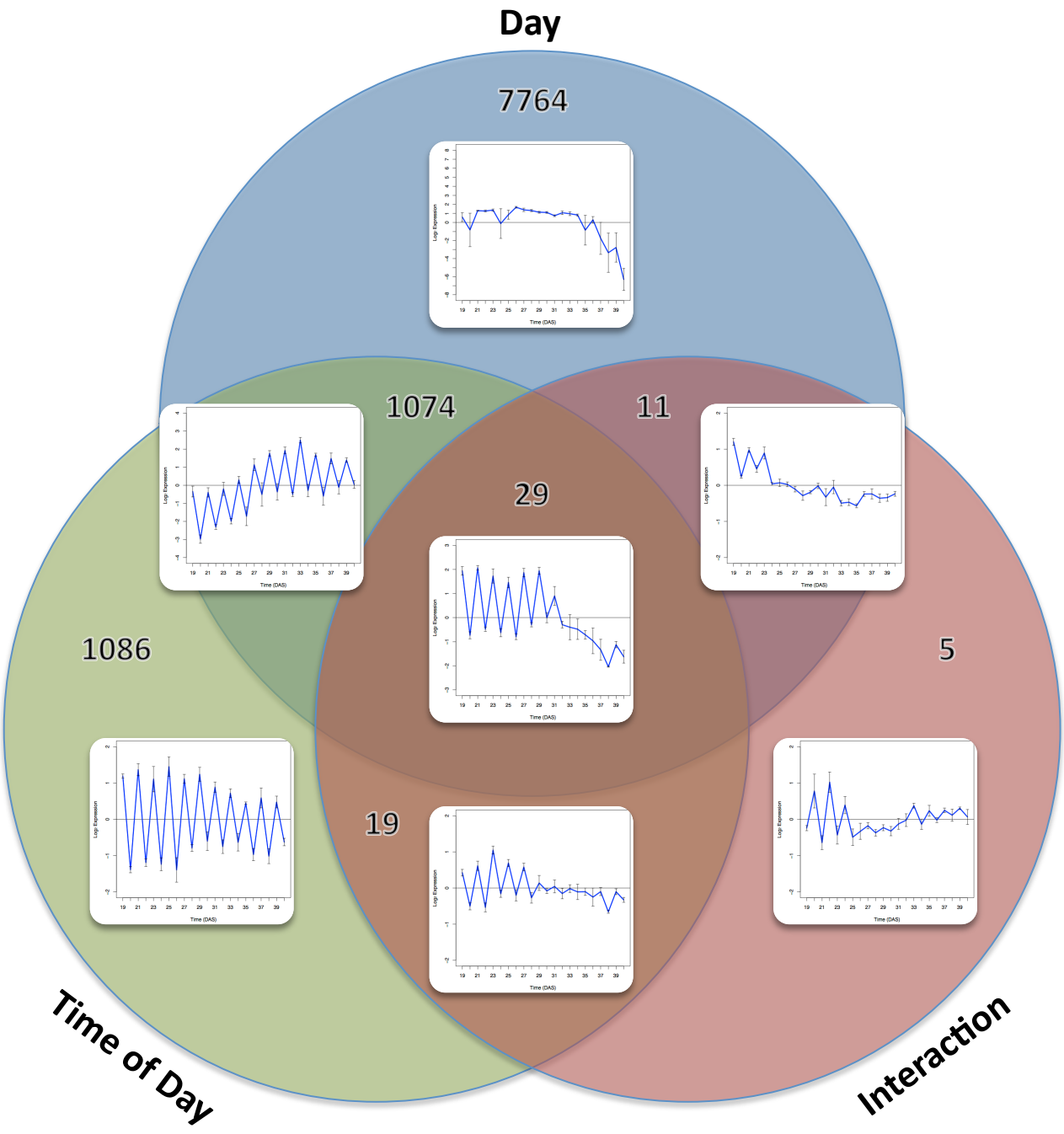
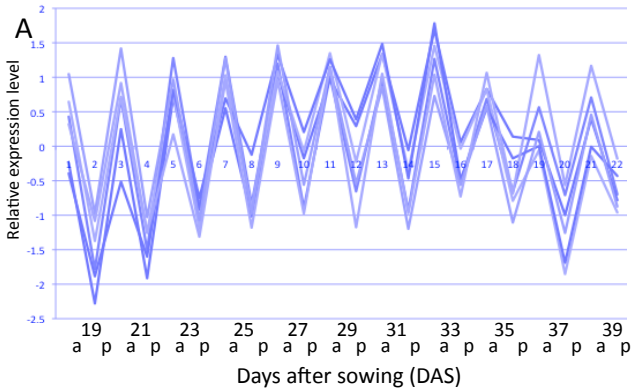
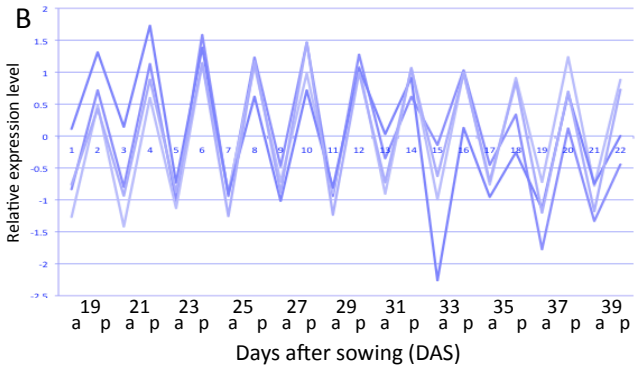


Supplemental Figure 1: Venn diagram summarising numbers of genes showing significant different expression for different combinations of treatment terms in the MAANOVA fixed model.
For each fixed term of the MAANOVA model – **Day**, **Time of Day**, and the **Interaction** between these factors, each gene was assessed for differential expression associated with that term relative to the biological (between-plant) variability using an F-test. After applying a false discovery rate multiple testing correction, the numbers of genes with significant test statistics for different combinations of terms (individual terms only, pairs of terms, all three terms) are shown. The expression profile for an exemplar gene has been plotted for each combination of significant terms. Those with just a significant effect of **Day** show a smooth pattern within each Day, but either a general trend across the complete time course or a strong change in expression levels at some point during the time course. Those with just a significant effect of **Time of Day** show a strong diurnal pattern, but no change in expression between days. Those with just a significant **Interaction** effect have a diurnal pattern for some period during the time course which is not present at other times. Genes with multiple significant terms show the appropriate combinations of these patterns.



AT5G02810	PSEUDO-RESPONSE REGULATOR 7 (PRR7)
AT4G08920	CRYPTOCHROME 1 (CRY1)
AT2G43010	phytochrome interacting factor 4 (PIF4)
AT1G09570	PHYTOCHROME A (PHYA)
AT2G46830	CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)
AT1G01060	LATE ELONGATED HYPOCOTYL (LHY)

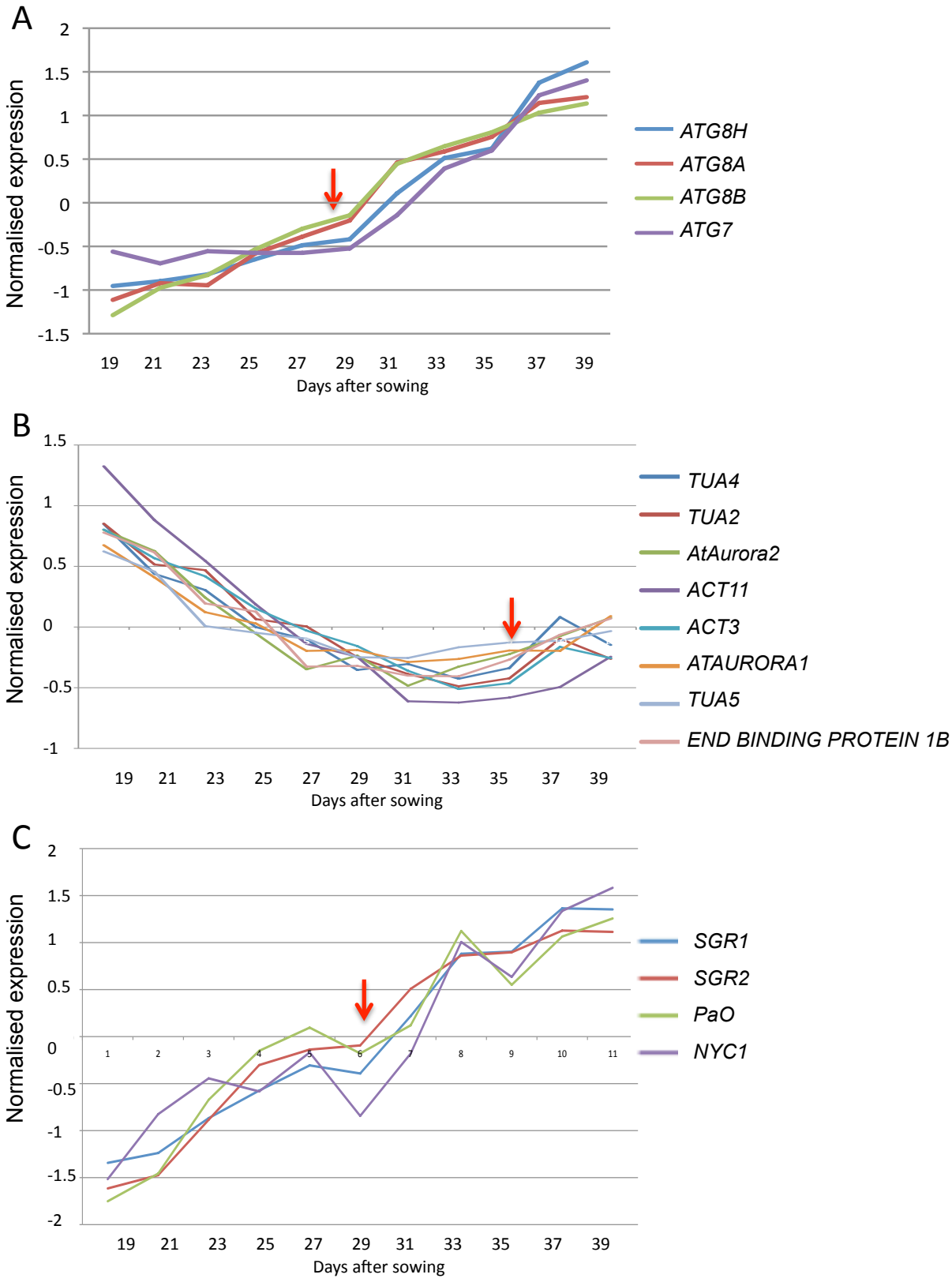


AT5G60100	ARABIDOPSIS PSEUDO-RESPONSE REGULAT
AT2G40080	EARLY FLOWERING 4 (ELF4)
AT3G22231	PATHOGEN AND CIRCADIAN CONTROLLED 1
AT3G46640	PHYTOCLOCK 1 (PCL1)

Supplemental Figure 2. Expression patterns of selected genes showing Time-of-Day changes only.

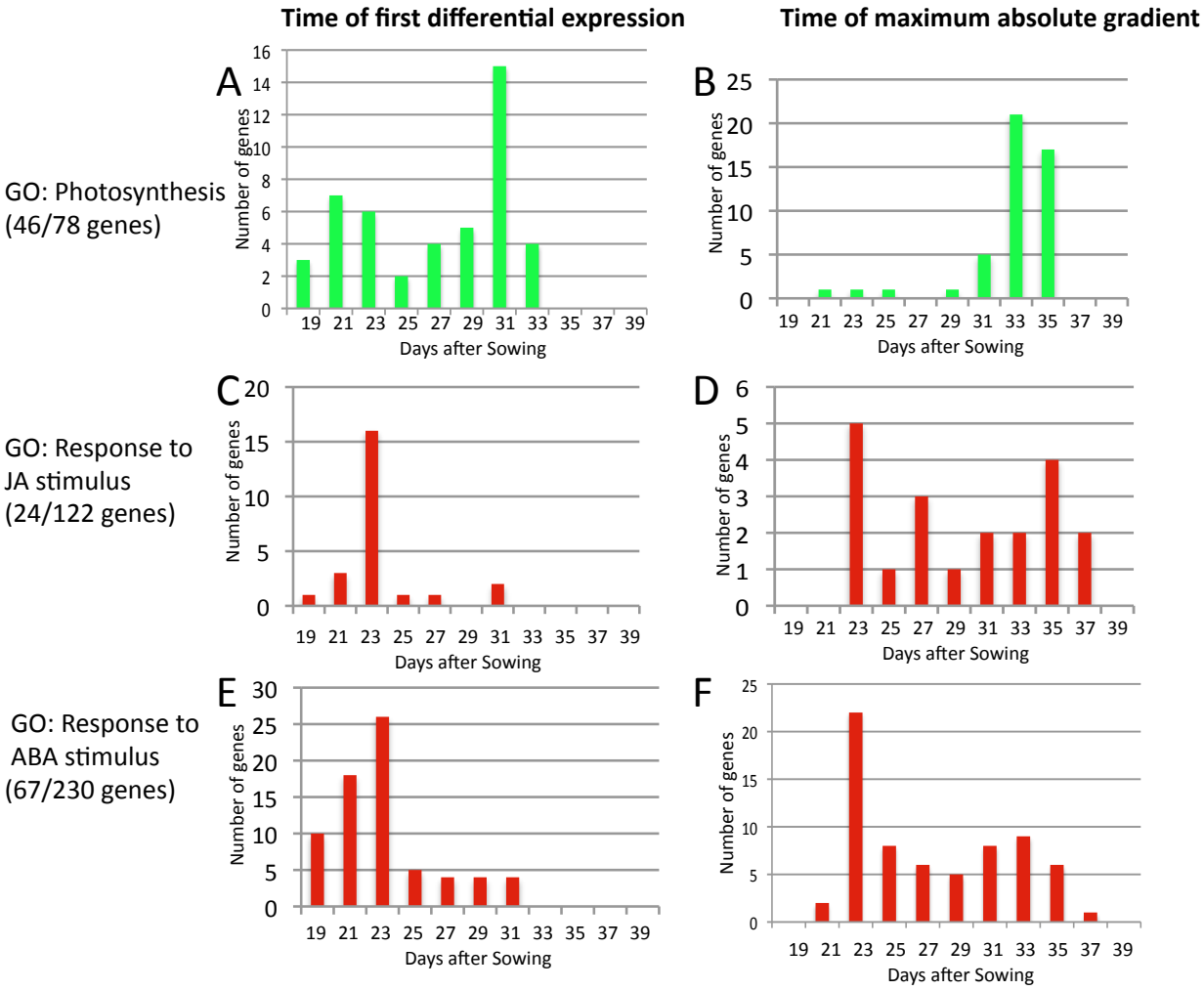
Genes that showed a significant Time-of-Day variation but no variation due to Day were clustered and selected examples are illustrated.

A. Morning genes showing higher expression in morning (a) samples. Examples shown have roles in the circadian clock and light signalling.
B. Afternoon genes showing higher expression in the afternoon (p) samples. Examples shown have a role in the circadian clock.



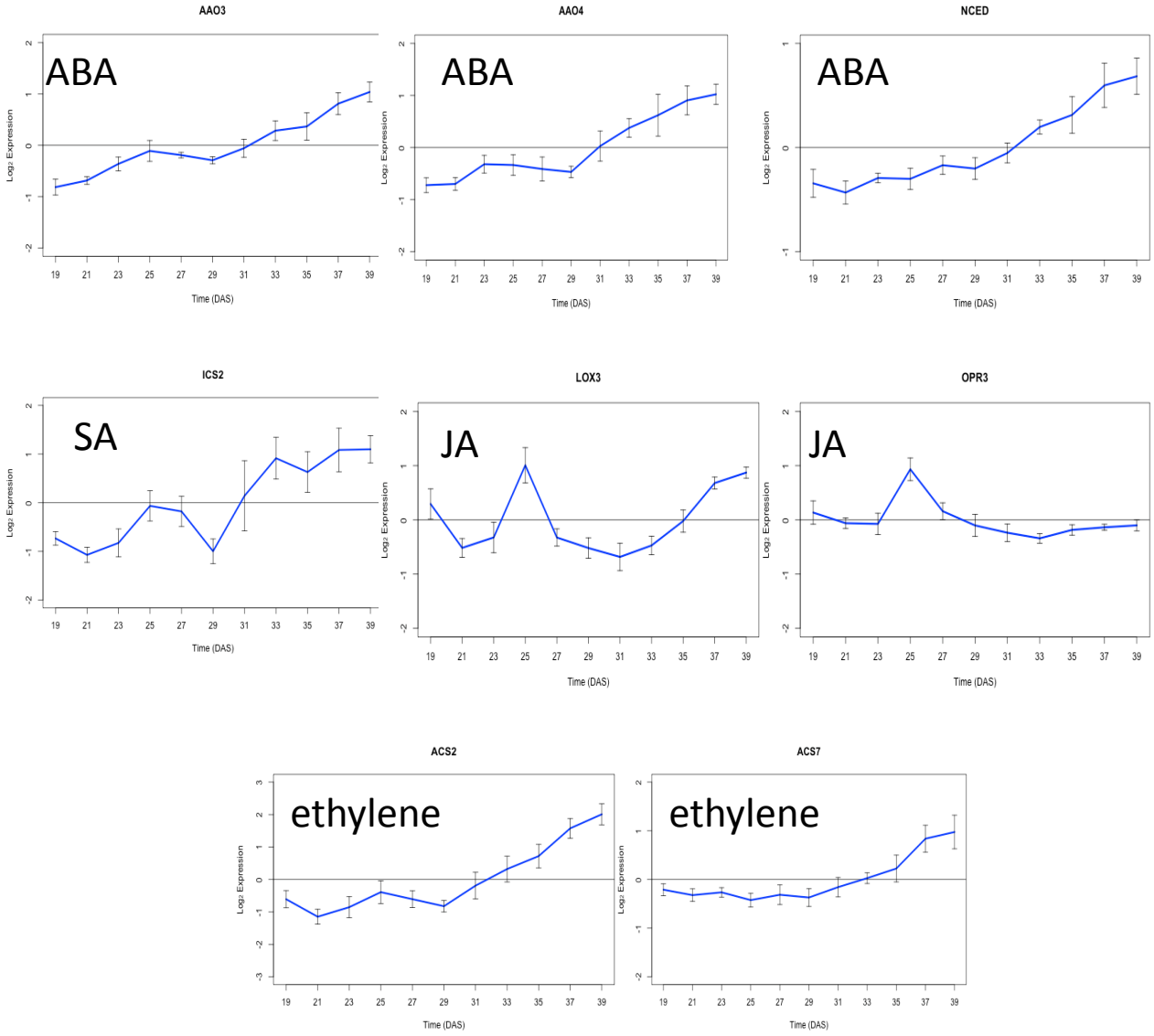
Supplemental Figure 3. Expression patterns of selected genes during leaf senescence. Data for the 11 time point series are shown.

A. Expression of selected autophagy genes during leaf senescence. Arrow shows parallel increase in expression at 29 DAS.
B. Cytoskeleton genes showing increased expression late in senescence shown by the arrow.
C. Expression patterns of chlorophyll degradation genes during senescence. Arrow shows the the parallel increase in expression of each gene at 29 DAS



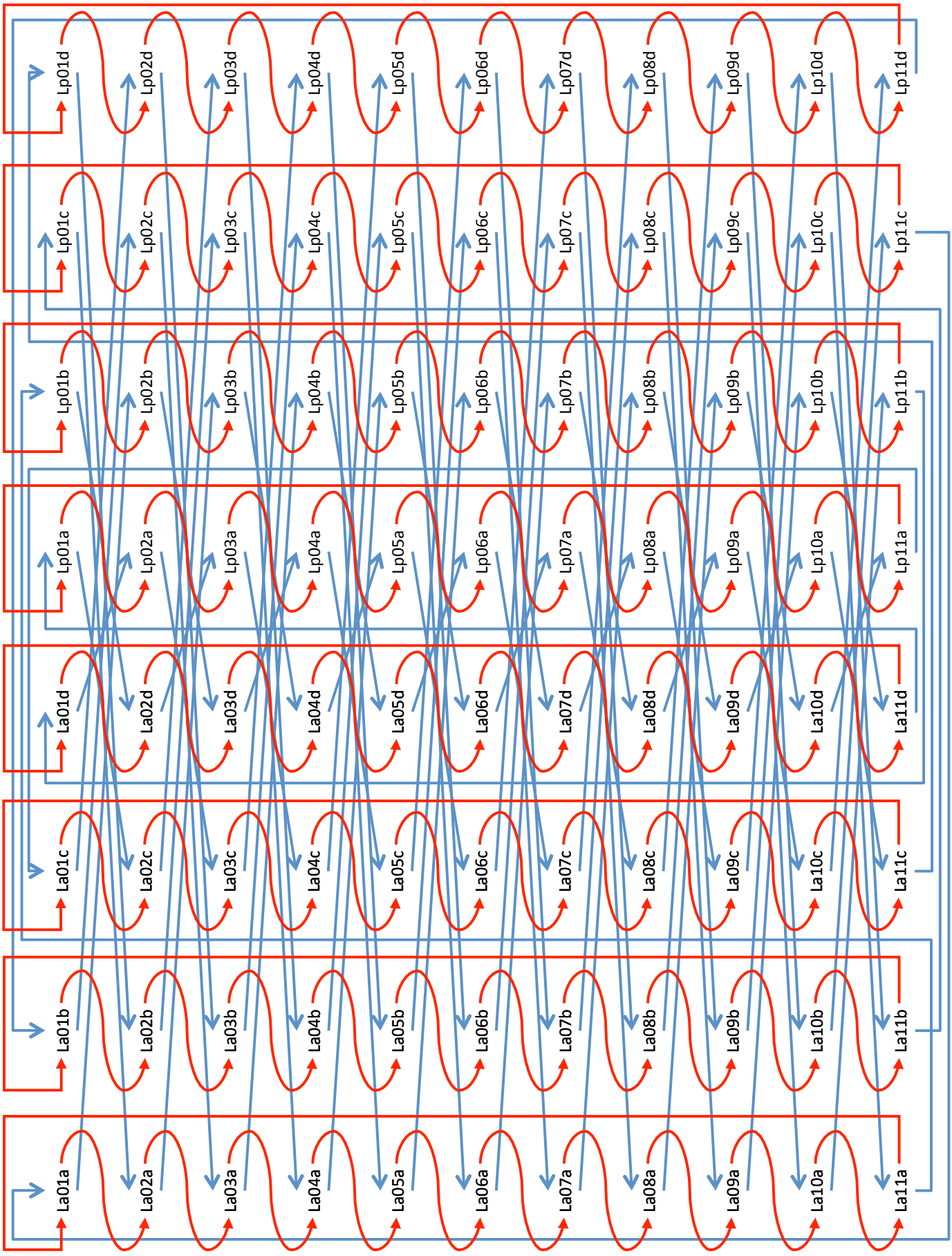
Supplemental Figure 4. Gradient analysis on selected groups of genes

Differentially expressed genes with GO term annotation 'photosynthesis' (**A & B**); 'Response to JA stimulus' (**C & D**) and 'Response to ABA stimulus' (**E & F**) were examined using the Gradient analysis tool for time of first differential expression (**A, C and E**) and time of maximum absolute gradient (**B, D and F**). (Number of differentially expressed genes versus total number of genes annotated for each GO term are shown in brackets)



Supplemental Figure 5. Expression patterns of selected hormone biosynthesis genes

Gene expression patterns (from GeneViewer, 11 timepoint data set) for selected genes involved in hormone biosynthesis.
ABA biosynthesis genes AAO3 (At2g27150), AAO4 (At1g04580, NCED (At3g14440); SA biosynthesis gene, ICS2 (At1g18870); JA biosynthesis genes LOX3 (At1g17420), OPR3 (At2g06050) and ethylene biosynthesis genes ACS2 (At1g01480), ACS7 (At4g26200).



Supplemental Figure 6. Representation of the experimental design for the microarray experiment.

Supplemental Figure 6. Representation of the complex loop design used for the microarray experiment.

Treatment codes indicate whether the sample was taken at 7h (a) or 14 h (p), the occasion on which the sample was taken (01 to 11; i.e. 19 to 39 DAS), and the arbitrary biological replicate label (a, b, c or d), so that code La04c indicates the sample collected on the fourth sampling occasion at the 7h time point from biological replicate c. Arrows link the pairs of samples appearing together on a microarray slide, with the arrow point indicating labelling with Cy5 and the arrow start indicating labelling with Cy3. Red arrows indicate arrays forming short (11 array) timecourse loops for each arbitrarily labelled biological replicate for either 7h or 14 time points. Blue arrows indicate arrays forming an 88 array loop providing comparisons between samples collected at different time points, for different arbitrarily labelled biological replicates and on different sampling occasions. Each of these arrays directly compares a sample collected at the 7h time point with one collected at the 14h time point.

Supplemental Table 1. Enriched GO terms in genes down or up regulated during senescence.

Enriched GO terms in the 2849 down regulated genes and the 3292 upregulated genes were identified using BiNGO and GO_full annotation (Maere *et al.* , 2005). Significantly overepresented terms with a p-value less than 0.05 following a Benjamini and Hochberg false discovery rate correction are shown.

GO ID	Clusters 1-24 (Down)	corrected p-value
9536	plastid	0.00E+00
5737	cytoplasm	0.00E+00
34357	photosynthetic membrane	1.00E-100
3824	catalytic activity	8.94E-32
15979	photosynthesis	2.74E-31
9543	chloroplast thylakoid lumen	9.58E-26
30090	photosystem	1.41E-22
9532	plastid stroma	5.70E-19
6082	organic acid metabolic process	1.36E-17
33014	tetrapyrrole biosynthetic process	2.37E-16
19684	photosynthesis, light reaction	3.27E-16
6520	amino acid metabolic process	6.19E-15
9523	photosystem II	7.46E-14
15995	chlorophyll biosynthetic process	6.51E-13
5975	carbohydrate metabolic process	2.12E-12
10287	plastoglobule	3.46E-11
9522	photosystem I	3.20E-08
9657	plastid organization and biogenesis	5.94E-08
9765	photosynthesis, light harvesting	8.75E-06
15977	carbon utilization	1.62E-05
9654	oxygen evolving complex	1.15E-04
30675	Rac GTPase activator activity	1.73E-04
6633	fatty acid biosynthetic process	3.54E-04
9085	lysine biosynthetic process	3.68E-04
43572	plastid fission	3.68E-04
9767	photosynthetic electron transport	7.37E-04
9628	response to abiotic stimulus	7.60E-04
19685	photosynthesis, dark reaction	8.05E-04
5840	ribosome	8.52E-04
16117	carotenoid biosynthetic process	8.65E-04
9538	photosystem I reaction center	1.38E-03
9416	response to light stimulus	1.44E-03

GO ID	Clusters 27-48 (Up)	corrected p-value
42221	response to chemical stimulus	2.89E-11
51869	response to stimulus	5.40E-11
9737	response to abscisic acid stimulus	6.76E-08
6950	response to stress	1.02E-07
46872	metal ion binding	1.04E-07
6914	autophagy	1.04E-07
9628	response to abiotic stimulus	5.81E-07
6970	response to osmotic stress	7.08E-07
9725	response to hormone stimulus	1.06E-05
9651	response to salt stress	1.63E-05
9415	response to water	2.11E-05
44248	cellular catabolic process	2.74E-05
9414	response to water deprivation	3.80E-05
9607	response to biotic stimulus	1.11E-04
30528	transcription regulator activity	1.14E-04
16602	CCAAT-binding factor complex	1.25E-03
6724	lipid catabolic process	1.32E-03
5478	transporter activity	2.27E-03
42594	response to starvation	2.50E-03
9723	response to ethylene stimulus	8.00E-03
10189	vitamin E biosynthetic process	9.91E-03

Supplemental Table 2. Genes included in the vBSSM model shown in Figure 9.

Gene name	AGI	22 time point cluster number	Upregulated in ANAC092 inducible OX line*	Function
RNS1	AT2G02990	38	yes	RIBONUCLEASE 1 (RNS1)
MT1C	AT1G07610	38	yes	MT1C (metallothionein 1C)
ARR16	AT2G40670	40	yes	ARABIDOPSIS RESPONSE REGULATOR 16 (ARR16)
PMI	AT1G62760	41	yes	invertase/pectin methylesterase inhibitor family protein
TBP1-1	At5g13820	41	yes	TELOMERIC DNA BINDING PROTEIN 1 (TBP1)
MYB78	At5g49620	41	no	myb domain protein 78 (AtMYB78)
SS3	AT1G74000	41	yes	STRICTOSIDINE SYNTHASE 3 (SS3)
NAC84	At5g14000	41	yes	Arabidopsis NAC domain containing protein 84 (ANAC084)
AMC6	AT1G79320	41	yes	metacaspase 6 (AtMC6)
AMC9	AT5G04200	41	yes	metacaspase 9 (AtMC9)
PMZ	At3g28210	41	no	PMZ; FUNCTIONS IN: zinc ion binding
BFN1	AT1G11190	41	yes	BIFUNCTIONAL NUCLEASE I (BFN1)
ATG8H	AT3G06420	42	yes	autophagy 8h (ATG8H)
NAC55	At3g15500	42	no	ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 55 (ANAC055)
GLY1	AT1G15380	42	yes	lactoylglutathione lyase family protein / glyoxalase I family protein
NAC102	At5g63790	42	no	ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 102 (ANAC102)
SEN	AT1G19200	42	yes	senescence-associated protein-related
MYB2	At2g47190	42	no	MYB DOMAIN PROTEIN 2 (MYB2)
ERF	At4g29100	42	no	ethylene-responsive family protein
SRG1	AT1G17020	42	yes	SENESCENCE-RELATED GENE 1 (SRG1)
PAP20	AT3G52780	42	yes	ATPAP20/PAP20; acid phosphatase/ protein serine/threonine phosphatase
ILR3	At5g54680	43	no	iaa-leucine resistant3 (ILR3)
STZ	At1g27730	43	no	salt tolerance zinc finger (STZ)
RD26	At4g27410	44	no	RESPONSIVE TO DESICCATION 26 (RD26)
NAC19	At1g52890	44	no	Arabidopsis NAC domain containing protein 19 (ANAC019)
NAC83	AT5G13180	45	yes	ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 83 (ANAC083)
NAC92	At5g39610	45	yes	ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 92 (ANAC092)
SINA	At3g13672	45	yes	seven in absentia (SINA) family protein
APG8A	AT4G21980	45	yes	AUTOPHAGY 8A (APG8A)

* data from Balazadeh et al., 2010

Supplemental Methods

Experimental design for microarray experiment

A novel experimental design strategy, based on the principle of the "loop design" (Kerr and Churchill, 2001), was developed to enable efficient extraction of information about key sample comparisons using a two-colour hybridisation experimental system. With 88 distinct samples (four biological replicates at each of 22 time points) to be compared, the experimental design included 176 two-colour microarray slides, allowing four technical replicates of each sample to be observed. Including both biological and technical replicates was essential given our interest in assessing the significance of differential expression relative to the biological variability, and therefore in identifying the relative sizes of these two sources of variation, and learning about the level of biological variability. The design takes account of the intended factorial structure (Day crossed with Time of Day) to be included in the analysis and, whilst not optimal, aims to provide good information about all fixed effects to be included in the model (including the effect of differences between biological replciates). Half of the slides were devoted to assessment of changes in gene expression between time points, using a simple loop design to link 11 samples from either the 7h time points or the 14h time points across the 11 sampling days, directly comparing samples collected on adjacent sampling days (i.e. 19 DAS with 21 DAS, 27 DAS with 29 DAS, etc.), and directly comparing the samples collected at 39 DAS with those collected at 19 DAS. Four separate loops were constructed for the 7h time points and four for the 14h time points, using the arbitrary biological replicate labelling to identify the samples to be included in each loop (Supplementary Figure S6– red arrows). The remaining slides provided assessment of differences between the 7h and 14h samples and between the arbitrarily labelled biological replicates, with some further assessment of changes between sampling days. All direct comparisons (pairs of samples hybridised together on a slide) were between 7h and 14h samples collected on adjacent sampling days (i.e. 19 DAS with 21 DAS, etc.), including comparisons between samples collected at 39 DAS and at 19 DAS, and between different arbitrarily labelled biological replicates. These 88 comparisons formed a single loop connecting all 88 treatments, therefore ensuring that the design was fully connected (allowing each sample to be compared with every other sample) (Figure S6– blue arrows). This design approach also removed any potential for dye bias, with each sample being labelled twice with Cy3 and twice with Cy5. The order in which the 176 slides were hybridized and scanned was randomised to minimise the impact on differences between samples of any potential variation in the processing conditions.

Data analysis and prediction methods using MAANOVA

A local adaptation of the MAANOVA (MicroArray ANalysis Of VAriance) package (Wu et al., 2003) was used to analyse the quantified microarray data, providing data quality assurance, within slide normalization through LOWESS data transformation, mixed model fitting and identification of genes showing significant differential expression via (approximate) F tests of fixed (treatment) terms included within the model. MAANOVA was selected to analyse the data because it is able to provide an accurate analysis of the effects on gene expression of multiple sources of variation (both fixed, treatment, terms, and random sources of background variation) in the experimental design, harnessing the power of direct comparisons between pairs of samples obtained using two-channel microarrays (Churchill, 2004). In order to prepare the data for mixed model fitting, the quality of the data is first maximised through a series of graphical comparisons, quantified using some simple statistical tests. We have made a number of improvements to the quality-checking functions of MAANOVA, including the quantification of effects using simple statistical tests, which have dramatically improved our ability to identify problems within the data and correct them at an experimental level prior to formal analysis. Types of artefacts which can be identified include: arrays with a dye bias; saturated or

underexposed spots; printing problems with specific arrays, such as probe smearing; non-specific binding regions of arrays; and location specific dye bias. With the additional functionality we have introduced we are also able to assess the quality of individual technical replicates for each sample by making comparisons between them, allowing the identification of suspect technical replicates for individual genes and the quantile estimation of more realistic values for these replicates, thus reducing the background variation (noise) and improving the ability to detect differential expression. Once the data had been thus cleaned, they were corrected for spatial variation within each slide, such as background noise and location specific dye bias, by a series of LOWESS transformations.

Fitting the mixed model to the observed data was done in two stages, as implemented in the MAANOVA package. The first stage effectively normalizes the data for overall variation of responses between arrays, between dyes and between combinations of arrays and dye, averaged across all the genes included on the arrays. The second stage takes the residuals from this first stage fitted model and fits a per gene model to assess for the effects of the different samples (treatment combinations, included as fixed terms in the model) on the response for each gene, allowing for gene-by-array and gene-by-dye variation (included as random terms in the model). The mixed model, fitted on a per gene basis can be represented as:

$$Y = \mu + Dye + Array + (Day * TimeOfDay) / BioRep + \varepsilon$$

in which Y is the residual gene expression value obtained from the first stage model fitting, μ is related to the average intensity of the gene throughout the experiment, Dye and Array are the random model terms, representing the gene-by-array and gene-by-dye variation, Day is the fixed model term representing variation in response between different sampling days, TimeOfDay is the fixed model term representing variation in response between different times of day (am and pm samples), “Day * TimeOfDay” indicates that both the main effects of these two terms and the interaction between them (allowing for different effects of time of day on different sampling days) are included in the model, BioRep is the fixed model term representing variation in response between biological replicates collected at each time point (with “/ BioRep” indicating that this fixed term is nested within the combinations of Day and TimeOfDay), and ε is measurement error which is unaccountable by other terms of the model (essentially variation between technical replicates for each sample, having allowed for variation between arrays and dyes for each gene). It can be argued that differences between biological replicates should be considered as a random effect in the mixed model, and, indeed, we were interested in assessing the variation between sampling times relative to the variability between biological replicates. However, subsequent modeling of gene networks required estimates of true biological replicates, so that the fitted model needed to allow estimation of the responses for each biological sample, and hence biological replicate needed to be included as a fixed effect term nested within the sampling time treatment combinations. Equally, it can be argued that differences between dyes should be considered as a fixed effect within the mixed model framework, because of the limited number of levels for this term. We took the pragmatic approach that dye represented a “nuisance” source of variability that we were not directly interested in, and so choose to include it as a random effect. In addition to the tests for significant differential expression due to each fixed effect term, output from the MAANOVA analysis can generate a single estimated expression score for each gene in each biological replicate at each time point, a necessary output for subsequent modeling of gene networks.

The resulting fitted mixed model allows extraction of fixed term estimates, which additively describe the expression levels of genes (based on the model formulae defined above). The amount of variation caused by each fixed term (Day, TimeOfDay, interaction between these terms, and nested biological replicates) was compared with the estimate of underlying (between technical replicate) variation using an F-test, with significant test statistics

Indicating significant variation in gene expression caused by these model terms. This provided a separate F statistic for each fixed model term for each spot of the array. Of more direct interest was to identify the significance of the time-course fixed model terms (Day, TimeOfDay and the interaction between these terms) relative to the pooled biological replicate variability rather than the underlying technical replicate variability. These approximate F statistics were easily obtained, based on the estimated variances calculated for each fixed effect term (including the nested between biological replicate term), by dividing the F statistic of each time-course fixed model term by the F statistic of the nested biological replicate term for each gene, cancelling out the estimated underlying (between technical replicate) variation included in the denominator of each original F statistic. Comparing these approximate F statistics to the appropriate F-distribution (based on the numerator degrees of freedom for the original F statistics) allows a significance level (p-value) to be associated with each F-test, representing the probability that the gene is showing a change in expression due to that term by chance alone, allowing for the level of variability between biological replicates for that gene. A multiple testing correction was applied to these F-tests using a step-down false discovery rate controlling procedure (Benjamini & Liu, 1999; Westfall et al., 1998) to calculate adjusted p-values under an overall false discovery rate of p=0.05. Having fitted the mixed model to each gene, predicted means were calculated for each of the 88 samples, either assuming the full treatment model (including effects of Day, Time of Day, the interaction between them and the nested biological replicates) to produce a 4-replicate 22 time point data set for each gene, or assuming a reduced treatment model (including just effects of Day and the nested biological replicates) to produce a 8-replicate 11 time point data set for each gene. These data sets were then used in subsequent analyses.

Supplemental Data. Breeze et al. (2011). Plant Cell 10.1105/tpc.111.083345.

Gaussian Process regression and gradient analysis.

Gaussian Processes

Using repeated microarray measurements, the gene expression of a particular gene over times $\mathbf{t} \in \{t_1, \ldots, t_N\}$ may be experimentally measured, denoted $\mathbf{y}$. For such experimental measurements it is often sensible to assume the underlying expression has been corrupted by some noise:

$$\mathbf{y} = f(\mathbf{t}) + \epsilon \tag{1}$$

where $f(\mathbf{t})$ represents the unobserved true gene-expression profile, and where ϵ represents independent Gaussian noise with zero-mean and covariance $\sigma_n^2\mathbb{I}$. If the function, $f(.)$ was explicitly known, then the conditional probability of the observation set $\mathbf{y}$ given the function, $\mathbf{f}$, may be calculated as:

$$\mathbf{y}|\mathbf{f} \quad \sim \quad \mathcal{N}(\mathbf{f}, \sigma_n^2\mathbb{I}) \tag{2}$$

In most cases, however, this latent function is unknown, with the function itself the object of interest. In these situations, it is often sensible to place a Gaussian process prior directly over the function. A Gaussian process is defined a collection of random variables, any finite number of which have a joint Gaussian distribution (Rasmussen and Williams. 2006), denoted,

$$f(t) \quad \sim \quad \mathcal{GP}(m(\mathbf{t}), k(t, t^*)) \tag{3}$$

where

$$m(ft) \quad = \quad \mathbb{E}[f(t)] \tag{4}$$
$$k(t, t^*) \quad = \quad \mathbb{E}[(f(t) - m(t)))(f(t^*) - m(t^*))] \tag{5}$$

denote the mean and covariance respectively.

A particular point of interest is making predictions about the process at times $\mathbf{t}^*$ given a set of training times $\mathbf{t} = \{t_1, \ldots, t_N\}$ and corresponding observed (noisy) gene-expression, $\mathbf{y} = \{y_1, \ldots, y_N\}$. This can be achieved by noting the full joint distribution is written,

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \quad \sim \quad \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{t}, \mathbf{t}) + \sigma_n^2\mathbb{I} & K(\mathbf{t}, \mathbf{t}^*) \\ K(\mathbf{t}^*, \mathbf{t}) & K(\mathbf{t}^*, \mathbf{t}^*) \end{bmatrix}\right) \tag{6}$$

which, by Bayes' rule suggests:

$$\mathbf{f}^*|\mathbf{y}, \mathbf{t}, \mathbf{t}^* \quad \sim \quad \mathcal{N}(\boldsymbol{\mu}, \bar{\mathbf{C}}) \tag{7}$$

where

$$\boldsymbol{\mu} \quad = \quad K(\mathbf{t}^*, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \sigma_n^2\mathbb{I}]^{-1}\mathbf{y}$$
$$\bar{\mathbf{C}} \quad = \quad K(\mathbf{t}^*, \mathbf{t}^*) - K(\mathbf{t}^*, \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \sigma_n^2\mathbb{I}]^{-1}K(\mathbf{t}*, \mathbf{t})$$

Consequently, it is possible to make predictions about the (unnoisy) behaviour of genes at times $\mathbf{t}^*$, conditioned upon experimental observations.

Covariance Function and Hyperparameter Selection

Making predictions about the underlying process at a particular time is dependent upon the choice of covariance function. For simplicity, the covariance function is chosen to be that of the isotropic squared-exponential covariance function (Rasmussen and Williams. 2006) such that the ith column of the jth row is calculated:

$$K(t_i, t_j) \quad = \quad \sigma_f^2 \exp\left(-\frac{1}{2l^2}(t_i - t_j)^2\right) \tag{8}$$

where $\mathcal{H} = \{l, \sigma_f^2, \sigma_n^2\}$ are tunable hyperparameters representing the length-scale of the process, the process variance and noise variance respectively. In this manner, the covariance and therefore the predictions are explicitly dependent upon the choice of hyperparmeters. By noting that the training data is drawn form the Gaussian distributions, $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, K + \sigma_n^2\mathbb{I})$, the marginal likelihood of the GP is computed for any set of hyperparameters as,

$$\log p(\mathbf{y}|\mathbf{t}, \mathcal{H}) \quad = \quad -\frac{1}{2}\mathbf{y}^T(K + \sigma_n^2\mathbb{I})^{-1}\mathbf{y} - \frac{1}{2}\log|K + \sigma_n^2\mathbb{I}| - \frac{n}{2}\log 2\pi. \tag{9}$$

The choice of hyperparameter may therefore be chosen to maximise this marginal likelihood

$$\hat{\mathcal{H}} \quad = \quad \arg_{\mathcal{H}}\max\, p(\mathbf{y}|\mathbf{t}, \mathcal{H}), \tag{10}$$

using e.g., a gradient based approach (Rasmussen and Williams. 2006).

Derivative Predictions using Gaussian Processes

The derivative of a Gaussian process is itself a Gaussian process (Rasmussen and Williams, 2006; Solak *et al.*, 2003). Given a set of training data, $\mathbf{y}$ observed for times $\mathbf{t}$ it should be possible to make predictions about the function derivative $\dot{\mathbf{f}}*$ at times $\mathbf{t}^*$ by noting the joint distribution can be written:

$$\begin{bmatrix} \mathbf{y} \\ \dot{\mathbf{f}}* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{t},\mathbf{t}) + \sigma_n^2\mathbb{I} & \frac{\partial K(\mathbf{t},\mathbf{t}^*)}{\partial \mathbf{t}^*} \\ \frac{\partial K(\mathbf{t}^*,\mathbf{t})}{\partial \mathbf{t}^*} & \frac{\partial^2 K(\mathbf{t}^*,\mathbf{t}^*)}{\partial \mathbf{t}^* \partial \mathbf{t}^*} \end{bmatrix}\right) \tag{11}$$

By Bayes' rule the conditional distribution is thus calculated:

$$\dot{\mathbf{f}}*|\mathbf{t},\mathbf{y},\mathbf{t}* \sim \mathcal{N}(\boldsymbol{\mu}, \bar{\mathbf{C}}) \tag{12}$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \frac{\partial K(\mathbf{t}^*,\mathbf{t})}{\partial \mathbf{t}^*}[K(\mathbf{t},\mathbf{t}) + \sigma_n^2\mathbb{I}]^{-1}\mathbf{y} \\ \bar{\mathbf{C}} &= \frac{\partial^2 K(\mathbf{t}^*,\mathbf{t}^*)}{\partial \mathbf{t}^* \partial \mathbf{t}^*} - \frac{\partial K(\mathbf{t}^*,\mathbf{t})}{\partial \mathbf{t}^*}[K(\mathbf{t},\mathbf{t}) + \sigma_n^2\mathbb{I}]^{-1}\frac{\partial K(\mathbf{t},\mathbf{t}^*)}{\partial \mathbf{t}^*} \end{aligned}$$

Thus, as in section it is possible to make predictions about the (unnoisy) derivative behaviour of genes at times $\mathbf{t}^*$, conditioned upon experimental observations.

Hypothesis Testing at time t

Using the methods described in section and it should be possible to calculate the (distribution over) gradients for each gene at each time point in the time course. In particular the conditional distribution of the gradient at time t^* given all experimental observations is a Gaussian distribution calculated according to Equation (12). Where temporal measurements are of sufficiently high resolution, a gene may be characterised as being switched-on (strong positive gradient), switched-off (strong negative gradient) or in a steady-state (zero gradient) depending upon the mean and variance of the marginal Gaussian distribution. Specifically, if the zero-point lies within a given number of standard-deviations of the posterior mean, the gene is determined as in steady-state, otherwise the gene is characterised as switched on or off according to the sign of the posterior mean. Where data is not sufficiently time-resolved to identify the gradient, the GP can nonetheless identify the derivative of global trends.

Supplemental References

Balazadeh, S., Siddiqui, H., Allu, A.D., Matallana-Ramirez, L.P., Caldana, C., Mehrnia, M., Zanor, M.I., Kohler, B., and Mueller-Roeber, B. (2010) A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ ORE1 during salt-promoted senescence. Plant J **62**: 250–264.

Benjamini, Y., and Liu, W. (1999). A step-down multiple testing procedure that controls the false discovery rate under independence. J. Statistical Planning and Inference, **82**: 163-170.

Churchill, G. (2004). Using ANOVA to analyze microarray data. BioTechniques, **37**: 173-175.

Kerr, M.K. and Churchill, G.A. (2001) Experimental design for gene expression microarrays. Biostatistics **2:** 183-201.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics **21**: 3448-3449.

Rasmussen, C.E. and Williams, C. K. I. (2006) Gaussian processes for Machine Learning. MIT Press, Second edition. ISBN 026218253X.

Solak, E., Murray-Smith, R., Leithead, W. E., Leith, D., and Rasmussen, C. E. (2003). Derivative Observations in Gaussian Process Models of Dynamic Systems. In Becker, S. S. T. and Obermayer, K., editors, Advances in Neural Information Processing Systems **15**: 1033–1040. MIT Press.

Westfall, P., Krishen, A., and Young, S.S. (1998). Using prior information to allocate significance levels for multiple endpoints. Statistics in Medicine **17:** 2107-2119.

Wu, H., Kerr, K., Cui, X., and Churchill, G. (2003). MAANOVA: a software package for the analysis of spotted cDNA microarray experiments. In: The Analysis of Gene Expression Data: Methods and Software. Editors: Parmigiani, G., Garett, E., Irizarry, R., and Zeger, S. pp 313–341.