iScience, Volume *23*

Supplemental Information

Translational Components Contribute to Acclimation Responses to High Light, Heat, and Cold in *Arabidopsis*

Antoni Garcia-Molina, Tatjana Kleine, Kevin Schneider, Timo Mühlhaus, Martin Lehmann, and Dario Leister

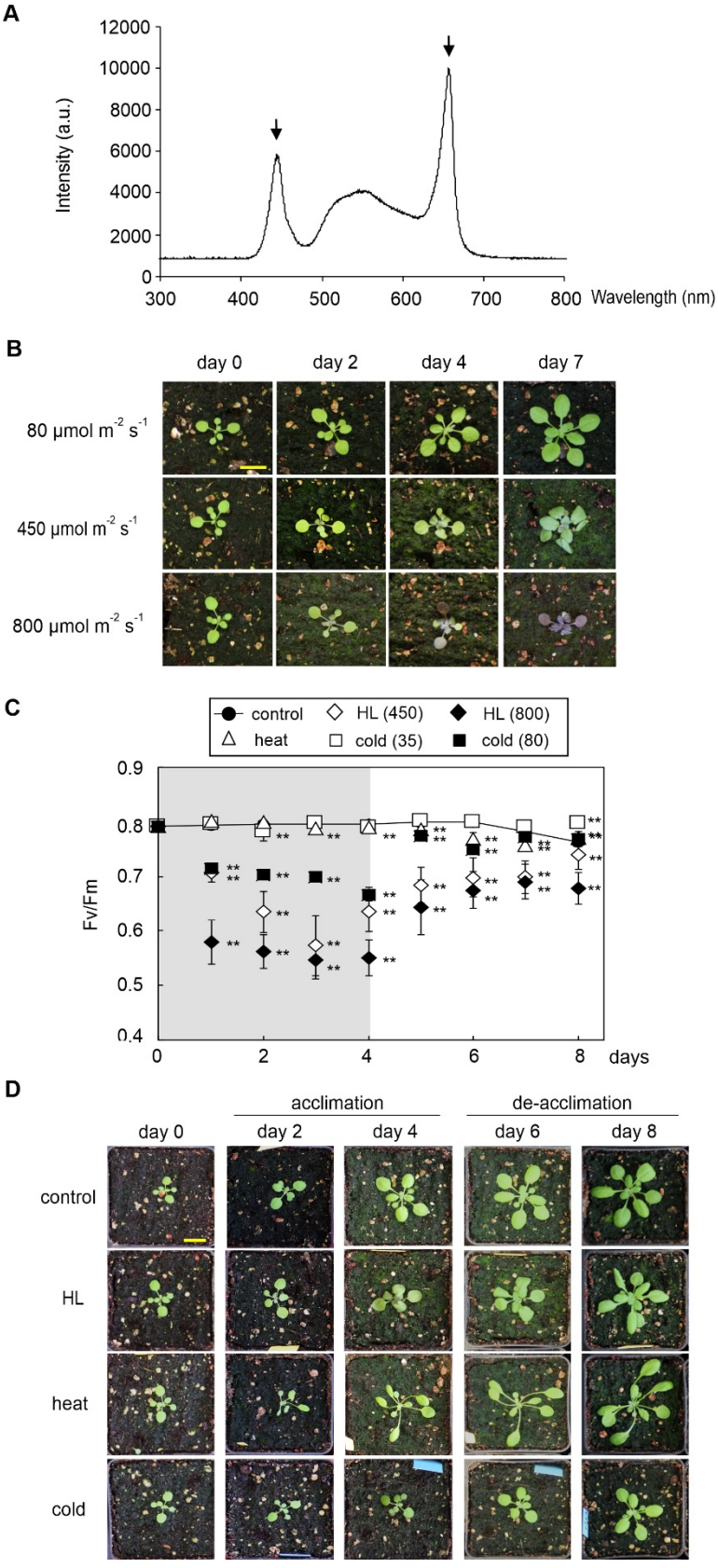

Figure S1. Set-up used to assay for (de-)acclimation to high light, heat and cold conditions, Related to Figure 1.

(A) Spectrum of the LED light source used for Arabidopsis growth. Arrows indicate blue (left) and red (right) peaks.

(B) Growth of Arabidopsis under three different light regimes. Plants were grown for 14 days under standard conditions and exposed to 80, 450 or 800 µmol photons m^{-1} s^{-2} for a further 7 days. Bar = 1 cm.

(**C**) Extent of photoinhibition during acclimation and de-acclimation. Maximum quantum yield of PSII (Fv/Fm) was recorded from plants exposed to standard conditions (control), high light (HL) at 450 or 800 µmol photons m^{-1} s^{-2} (450, 800), heat and cold at 35 or 80 µmol photons m^{-1} s^{-2} (35, 80). Values correspond to the mean ± SD of n ≥ 4 independent experiments. $^{**}P < 0.01$; $^{*}P < 0.05$. See Table S1 for standard deviations and statistics.

(D) Growth of Arabidopsis during exposure (acclimation) to high light (HL), heat and cold, followed by return to control conditions (de-acclimation). Representative plants are shown after 2 and 4 days of acclimation, and a further 2 (day 6) and 4 days (day 8) of de-acclimation. Bar = 1 cm.

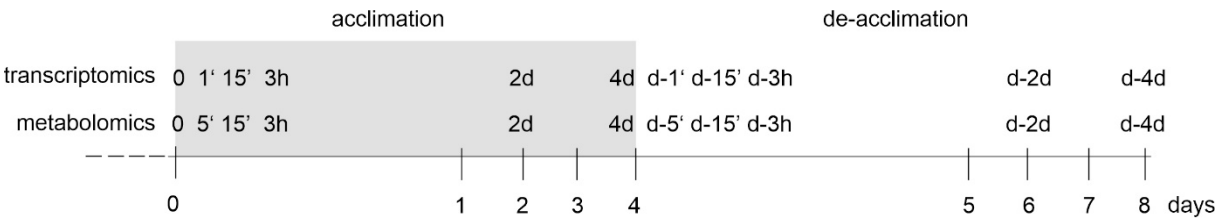

Figure S2. Time-points at which metabolites and transcripts were analyzed, Related to Figures 1, 2 and 3.

Equivalent time-points in the acclimation and de-acclimation phases were selected, and the latter are indicated by the prefix “d-“.

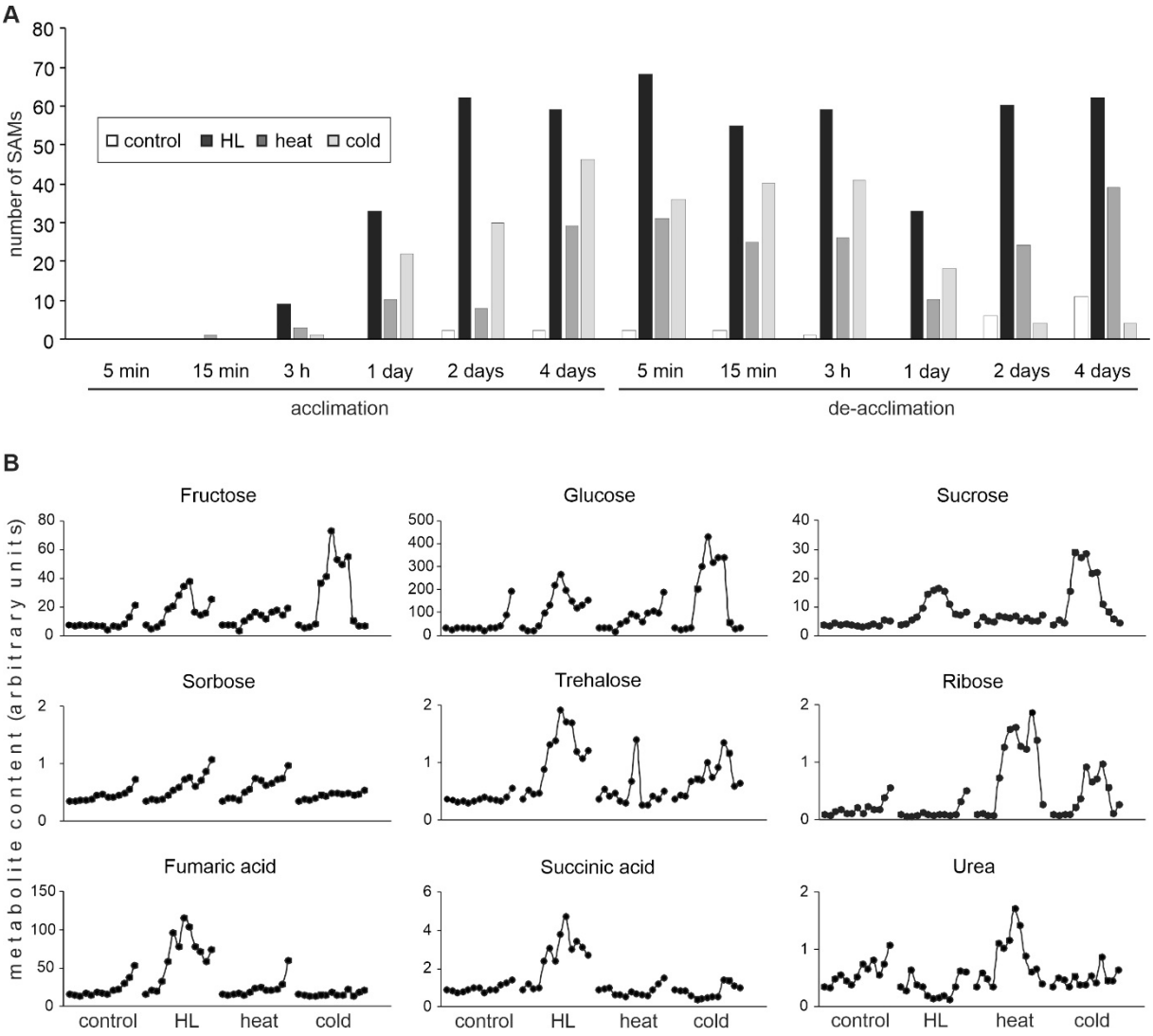

Figure S3. Metabolite changes during acclimation and de-acclimation to high light (HL), heat and cold, Related to Figure 2.

(A) Numbers of significantly altered metabolites (SAMs). SAMs are defined as metabolites that showed at least a twofold change in concentration relative to the initial value at 0 min according to the Student's *t*-test (FDR ≤ 0.05).

(B) Contents of representative metabolites (relative to a ^{13}C sorbitol standard) throughout the time-course under the four conditions. Note that differences between control conditions and treatments are not attributable to changes in periodicity caused by the circadian clock.

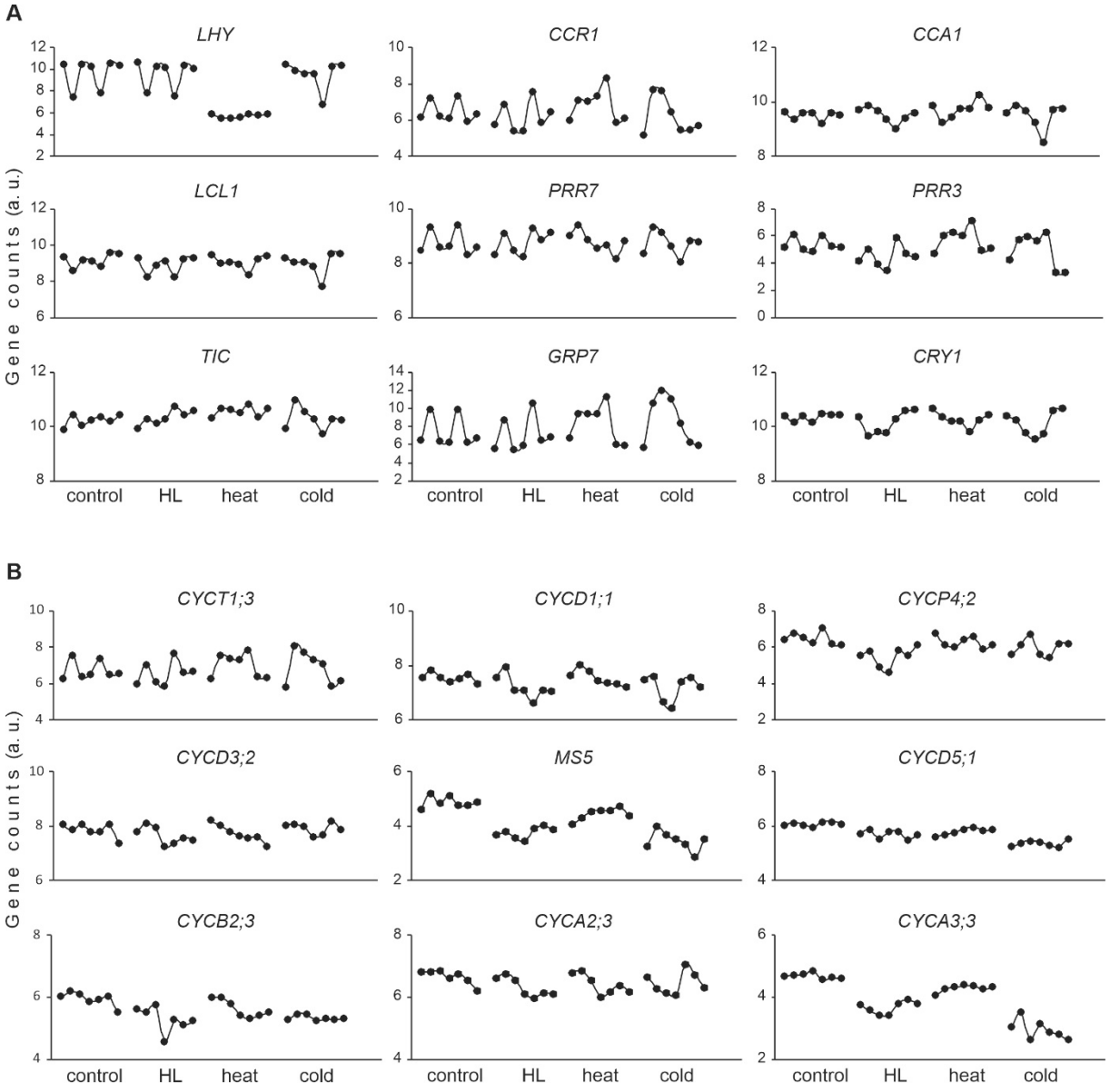

Figure S4. Levels of marker-gene transcripts plotted against circadian-clock time or cell-cycle stage, Related to Figure 3.

(**A**, **B**) The means of normalized counts were used to plot the mRNA expression pattern of *bona-fide* markers for circadian periodicity (**A**) and cell-cycle phase (**B**) at the initial time-point (0 min) and after 3 h, 2 days and 4 days of acclimation and de-acclimation.

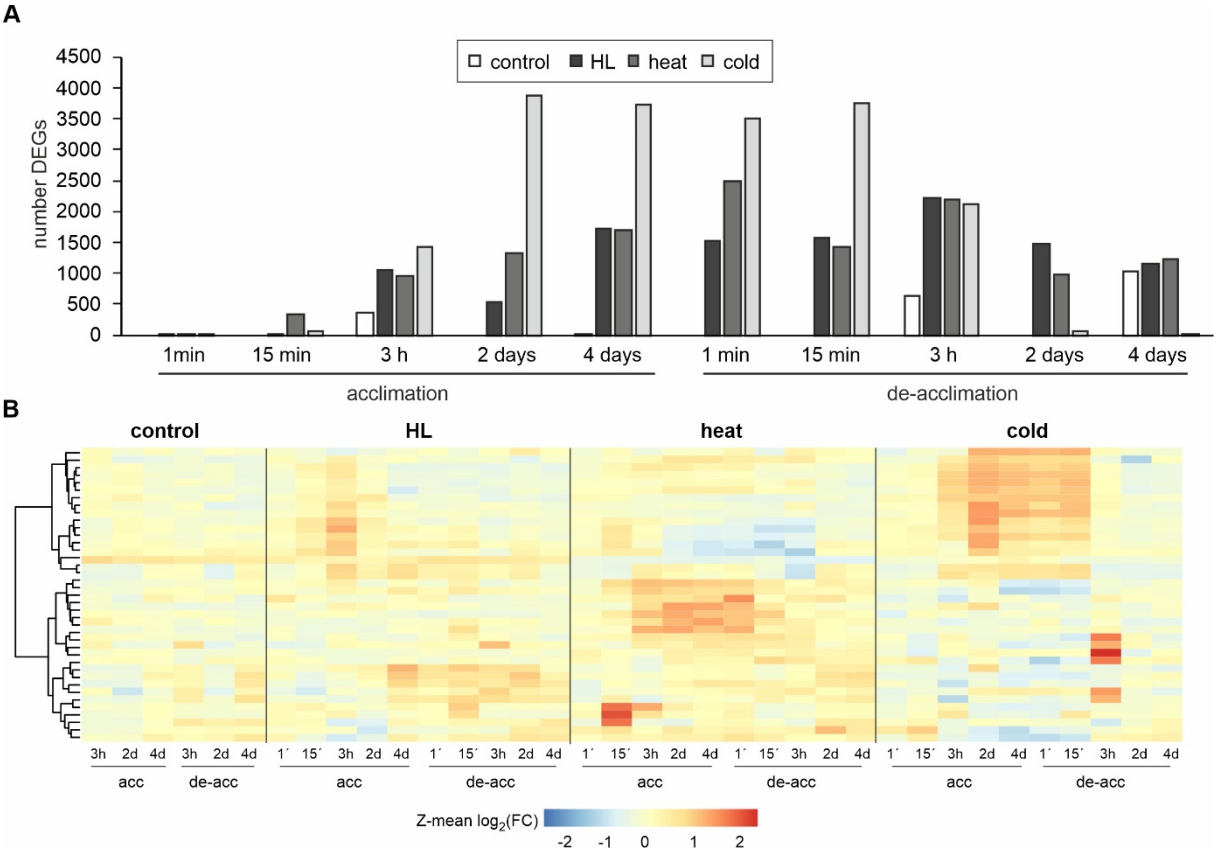

Figure S5. Changes in transcript levels during (de-)acclimation, Related to Figure 3.

(**A**) Differentially expressed genes (DEGs, showing fold changes (FC) ≥ 2 relative to the initial time-point t = 0 (with a false discovery rate [FDR] ≤ 0.05) during control conditions and phases of acclimation and de-acclimation to high light (HL), heat and cold.

(**B**) Heatmap constructed by hierarchical clustering according to the Ward D2 method for DEGs coding for heat-shock proteins and heat-shock factors. Z-means for log_2-converted FC scores under standard conditions, and exposure to HL, heat and cold during acclimation (acc) and de-acclimation (de-acc) phases, were used.

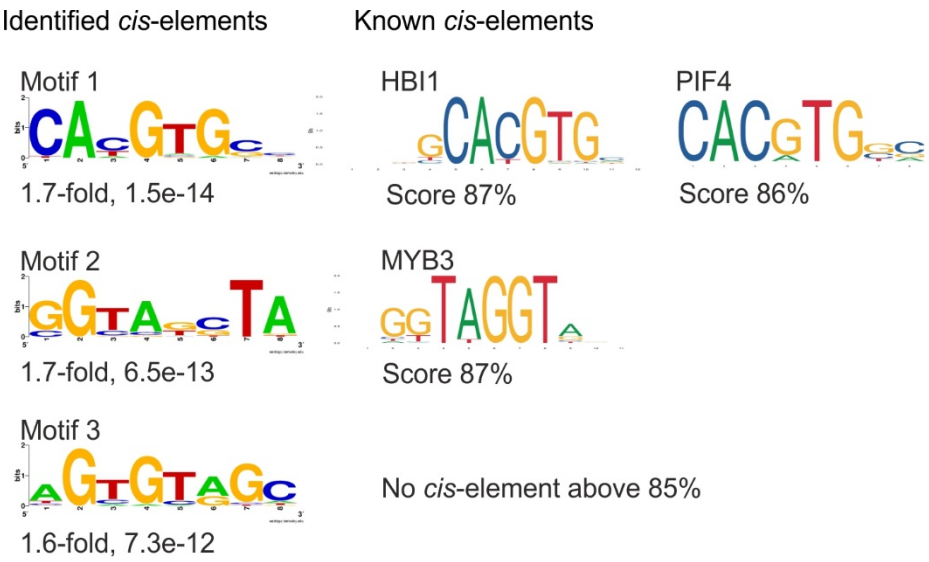

Figure S6. Sequence logos of the most significantly identified *cis*-elements of genes whose expression was regulated under all investigated conditions, Related to Figure 3.

The names of the putative transcription factors binding to the identified *cis*-elements, together with their respective logos, are also shown. Note that motifs can be read in opposite or same direction.

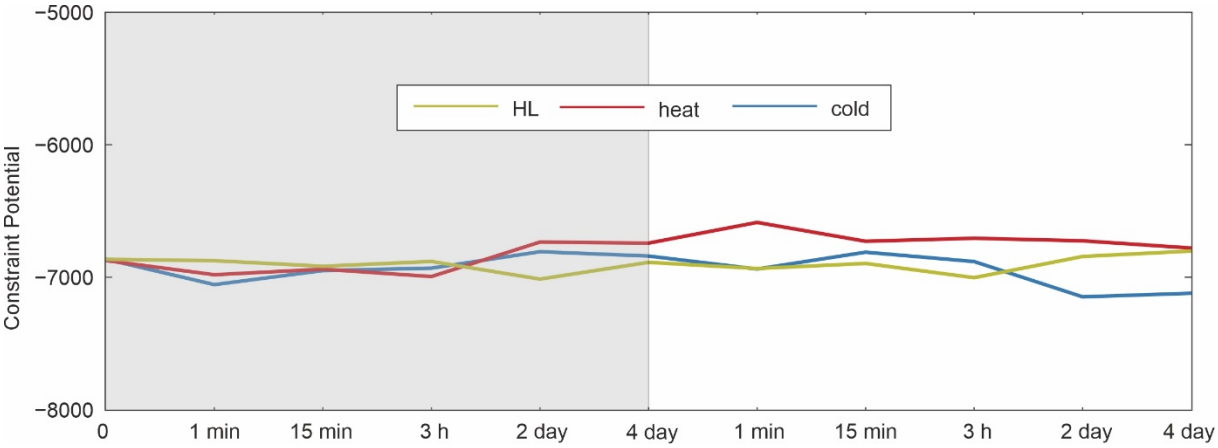

Figure S7. Baseline state for transcriptome changes during (de-)acclimation derived from surprisal analysis, Related to Figure 4.

Time course of the baseline state as determined by surprisal analysis of the transcriptome profiles measured during exposure to high light (HL), heat, or cold treatment and recovery. The acclimation phase is indicated by the gray shading.

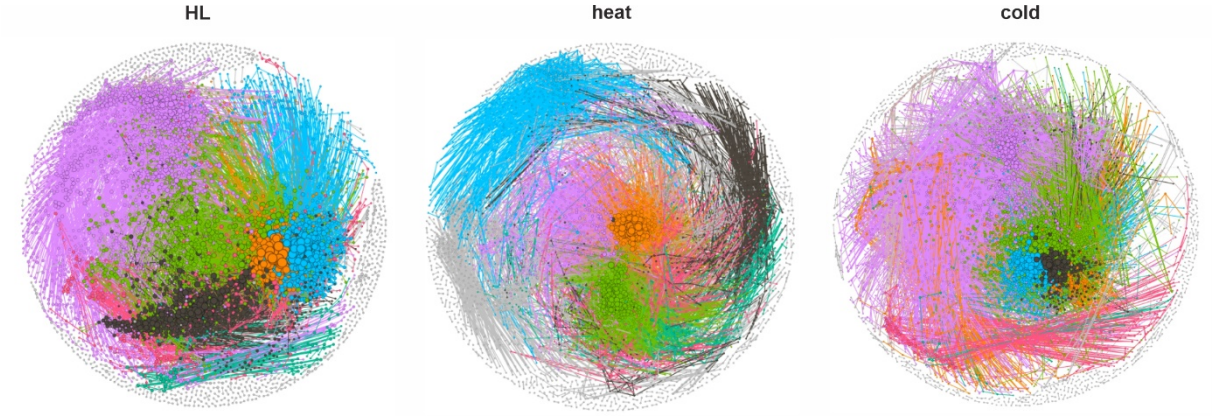

Figure S8. Conditional networks underlying transcriptome changes during (de-)acclimation, Related to Figure 5.

Co-expression of transcriptome changes during acclimation and de-acclimation was assessed by Pearson correlation (R) and mapped onto a reference network. The resulting networks encompass nodes corresponding to transcripts connected by edges ($R \geq 0.9$) as defined in **Experimental Procedures**. The node size is proportional to its degree. Colors indicate network communities. The topological properties of the conditional networks are summarized in **Table S3**.

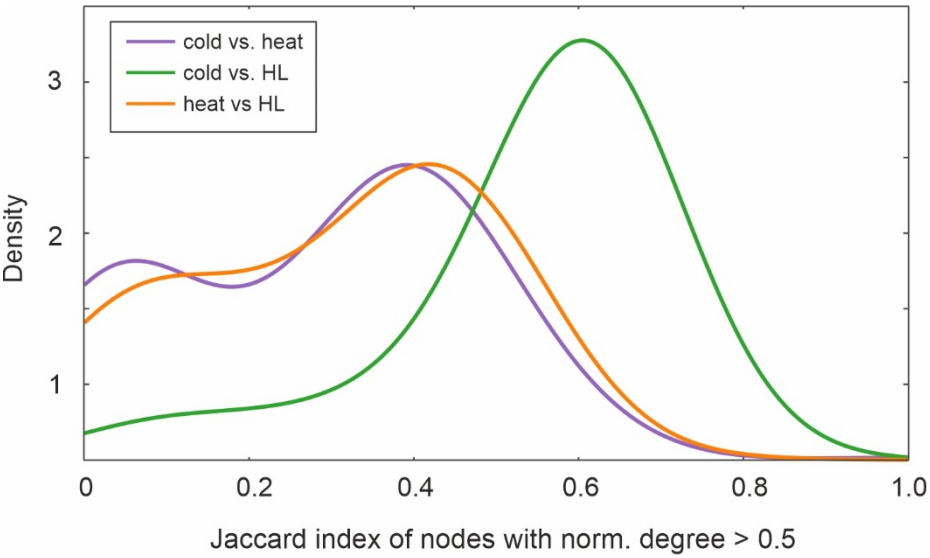

Figure S9. Additional comparison of network topology for transcriptome changes, Related to Figure 5.

Density plot of Jaccard index ($J(A,B)=(|A\cap B|)/(|A|+|B|-|A\cap B|)$, where A is the neighborhood of a node in network A, and B the neighborhood of the same node in network B calculated for the neighborhoods of all nodes in the conditional networks with a normalized degree > 0.5.

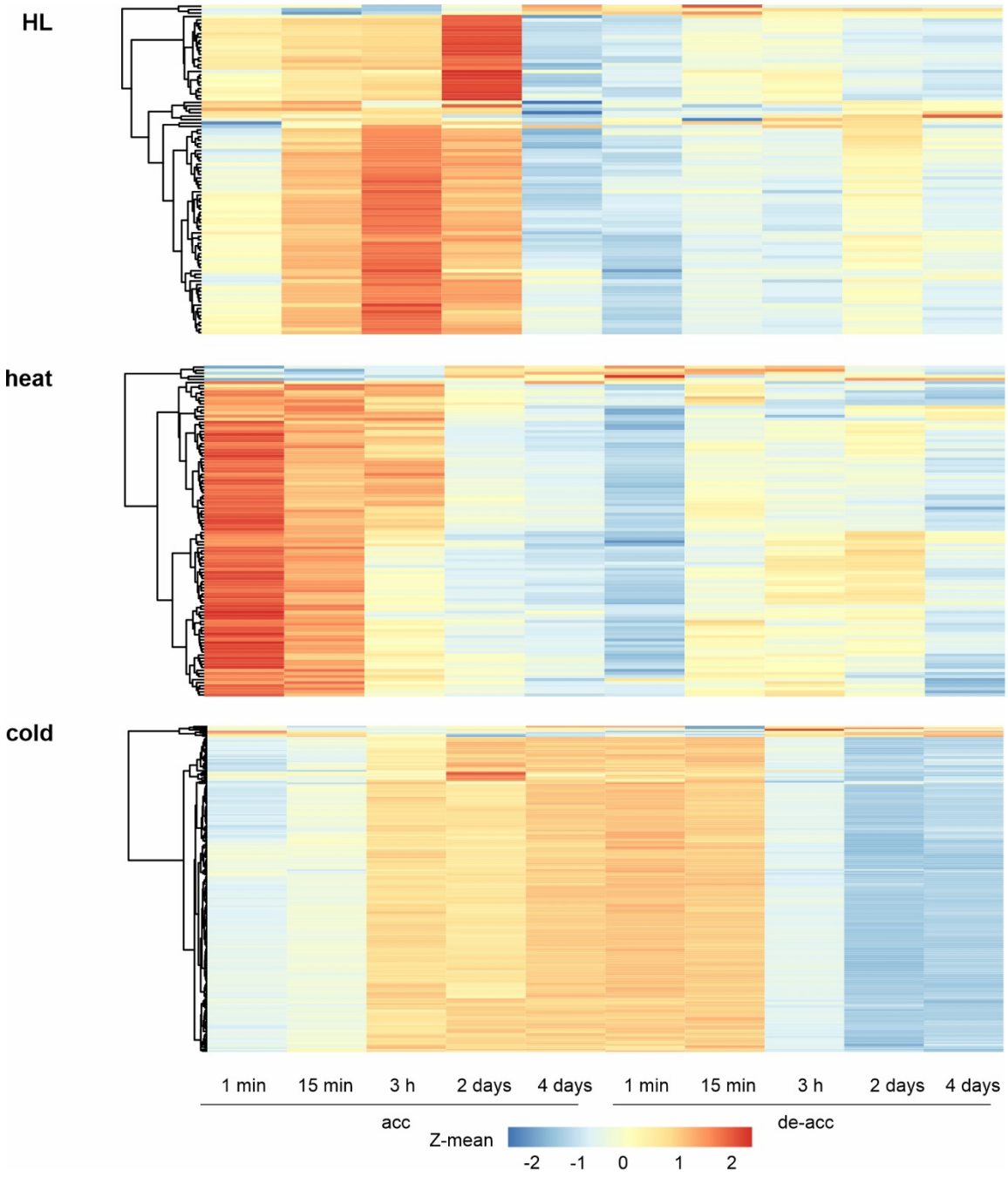

Figure S10. Expression patterns of transcripts for ribosomal proteins (RPs) identified as super-hubs in conditional networks, Related to Figure 6.

Heatmaps obtained by hierarchical clustering according to the Ward D2 method for the RP super-hubs listed in **Table 1**. Z-means were calculated from $\log_2$-transformed fold changes compared to the initial time-point (0 min). acc, acclimation; de-acc, de-acclimation; HL, high light.

Supplemental Tables (Tables S1 to S6, and S9 to S10 are provided as separate Excel files)

Table S7. Overview of DEGs encoding chloroplast proteins, Related to Figure 3. DEGs that behaved similarly in response to high light, heat and cold (de-)acclimation during the central part of the experimental time-courses (from 2 d acclimation to 15 min de-acclimation) and are related to chloroplast functions were sorted according to their molecular roles.

Gene ID	Symbol	Annotation
		Photosynthesis
AT1G29910	Lhcb1.2	Light harvesting chlorophyll a/b binding protein 1.2
AT2G34430	Lhcb1.4	Light-harvesting chlorophyll a/b binding protein 1.4
AT3G27690	Lhcb2.3	Light-harvesting chlorophyll a/b binding protein 2.3
AT5G54270	Lhcb3	Light-harvesting chlorophyll a/b protein 3
AT4G10340	Lhcb5	Light-harvesting complex of photosystem II subunit 5
AT1G15820	Lhcb6	Light-harvesting complex photosystem II subunit 6
AT4G09650	AtpD	F-type H -transporting ATPase subunit delta
AT1G03600	Psb27	Photosystem II family protein
AT3G50820	PsbO2	Photosystem II subunit O2
AT4G21280	PsbQ1	Photosystem II subunit Q1
		Redox
AT1G76100	PETE1	Plastocyanin isoform 1
AT1G20340	PETE2	Plastocyanin isoform 2
AT4G09010	TL29	Ascorbate peroxidase 4
AT1G77490	tAPX	Thylakoidal ascorbate peroxidase
AT3G09580	-	FAD/NAD(P)-binding oxidoreductase family protein
AT1G77510	PDIL1-2	PDI-like 1-2
		Chromophor
AT1G44446	CH1	Pheophorbide a oxygenase family protein
AT1G58290	HEMA1	Glutamyl-tRNA reductase family protein
AT3G14930	HEME1	Uroporphyrinogen decarboxylase
AT4G25080	CHLM	Magnesium-protoporphyrin IX methyltransferase
		Metabolism
AT5G35790	G6PD1	Glucose-6-phosphate dehydrogenase 1
AT1G12900	GAPA-2	Glyceraldehyde 3-phosphate dehydrogenase A subunit 2
AT1G61800	GPT2	Glucose-6-phosphate/phosphate translocator 2
		Others
AT5G57560	TCH4	Xyloglucan endotransglucosylase/hydrolase
AT2G47450	CAO	Chloroplast signal recognition particle component
AT1G31690	-	Copper amine oxidase family protein
AT2G29090	CYP707A2	Cytochrome P450
AT3G09200	-	Ribosomal protein L10 family protein
AT3G50480	HR4	Homolog of RPW8 4
AT1G74710	EDS16	ADC synthase superfamily protein
AT5G01600	FER1	Ferretin 1

Table S8. Topological properties of conditional networks for transcripts, Related to Figure 5. Structural properties of each conditional network, including the numbers of nodes, connections (edges), average degree, modularity and super-hubs (all nodes with more than 100 edges). Note that, due to network complexity, hubs behave as highly interconnected nodes, i.e. *party hubs* (Han et al., 2004) and therefore the product "number hubs x 100" can be higher than the total number of nodes in some networks.

	HL	Heat	Cold
Nodes	5,082	4,634	5,648
Edges	61,800	35,706	110,446
Average degree	16.89	15.41	39.11
Modularity	0.48	0.60	0.40
Super-hubs	371	223	657

Transparent Methods

Plant cultivation and sampling

Arabidopsis thaliana Col-0 seeds were stratified at 4ºC for 2 days, sown in 9x9-cm pots at a density of approximately 50 seeds per pot and cultivated in LED-41 HIL2 cabinets (Percival Scientific, Perry, Iowa, USA) under standard long-day conditions ("control conditions") [LD; 16 h light (80 µmol photons m^{-2} s^{-1}) using 18% white and red LED light intensities (see **Figure S1A** for spectrum) at 22ºC and 8 h darkness at 18ºC]. After 14 days the light intensity was increased to 450 µmol photons m^{-2} s^{-1} (corresponding to 80 % of the white and red LED light intensities) to investigate acclimation to high light (HL). For heat treatment, plants were grown at a constant temperature of 32ºC. In the case of cold treatment, temperature and light intensity were reduced to 4ºC and 35 µmol photons m^{-2} s^{-1}, respectively. In all cases, light was supplied for 18 h per day. Acclimation treatments were started 4 h after light onset and applied for 4 days. Finally, plants were exposed to control conditions for 4 additional days for de-acclimation. Sampling was carried out at the selected time-points (see **Figure S2**) by harvesting of entire shoots following immersion of plantlets in liquid nitrogen. Samples were then ground and stored at -80ºC prior to characterization.

Measurements of physiological parameters

Fresh weight was determined as the mean of groups of 10 plants. Anthocyanin and chlorophyll content was determined as described previously (Mita et al., 1997; Parsons and Strickland, 1963). Chlorophyll *a* fluorescence measurements were conducted using the IMAGING-PAM M-Series instrument (Walz, Effeltrich, Germany) on plants that had been dark-adapted for 30 min. After determining the minimal fluorescence (F_o), a saturation pulse of actinic light (0.5 s; 2700 µmol photons m^{-2} s^{-1}; 450 nm) was applied to determine the maximum fluorescence (F_m) and the maximum quantum yield of photosystem II (PSII) (F_v/F_m; calculated as $(F_m - F_o)/F_m$).

The effective quantum yield of PSII ($\Phi_{II} = (F_m' - F_0')/F_m']$) and non-photochemical quenching (NPQ; calculated as (Fm-Fm')/Fm')) were monitored at 80 μmol photons m^{-2} s^{-1} every 2 sec for 5 min (Armbruster et al., 2010) and the value at 260 s is depicted in the Figures.

RNA isolation, transcriptome profiling and data analysis

Total RNA from aerial parts of plants was isolated using *Trizol* (Invitrogen, Carlsbad, Calif., USA) using 1 part plant to 15 parts reagent (w/v), and purified on *Direct-zol™ RNA MiniPrep Plus* columns (Zymo Research, Irvine, Calif., USA) according to the manufacturer's instructions. RNA integrity and quality were assessed by gel electrophoresis using the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, Calif., USA). Only those samples with an RNA Integrity Number (RIN) ≥ 7 were further processed. Ribosomal RNA depletion, generation of RNA-Seq libraries and 150-bp sequencing of long-non-coding (lnc) RNAs using the paired-end mode were conducted by *Novogene Biotech* (Beijing, China) with standard Illumina protocols. The RNA-Seq libraries were sequenced on an Illumina HiSeq 2500 system (Illumina, San Diego, Calif. USA). Three independent biological replicates were used per time-point and treatment.

RNA-Seq datasets were analyzed as follows: adaptor removal and sequencing quality was carried out with *Trimmomatic* (Bolger et al., 2014) and *FastQC* (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), respectively. Reads were mapped to the Arabidopsis genome (TAIR10) with *Tophat 2.1.1* (Kim et al., 2013) for First Read (FR) unstranded libraries, adjusting the maximum intron length to 3000 bp. Reads were counted with *featureCounts* (Liao et al., 2014) according to the gene annotation in Araport11 (www.araport.org/data/araport11). Differentially expressed genes (DEGs) were obtained with *DESeq2* (Love et al., 2014) by comparing each time-point of the treatments with the initial time point (0 min).

All transcriptome datasets were deposited at the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE125950.

Metabolite isolation, metabolome profiling and data analysis

Polar primary metabolites were identified in 14-day-old plants and at the indicated times after transfer to acclimation and de-acclimation conditions. To this end, metabolites were extracted from 50-mg samples of frozen rosettes (n = 6 plants from 6 independent experiments) and derivatized as described previously (Erban et al., 2007; Lisec et al., 2006; Rossel et al., 2002). Ribitol (0.2 mg mL^{-1} in water) and ^{13}C-labeled sorbitol (0.2 mg mL^{-1} in water) served as internal standards for relative quantification. The derivatized samples were injected into a gas chromatograph coupled to a time-of-flight mass spectrometer (GC-TOF-MS) system (Pegasus HT, Leco, St Joseph, Mich., USA) and chromatographic separation was performed on an Agilent GC 7890A, using a 30 m VF-5ms column with 10 m EZ-Guard column. Mass spectra were recorded at 20 scans s^{-1} with an 50-800 m/z scan range and evaluated using ChromaTOF 4.5 and TagFinder 4.1 (Luedemann et al., 2008) and the compounds were manually annotated based on the Golm Metabolome Database (Kopka et al., 2005).

Surprisal analysis

Surprisal analysis is a thermodynamic approach that yields a biophysicochemical understanding and quantitative characterization of biological systems using a molecule-centered approach. The key step in surprisal analysis is the definition of a balanced state, i.e., the steady-state of the system that has the maximum entropy. Surprisal analysis then allows one to identify deviations of molecule levels with respect to the balanced state, which are quantified by constraints that characterize their responses. The surprisal $I(x)$ of each individual transcript $x_1 \ldots x_i$ at the time-point t is defined as the deviation from the individual transcript's

contribution to the baseline state x_i^0: $I(x) = -\ln(\frac{x_i(t)}{x_i^0})$. This is achieved by fitting the surprisal using a number of terms, $-\sum_{\alpha=1} G_{i\alpha}\,\lambda_\alpha(t)$, where α is the index of the constraint, $G_{i\alpha}$ is the weight of the transcript x_i in constraint G_α, and $\lambda_\alpha(t)$ is the Lagrange multiplier for G_α that is being varied to find the best fit. This is achieved as described by Remacle et al. (2010) using the singular value decomposition. Raw count data from the transcriptome datasets for each experimental condition were averaged across replicates. To deal with counts equal to 0, all counts were increased by 1. Subsequently, the resulting count means were normalized using the FPKM (Fragment Per Kilobase Million) method. Surprisal analysis was performed as described by Remacle et al. (2010) on the preprocessed datasets transformed into the natural logarithm space using the F# implementation available in the *FSharp.Stats* package (https://github.com/CSBiology/FSharp.Stats @ v0.1.1).

For constraint-based time-course comparisons between experimental conditions, linear regression of each constraint potential was calculated against all conditions using the *FSharp.Stats* package. The coefficient of determination (R^2) of these regressions was then weighted using the mean of the respective constraint weights (the singular values obtained by surprisal analysis).

Construction of conditional networks

Correlation Networks of the respective conditions were elaborated from the transcription profiles. To this end, pairwise Pearson correlations (R) among the mean FPKM values for each transcript were computed, and the subsequent correlation matrix was used to construct the network. Subsequently, Random Matrix Theory (RMT) (Luo et al., 2007) implemented in the *FSharp.Stats* package was used to find an optimal threshold to filter out spurious correlations and noise. Accordingly, absolute values of R = 0.9003, 0.9106 and 0.9116 were used for the data from HL, heat and cold treatments, respectively. To further constrain the correlation

network, the topological overlap of the filtered correlation networks with the structural and functional reference network ARANet v2 (Lee et al., 2015) (https://www.inetbio.org/aranet/downloadnetwork.php) was computed by filtering edges that were not present in the reference network. The resulting networks contain specific information about the experimental condition and general information about structure- function relations in *Arabidopsis thaliana* and are therefore termed conditional networks. Networks were visualized and analyzed in Gephi (https://gephi.org/).

To assess conditional network similarity between experimental conditions, all node degrees were normalized to the highest degree and filtered for overlapping nodes for each comparison. Pearson correlations of normalized node degrees of these overlapping nodes were calculated using the *FSharp.Stats* package.

Further bioinformatic analyses

Partial Least Squares (PLS) regression analysis was performed using the *plsdepot* R package (https://CRAN.R-project.org/package=plsdepot @ v0.1.17), extracting the first two components without crossvalidation, and using the constraint potentials obtained by surprisal analysis as predictors for the physiological parameters of the respective condition (responses).

Differences in transcripts were estimated according to DESeq2, as mentioned above. Metabolome datasets were filtered by one-way ANOVA (FDR ≤ 0.05) to exclude differences due to sample variability, and Student´s t-test was applied. In all cases, pairwise comparisons of each time-point for the treatments against the respective control were considered. Significant differences in transcripts (DEGs) and metabolites (SAMs) were defined by applying a cut-off of an absolute $\log_2$-fold change (FC) ≥ 1 and a false discovery rate (FDR) of ≤ 0.05 after applying the multiple testing Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Heat maps with hierarchical clustering according to the Ward D2 method were elaborated using Z-mean values using the *pheatmap* package integrated in RStudio

(https://www.bioconductor.org). Venn diagrams were elaborated in the interface Venny 2.1 (Oliveros, 2007-2015) and the significance of overlaps was calculated with the RStudio package *SuperExactTest* (Wang et al., 2015). Gene Ontology (GO) enrichments were obtained from the *Database for Annotation, Visualization and Integrated Discovery* (DAVID) (Huang da et al., 2009a, b), applying a cut-off of 2-fold enrichment compared to the expected frequency in the Arabidopsis genome and an FDR (Benjamini-Hochberg) of ≤ 0.05. Non-redundant GO terms were selected in the interface REVIGO using the small similarity (0.5) parameter (Supek et al., 2011). Significance of enrichments were calculated with Fisher's exact test with the *stat* and package (RStudio).

Supplemental References

Armbruster, U., Zühlke, J., Rengstl, B., Kreller, R., Makarenko, E., Rühle, T., Schünemann, D., Jahns, P., Weisshaar, B., Nickelsen, J., *et al.* (2010). The Arabidopsis thylakoid protein PAM68 is required for efficient D1 biogenesis and photosystem II assembly. Plant Cell *22*, 3439-3460.

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

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Erban, A., Schauer, N., Fernie, A.R., and Kopka, J. (2007). Nonsupervised construction and application of mass spectral and retention time index libraries from time-of-flight gas chromatography-mass spectrometry metabolite profiles. Methods Mol Biol *358*, 19-38.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., *et al.* (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature *430*, 88-93.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmuller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M., *et al.* (2005). GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics *21*, 1635-1638.

Lee, T., Yang, S., Kim, E., Ko, Y., Hwang, S., Shin, J., Shim, J.E., Shim, H., Kim, H., Kim, C., *et al.* (2015). AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species. Nucleic Acids Res *43*, D996-1002.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A.R. (2006). Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc *1*, 387-396.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Luedemann, A., Strassburg, K., Erban, A., and Kopka, J. (2008). TagFinder for the quantitative analysis of gas chromatography--mass spectrometry (GC-MS)-based metabolite profiling experiments. Bioinformatics *24*, 732-737.

Luo, F., Yang, Y., Zhong, J., Gao, H., Khan, L., Thompson, D.K., and Zhou, J. (2007). Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. BMC Bioinformatics *8*, 299.

Mita, S., Murano, N., Akaike, M., and Nakamura, K. (1997). Mutants of *Arabidopsis thaliana* with pleiotropic effects on the expression of the gene for beta-amylase and on the accumulation of anthocyanin that are inducible by sugars. Plant J *11*, 841-851.

Oliveros, J.C. (2007-2015). Venny. An interactive tool for comparing lists with Venn's diagrams. (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

Parsons, T.R., and Strickland, J.D.H. (1963). Discussion of spectrophotometric determination of marine-plant pigments, with revised equations for ascertaining chlorophyll-a and carotenoids. J Marine Res *21*, 105-156.

Remacle, F., Kravchenko-Balasha, N., Levitzki, A., and Levine, R.D. (2010). Information-theoretic analysis of phenotype changes in early stages of carcinogenesis. Proc Natl Acad Sci U S A *107*, 10324.

Rossel, J.B., Wilson, I.W., and Pogson, B.J. (2002). Global changes in gene expression in response to high light in Arabidopsis. Plant Physiol *130*, 1109-1120.

Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800.

Wang, M., Zhao, Y., and Zhang, B. (2015). Efficient test and visualization of multi-set intersections. Sci Rep *5*, 16923.