Catinot, J. & Zimmerli, L. Ethylene response factors in *Arabidopsis* immunity. *J. Exp. Bot*. **67**, 1231–1241 (2016).

35. Xu, J. & Zhang, S. Regulation of ethylene biosynthesis and signaling by protein kinases and phosphatases. *Mol. Plant* **7**, 939–942 (2014).

36. Xu, J. et al. Activation of *MAPK kinase 9* induces ethylene and camalexin biosynthesis and enhances sensitivity to salt stress in *Arabidopsis*. *J. Biol. Chem*. **283**, 26996–27006 (2008).

37. Shin, K. et al. Genetic identification of *ACC-RESISTANT2* reveals involvement of *LYSINE HISTIDINE TRANSPORTER1* in the uptake of 1-aminocyclopropane-1-carboxylic acid in *Arabidopsis thaliana*. *Plant Cell Physiol*. **56**, 572–582 (2015).

38. Zhu, T. et al. Mitochondrial alternative oxidase-dependent autophagy involved in ethylene-mediated drought tolerance in *Solanum lycopersicum*. *Plant Biotechnol. J*. **16**, 2063–2076 (2018).

39. Birkenbihl, R. P., Kracher, B., Roccaro, M. & Somssich, I. E. Induced genome-wide binding of three *Arabidopsis* WRKY transcription factors during early MAMP-triggered immunity. *Plant Cell* **29**, 20–38 (2017).

40. Shi, J., Drummond, B. J., Wang, H., Archibald, R. L. & Habben, J. E. Maize and Arabidopsis ARGOS proteins interact with ethylene receptor signaling complex, supporting a regulatory role for ARGOS in ethylene signal transduction. *Plant Physiol*. **171**, 2783–2797 (2016).

41. Samalova, M. et al. Hormone-regulated expansins: expression, localization, and cell wall biomechanics in *Arabidopsis* root growth. *Plant Physiol*. **194**, 209–229 (2024)

42. Samalova, M., Gahurova, E. & Hejatko, J. Expansin-mediated developmental and adaptive responses: a matter of cell wall biomechanics? *Quant. Plant Biol*. **3**, e11 (2022).

43. Xie, Z., Nolan, T. M., Jiang, H. & Yin, Y. *AP2/ERF* transcription factor regulatory networks in hormone and abiotic stress responses in *Arabidopsis*. *Front. Plant Sci*. **10**, 228 (2019).

44. Nakano, T., Suzuki, K., Fujimura, T. & Shinshi, H. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol*. **140**, 411–432 (2006).

45. Zhang, M. & Zhang, S. Mitogen-activated protein kinase cascades in plant signaling. *J. Integr. Plant Biol*. **64**, 301–341 (2022).

46. Kvint, K., Nachin, L., Diez, A. & Nyström, T. The bacterial universal stress protein: function and regulation. *Curr. Opin. Microbiol*. **6**, 140–145 (2003).

47. Shaik, R. & Ramakrishna, W. Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. *Plant Physiol*. **164**, 481–495 (2014).

48. Ma, C., Xin, M., Feldmann, K. A. & Wang, X. Machine learning-based differential network analysis: a study of stress-responsive transcriptomes in *Arabidopsis*. *Plant Cell* **26**, 520–537 (2014).

49. English, P. J., Lycett, G. W., Roberts, J. A. & Jackson, M. B. Increased 1-aminocyclopropane-1-carboxylic acid oxidase activity in shoots of flooded tomato plants raises ethylene production to physiologically active levels. *Plant Physiol*. **109**, 1435–1440 (1995).

50. Depaepe, T., Vanhaelewyn, L. & Van Der Straeten, D. UV-B responses in the spotlight: dynamic photoreceptor interplay and cell-type specificity. *Plant Cell Environ*. **46**, 3194–3205 (2023).

51. Xiong, L. & Zhu, J. K. Molecular and genetic aspects of plant responses to osmotic stress. *Plant Cell Environ*. **25**, 131–139 (2002).

52. Kim, J. S., Kidokoro, S., Yamaguchi-Shinozaki, K. & Shinozaki, K. Regulatory networks in plant responses to drought and cold stress. *Plant Physiol*. **195**, 170–189 (2024).

53. Lewandowska, M. et al. Wounding triggers wax biosynthesis in *Arabidopsis* leaves in an abscisic acid-dependent and jasmonoyl-

isoleucine-dependent manner. *Plant Cell Physiol.* **65**, 928–938 (2024).

54. Prasad, A., Sedlářová, M., Balukova, A., Rác, M. & Pospíšil, P. Reactive oxygen species as a response to wounding: imaging in *Arabidopsis thaliana*. *Front. Plant Sci.* **10**, 1660 (2019).

55. Lee, S. & Park, C. M. Regulation of reactive oxygen species generation under drought conditions in *Arabidopsis*. *Plant Signal Behav.* **7**, 599–601 (2012).

56. Sato, H., Mizoi, J., Shinozaki, K. & Yamaguchi-Shinozaki, K. Complex plant responses to drought and heat stress under climate change. *Plant J.* **117**, 1873–1892 (2024).

57. Jahed, K. R., Saini, A. K. & Sherif, S. M. Coping with the cold: unveiling cryoprotectants, molecular signaling pathways, and strategies for cold stress resilience. *Front. Plant Sci.* **14**, 1246093 (2023).

58. Taji, T. et al. Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. *Plant J.* **29**, 417–426 (2002).

59. Tan, W.-J. et al. *DIACYLGLYCEROL ACYLTRANSFERASE* and *DIACYLGLYCEROL* KINASE modulate triacylglycerol and phosphatidic acid production in the plant response to freezing stress. *Plant Physiol.* **177**, 1303–1318 (2018).

60. Baez, L. A., Tichá, T. & Hamann, T. Cell wall integrity regulation across plant species. *Plant Mol. Biol.* **109**, 483–504 (2022).

61. Lamers, J., van der Meer, T. & Testerink, C. How plants sense and respond to stressful environments. *Plant Physiol.* **182**, 1624–1635 (2020).

62. Kudla, J. et al. Advances and current challenges in calcium signaling. *N. Phytol.* **218**, 414–431 (2018).

63. Wang, C., Teng, Y., Zhu, S., Zhang, L. & Liu, X. NaCl- and cold-induced stress activate different $Ca^{2+}$-permeable channels in *Arabidopsis thaliana*. *Plant Growth Regul.* **87**, 217–225 (2019).

64. Corso, M., Doccula, F. G., de Melo, J. R. F., Costa, A. & Verbruggen, N. Endoplasmic reticulum-localized CCX2 is required for osmotolerance by regulating ER and cytosolic Ca dynamics. *Proc. Natl Acad. Sci. USA* **115**, 3966–3971 (2018).

65. Li, Z. et al. CCX1, a putative cation/$Ca^{2+}$ exchanger, participates in regulation of reactive oxygen species homeostasis and leaf senescence. *Plant Cell Physiol.* **57**, 2611–2619 (2016).

66. Dodd, A. N., Kudla, J. & Sanders, D. The language of calcium signaling. *Annu. Rev. Plant Biol.* **61**, 593–620 (2010).

67. Zhu, X. et al. Calmodulin-like protein CML24 interacts with CAMTA2 and WRKY46 to regulate ALMT1-dependent Al resistance in *Arabidopsis thaliana*. *N. Phytol.* **233**, 2471–2487 (2022).

68. Menke, F. L. H., van Pelt, J. A., Pieterse, C. M. J. & Klessig, D. F. Silencing of the mitogen-activated protein kinase MPK6 compromises disease resistance in *Arabidopsis*. *Plant Cell* **16**, 897–907 (2004).

69. Shi, Y. et al. Ethylene signaling negatively regulates freezing tolerance by repressing expression of *CBF* and *type-A ARR* genes in *Arabidopsis*. *Plant Cell* **24**, 2578–2595 (2012).

70. Hartman, S. et al. Ethylene-mediated nitric oxide depletion pre-adapts plants to hypoxia stress. *Nat. Commun.* **10**, 4020 (2019).

71. Vaseva, I. I. et al. Ethylene signaling in salt-stressed *Arabidopsis thaliana ein2-1* and *ctr1-1* mutants—a dissection of molecular mechanisms involved in acclimation. *Plant Physiol. Biochem.* **167**, 999–1010 (2021).

72. Zhang, X. et al. MAMP-elicited changes in amino acid transport activity contribute to restricting bacterial growth. *Plant Physiol.* **189**, 2315–2331 (2022).

73. Batista-Silva, W. et al. The role of amino acid metabolism during abiotic stress release. *Plant Cell Environ.* **42**, 1630–1644 (2019).

74. Chen, Y. & Zhang, J. Multiple functions and regulatory networks of *WRKY33* and its orthologs. *Gene* **931**, 148899 (2024).

75. De Paepe, A., Vuylsteke, M., Van Hummelen, P., Zabeau, M. & Van Der Straeten, D. Transcriptional profiling by cDNA-AFLP and microarray analysis reveals novel insights into the early response to ethylene in *Arabidopsis*. *Plant J.* **39**, 537–559 (2004).

76. Kilian, J. et al. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* **50**, 347–363 (2007).

77. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

78. Kauffmann, A., Gentleman, R. & Huber, W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415–416 (2009).

79. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).

80. Gower, J. C. A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857 (1971).

81. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2018).

82. Yoon, S., Baik, B., Park, T. & Nam, D. Powerful *p*-value combination methods to detect incomplete association. *Sci. Rep.* **11**, 6980 (2021).

83. Sharifi, S. et al. Integration of machine learning and meta-analysis identifies the transcriptomic bio-signature of mastitis disease in cattle. *PLoS ONE* **13**, e0191227 (2018).

84. Roman, I., Santana, R., Mendiburu, A. & Lozano, J. A. In-depth analysis of SVM kernel learning and its components. *Neural Comput. Appl.* **33**, 6575–6594 (2021).

85. Lameski, P., Zdravevski, E., Mingov, R. & Kulakov, A. SVM parameter tuning with grid search and its impact on reduction of model over-fitting. In: Yao, Y., Hu, Q., Yu, H., Grzymala-Busse, J.W. (eds) *Lecture Notes in Computer Science* 464–474 (Springer, Cham, 2015).

86. Luesse, D. R., Wilson, M. E. & Haswell, E. S. RNA sequencing analysis of the msl2msl3, crl, and ggps1 mutants indicates that diverse sources of plastid dysfunction do not alter leaf morphology through a common signaling pathway. *Front. Plant Sci.* **6**, 1148 (2015).

87. Bedre, R. & Mandadi, K. GenFam: a web application and database for gene family-based classification and functional enrichment analysis. *Plant Direct* **3**, e00191 (2019).

88. Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

89. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

90. Vanderstraeten, L., Sanchez-Muñoz, R., Depaepe, T., Auwelaert, F. & Van Der Straeten, D. Mix-and-match: an improved, fast and accessible protocol for hypocotyl micrografting of *Arabidopsis* seedlings with systemic ACC responses as a case study. *Plant Methods* **18**, 24 (2022).

91. De Vylder, J., Vandenbussche, F., Hu, Y., Philips, W. & Van Der Straeten, D. Rosette tracker: an open source image analysis tool for automatic quantification of genotype effects. *Plant Physiol.* **160**, 1149–1159 (2012).

## Acknowledgements

## Author contributions

D.V.D.S. was responsible for the design and supervision of the biological aspects of this work, while I.Z. was responsible for the design and supervision of the computational aspects. R.S.-M. performed the data-mining, the meta-analysis and the machine-learning analysis. R.S.-M., together with T.D., performed the functional validation of *WRKY33, MKK9* and *LHT1*. M.S. and J.H. provided the *EXPA* translational-reporter lines and guided their analysis. T.D. performed the *EXPA* translational-reporter-line analysis. R.S.-M., T.D., I.Z. and D.V.D.S. prepared the manuscript. All authors reviewed the manuscript and agreed to its submission.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at
https://doi.org/10.1038/s41467-025-59542-3.

**Correspondence** and requests for materials should be addressed to Isiah Zaplana or Dominique Van Der Straeten.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at
http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.