*Journal of Experimental Botany*, Vol. 76, No. 10 pp. 2743–2762, 2025
https://doi.org/10.1093/jxb/eraf005 Advance Access Publication 8 January 2025

RESEARCH PAPER

# Chronology of transcriptome and proteome expression during early Arabidopsis flower development

**Raquel Álvarez-Urdiola[1], José Tomás Matus[1,2], Víctor Manuel González-Miguel[1], Martí Bernardo-Faura[1,†], and José Luis Riechmann[1,3,*]**

[1] Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Cerdanyola del Vallès, 08193 Barcelona, Spain
[2] Institute for Integrative Systems Biology (I2SysBio), Universitat de València-CSIC, Paterna, 46908 Valencia, Spain
[3] Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain
[†] Present address: Discovery and Translational Sciences, Clarivate Analytics, 08025 Barcelona, Spain.

* Correspondence: joseluis.riechmann@cragenomica.es

Received 18 September 2024; Editorial decision 3 January 2025; Accepted 6 January 2025

Editor: Rainer Melzer, University College Dublin, Ireland

## Abstract

**The complex gene regulatory landscape underlying early flower development in Arabidopsis has been extensively studied through transcriptome profiling, and gene networks controlling floral organ development have been derived from the analyses of genome-wide binding of key transcription factors. In contrast, the dynamic nature of the proteome during the flower development process is much less understood. In this study, we characterized the floral proteome at different stages during early flower development and correlated it with unbiased transcript expression data. Shotgun proteomics and transcript profiling were conducted using an APETALA1 (AP1)-based floral induction system. A specific analysis pipeline to process the time-course proteomics data was developed. In total, 8924 proteins and 23 069 transcripts were identified. Co-expression analysis revealed that RNA–protein pairs clustered in various expression pattern modules. An overall positive correlation between RNA and protein level changes was observed, but subgroups of RNA–protein pairs with anti-correlated gene expression changes were also identified and found to be enriched in hormone-responsive pathways. In addition, the RNA-seq dataset reported here further expanded the identification of genes whose expression changes during early flower development, and its combination with previously published AP1 ChIP-seq datasets allowed the identification of additional direct and high-confidence targets of AP1.**

**Keywords:** APETALA1, Arabidopsis, co-expression analysis, flower development, proteome, target genes, transcriptome.

## Introduction

The onset of flower formation is a key regulatory event that is under a tight and widely conserved genetic control, and the identification of regulators of floral initiation and development showed that many of them encode transcription factors (Theissen *et al.*, 2016; Thomson and Wellmer, 2019). The downstream targets of some of these factors have been elucidated and the architecture of gene regulatory networks inferred, especially in Arabidopsis (Wils and Kaufmann, 2017; Chen *et al.*, 2018). However, the characterization of the proteome—to complement the transcriptome and provide a

© The Author(s) 2025. Published by Oxford University Press on behalf of the Society for Experimental Biology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

more comprehensive view of the process (e.g. Kumar *et al.*, 2016; Mergner *et al.*, 2020a; Song *et al.*, 2023)—is still lacking. Proteome composition can vary rapidly during developmental stages or in response to changes in growth conditions (Mergner and Kuster, 2022).

The integration of MS and RNA-sequencing (RNA-seq) has been used to study the correlation, or lack thereof, between transcriptome and proteome data in organisms such as human and mouse (e.g. Edfors *et al.*, 2016; Hoogendijk *et al.*, 2019; Wang *et al.*, 2019; Jiang *et al.*, 2020; Gerdes Gyuricza *et al.*, 2022; Sidhaye *et al.*, 2023). The extent to which transcript and protein levels—and changes in them—correlate is still under debate (Liu *et al.*, 2016; Bishop and Hawley, 2022), with some studies concluding that there is not a strong correlation, whereas in others such correlation is more apparent. Few studies have specifically addressed this issue in developmental processes in plants, although there are examples of the combination of proteomics and transcriptomics to analyze plant development. In the case of Arabidopsis, these include embryogenesis (Huang *et al.*, 2022), seed germination (Bai *et al.*, 2021), leaf development (Mergner *et al.*, 2020b; Omidbakhshfard *et al.*, 2021), and late-stage flower development (Mergner *et al.*, 2020b). In other plants, combined transcriptome–proteome analyses have been used to study, for instance, fruit development and ripening in pear (Wang *et al.*, 2023).

A lack of correlation between transcript and protein levels may result from missing values in label-free quantitative proteomics; the fact that the sensitivity of signal extraction and the quantification techniques differ between proteomic and transcriptomic analyses; or may stem from difficulties in obtaining truly comparable samples and datasets. However, observed differences in transcript and protein abundance might also be caused by post-translational regulation of protein levels (Vogel and Marcotte, 2012), or by their different expression and degradation kinetics, as longer protein half-lives would buffer changes in mRNA levels (Raj *et al.*, 2006; Taniguchi *et al.*, 2010; Csardi *et al.*, 2015; Oliva-Vilarnau *et al.*, 2020). In this respect, time-course studies may be an approach for addressing this gap, as there might be a temporal shift between mRNA and protein level changes, and successive analyses at different time points could allow the discovery of correlative behaviors of protein and mRNA levels through time (Tarazona *et al.*, 2018; Hoogendijk *et al.*, 2019; Bai *et al.*, 2021; Omidbakhshfard *et al.*, 2021; Huang *et al.*, 2022; Wang *et al.*, 2023; Fan *et al.*, 2024).

APETALA1 (AP1)-based floral induction systems have been used to study the early stages of flower development in Arabidopsis through genomic approaches (O'Maoiléidigh *et al.*, 2023). The MADS-domain transcription factor AP1 is a key regulator of floral meristem identity and activates flower development (Mandel *et al.*, 1992; Liljegren *et al.*, 1999; Ng and Yanofsky, 2001). The integration of the transcriptome, cistrome, and epigenome associated with this transcription factor led to a better understanding of the early flower development regulatory network (Wellmer *et al.*, 2006; Kaufmann *et al.*, 2010; Wuest *et al.*, 2012; Ó'Maoiléidigh *et al.*, 2013; Pajoro *et al.*, 2014; Ryan *et al.*, 2015; Chen *et al.*, 2018). In this work, the floral induction system pAP1:AP1-GR *ap1 cal* (Ó'Maoiléidigh *et al.*, 2013, 2015) was used to characterize non-biased abundance changes of transcripts and proteins in a temporal sequence following the activation of the early flower development program. To facilitate the transcript–protein comparison, an imputation approach was developed for dealing with missing data in the quantitative proteomics dataset. Gene and protein expression was analyzed on a genome-wide scale, identifying transcript–protein pairs with significant expression changes for both molecules at different stages of flower development. In addition, new genes differentially expressed during early flower development were identified, and the combined analysis of the transcriptome and published AP1 ChIP-seq datasets allowed the identification of additional AP1 direct target genes.

## Materials and methods

### Plant lines and plant growth conditions

The pAP1:AP1-GR *ap1 cal* (O'Maoiléidigh *et al.*, 2023) plants were grown on a soil:vermiculite:perlite mixture in growth chambers at 21 °C under long-day conditions (16 h light, 8 h darkness), after a 4 d period of stratification at 4 °C in darkness.

### Tissue collection

For RNA-seq and LC-MS/MS experiments, 4-week-old pAP1:AP1-GR *ap1 cal* plants were used. Four biological replicates were generated for each time point. For each replicate, from ~80 (D0) to 40 plants (D5) were used to obtain 300–500 µg of total protein. Inflorescence tissue was collected using jeweler's forceps as previously described (Wellmer *et al.*, 2006). For induction, inflorescences were treated with a dexamethasone (DEX) induction solution [2 µM DEX, 0.01% (v/v) ethanol, and 0.01% (v/v) Silwet L-77]. Using plastic pipettes, the solution was directly applied onto the inflorescences so that the cauliflower-like structures were completely drenched. First induction was performed 8 h after lights on, and daily inductions at 4 h after lights on. Samples were collected immediately after solution application (D0), as well as at 1, 2, 3, 4, and 5 d (D1–D5) after the first treatment (Fig. 1A).

### Protein extraction

Protein and RNA extractions had common initial steps (as described in Alvarez-Urdiola *et al.*, 2023). Tissue was ground in liquid nitrogen. For each time point, ~0.25 g of plant material was used. Ground material was resuspended in 1 ml of Trizol and incubated on ice for 5 min. Then, 200 µl of chloroform were added and thoroughly mixed by vortexing. After 5 min incubation on ice, samples were centrifuged at 4 °C for 15 min at maximum speed. Upon centrifugation, three phases are formed, the aqueous phase contains RNA (~550 µl, transparent), the interphase, DNA (white), and the organic phase, protein and lipids (~450 µl, pink). After the aqueous phase was transferred to a new microcentrifuge tube (see RNA extraction), 300 µl of 100% ethanol were added to the organic phase to continue with protein extraction, and the mix was incubated on ice. Samples were centrifuged for 10 min at 2000 *g* to separate DNA from proteins. The supernatant was placed in a clean 2 ml microcentrifuge tube, 1 ml of pure isopropanol was added, and samples were incubated at room temperature for 10 min. After a 10 min centrifugation at 12 000 *g*

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

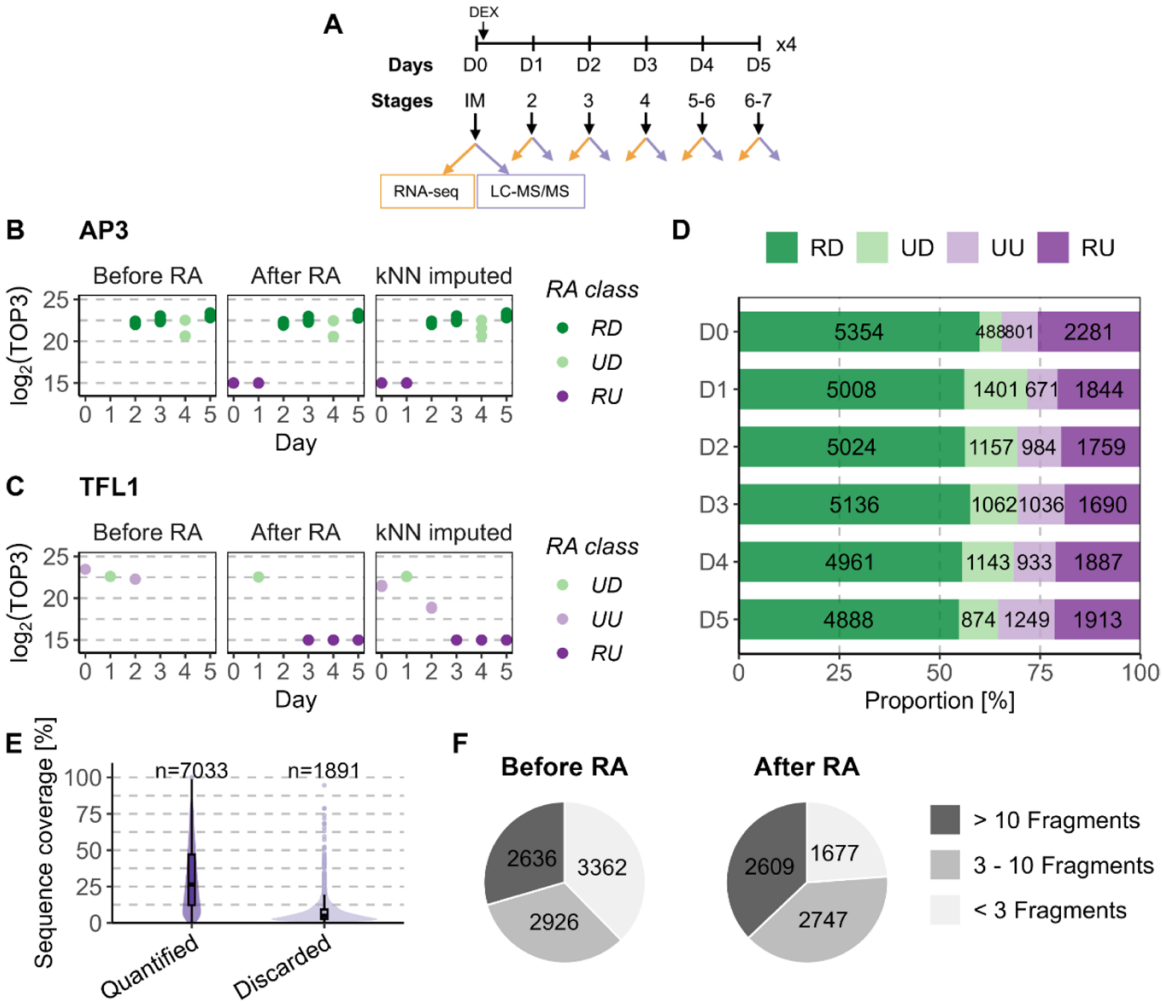


**Fig. 1.** Overview of experimental design and the reliability analysis pipeline for addressing missing values in proteomics. (A) Experimental setup for the temporal transcriptomic and proteomic analysis of flower development. Inflorescence samples of four biological replicates of 40–80 pAP1:AP1-GR *ap1 cal* plants each were collected immediately after DEX application (D0), and at 1, 2, 3, 4, and 5 d (D1–D5) after treatment. Imputation of missing values considering their biological context: time-series abundances ($log_2$TOP3) of AP3 (B) and TFL1 (C) before and after the reliability analysis (RA), and after k-nearest neighbor (kNN) imputation. (D) Proportion of proteins considered as reliably detected (RD), unreliably detected (UD), unreliably undetected (UU), and reliably undetected (RU) for each time point. (E) Distribution of peptide-based sequence coverage of proteins that were reliably or unreliably detected in at least one time point—day—(quantified), and of those that were reliably or unreliably undetected at every time point (discarded). (F) Pie charts displaying percentage and number of proteins identified by <3, 3–10, or >10 peptidic fragments before and after the reliability analysis (the peptide-based coverage of 1685 of the 1891 discarded proteins was lower than three peptides per protein).

at 4 °C the supernatant was discarded. The pellet was resuspended in 2 ml of a solution of 0.3 M guanidine in 95% (v/v) ethanol for washing, and sonicated during 5 min. Samples were centrifuged 8000 *g* for 5 min at 4 °C. This washing procedure was repeated twice. The final pellet was stored at –20 °C and washed with 90% (v/v) ethanol before the final resuspension in a 70% (v/v) acetonitrile buffer for LC-MS/MS.

## RNA extraction

The organic phase (see Protein extraction) was transferred to a clean microcentrifuge tube and mixed vigorously with one volume of pure isopropanol. After 15 min of incubation on ice, samples were centrifuged at 4 °C for 10 min and the supernatant was discarded. Each pellet was resuspended in 750 µl of 3 M LiCl, incubated on ice for 10 min, and centrifuged at 4 °C for 10 min at maximum speed. The supernatant was discarded, and each pellet was gently washed with 500 µl of 85% (v/v) ethanol. The last centrifugation was performed at 4 °C for 10 min at maximum speed, and the supernatant was discarded. Each pellet was resuspended in 21 µl of diethylpyrocarbonate (DEPC)-treated water after drying. RNA concentrations were determined with a NanoDrop 1000 spectrophotometer.

## Liquid chromatography–tandem mass spectrometry

### *Sample preparation*

Samples were reduced with DTT (30 nmol, 37 °C, 60 min) and alkylated in the dark with iodoacetamide (60 nmol, 25 °C, 30 min). The resulting protein extract was first diluted to 2 M urea with 200 mM ammonium bicarbonate for digestion with endoproteinase LysC (1:10 w:w, 37 °C, overnight, Wako, cat # 129-02541), and then diluted 2-fold with

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

200 mM ammonium bicarbonate for trypsin digestion (1:10 w:w, 37 °C, 8 h, Promega cat # V5113). After digestion, peptide mix was acidified with formic acid and desalted with a MicroSpin C18 column (The Nest Group, Inc.) prior to LC-MS/MS analysis.

*Chromatographic and mass spectrometric analysis*

Samples were analyzed using an LTQ-Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an EASY-nLC 1000 [Thermo Fisher Scientific (Proxeon), Odense, Denmark]. Peptides were loaded directly onto the analytical column and were separated by reversed-phase chromatography using a 50 cm column with an inner diameter of 75 μm, packed with 2 μm C18 particles. Chromatographic gradients started at 95% buffer A and 5% buffer B with a flow rate of 300 nl $min^{-1}$ for 5 min, and gradually increased to 22% buffer B and 78% A in 79 min and then to 35% buffer B and 65% A in 11 min. After each analysis, the column was washed for 10 min with 10% buffer A and 90% buffer B. Buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in acetonitrile.

The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.4 kV and source temperature at 275 °C. Ultramark 1621 was used for external calibration of the Fourier-transform mass analyzer prior to the analyses, and an internal calibration was performed using the background polysiloxane ion signal at $m/z$ 445.1200. The acquisition was performed in data-dependent acquisition (DDA) mode and full MS scans with 1 micro scan at a resolution of 120 000 were used over a mass range of $m/z$ 350–1500 with detection in the Orbitrap mass analyzer. Auto gain control (AGC) was set to 1E5 and charge state filtering disqualifying singly charged peptides was activated. In each cycle of data-dependent acquisition analysis, following each survey scan, the most intense ions above a threshold ion count of 10 000 were selected for fragmentation. The number of selected precursor ions for fragmentation was determined by the 'Top Speed' acquisition algorithm and a dynamic exclusion of 60 s. Fragment ion spectra were produced via high-energy collision dissociation (HCD) at a normalized collision energy of 28% and they were acquired in the ion trap mass analyzer. AGC was set to 1E4, and an isolation window of 1.6 $m/z$ and a maximum injection time of 200 ms were used. All data were acquired with Xcalibur software v4.1.31.9. Digested BSA (New England Biolabs cat # P8108S) was analyzed between each sample to avoid sample carry-over and to ensure stability of the instrument, and QCloud (Chiva *et al.*, 2018) was used to control instrument longitudinal performance during the project.

*Data processing*

Acquired spectra were analyzed using the Proteome Discoverer software suite (v2.0, Thermo Fisher Scientific) and the Mascot search engine (v2.5 Matrix Science) (Perkins *et al.*, 1999). The data were searched against a UniProt Arabidopsis database plus a list of common contaminants (Beer *et al.*, 2017) and all the corresponding decoy entries. For peptide identification, a precursor ion mass tolerance of 7 ppm was used for MS1 level, trypsin was chosen as enzyme, and up to three missed cleavages were allowed. The fragment ion mass tolerance was set to 0.5 Da for MS2 spectra. Oxidation of methionine and N-terminal protein acetylation were used as variable modifications, whereas carbamidomethylation on cysteines was set as a fixed modification.

The false discovery rate (FDR) in peptide identification was set to a maximum of 5%. Peptide quantification data were retrieved from the 'Precursor ion area detector' node from Proteome Discoverer (v2.0) using 2 ppm mass tolerance for the peptide extracted ion current (XIC). Protein abundance in each condition was estimated using the average of the three most intense peptides per protein group (TOP3) (Silva *et al.*, 2006). The raw proteomics data have been deposited in the PRIDE repository (Perez-Riverol *et al.*, 2022) with the dataset identifier PXD038980. For subsequent statistical analysis, median normalization was performed by subtracting from each logged value the sample median and adding the global dataset median. Replicate 1 of Day 0 (R1D0) highly differed from the rest (Supplementary Fig. S1), so it was removed from the dataset, as were the 165 proteins that were only detected in this sample.

*Not assigned values: reliability analysis*

For the reliability analysis, each time point for a protein was classified as reliably or unreliably detected or undetected depending on its number of not assigned (NA) values and the number of NA values of its immediately adjacent days (neighbors). Days 0 and 5 were considered as reliably undetected when all replicates were NA, and days 1–4, besides that, must have at least one neighbor with two or more NA values. Those were considered as MNAR (missing not at random) values, and NA values were replaced by the minimum of detection of the dataset (deterministic minimum imputation method; Meleth *et al.*, 2005). Days with one or no NA values were defined as reliably detected and their abundance values were kept. Finally, days with ≥2 NA values were classified as unreliably detected when they had at least one neighbor with ≤2 NA values, keeping their quantification values; otherwise, they were classified as unreliably undetected, and quantification values were replaced by NA values in all replicates. All those proteins which were reliably or unreliably undetected in every time point were discarded. The remaining NA values were estimated by k-nearest neighbors (kNN) imputation (k=10) (Troyanskaya *et al.*, 2001) (Figs 1B, C, 2).

## RNA-seq experiments

The 24 samples were sequenced on an Illumina HighSeq 2000 instrument. Cleaned reads together with the transcriptome of Arabidopsis (TAIR10) were used to quantify gene expression at the transcript level, in counts and transcripts per million (TPMs) using the software Salmon (v0.12.0). The quantification data were grouped so that genes instead of transcripts were analyzed, using the tximport package in R. Genes that had <10 counts across all the samples were removed to facilitate further analyses (DESeq2 package in R). Library size normalization factors and dispersions were estimated using the DEseq function from the package DESeq2 in R. Dispersion estimates for all genes were obtained considering the information for each gene separately.

## Representative proteins and genes: markers, 'supermarkers', and AP1 targets

A group of 69 proteins were selected as markers on the basis of the detection of expression of their corresponding genes in previous time-course experiments performed using AP1 floral induction systems, including gene expression profiling using DNA microarrays (Wellmer *et al.*, 2006; Kaufmann *et al.*, 2010) and unpublished RNA-seq data from our laboratory. Marker proteins corresponded with up- or down-regulated genes in the microarray experiments [absolute fold change (FC) ≥2 for the first replicate when comparing days 1 and 0, and Benjamini and Hochberg (BH)-adjusted *P*-value ≤0.05; or in the RNA-seq [absolute FC ≥2 for all replicates when comparing day 2 and 0, and day 4 and 0, and fragments per kilobase of transcript per million mapped reads (FPKM) >1]. The seven 'supermarker' proteins, with similar characteristics to the markers, are transcription factors controlling different aspects of flower development (AP2, AP3, PI, TFL1, CRC, LFY, and FIL-YABI1). Out of the 249 AP1 high-confidence targets defined in Kaufmann *et al.* (2010), 247 were quantified in the RNA-seq experiment reported in our work.

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

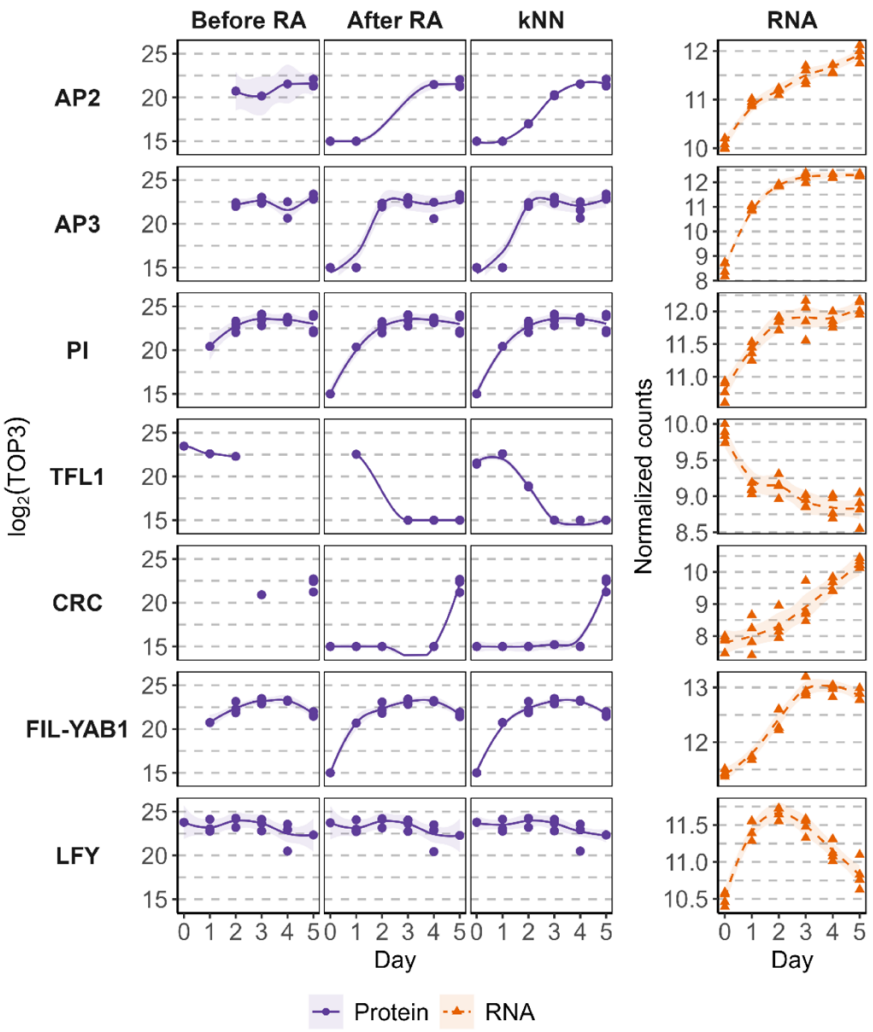


**Fig. 2.** Expression of the 'supermarker' proteins throughout the time-course and effects of data imputation. Protein abundance is shown as $\log_2$(TOP3), before and after the reliability analysis (RA), and after kNN imputation. mRNA abundance as normalized counts is also shown.

## Data analysis

### *Genome and proteome annotations*

Araport11 gene identifiers were mapped to the UniProt *Arabidopsis thaliana* reference proteome (taxon identifier 3702; UP000006548) based on protein sequence. N- and C-terminal peptide sequences were extracted from the Mascot.txt file and filtered for zero missed cleavages (Supplementary Fig. S2A). N-terminal peptides were divided into groups with ($n$=1203) or without ($n$=769) cleavage of the initiator methionine. Then, the frequency of the 20 genetically encoded amino acids at the position after the start codon was calculated and displayed as a pie chart (Supplementary Fig. S2B) for both groups. The percentage of acetylated N-terminal peptides with the same amino acids in the second position was calculated for both groups and represented as a bar plot (Supplementary Fig. S2C).

### *Protein and RNA level variation through time*

An ANOVA was performed for the normalized proteome dataset, followed by a Tukey post-hoc test. Proteins were considered as stage-variant proteins (SVPs) if their BH-adjusted $P$-value, which can be interpreted as the FDR (Benjamini and Hochberg, 1995), was <0.05. For the RNA-seq dataset, a moderated likelihood ratio test (LRT) was applied to get a statistic for ranking genes according to the difference in expression profiles among time points. LRT is a test of significance for differences of any level of the factor. Genes with an adjusted $P$-value (FDR) lower than 0.01 were considered as stage-variant genes (SVGs). The $\log_2$ FC (LFC) in expression between subsequent stages were calculated for all transcripts. The $P$-values were adjusted for multi-hypothesis testing using the BH procedure (FDR). Transcripts with an LFC with an adjusted $P$-value <0.05 at any day to previous day comparison were considered as differentially expressed genes (DEGs). No LFC cut-off was applied.

### *Gene–protein correlations*

A total of 7003 quantified proteins in the MS proteome dataset had their corresponding gene in the RNA-seq transcriptome dataset. Protein–gene pairs were grouped in four subsets: stage variant at RNA and protein levels (SVG–SVP, $n$=973); non-variant genes–SVPs (NVG–SVP, $n$=1006); SVGs–non-variant proteins (SVG–NVP, $n$=1808); and non-variant pairs (NVG–NVP, $n$=3216). Pearson's correlation coefficient ($r_s$; $P$-value ≤0.05) was used to find correlations between protein levels and corresponding genes within and between all time points, using a square matrix. The Spearman's rank correlation coefficient (ρ) of each gene–protein pair

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

individually was used for correlating transcriptome and proteome levels in each subset (SVG–SVP, SVG–NVP, NVP–SVP, and NVG–NVP). The slopes were estimated by ranged major-axis (RMA) regression, which allows errors in both variables and is symmetric, using the R package lmodel2 (Csardi *et al.*, 2015).

Principal component analysis (PCA) of proteome and transcriptome data was performed in R for each normalized dataset separately (proteome before and after the reliability analysis, and transcriptome), but also for their intersection ($n$=7003 transcript–protein pairs), and for the SVG–SVP subset ($n$=973).

#### *Transcript–protein co-expression network analysis*

Transcriptome and proteome dynamics were evaluated by means of weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008). Normalized RNA-seq counts and protein abundance data (after reliability analysis) were separately $z$-score transformed for each subgroup, and WGCNA was performed with a soft-power of 6 signed network. Modules were defined by dynamic tree cut with a minimum size of 10 and deep split of 4. To reduce the final number of modules, those with a similitude superior to 0.9 were merged, leading to the final number of modules that were considered.

#### *Protein–protein interaction network*

Arabidopsis protein–protein interactions were downloaded from STRING (https://stringdb-static.org/download/protein.links.detailed.v11.0/) and IntAct (https://www.ebi.ac.uk/intact/). In addition, we checked which protein–protein interactions between the proteins quantified by MS were annotated in The Arabidopsis Information Resource (TAIR, https://arabidopsis.org), finding 598 interactions. Interaction clusters were determined using the GLay clustering tool (Su *et al.*, 2010) of the ClusterMaker package (Morris *et al.*, 2011) for Cytoscape (Shannon *et al.*, 2003).

Proteins from transcript–protein pairs with similar expression patterns (from the same interaction module), on average, showed higher STRING co-expression scores. To determine if among the protein dataset there were high confidence previously described physical interactions, the STRING data from interactions with high co-expression (above median) were combined with IntAct-registered (Hermjakob *et al.*, 2004) protein–protein interactions. There were 129 self-interacting proteins, 70 protein–protein interactions between proteins in the same module, and 2656 protein–protein interactions between proteins from different trajectory modules.

#### *Function data analysis*

Gene ontology (GO) (Yu *et al.*, 2012) and Kyoto Encyclopedia of Genes and Genomes (KEGG) term enrichments (Kanehisa and Goto, 2000) were performed using clusterProfiler (Yu *et al.*, 2012). Enrichment was determined with Fisher's exact tests followed by Bonferroni–Yekutieli multiple hypotheses test correction. The composition of the interaction clusters and the trajectory modules was also analyzed in the sense of whether they contained proteins from the same annotated family according to TAIR, but there was no significant enrichment in members of any specific family for the clusters. Family annotations were downloaded from TAIR (gene_families_sep_29_09_update.txt).

### Comparison with previous studies

RNA-seq differential expression results were compared with those found in Wellmer *et al.* (2006), Kaufmann *et al.* (2010), Pajoro *et al.* (2014), Ryan *et al.* (2015), and Chen *et al.* (2018). A filter on adjusted $P$-values (<0.05) was applied to the list of common genes to keep only those with significant values.

Novel AP1 high-confidence targets were defined using the RNA-seq data (D1 versus D0 DEGs, adjusted $P$-value <0.05, absolute LFC >0.29) and ChIP-seq data from Kaufmann *et al.* (2010). The criteria for classifying a gene as an AP1 high-confidence target in Kaufmann *et al.* (2010) included (i) containing one or more AP1 ChIP-seq binding sites within 3 kb upstream of the 5' end and 1 kb downstream of the 3' end of the gene, which defined a set of 2298 putative AP1 targets, and (ii) showing robust differential expression in the time-course (absolute FC >1.8). All the 81 high-confidence targets from the set of 249 high-confidence targets in Kaufmann *et al.* 2010) that were detected in the RNA-seq D1 versus D0 comparison were above an absolute LFC of 0.29; therefore, this LFC value was used as a threshold in the search for novel high-confidence targets (the use of two different floral induction systems, i.e. p35:AP1-GR and pAP1:AP1-GR; and technological platforms, i.e. oligonucleotide microarrays versus RNA-seq, in Kaufmann *et al.* and in this study argued against the simple use of the same numerical cut-off values for the comparison, favoring instead the empirical approach indicated above).

For identification of novel AP1 confidence targets, the list of 2710 putative targets identified in Chen *et al.* (2018) (putative targets defined by having ChIP-seq binding sites within ±3 kb of the start of the gene) was compared with the list of all DEGs from all pairwise comparisons that were above an absolute LFC of 0.29.

## Results

### Integrated transcriptome and proteome analyses for early Arabidopsis flower development

An AP1-based floral induction system (pAP1:AP1-GR *ap1 cal*) (Ó'Maoiléidigh *et al.*, 2013, 2023) was used to characterize proteomic (LC-MS/MS) and transcriptomic (RNA-seq) changes during the initial stages of Arabidopsis flower development. In this system, DEX treatment causes the simultaneous transformation of the inflorescence-like meristems of *ap1 cauliflower* double-mutant plants into floral primordia and triggers flower development (O'Maoiléidigh *et al.*, 2023). To study whether mRNA levels were correlated to proteome changes, tissue from DEX-treated plants was collected as previously described (Wellmer *et al.*, 2006), immediately after the treatment and at 1 d intervals up to day 5 afterwards (corresponding to stages 6–7 of flower development, Smyth *et al.*, 1990) (Fig. 1A). Total RNA and protein were extracted from four biologically independent replicates and analyzed through RNA-seq and LC-MS/MS, respectively. In these analyses, a total of 75 244 peptidic fragments from 8924 proteins were identified in at least one sample, and 23 069 transcripts (corresponding to 84% of the Arabidopsis annotated genes) were quantified in the RNA-seq experiments. Overall, 8708 protein-coding genes were identified at both the transcript and protein level. The proteomics data corroborated a substantial number of annotated ORF borders (Supplementary Fig. S2A). The mass spectrometry data covered, on average, ~21% of each protein sequence.

### Development of a reliability analysis pipeline to address missing values in the proteomics dataset

Missing values in mass spectrometry detection data are commonly observed and might stem from low abundance proteins

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

that are below the detection threshold or simply reflect the absence of a protein in a sample [i.e. missing not at random (MNAR) values], instead of technical artifacts [i.e. missing at random (MAR) values]. To compensate for absent data points [not-assigned (NA) values) in the dataset, a reliability analysis pipeline was developed that took advantage of the characteristics of the experimental design (successive time points and replicates). In this pipeline, it was considered that the reliability of protein detection would depend on the number of missing values per time point ($n$=4 biological replicates) in the dataset. A protein at a given time point (i.e. day) was classified as reliably detected (RD), unreliably detected (UD), unreliably undetected (UU), or reliably undetected (RU) depending on the number of replicates for that time point in which the protein showed NAs, and the number of NAs in the immediately adjacent time points (Table 1; Fig. 1B, C). In total, 7033 proteins (out of the initial set of 8924) were considered as 'quantified' (RD or UD for at least one time point), whereas the remaining 1891 proteins were discarded for further analyses because they were classified as either RU or UU at all time points (Fig. 1E, F). At each time point, ~5000 proteins were classified as RD, and a total of 3176 proteins were classified as RD or UD across all time points (Fig. 1D). In the final step of the pipeline, all the remaining NA values were imputed using the kNN algorithm (Fig. 1B, C).

To validate the reliability analysis pipeline, 69 'marker' proteins were selected on the basis that their corresponding genes are known to be up- or down-regulated in floral organs and/or throughout flower development (Wellmer *et al.*, 2006; Kaufmann *et al.*, 2010; Pajoro *et al.*, 2014). In addition, seven 'supermarker' proteins were chosen because they are well-characterized transcription factors related to flower initiation and development (Supplementary Table S1). Approximately 60% of the marker and all of the 'supermarker' proteins were retained as 'quantified' after the reliability analysis (Fig. 2).

The effect of the reliability analysis and kNN imputation was also evaluated by PCA and hierarchical clustering. PCA was performed with the subset of proteins that did not have missing values before performing the reliability analysis and the imputation, and with all the proteins classified as 'quantified' after those data processing steps. After the processing, the variability observed was discretely grouped by time point (Supplementary Fig. S3A, B). Hierarchical clustering performed before and after the reliability analysis and kNN imputation showed that after those steps, replicates all clustered together by day, with adjacent days also clustering together (Supplementary Fig. S3C). This clearer separation through time and chronologically according to the flower developmental stages demonstrated the robustness of the employed method of LC-MS/MS data generation followed by the reliability analysis and imputation approach.

## Stage-variant proteins showed different abundance patterns over time

To determine which of the 7033 quantified proteins showed significant differences in expression levels in the time-course experiment, an ANOVA was performed, resulting in the classification of a total of 2037 proteins as SVPs (FDR=5%), of which 1430 were considered as RU at least at one time point (Table 2; Supplementary Table S2). SVPs presented different expression patterns that can be summarized as: (i) increased expression over time (groups A.1 and A.2); (ii) reduced expression over time (B.1 and B.2); (iii) transient expression at middle time points (C.1 and C.2); and (iv) gap of expression at middle time points (D.1 and D.2) (Fig. 3A; Supplementary Table S2).

**Table 1.** Data classification in the reliability analysis

| | Initial/final time point (D0/D5) | Intermediate time point (D1–D4) |
|---|---|---|
| **Reliably detected (RD)** | 0 or 1 NA | |
| **Unreliably detected (UD)** | 2 or 3 NAs+supporting neighbor | |
| **Unreliably undetected (UU)** | 2 or 3 NAs+unsupporting neighbor | 2 or 3 NAs+unsupporting neighbor<br>OR<br>4 NAs+supporting neighbor |
| **Reliably undetected (RU)** | 4 NAs | 4 NAs+unsupporting neighbor |

Supporting neighbors are those adjacent time points with 0, 1, or 2 NAs, while unsupporting neighbors have 3 or 4 NAs, in the corresponding biological replicates.

**Table 2.** Summary of stage-variant proteins according to their classification in the reliability analysis

| | RD or UD in all time points | RU or UU at some time points | Total |
|---|---|---|---|
| **SVPs** | 490 | 1547 | 2037 |
| **NVPs** | 2686 | 2310 | 4996 |
| Total | 3176 | 3857 | **7033** |

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

## Patterns of gene expression changes throughout the time-course experiment

A PCA comprising all the 23 069 transcripts identified in the RNA-seq experiment separated the early flower developmental stages by time point (Supplementary Fig. S4). A moderated LRT was applied to obtain a statistical metric for ranking genes according to the differences in their expression profiles. This analysis (LRT with adjusted *P*-value ≤0.01) led to the identification of 8125 transcripts with a significant variation through time, herein referred to as SVGs. These SVGs can be considered as 'related to' or 'influenced by' AP1 activity (Supplementary Table S3).

The 8125 genes defined as SVGs showed, like the proteins defined as SVPs, four main different transcript accumulation patterns: (i) increment in expression through time (group A); (ii) higher expression during mid-term stages (group C); (iii) down-regulated expression during mid-term stages with higher expression levels at D0 and D5 (group D); and (iv) decrease in expression over time (group B) (Fig. 3B).

## RNA-seq results expand previously published transcriptome data and identify novel AP1 targets

The RNA-seq results were compared with those obtained in previous analyses conducted using either microarrays or RNA-seq in similar experimental contexts, namely time-course designs with AP1-based floral induction systems (Wellmer *et al.*, 2006; Kaufmann *et al.*, 2010; Pajoro *et al.*, 2014; Ryan *et al.*, 2015; Chen *et al.*, 2018) (Table 3). To make our RNA-seq results more comparable with those previously obtained using microarrays, a direct differential expression analysis between subsequent stages (i.e. D1 versus D0, D2 versus D1, D3 versus D2, D4 versus D3, and D5 versus D4) was performed (Table 3; Supplementary Table S4). In this analysis, and using the same criteria for differential expression as in Wellmer *et al.* (2006) (adjusted *P*-value <0.05 and no FC cut-off), a total of 5150 genes were classified as DEGs, compared with the 8125 identified as SVGs with the LRT approach (which is more sensitive than pairwise comparisons to slight changes in expression levels between subsequent days). This, together with the 7416 DEGs identified

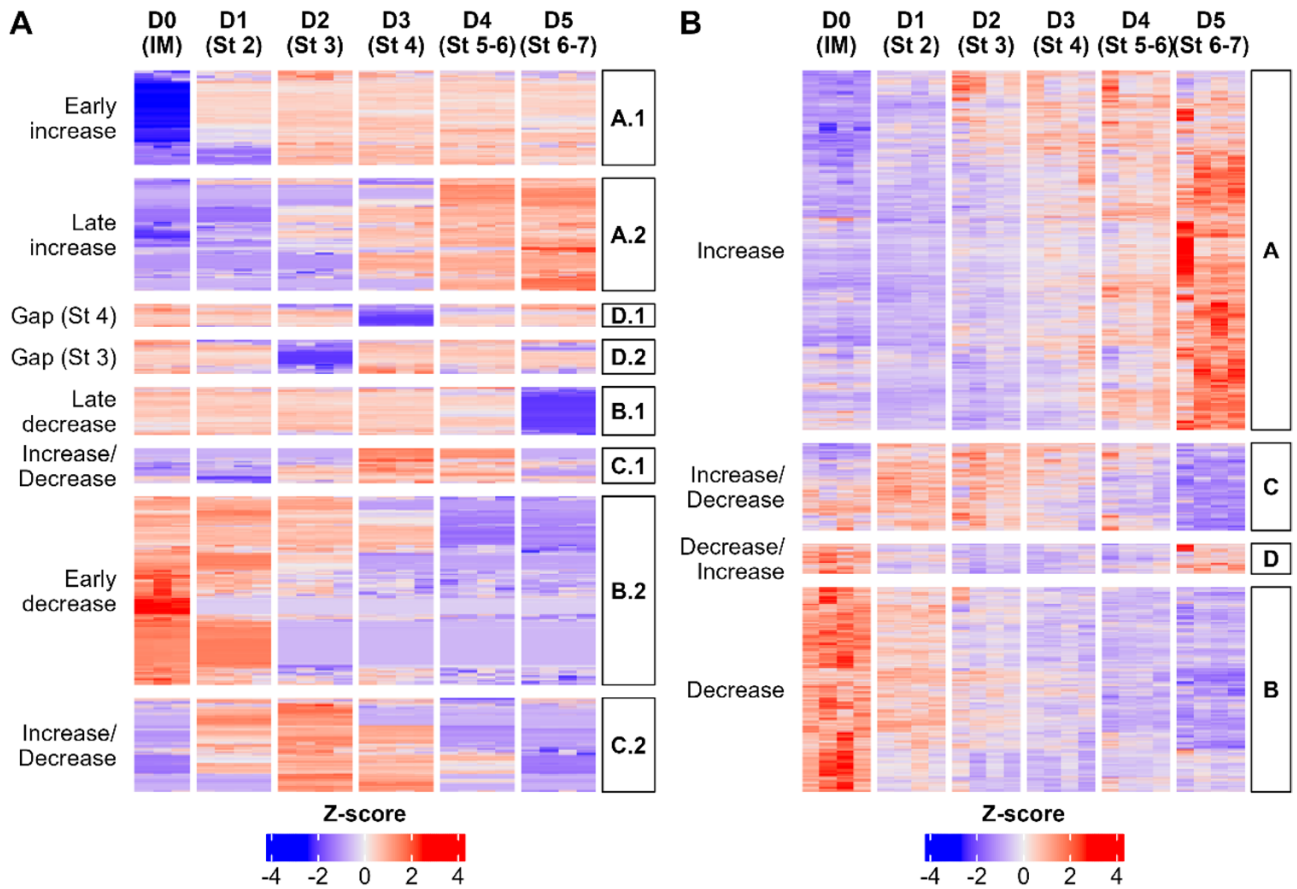


**Fig. 3.** Expression behavior of stage-variant proteins and genes (SVPs and SVGs). (A) Heatmap displaying the abundance levels of the 2037 SVPs throughout the time-course. Color scale represents *Z*-scored TOP3 values. (B) Gene expression patterns of the 8125 SVGs. Color scale represents *Z*-scored normalized RNA counts.

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

by Chen *et al.* (2018) and the other transcriptomic datasets indicated above (Table 3), adds up to a total of 16 613 genes (~60% of all annotated genes) that are differentially expressed during the first 13 d after floral induction (Fig. 4A; Supplementary Fig. S5; Supplementary Table S5). Thus, the RNA-seq dataset reported here complemented previous studies and expanded our understanding of transcriptome changes during early flower development.

As reported previously (Wellmer *et al.*, 2006; Ryan *et al.*, 2015), our RNA-seq results also showed a higher number of up-regulated genes than of down-regulated genes between the time points D1 and D5, probably as a consequence of the activation of genes involved in floral organ initiation and development; and that for the first time point after the induction (D1 versus D0) a preponderance of gene repression was detected (Fig. 4B), probably reflecting the suppression of the genetic program for shoot development during flower initiation. Substantially more gene expression changes were detected by RNA-seq at the first and last time points (D1 versus D0, and D5 versus D4 comparisons, respectively) than in the intermediate time points (D2 versus D1, D3 versus D2, and D4 versus D3 comparisons) (Fig. 4B).

We previously used a combination of gene expression profiling and genome-wide DNA binding experiments to identify AP1 direct target genes (Kaufmann *et al.*, 2010). The criteria for classifying a gene as an AP1 high-confidence target included the presence of one or more AP1 ChIP-seq binding sites, and the gene showing robust differential expression in response to AP1 activation. One of the surprising discoveries of that study was that the vast majority of genes bound by AP1 did not respond transcriptionally to AP1 in a robust manner. Since the previous study employed microarrays for transcriptome profiling, which are less sensitive when compared with RNA-seq analysis (Sirbu *et al.*, 2012; Zhao *et al.*, 2014), we used our RNA-seq dataset to test whether the transcriptional response of AP1-bound genes is more widespread than previously thought, and therefore if additional AP1 high-confidence direct targets could be identified. To this end, we focused on the D1 versus D0 comparison as

**Table 3.** Transcriptomic studies conducted in AP1-based floral induction systems compared in this work

| Line | Time-course | Experiment | Differential expression | Reference |
|---|---|---|---|---|
| 35S:AP1-AP1-GR | D0–D1–D2–D3–D4–D5 | Microarray | 1653 DEGs ($P$<0.05) | Wellmer *et al.* (2006) |
| 35S:AP1-AP1-GR | 0 h–2 h–4 h–8 h–12 h | Microarray | 1366 DEGs (FC>1.8, $P$<0.05) | Kaufmann *et al.* (2010) |
| pAP1:AP1-AP1-GR | D0–D2–D4–D8 | Microarray | 6010 DEGs (FC>1.8, $P$<0.05) | Pajoro *et al.* (2014) |
| pAP1:AP1-AP1-GR | D0–D1–D1.5–D2–D2.5–D3–D3.5–D4–D4.5–D5–D7–D9–D11–D13 | Microarray | 7405 DEGs (FC>1.7, $P$<0.01) | Ryan *et al.* (2015) |
| pAP1:AP1-AP1-GR | D0–D2–D4–D8 | RNA-seq | 7416 DEGs (ANOVA, FC>2, FDR<0.05) | Chen *et al.* (2018) |
| pAP1:AP1-AP1-GR | D0–D1–D2–D3–D4–D5 | RNA-seq | 8125 SVGs (LRT, $P$≤0.01)<br>5150 DEGs ($P$<0.05) | This study |

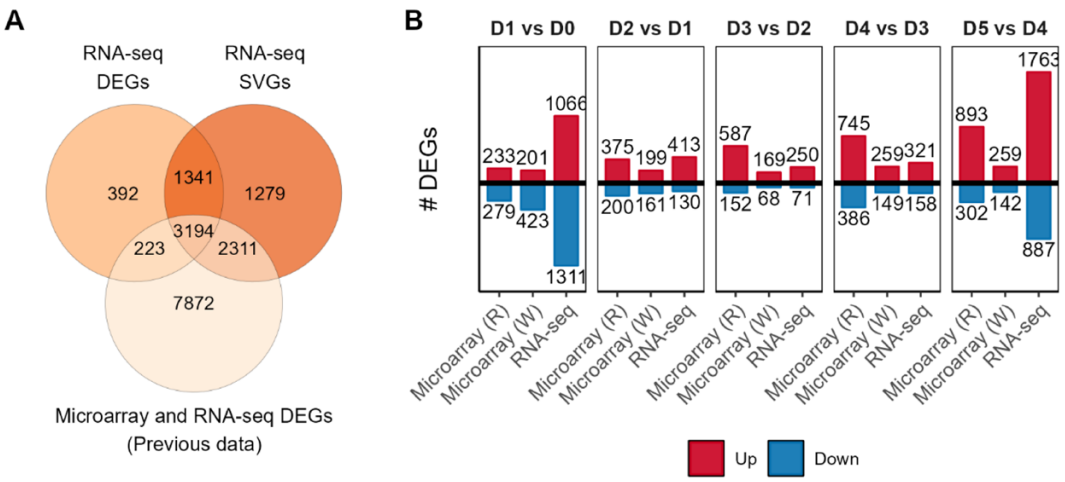


**Fig. 4.** Differentially expressed genes during early flower development. Comparison of the RNA-seq data from this study with previous microarray and RNA-seq data (Table 3). (A) Venn diagram showing the number of RNA-seq DEGs and SVGs identified in this study and of DEGs from previous reports, and the overlap between the datasets. (B) Number of up- and down-regulated DEGs identified in this study (RNA-seq) compared with those in Wellmer *et al.* (2006) (W) and in Ryan *et al.* (2015) (R) (microarray data) at each day-to-previous-day comparison (adj. $P$-value <0.05 for our data and Wellmer *et al.*, and adj. $P$-value <0.01 for Ryan *et al.*).

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

it was the closest to the 12 h time-course used in Kaufmann *et al.* (2010). Among the DEGs identified in this comparison (Fig. 4B; Supplementary Table S4) were many of the high-confidence targets described in Kaufmann *et al.* (2010), including genes both up- and down-regulated by AP1 (Fig. 5A). Thus, to identify novel AP1 high-confidence targets, we compared the 1782 DEGs from the D1 versus D0 comparison that showed robust expression changes with the list of 2298 AP1-bound genes (putative targets) described in Kaufmann *et al.* (2010). This analysis resulted in the identification of 230 novel high-confidence AP1 targets (Fig. 5A; Supplementary Table S6).

Other AP1 ChIP-seq studies using the floral induction system have been conducted to identify AP1 targets and, in particular, a subsequent integrative re-analysis of previous datasets has been reported (Chen *et al.*, 2018) (datasets obtained at 2 h and at 2, 4, and 8 d after DEX induction, respectively: Kaufmann *et al.*, 2010; Pajoro *et al.*, 2014). This comprehensive integrative re-analysis identified a total of 2710 putative targets of AP1 (defined by having ChIP-seq binding sites in their vicinity), and 1079 genes of those were also detected as differentially expressed in a specific D0–D8 RNA-seq dataset of that study (Chen *et al.*, 2018) (these 1079 genes are herein referred to as 'confidence target' genes) (Supplementary Table S6). That is, confidence targets would be genes that are bound by AP1 and show expression changes that might be robust but that might not be temporally tightly associated with such binding. Therefore, to uncover additional confidence targets, we also compared the DEGs that showed robust expression changes in any of the day-to-previous day comparisons reported here (i.e. along the entire time-course) with the list of 2710 putative targets, which led to the identification of 157 novel AP1 confidence targets (Fig. 5A; Supplementary Table S6).

In total, these comparisons allowed the identification of 262 novel AP1 direct targets, 105 of which were classified as high-confidence targets and 157 as confidence targets (Fig. 5A). A GO enrichment analysis of this gene set revealed a preponderance of genes associated with auxin responses, meristem maintenance, and shoot morphogenesis, befitting of direct AP1 targets (Fig. 5B). In addition, among these newly identified AP1 high-confidence targets and confidence targets there is a substantial enrichment of transcription factor-coding genes (22% of high-confidence targets, 23 genes; and 13% of confidence targets, 21 genes; Fig. 5C). These include genes known to be involved in flowering or flower development (e.g. *LMI1*, *ERF12*, *FBH3*, *BBX7*, and *SEP4*) as well as many others that have not been linked to those processes yet (such as *DEWAX2* or *ATHB3*) (Fig. 5C).

In summary, our results have expanded the identification of genes whose expression varies during early flower development and the set of identified AP1 high-confidence and confidence targets, that now add up to 479 and 952, respectively (Fig. 5A; Supplementary Table S6).

## Correlation between RNA and protein levels during early flower development

The dynamic range of protein and transcript expression, as determined by LC-MS/MS and RNA-seq, spanned six and four orders of magnitude, respectively (Fig. 6A; Supplementary Fig. S2D). Protein evidence was under-represented for low-abundance transcripts (ANOVA with Tukey post-hoc test; $P$-value ≤0.001), as described in other RNA–protein comparison studies (Hoogendijk *et al.*, 2019; Mergner *et al.*, 2020a, b), and the median expression levels for transcripts were similar within days (Fig. 6A).

Amongst the 7003 quantified transcript–protein pairs were (i) 973 pairs with stage-dependent variation at both RNA and protein levels (SVG–SVP); (ii) 1006 pairs non-variant at the RNA level, but stage-variant for proteins (NVG–SVP); (iii) 1808 pairs stage-variant at the transcript level, and non-variant for proteins (SVG–NVP); and (iv) 3216 pairs with non-variable levels for both molecules (NVG–NVP) (Fig. 6B; Supplementary Table S7). The seven 'supermarker' proteins, 37 of the marker proteins, and 43 of the AP1 high-confidence targets defined in Kaufmann *et al.* (2010) were found as 'quantified' at both transcript and protein levels. These three subsets were significantly enriched in SVG–SVP pairs (Fisher's exact test, $P$-value ≤0.001) (Supplementary Fig. S6).

To provide a measure of similarity among developmental stages and to test whether there is a shift between mRNA and protein levels at different time points, Pearson's correlation coefficient ($r$) was calculated for gene expression and protein abundance of all pairwise time point combinations (e.g. protein D0 versus protein D0–D5, protein D0 versus RNA D0–D5, etc.). Correlations were computed on the protein and transcript level for all RNA–protein pairs and the SVG–SVP pairs separately (Supplementary Fig. S7A). In general, there is a relatively good correlation between RNA levels and protein abundances during the early stages of flower development, although Pearson's coefficients were slightly higher for the SVG–SVP group. For both SVG–SVP pairs and all transcript–protein pairs, the RNA–protein correlation seemed to be moderately shifted throughout the time-course and particularly at D3 and afterwards, as protein levels at D4 correlated equally well with RNA levels at D3, and protein levels at D5 correlated equally well with RNA levels at D4. PCA for averaged $Z$-scored values (i.e. independently scored) showed that RNA–protein levels clustered according to the time point, with this correlation being more obvious at D2, D3, and D4 (Supplementary Fig. S7B). D0, D1, and D5 presented the greater differences between RNA and protein levels; in fact, D5 protein levels correlated better with RNA levels at D4 than at D5. These differences, especially in the initial and end points of the time-course (i.e. D0 and D5), can be partly explained by the time lag and are further accentuated by the fact that D1 (versus D0) and D5 (versus D4) were the time points with the largest number

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

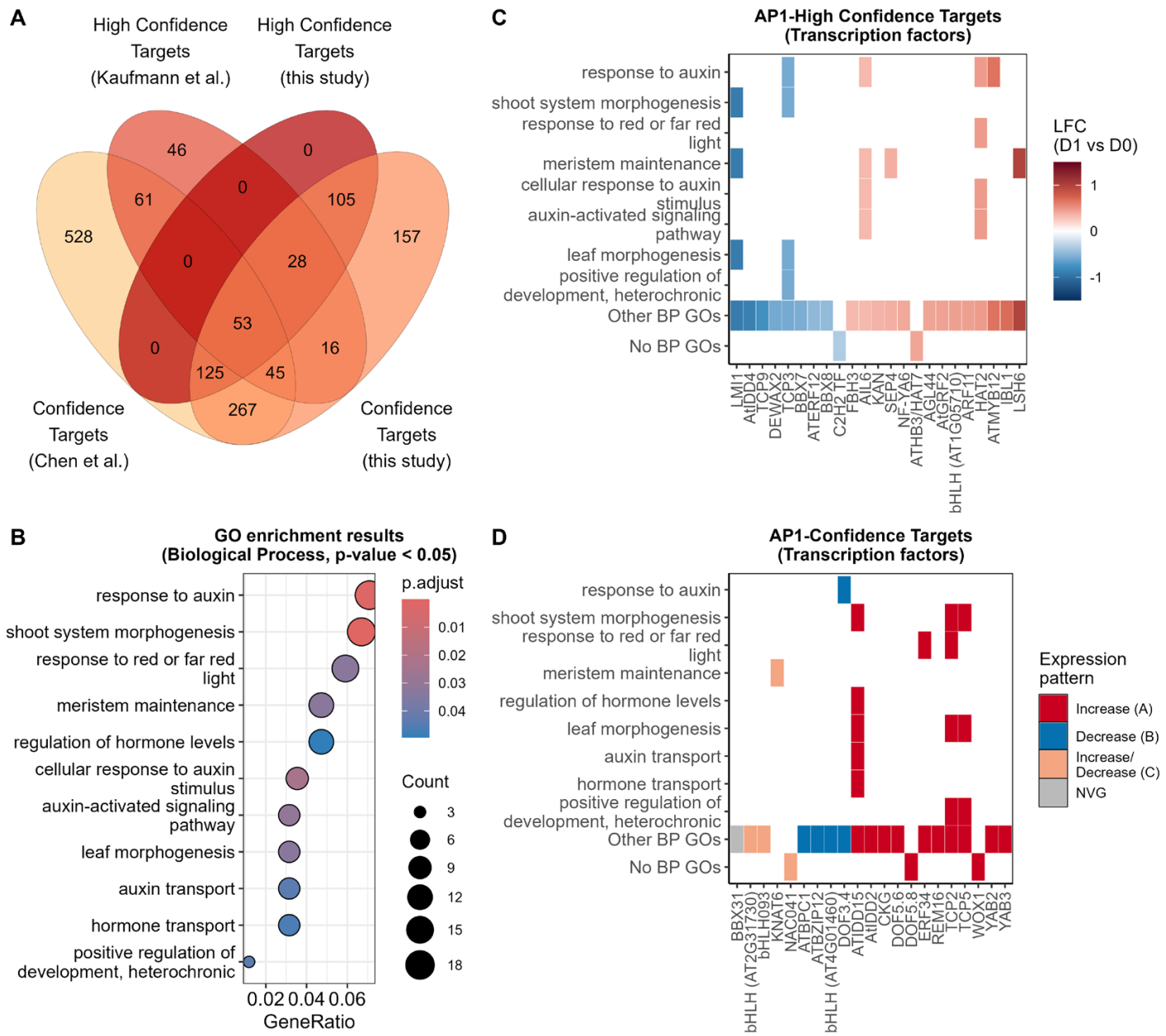


**Fig. 5.** Identification of AP1 target genes. (A) Overlap between AP1 target genes identified in different studies. AP1 high-confidence target genes were as defined in Kaufmann *et al.* (2010), or in this study if they appeared on the list of putative targets from Kaufmann *et al.* (2010) and were detected as robustly differentially expressed in the D1 versus D0 RNA-seq data comparison. Genes were defined as AP1 confidence targets if they appeared on the target list from Chen *et al.* (2018) and were also detected as differentially expressed in that study, or also detected as robustly differentially expressed in at least one day-to-previous day comparison of the RNA-seq dataset reported here. See the Materials and methods. Data are from Supplementary Table S6. (B) Gene Ontology (GO) enrichment analysis of the novel AP1 targets identified in this study. The set of 262 novel AP1 targets includes 105 high-confidence targets and 157 confidence targets. Dot plot representing the most enriched biological process GO categories (adjusted *P*-value <0.05) for the novel AP1 targets, colored by their adjusted *P*-value. Dot size represents the number of genes identified in each category. Gene ratio is the proportion between the number of identified genes for each category and the total number of genes registered for that category. (C) Enriched biological process GO categories for the AP1 high-confidence targets that code for transcription factors, colored by their log fold change (LFC) in the D1 versus D0 comparison. (D) Enriched biological process GO categories for the AP1 confidence targets that code for transcription factors, colored by their gene expression pattern (as described in Fig. 3).

of DEGs (Fig. 4B). This is because on D1 there was a preponderance of transcript down-regulation, which probably does not correlate as well with protein levels as when it is up-regulation, and on D5 there was new up-regulation than would be translated into proteins partly on D6, according to the detected time lag.

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

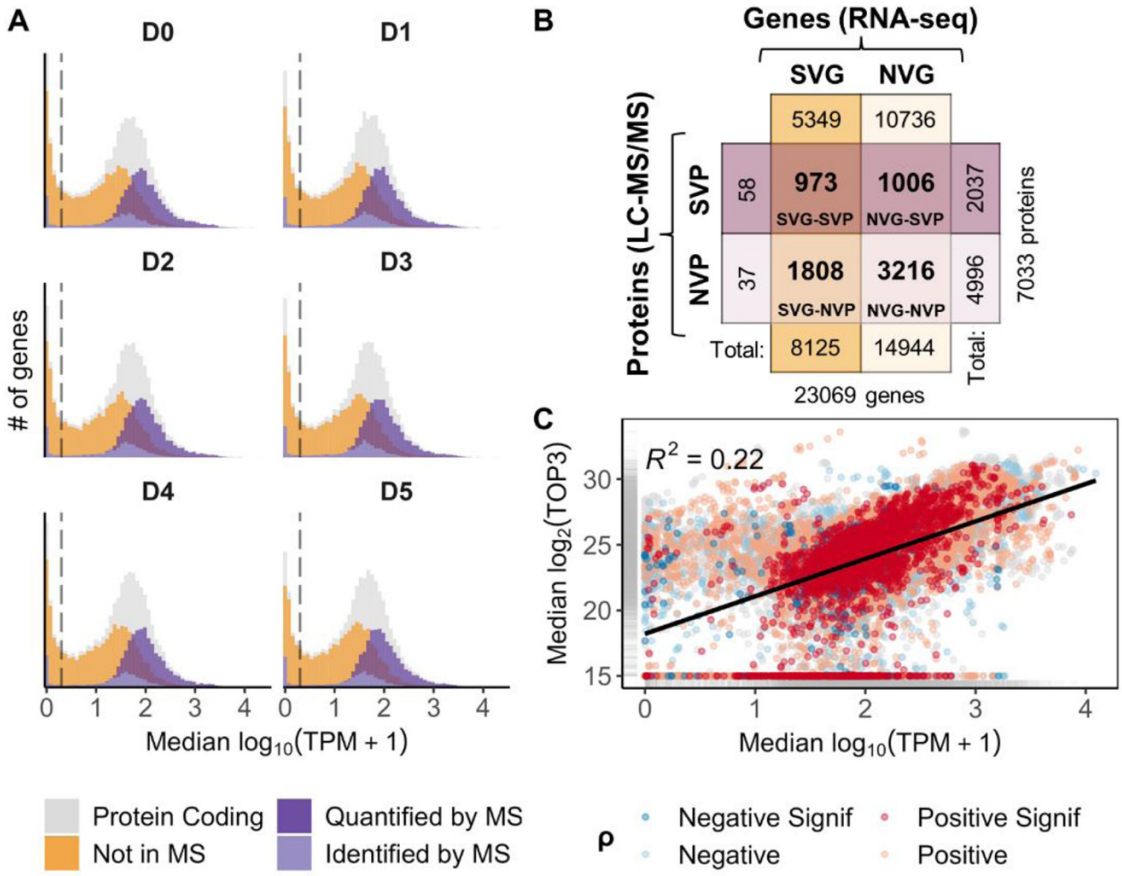


**Fig. 6.** Gene and protein classification according to abundance through time. (A) Histogram of RNA expression range. Gray, all detected protein-coding transcripts; orange, protein-coding transcripts not detected as a protein by LC-MS/MS; dark purple, protein-coding transcripts quantified as a protein; light purple, transcripts corresponding to a protein identified by LC-MS/MS but that was discarded because it was classified as undetected (UU or RU) at every time point (i.e. not quantified). Dashed line indicates TPM=1. (B) Schema illustrating the number of expressed genes, stage-variant genes (SVGs), quantified proteins, and stage-variant proteins (SVPs). The sum of gene–protein pairs differs from the number of genes and proteins identified separately as there are cases of a single gene ID (AGI) associated with more than one Uniprot code, and vice versa. (C) Scatter plot of protein abundance and RNA expression level for all RNA–protein pairs at every time point. Colored by RNA–protein correlation (Spearman's rank coefficient, ρ). Positive if ρ≥0.4. Negative if ρ≤ –0.4. Significant if *P*-value <0.05. The coefficient of determination $R^2$ was calculated for the data points represented in the graph.

The correlation of mRNA and protein levels through time was also measured by calculating the Spearman's rank correlation coefficient (ρ) for each RNA–protein pair (i.e. within-gene correlation, Buccitelli and Selbach, 2020). In total, there were 2540 RNA–protein pairs with a positive correlation (ρ≥0.4, as defined in Akoglu, 2018), and almost 17% of these pairs had a significant and highly positive correlation (ρ≥0.8 and *P*-value ≤0.05). In contrast, 975 RNA–protein pairs presented a negative correlation (ρ≤ –0.4), and ~10% of them showed a significant and highly negative correlation (ρ≤ –0.8 and *P*-value ≤0.05) (Table 4; Supplementary Table S7). Moreover, the mRNA-to-protein abundance correlation was very different for the SVG–SVP, SVG–NVP, NVG–SVP, and NVG–NVP subsets (Fig. 7A; Table 4). Accordingly, the global correlation for the complete dataset was moderate (Fig. 6C; Supplementary Fig. S8).

RNA–protein pairs that varied at both molecule levels (SVG–SVP) presented the strongest positive correlation, with a median ρ of 0.6, and 63% of pairs with positive correlation and <8% of pairs with negative correlation (Table 4; Fig. 7A). In summary, ~36% of the total RNA–protein pairs presented a positive correlation between their RNA and protein expression levels, and the correlation between RNA and protein expression levels was higher for those RNA–protein pairs with differential expression over time for both molecules (Table 4; Fig. 7A).

A total of 80% of both flower markers and AP1 targets (high-confidence targets) in the SVG–SVP subset presented a positive correlation, although there were some exceptions with a significant and highly negative correlation [e.g. *WUSCHEL RELATED HOMEOBOX13* (*WOX13*) and *AT4G27980*] or without positive or negative ρ [e.g. *LUTEIN-DEFICIENT1* (*LUT1*) and *PYRABACTIN RESISTANCE1-LIKE1* (*PYL1*)] (Fig. 7A, B). In addition, the time-course trajectories (mRNA and protein) of the seven 'supermarkers' showed a positive ρ above the median and were in good agreement with their previously published expression patterns (Wellmer *et al.*, 2006)

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

**Table 4.** Global protein–mRNA correlation among variant and non-variant subsets

| | Median ρ | Positive (ρ>0.4) | Negative (ρ< –0.4) | Uncorrelated | Total |
|---|---|---|---|---|---|
| **SVG–SVP** | 0.6 | 615 (217 sig.) | 76 (15 sig.) | 282 | 973 |
| **SVG–NVP** | 0.31 | 794 (96 sig.) | 197 (18 sig.) | 817 | 1808 |
| **NVG–SVP** | 0.13 | 300 (42 sig.) | 167 (28 sig.) | 539 | 1006 |
| **NVG–NVP** | 0.08 | 831 (74 sig.) | 535 (36 sig.) | 1850 | 3216 |
| **Total** | 0.2 | 2540 (429 sig.) | 975 (97 sig.) | 3488 | **7003** |

For each subset, Spearman's rank coefficients (ρ) between RNA and protein levels and the number of RNA–protein pairs are shown. Significant (sig.): adjusted *P*-value (BH) <0.05

(Figs 2, 7B). *AP1* was also included in the SVG–SVP subset, but nevertheless the variation in RNA and protein levels during the time-course was not substantial (Supplementary Fig. S9).

## RNA–protein pairs clustered in various expression pattern modules

To elucidate transcript–protein dynamics of the complete dataset, unbiased clustering based on the correlation of mRNA and protein expression patterns was performed. WGCNA using the SVG–SVP, NVG–SVP, and SVG–NVP transcript–protein pairs separately resulted in 18, 18, and 25 co-expression eigen-modules (MEs), respectively, ranging in size from 10 to 485 gene–protein pairs (Fig. 8; Supplementary Fig. S10).

Combined expression patterns of SVG–SVP pairs were categorized in four groups: (i) increasing mRNA and protein levels (A.1); (ii) increasing mRNA and decreasing protein levels (A.2); (iii) decreasing mRNA and increasing protein levels (B.1); and (iv) decreasing mRNA and protein levels (B.2), with groups A.1 and B.2 (i.e. those in which RNA and protein levels change in the same direction) comprising the vast majority of SVG–SVP pairs (and of MEs) and with most of the different MEs in those groups showing a high correlation (i.e. substantial ρ values) (Fig. 8A). In groups A.2 and B.1 in which mRNA and protein levels were anti-correlated, a few MEs also showed relevant ρ values (ME11, 13, and 17 in A.2, and ME19 in B.1), although the number of gene–protein pairs encompassed by those MEs is small (64 versus 824 in the most significant MEs of A.1 and B.2). These observations support the idea that there is a relatively good correlation between RNA and protein level changes during early flower development, and identify a few specific and small subgroups of genes in which the changes are anti-correlated.

NGV–SVP pairs grouped in clusters with patterns that were similar to those observed for the complete list of SVPs (Fig. 3A), that is: (i) increased protein abundance over time (ME01, 05, 07, 09); (ii) reduced levels over time (ME02, 03, 04, 11, 14); (iii) transient protein expression at middle time points (ME06, 10, 12, 13, 15, 16, 17); and (iv) transient proteins with a gap in their expression at intermediate time points (ME08, 18). Last, SVG–NVP grouped pairs also showed the same trajectory patterns as the complete set of SVGs (Fig. 3B): (i) increased expression through time (ME02, 05, 06, 10, 13, 25, 26, 27, 29, 30); (ii) higher expression during mid-term stages (D1–D4) (ME09, 14, 17, 29); (iii) high expression at D0 and D5, but down-regulated expression during mid-term stages (ME15, 22, 24, 28); and (iv) expression reduction over time (ME01, 04, 07, 08, 11, 12, 23). The correlation between protein and RNA expression levels for all these NVG–SVP and SVG–NVP MEs was variable but was not as high as that observed for the majority of the SVG–SVP modules (Fig. 8A).

## Modules with opposite patterns for mRNA and protein levels are enriched in hormone-responsive pathways

GO and KEGG enrichment analyses were performed to retrieve the functional biological processes that accompany early flower development and to investigate the few specific and small subgroups of genes showing anti-correlated RNA and protein abundance changes (Supplementary Tables S8, S9). Interestingly, a high percentage of gene–protein pairs with decreasing levels of RNA combined with increasing protein abundance are known to correspond to proteins localized to chloroplasts. in particular, six out of 10 genes included in SVG–SVP ME19 (*AT1G79460*, *AT3G07310*, *AT1G29070*, *AT4G17300*, *AT2G29180*, and *AT5G23040*). Among these, GA2 (AT1G79460) and a putative phosphoserine aminotransferase (AT3G07310) are included in the gibberellic acid signaling pathway (although the module itself was not significantly enriched in gibberellin pathway genes; adjusted *P*-value >0.05). On the other hand, gene–protein pairs with increasing levels of RNA combined with decreasing protein abundance (i.e. SVG–SVP ME11, 13, 17, and 18) contain proteins involved in fatty acid metabolic processes related to acetyl-CoA and jasmonic acid (JA) pathways.

Other modules with increasing levels of mRNAs (SVG–NVP ME02 and ME06), or of both mRNA and proteins (SVG–SVP ME09), were enriched in auxin metabolic processes, whereas modules with decreasing levels of mRNA (SVG–NVP ME01), or of mRNA and protein (SVG–SVP ME01) were enriched in cytokinin-responsive processes.

In summary, the main result from the GO analysis was that modules with opposite patterns for mRNA and protein levels were enriched in hormone-responsive pathways.

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

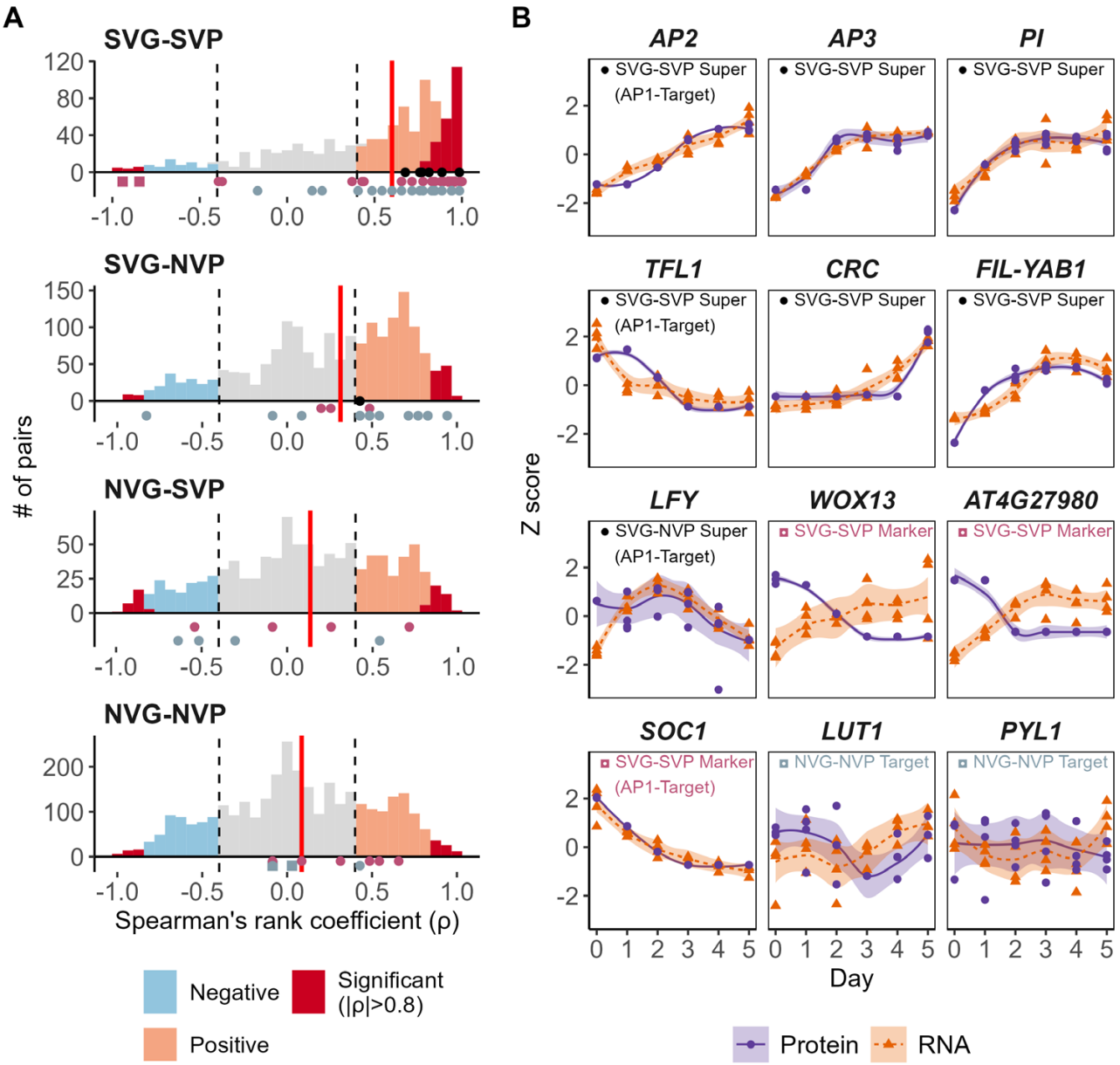


**Fig. 7.** Correlation and trajectory patterns for gene–protein pairs. (A) Spearman's rank correlation coefficient (ρ) between RNA and protein levels of each pair depending on the SV–NV classification. Red lines, median ρ of each subset. Dashed lines indicate the limits considered for positive and negative correlation. Circles represent ρ values of 'supermarkers' (black), markers (pink), and AP1 targets (gray). Squares represent the ρ for the markers and AP1 targets depicted in (B). (B) *Z*-scored RNA and protein abundances (*Z*-scored separately) of selected genes: the seven 'supermarkers' (ρ≥0.4) (SVG–SVP: *AP2*, *AP3*, *PI*, *TFL1*, *CRC*, *FIL-YAB1*; SVG–NVP: *LFY*), two markers with ρ≤ –0.8 (SVG–SVP: *WOX13*, *AT4G27980*), one marker and AP1 target with ρ≥0.8 (SVG–SVP: *SOC1*), and two AP1 targets with non-significant ρ (NVG–NVP: *LUT1*, *PYL1*).

## Physically interacting proteins had different RNA–protein expression levels through time

To investigate possible functions associated with the different MEs, or functional relationships within each module, the interaction network of all MS-detected proteins among themselves and with all other Arabidopsis proteins was analyzed using information about known and predicted protein–protein interactions from the STRING database (Szklarczyk *et al.*, 2017) (Supplementary Fig. S11).

The final protein–protein interaction network, with a total of 6403 nodes (proteins) and 66 350 edges (protein–protein interactions), was divided into five main clusters with >30 proteins each, and 55 smaller protein groups, ranging from two to 23 proteins (Fig. 9; Supplementary Tables S10–S13). No clear association between RNA–protein trajectory

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

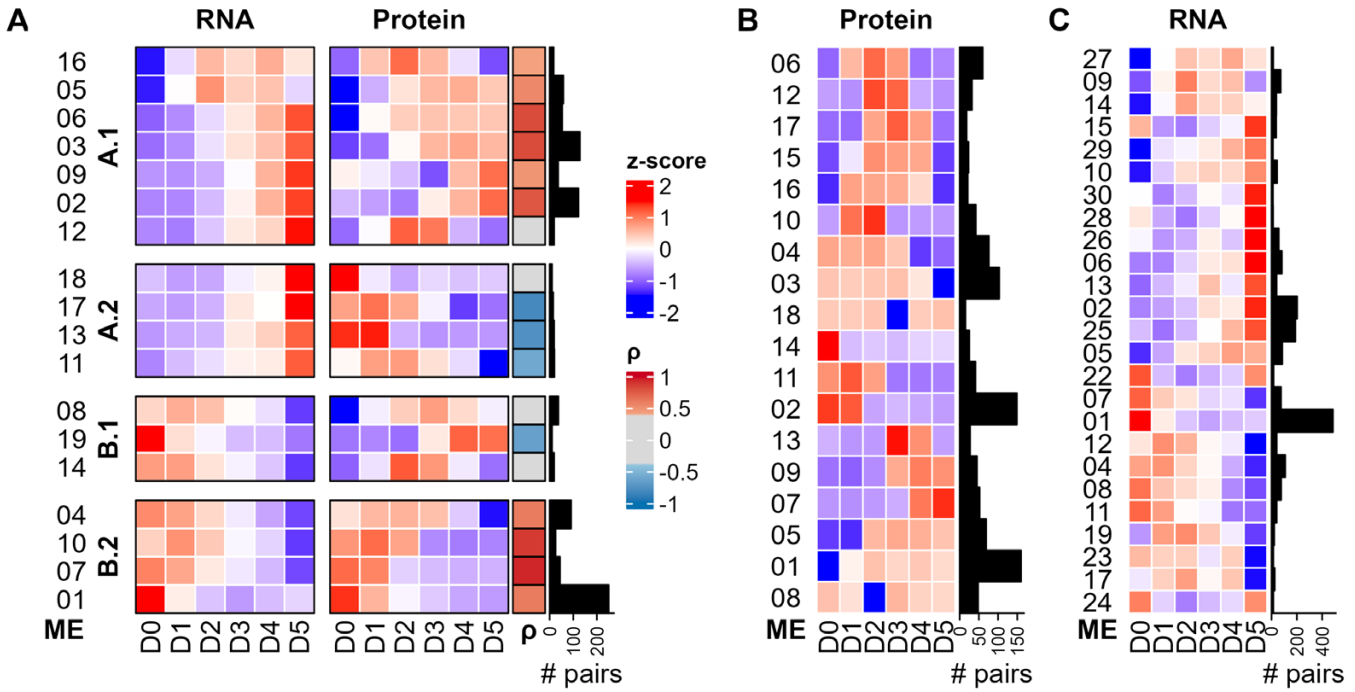


**Fig. 8.** Trajectory patterns for gene–protein pairs. Trajectory clustering (WGCNA) for SVG–SVP (18 modules) (A), NVG–SVP (18 modules) (B), and SVG–NVP (25 modules) (C) gene–protein pairs. Black bar graphs on the right side of each heatmap indicate the number of gene–protein pairs included in each module. The average ρ value for gene–protein pairs included in each SVG–SVP (A) module is shown; for the NVG–SVP (B) and SVG–NVP (C) modules, this value was between –0.4 and 0.4 in all cases (no-correlation, 'gray').

patterns and the interaction clusters was found, as each one of the interaction clusters contained proteins that presented different RNA–protein patterns (Fig. 9). The only exception was interaction cluster 2, which had a central hub composed of proteins whose RNA–protein levels were down-regulated during the time-course and whose main KEGG pathway was 'photosynthesis and carbon fixation' (Supplementary Table S14).

## Discussion

The process of early flower development in Arabidopsis was analyzed by comparing gene and protein expression profiles through a time-course conducted using a pAP1:AP1-GR *ap1 cal* inducible line. A major concern in label-free quantitative proteomics is the high rate of missing values that can hinder data analyses and the comparison with other omics data. To address this issue, a specific data processing pipeline was developed that takes advantage of the experimental design (i.e. time-course and multiple biological replicates) and that combines logic assumptions and value imputations. Thanks to this 'reliability analysis' workflow, it was possible to distinguish the nature of the data missingness and to treat the NA values of the LC-MS/MS results accordingly. The appropriateness of this method was determined through various approaches, and it was validated using a substantial number of marker genes. After the reliability analysis and NA imputation, replicates clustered better together according to time (day), demonstrating the robustness and reproducibility of the results obtained through the combination of LC-MS/MS followed by the reliability analysis approach. Because many omics experimental designs in plant research follow the same principles employed here (i.e. temporal analyses and the use of biological replicates), the reliability analysis workflow might prove useful in studies addressing many different areas of plant biology. Given the inherent limitations and stochasticity in detecting proteins by MS, and in particular low abundance proteins, this workflow may provide a more complete view of proteome changes across dynamic processes than the alternative and frequently used approach of only considering as identified those proteins that are detected in all or most of the replicates of a given time point or developmental stage, while disregarding the information provided by adjacent time points (e.g. Buccitelli and Selbach, 2020; Huang *et al.*, 2022).

The overall across-gene correlation between mRNA and protein levels (i.e. across all gene–protein pairs) was moderate for the complete dataset: it was within the range of what has been observed for other organisms and processes (Gao *et al.*, 2017; Buccitelli and Selbach, 2020) (Fig. 6C; Supplementary Fig. S8), but lower than for instance in Arabidopsis early embryogenesis (Huang *et al.*, 2022) or leaf development (Omidbakhshfard *et al.*, 2021), or in general in different Arabidopsis tissues (Mergner *et al.*, 2020a). This effect was largely due to the

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

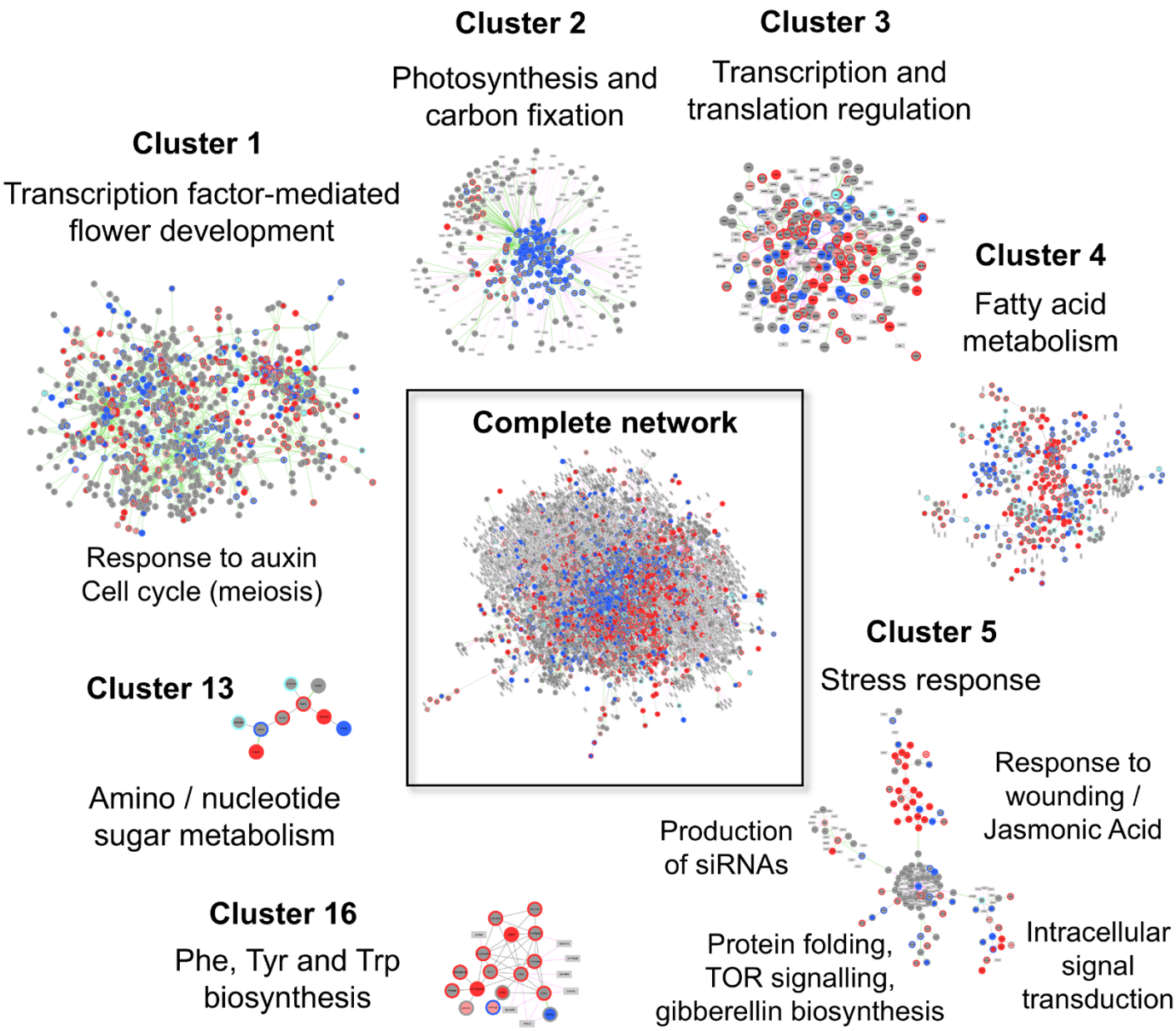


**Fig. 9.** Protein–protein interaction clusters. Network depicting physical interactions and co-expression in interaction clusters between proteins included in the LC-MS/MS dataset and other proteins in Arabidopsis (IntAct, STRING). The five largest interaction clusters (Clusters 1–5) are shown, as well as two interaction clusters that only contain proteins from a specific metabolic pathway (Clusters 13 and 16). The main KEGG pathways for each cluster are shown (see also Supplementary Table S10). Clusters 1, 3, and 5 contain proteins involved in developmental processes and stress responses, whereas clusters 2, 4, 13, and 16 are enriched in proteins related to metabolic pathways. Node border represents RNA levels, and node inner color represents protein levels. Blue, decreasing trajectories; red, increasing trajectories; salmon, trajectories with a maximum peak (increase–decrease); light blue, trajectories with a minimum (decrease–increase); gray, non-variant. Proteins not included in the LC-MS/MS dataset are depicted with rectangles.

non-variant (NVG–NVP) subset, which comprised almost half of the complete dataset and, as expected, showed a very low or non-existent correlation. Examples of this category included AP1 targets such as *LUTEIN-DEFICIENT 1* (*LUT1*) (Tian *et al.*, 2004) and *PYRABACTIN RESISTANCE 1-LIKE 1* (*PYL1*) (Yin *et al.*, 2009) (Fig. 7B). In addition, the fact that the predominant initial response upon AP1 activation is gene repression and that the lag effect for transcript and protein accumulation was detected in the time-course also contribute to this overall moderate correlation, as reflected in the PCA of all RNA–protein pairs (Supplementary Fig. S7B). On the other hand, the across-gene correlation for the smaller SVG–SVP subset was substantially high (Fig. 7A).

Similarly, the within-gene correlation (i.e. how each gene's protein levels track with those of the corresponding transcript across different time points, Buccitelli and Selbach, 2020) was

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

high for those genes in the SVG–SVP group (Fig. 7). These observations indicate that translation is the primary determinant for protein abundance for those genes that change their expression during early flower development. In agreement with this, the correlation between RNA and protein seemed to be moderately shifted throughout the time-course, as protein abundance on a given day would correlate equally well with RNA levels of the same and the previous day. Considering the complete dataset of 7003 gene–protein pairs, a positive and significant within-gene correlation was observed for ~6% of the pairs (Table 4), a similar percentage to that detected in Arabidopsis early embryogenesis (Huang *et al.*, 2022). In addition, a small number of genes showing a clear and significant negative correlation were also identified through the Spearman correlation analysis (Fig. 7) and/or the WGCNA trajectory clustering (Fig. 8). This was the case, for instance, for WUSCHEL RELATED HOMEOBOX13 (WOX13) (Costanzo *et al.*, 2014) and AT4G27980 (Wang *et al.*, 2008), two marker proteins (Wellmer *et al.*, 2006) whose mRNA levels increased whereas their protein levels decreased across the time-course ($\rho \leq -0.8$, *P*-value ≤0.05) (Fig. 7B).

Our analysis also highlighted the extensive remodeling of the proteome that takes place during early flower development, reflected by the number of SVPs compared with the total number of gene–protein pairs (~30% of the 7003 pairs included SVPs; Table 4) and by the diversity of protein expression patterns during the time-course (Fig. 3A). For SVGs, the proportion was even higher (~40%), in agreement with the observation that the highest number of DEGs was detected at D5 (D5 versus D4 comparison) and that the corresponding proteins would be in part detected at a later (not analyzed) time point because of the time lag for protein accumulation.

As seen, in some cases, there was an association between the behavior in expression of gene–protein pairs and their functions. Six out of 10 genes included in the SVG–SVP module 19 (ME19), with decreasing levels of RNA and increasing protein abundance, are related to chloroplasts (*AT1G79460*, *AT3G07310*, *AT1G29070*, *AT4G17300*, *AT2G29180*, and *AT5G23040*). Among these, GA2 (AT1G79460) and a putative phosphoserine aminotransferase (AT3G07310) are included in the gibberellic acid signaling pathway; PRPL34 (AT1G29070) is a structural constituent of the ribosome; AT2G29180 (a thylakoid membrane protein) positively regulates transcription; NS1 (AT4G17300) is also related to chloroplast transcription; and CDF1 (AT5G23040) is a thylakoid membrane chaperone required for chloroplast biogenesis and development. All these proteins are related to the cellular response to lipids and gibberellins. On the other hand, gene–protein pairs with increasing RNA levels but decreasing protein abundance (i.e. SVG–SVP ME11, 13, 17, and 18) were enriched in proteins involved in fatty acid metabolic processes related to acetyl-CoA and JA pathways. These pathways are required for the proper execution of flower developmental processes, such as flower maturation (Reeves *et al.*, 2012). Interestingly, JA induces the expression of *BpWOX9* and *BpWOX10* in *Broussonetia papyrifera* (Tang *et al.*, 2017), orthologs of Arabidopsis *WOX13*, which is in fact part of SVG–SVP ME13. In summary, the main result from the GO analysis was that modules with opposite patterns for mRNA and protein levels were enriched in hormone-responsive pathways, identifying candidate genes that could be used to explore the molecular mechanisms and functional consequences of these divergent RNA and protein trajectories.

The transcriptome profiling reported here expanded the identification of genes whose expression changes during early flower development. The resulting combined set of DEGs now extends to >16 000 genes, representing ~60% of those in the genome and of which ~18% (3012 genes) were added by this study. This expanded coverage of transcriptional changes could also be useful for other genome-wide analyses of flower development. Accordingly, we used the newly generated transcriptome to revisit the identification of AP1 direct targets. Among the newly identified direct targets (105 high-confidence targets and 157 confidence targets), there were a substantial number of transcription factors (44; Fig. 5), including those related to flowering or flower development and that are now directly associated with AP1, as well as others that have not been linked to those processes yet (in particular, more than half of the 44 transcription factors), together expanding the flower development gene regulatory network.

For instance, ALOG [Arabidopsis LIGHT-DEPENDENT SHORT HYPOCOTYLS 1 (LSH1) and Oryza G1] transcription factor family genes play important roles in organ development in various plant species, and it has recently been shown that in Arabidopsis *LSH1*, *LSH3*, and *LSH4* act redundantly to repress bract development in the pedicel (Rieu *et al.*, 2024). These three genes as well as *LSH2* had been previously identified as AP1 high-confidence targets (Kaufmann *et al.*, 2010) and, in this study, *LSH6* is added to that group, implying that the five family members are expressed during early flower development and that they may play additional roles in the process, not necessarily acting in a redundant fashion. Whereas *LSH1–LSH4* comprise a specific clade in the Arabidopsis ALOG phylogenetic tree (Naramoto *et al.*, 2020) and are repressed by AP1 activation, *LSH6* belongs to a different ALOG subgroup and is activated. Another example is provided by *ETHYLENE RESPONSE FACTOR12* (*EFR12*), which promotes the floral transition in addition to pleiotropically affecting flower development, for instance by counteracting the promotion of floral organ initiation (Chandler and Werr, 2020). *ERF12* is now identified as a high-confidence target that is down-regulated by AP1, as several other flowering time genes are (Kaufmann *et al.*, 2010). Another newly identified AP1 high-confidence target known to be involved in flower development is the homeodomain leucine-zipper transcription factor LATE MERISTEM IDENTITY1 (LMI1), a meristem identity regulator that activates *CAL* expression together with LEAFY (LFY) (Saddic *et al.*, 2006). Likewise, *SEPALLATA4* (*SEP4*), which functions in floral organ and

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

meristem identity (Ditta *et al.*, 2004), had not been previously categorized as a high-confidence target that is positively regulated by AP1.

In summary, the work reported here has advanced our understanding of the relationship between the transcriptome and proteome during flower development, identified novel genes of interest in the process, including AP1 targets that expand the corresponding gene regulatory network, and provided useful datasets and a data analysis method for further investigations.

## Supplementary data

The following supplementary data are available at *JXB* online.

Fig. S1. Inter-sample variability in the proteomics data before D0R1 removal.

Fig. S2. Proteomic results overview.

Fig. S3. Effects of the reliability analysis in the proteome dataset.

Fig. S4. PCA of stage-variant genes (SVGs).

Fig. S5. Overlap between genes with differential expression identified in different studies that used AP1 flower induction systems.

Fig. S6. Relative distribution of absolute numbers of transcript–protein pairs in selected classes across the four expression categories (SVG–SVP, SVG–NVP, NVG–SVP, and NVG–NVP).

Fig. S7. RNA–protein comparisons.

Fig. S8. Correlation between each RNA–protein pair for the complete dataset.

Fig. S9. AP1 RNA and protein expression throughout the time-course.

Fig. S10. Correlation between RNA and protein levels for all different MEs.

Fig. S11. STRING and IntAct interaction scores.

Table S1. List of flower-related 'marker' and 'supermarker' genes/proteins.

Table S2. List of proteins classified as stage-variant proteins (SVPs).

Table S3. List of genes classified as stage-variant genes (SVGs).

Table S4. Pairwise comparisons (day-to-previous day) of the RNA-seq data.

Table S5. List of differentially expressed genes identified in different studies performed with the AP1 flower induction system.

Table S6. List of AP1 confidence targets and high-confidence targets.

Table S7. Summary of information for the 7003 gene–protein pairs.

Table S8. KEGG pathways found enriched in the protein interaction clusters.

Table S9. Summary of enriched Gene Ontologies for each trajectory module.

Table S10. Node table of the protein interaction clusters.

Table S11. Edge table of the protein interaction clusters.

Table S12. Number of nodes and edges included in each interaction cluster.

Table S13. Protein–protein interactions among the proteins included in the RNA–protein pairs of this study, according to the TAIR database.

Table S14. KEGG enriched pathways for each trajectory module.

Table S15. Proteomics data (LC-MS/MS results) after the reliability analysis and the kNN imputation.

Table S16. RNA data (RNA-seq results) in the form of normalized counts.

## Acknowledgements

We are grateful to Eva Borràs and Eduard Sabidó from the Proteomics Unit of the Centre for Genomic Regulation (CRG, Barcelona) for their advice and support for the proteomics experiments, which were performed at the CRG facility. We are also grateful to Laura Ossorio for her initial experimental contributions to the project.

## Author contributions

JLR: conceived the project and obtained funding; JLR, RAU, and JTM: planned and designed the research, and wrote the manuscript; RAU and JTM: performed the experiments; RAU, JTM, VMGM, MBF, and JLR: analyzed the data; all authors reviewed and approved the manuscript.

## Conflict of interest

The authors declare that they have no conflicts of interests.

## Funding

This work was supported by grants BFU2014-58289-P and PID2022-139578NB-I00 (funded by MICIN/AEI/10.13039/501100011033 and by 'ERDF A way of making Europe') and 2017SGR718 and 2021SGR00792 (from the Agencia de Gestió d'Ajuts Universitaris i de Recerca) to JLR; and by institutional grants SEV-2015-0533 and CEX2019-000902-S (funded by MICIN/AEI/10.13039/501100011033) and by the CERCA Programme/Generalitat de Catalunya. RA was supported by fellowship PRE2018-084278 funded by MICIN/AEI/10.13039/501100011033 and by 'ESF Investing in your future'.

## Data availability

The LC-MS/MS proteomics data from this study have been deposited at the ProteomeXchange Consortium (https://www.proteomexchange.org) under dataset identifier PXD038980. The RNA-seq data are available at NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/) under dataset identifier GSE217606. The code and scripts used for the reliability analysis pipeline have been deposited in GitHub (https://github.com/Alv-Urd/Proteomics-Reliability-Analysis). Supplementary Table S15

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

contains all the proteomics data (LC-MS/MS results) after the reliability analysis and the kNN imputation, and Supplementary Table S16 the RNA data (RNA-seq results) in the form of normalized counts.

## References


**Akoglu H.** 2018. User's guide to correlation coefficients. Turkish Journal of Emergency Medicine **18**, 91–93.

**Alvarez-Urdiola R, Matus JT, Riechmann JL.** 2023. Multi-omics methods applied to flower development. Methods in Molecular Biology **2686**, 495–508.

**Bai B, van der Horst N, Cordewener JH, America AHP, Nijveen H, Bentsink L.** 2021. Delayed protein changes during seed germination. Frontiers in Plant Science **12**, 735719.

**Beer LA, Liu P, Ky B, Barnhart KT, Speicher DW.** 2017. Efficient quantitative comparisons of plasma proteomes using label-free analysis with MaxQuant. Methods in Molecular Biology **1619**, 339–352.

**Benjamini Y, Hochberg Y.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B: Statistical Methodology **57**, 289–300.

**Bishop DJ, Hawley JA.** 2022. Reassessing the relationship between mRNA levels and protein abundance in exercised skeletal muscles. Nature Reviews. Molecular Cell Biology **23**, 773–774.

**Buccitelli C, Selbach M.** 2020. mRNAs, proteins and the emerging principles of gene expression control. Nature Reviews. Genetics **21**, 630–644.

**Chandler JW, Werr W.** 2020. A phylogenetically conserved APETALA2/ETHYLENE RESPONSE FACTOR, ERF12, regulates Arabidopsis floral development. Plant Molecular Biology **102**, 39–54.

**Chen D, Yan W, Fu LY, Kaufmann K.** 2018. Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana*. Nature Communications **9**, 4534.

**Chiva C, Olivella R, Borras E, Espadas G, Pastor O, Sole A, Sabido E.** 2018. QCloud: a cloud-based quality control system for mass spectrometry-based proteomics laboratories. PLoS One **13**, e0189209.

**Costanzo E, Trehin C, Vandenbussche M.** 2014. The role of WOX genes in flower development. Annals of Botany **114**, 1545–1553.

**Csardi G, Franks A, Choi DS, Airoldi EM, Drummond DA.** 2015. Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. PLoS Genetics **11**, e1005206.

**Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF.** 2004. The SEP4 gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. Current Biology **14**, 1935–1940.

**Edfors F, Danielsson F, Hallstrom BM, Kall L, Lundberg E, Ponten F, Forsstrom B, Uhlen M.** 2016. Gene-specific correlation of RNA and protein levels in human cells and tissues. Molecular Systems Biology **12**, 883.

**Fan D, Cong Y, Liu J, Zhang H, Du Z.** 2024. Spatiotemporal analysis of mRNA–protein relationships enhances transcriptome-based developmental inference. Cell Reports **43**, 113928.

**Gao Y, Liu X, Tang B, *et al*.** 2017. Protein expression landscape of mouse embryos during pre-implantation development. Cell Reports **21**, 3957–3969.

**Gerdes Gyuricza I, Chick JM, Keele GR, Deighan AG, Munger SC, Korstanje R, Gygi SP, Churchill GA.** 2022. Genome-wide transcript and protein analysis highlights the role of protein homeostasis in the aging mouse heart. Genome Research **32**, 838–852.

**Hermjakob H, Montecchi-Palazzi L, Lewington C, *et al*.** 2004. IntAct: an open source molecular interaction database. Nucleic Acids Research **32**, D452–D455.

**Hoogendijk AJ, Pourfarzad F, Aarts CEM, Tool ATJ, Hiemstra IH, Grassi L, Frontini M, Meijer AB, van den Biggelaar M, Kuijpers TW.** 2019. Dynamic transcriptome–proteome correlation networks reveal human myeloid differentiation and neutrophil-specific programming. Cell Reports **29**, 2505–2519.

**Huang Y, Zhou L, Hou C, Guo D.** 2022. The dynamic proteome in *Arabidopsis thaliana* early embryogenesis. Development **149**, dev200715.

**Jiang L, Wang M, Lin S, *et al*.** 2020. A quantitative proteome map of the human body. Cell **183**, 269–283.

**Kanehisa M, Goto S.** 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research **28**, 27–30.

**Kaufmann K, Wellmer F, Muino JM, *et al*.** 2010. Orchestration of floral initiation by APETALA1. Science **328**, 85–89.

**Kumar D, Bansal G, Narang A, Basak T, Abbas T, Dash D.** 2016. Integrating transcriptome and proteome profiling: strategies and applications. Proteomics **16**, 2533–2544.

**Langfelder P, Horvath S.** 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9**, 559.

**Liljegren SJ, Gustafson-Brown C, Pinyopich A, Ditta GS, Yanofsky MF.** 1999. Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. The Plant Cell **11**, 1007–1018.

**Liu Y, Beyer A, Aebersold R.** 2016. On the dependency of cellular protein levels on mRNA abundance. Cell **165**, 535–550.

**Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF.** 1992. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. Nature **360**, 273–277.

**Meleth S, Deshane J, Kim H.** 2005. The case for well-conducted experiments to validate statistical protocols for 2D gels: different pre-processing = different lists of significant proteins. BMC Biotechnology **5**, 7.

**Mergner J, Frejno M, List M, *et al*.** 2020a. Mass-spectrometry-based draft of the Arabidopsis proteome. Nature **579**, 409–414.

**Mergner J, Frejno M, Messerer M, Lang D, Samaras P, Wilhelm M, Mayer KFX, Schwechheimer C, Kuster B.** 2020b. Proteomic and transcriptomic profiling of aerial organ development in Arabidopsis. Scientific Data **7**, 334.

**Mergner J, Kuster B.** 2022. Plant proteome dynamics. Annual Review of Plant Biology **73**, 67–92.

**Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, Bader GD, Ferrin TE.** 2011. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics **12**, 436.

**Naramoto S, Hata Y, Kyozuka J.** 2020. The origin and evolution of the ALOG proteins, members of a plant-specific transcription factor family, in land plants. Journal of Plant Research **133**, 323–329.

**Ng M, Yanofsky MF.** 2001. Activation of the Arabidopsis B class homeotic genes by APETALA1. The Plant Cell **13**, 739–753.

**Oliva-Vilarnau N, Vorrink SU, Ingelman-Sundberg M, Lauschke VM.** 2020. A 3D cell culture model identifies Wnt/beta-catenin mediated inhibition of p53 as a critical step during human hepatocyte regeneration. Advanced Science **7**, 2000248.

**O'Maoiléidigh DS, Thomson B, Raganelli A, Wuest SE, Ryan PT, Kwasniewska K, Carles CC, Graciet E, Wellmer F.** 2015. Gene network analysis of *Arabidopsis thaliana* flower development through dynamic gene perturbations. The Plant Journal **83**, 344–358.

**O'Maoiléidigh DS, Thomson B, Wellmer F.** 2023. Floral induction systems for the study of Arabidopsis flower development. Methods in Molecular Biology **2686**, 285–292.

**Ó'Maoiléidigh DS, Wuest SE, Rae L, *et al*.** 2013. Control of reproductive floral organ identity specification in Arabidopsis by the C function regulator AGAMOUS. The Plant Cell **25**, 2482–2503.

**Omidbakhshfard MA, Sokolowska EM, Di Vittori V, Perez de Souza L, Kuhalskaya A, Brotman Y, Alseekh S, Fernie AR, Skirycz A.** 2021. Multi-omics analysis of early leaf development in *Arabidopsis thaliana*. Patterns **2**, 100235.

**Pajoro A, Madrigal P, Muino JM, *et al*.** 2014. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biology **15**, R41.


Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026

**Perez-Riverol Y, Bai J, Bandla C, *et al***. 2022. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Research **50**, D543–D552.

**Perkins DN, Pappin DJ, Creasy DM, Cottrell JS.** 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20**, 3551–3567.

**Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S.** 2006. Stochastic mRNA synthesis in mammalian cells. PLoS Biology **4**, e309.

**Reeves PH, Ellis CM, Ploense SE, *et al***. 2012. A regulatory network for coordinated flower maturation. PLoS Genetics **8**, e1002506.

**Rieu P, Beretta VM, Caselli F, *et al***. 2024. The ALOG domain defines a family of plant-specific transcription factors acting during Arabidopsis flower development. Proceedings of the National Academy of Sciences, USA **121**, e2310464121.

**Ryan PT, O'Maoiléidigh DS, Drost HG, Kwasniewska K, Gabel A, Grosse I, Graciet E, Quint M, Wellmer F.** 2015. Patterns of gene expression during Arabidopsis flower development from the time of initiation to maturation. BMC Genomics **16**, 488.

**Saddic LA, Huvermann B, Bezhani S, Su Y, Winter CM, Kwon CS, Collum RP, Wagner D.** 2006. The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. Development **133**, 1673–1682.

**Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T.** 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research **13**, 2498–2504.

**Sidhaye J, Trepte P, Sepke N, Novatchkova M, Schutzbier M, Durnberger G, Mechtler K, Knoblich JA.** 2023. Integrated transcriptome and proteome analysis reveals posttranscriptional regulation of ribosomal genes in human brain organoids. eLife **12**, e85135.

**Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ.** 2006. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Molecular & Cellular Proteomics **5**, 144–156.

**Sirbu A, Kerr G, Crane M, Ruskin HJ.** 2012. RNA-Seq vs dual- and single-channel microarray data: sensitivity analysis for differential expression and clustering. PLoS One **7**, e50986.

**Smyth DR, Bowman JL, Meyerowitz EM.** 1990. Early flower development in Arabidopsis. The Plant Cell **2**, 755–767.

**Song YC, Das D, Zhang Y, Chen MX, Fernie AR, Zhu FY, Han J.** 2023. Proteogenomics-based functional genome research: approaches, applications, and perspectives in plants. Trends in Biotechnology **41**, 1532–1548.

**Su G, Kuchinsky A, Morris JH, States DJ, Meng F.** 2010. GLay: community structure analysis of biological networks. Bioinformatics **26**, 3135–3137.

**Szklarczyk D, Morris JH, Cook H, *et al***. 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Research **45**, D362–D368.

**Tang F, Chen N, Zhao M, Wang Y, He R, Peng X, Shen S.** 2017. Identification and functional divergence analysis of WOX gene family in paper mulberry. International Journal of Molecular Sciences **18**, 1782.

**Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS.** 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science **329**, 533–538.

**Tarazona S, Balzano-Nogueira L, Conesa A.** 2018. Multiomics data integration in time series experiments. In: Jaumot J, Bedia C, Tauler R, eds. Comprehensive analytical chemistry. Amsterdam: Elsevier, 505–532.

**Theissen G, Melzer R, Rumpler F.** 2016. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. Development **143**, 3259–3271.

**Thomson B, Wellmer F.** 2019. Molecular regulation of flower development. Current Topics in Developmental Biology **131**, 185–210.

**Tian L, Musetti V, Kim J, Magallanes-Lundback M, DellaPenna D.** 2004. The Arabidopsis LUT1 locus encodes a member of the cytochrome p450 family that is required for carotenoid epsilon-ring hydroxylation activity. Proceedings of the National Academy of Sciences, USA **101**, 402–407.

**Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB.** 2001. Missing value estimation methods for DNA microarrays. Bioinformatics **17**, 520–525.

**Vogel C, Marcotte EM.** 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature Reviews. Genetics **13**, 227–232.

**Wang D, Eraslan B, Wieland T, *et al***. 2019. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Molecular Systems Biology **15**, e8503.

**Wang P, Wu X, Shi Z, *et al***. 2023. A large-scale proteogenomic atlas of pear. Molecular Plant **16**, 599–615.

**Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH.** 2008. Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in Arabidopsis. Plant Physiology **148**, 1201–1211.

**Wellmer F, Alves-Ferreira M, Dubois A, Riechmann JL, Meyerowitz EM.** 2006. Genome-wide analysis of gene expression during early arabidopsis flower development. PLoS Genetics **2**, e117.

**Wils CR, Kaufmann K.** 2017. Gene-regulatory networks controlling inflorescence and flower development in *Arabidopsis thaliana*. Biochimica et Biophysica Acta **1860**, 95–105.

**Wuest SE, O'Maoiléidigh DS, Rae L, Kwasniewska K, Raganelli A, Hanczaryk K, Lohan AJ, Loftus B, Graciet E, Wellmer F.** 2012. Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. Proceedings of the National Academy of Sciences, USA **109**, 13452–13457.

**Yin P, Fan H, Hao Q, Yuan X, Wu D, Pang Y, Yan C, Li W, Wang J, Yan N.** 2009. Structural insights into the mechanism of abscisic acid signaling by PYL proteins. Nature Structural & Molecular Biology **16**, 1230–1236.

**Yu G, Wang LG, Han Y, He QY.** 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS **16**, 284–287.

**Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X.** 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One **9**, e78644.

Downloaded from https://academic.oup.com/jxb/article/76/10/2743/7945333 by guest on 23 June 2026