Floral pathway integrator gene expression mediates gradual transmission of environmental and endogenous cues to flowering time

Aalt D.J. van Dijk [Corresp., 1, 2, 3] , **Jaap Molenaar** [1]

[1] Biometris, Department for Mathematical and Statistical Methods, Wageningen University, Wageningen, The Netherlands

[2] Laboratory of Bioinformatics, Wageningen University, Wageningen, The Netherlands

[3] Bioscience, Wageningen University and Research, Wageningen, The Netherlands

Corresponding Author: Aalt D.J. van Dijk
Email address: aaltjan.vandijk@wur.nl

The appropriate timing of flowering is crucial for the reproductive success of plants. Hence, intricate genetic networks integrate various environmental and endogenous cues such as temperature or hormonal statues. These signals integrate into a network of floral pathway integrator genes. At a quantitative level, it is currently unclear how the impact of genetic variation in signaling pathways on flowering time is mediated by floral pathway integrator genes. Here, using datasets available from literature, we connect *Arabidopsis thaliana* flowering time in genetic backgrounds varying in upstream signalling components with the expression levels of floral pathway integrator genes in these genetic backgrounds. Our modelling results indicate that flowering time depends in a quite linear way on expression levels of floral pathway integrator genes. This gradual, proportional response of flowering time to upstream changes enables a gradual adaptation to changing environmental factors such as temperature and light.

1 **Short title: Linear response of flowering time**

2 **Floral Pathway Integrator Gene Expression mediates Gradual Transmission of**

3 **Environmental and Endogenous Cues to Flowering Time**

4 A.D.J. van Dijk[1,2*] and J. Molenaar[1]

5 [1] Biometris, Wageningen UR, Wageningen, The Netherlands
6 [2] Bioscience, Wageningen UR, Wageningen, The Netherlands
7
8 * Corresponding author: aaltjan.vandijk@wur.nl

9 **Abstract**

10 The appropriate timing of flowering is crucial for the reproductive success of plants. Hence,

11 intricate genetic networks integrate various environmental and endogenous cues such as

12 temperature or hormonal statues. These signals integrate into a network of floral pathway

13 integrator genes. At a quantitative level, it is currently unclear how the impact of genetic

14 variation in signaling pathways on flowering time is mediated by floral pathway integrator genes.

15 Here, using datasets available from literature, we connect *Arabidopsis thaliana* flowering time in

16 genetic backgrounds varying in upstream signalling components with the expression levels of

17 floral pathway integrator genes in these genetic backgrounds. Our modelling results indicate that

18 flowering time depends in a quite linear way on expression levels of floral pathway integrator

19 genes. This gradual, proportional response of flowering time to upstream changes enables a

20 gradual adaptation to changing environmental factors such as temperature and light.

21

Introduction

The reproductive success of flowering plants depends on flowering at the right moment. Hence, plants have evolved genetic and molecular networks integrating various environmental cues with endogenous signals in order to flower under optimal conditions [1]. The signal transduction pathways that receive and transmit input signals include the photoperiod pathway, the vernalization pathway, the ambient temperature pathway, and the autonomous pathway [2]. The input from these pathways is integrated by a core set of floral pathway integrator genes [3]. The regulation of flowering time by these various factors has been extensively studied experimentally in the plant model species *Arabidopsis thaliana*. Substantial qualitative information is available about the factors involved and how these interact genetically, both for the signal transduction pathways and the floral pathway integrator genes [4]. Activation of the photoperiodic flowering pathway leads to transcriptional activation of *FLOWERING LOCUS T (FT)*, an activator of flowering. FT is produced in the leaves and moves to the shoot apical meristem [5], leading to activation of *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* [6] and *APETALA1 (AP1)* expression [7, 8]. The vernalization (winter cold) pathway inhibits the transcription of *FLOWERING LOCUS C (FLC)*. FLC, together with SHORT VEGETATIVE PHASE (SVP), represses the transcription of *SOC1* and *FT*. Thus FLC acts as a flowering repressor by blocking the photoperiodic flowering pathway. In the ambient temperature pathway, which involves amongst other *FLOWERING LOCUS M (FLM)* and *SVP*, small fluctuations in temperature influence flowering time via floral pathway integrators including *FT* and *SOC1* [9, 10]. *SOC1* integrates signals from multiple pathways and transmits the outcome to *LEAFY (LFY)* [11, 12]; *SOC1* is supposed to act at least partially via a positive feed-back loop in which AGAMOUS-LIKE 24 (AGL24) is involved upon dimerizing with SOC1 [13]. Autonomous pathway mutants are characterized by delayed flowering irrespective of day length. The proteins

46   encoded by the genes in the autonomous pathway generally fall into two broad functional

47   categories: general chromatin remodelling or maintenance factors, and proteins that affect RNA

48   processing [1]. Gibberellins influence the floral transition through the regulation of *SOC1* and

49   *LFY* [14]. LFY is a positive regulator of *AP1* [15] and the commitment to flower is ascertained

50   by a direct positive feed-back interaction between *AP1* and *LFY*. Once the expression of *AP1* has

51   been initiated, this transcription factor orchestrates the floral transition by specifying floral

52   meristem identity and regulating the expression of genes involved in flower development [16].

53   In addition to qualitative information on pathways involved in flowering time regulation,

54   recently quantitative information has become available. This includes flowering time

55   measurements under various conditions and in different genetic backgrounds [17, 18], and time

56   series of expression for key floral pathway integrator genes [18]. Such quantitative information

57   has enabled construction of a set of models describing flowering time regulation at the molecular

58   level [18-21]. Given the above-described complexity, computational models are useful tools to

59   comprehend flowering time regulation. One example of a quantitative finding from our model

60   [18] for the network of floral pathway integrator genes is that a disturbance in a particular gene

61   has not necessarily the largest impact on directly connected genes. For example, the model

62   predicts that *SOC1* mutation has a larger impact on *AP1*, which is not directly regulated by

63   SOC1, compared to its effect on *LFY* which is under direct control of SOC1. This prediction was

64   confirmed by expression data.

65   Flowering time regulation facilitates the successful dispersion of flowering plants over the world

66   [2] by contributing to the adaptation of plants to different environmental conditions. In this

67   context, it is an important question how genetic variation in the various signaling pathways

68   influences flowering time regulation. Can we describe the effect of genetic variation in these

69 signaling pathways by linking the magnitude of flowering time change to the magnitude of

70 expression change of floral pathway integrator genes? If so, what type of relationship exists

71 between expression levels of floral pathway integrator genes and flowering time in genetic

72 backgrounds which differ in signaling components?

73 The above-mentioned quantitative analyses focus on one specific Arabidopsis genetic

74 background, without genetic difference in signaling pathways being taken into account, leaving

75 these questions so far unanswered. In principle, one could imagine answering these questions by

76 extending these models to include a large number of signaling pathway components. However,

77 construction of such large models would lead to serious complications in terms of e.g. parameter

78 estimation. Here we follow a different route to investigate how the effect of genetic variation in

79 components of upstream signalling pathways on flowering time is mediated by floral pathway

80 integrator genes. We establish a quantitative connection between expression levels of floral

81 pathway integrator genes, and flowering times in various genetic backgrounds differing in

82 upstream signal components. This demonstrates that in many cases, floral pathway integrator

83 genes transmit perturbations to flowering time via gradual, proportional changes in their

84 expression levels. Our current study is complementary to our previous modelling approach

85 which focused on the floral pathway integrator gene network, and not on the input to this

86 network by upstream signalling components. This analysis provides a quantitative understanding

87 of the effect of variation in the various input pathways on flowering time, which will ultimately

88 enable us to better understand plant adaptation.

89   **Methods**

90   *Simulations*

91   Predictions from the dynamic flowering time model were obtained using the model as presented

92   in ref. [18]. This consists of a set of Ordinary Differential Equations (ODEs) for the dynamics of

93   *AP1*, *LFY*, *SOC1*, *FD*, *FT* and *AGL24*; *SVP* and *FLC* are present as external inputs in the model.

94   In each of the six ODEs, regulation of gene expression is described by one or more terms of the

95   form $\beta*f(x)$, where f is a function of concentrations *x* of one or more regulators. To simulate the

96   effect of genetic variation in upstream signalling pathways influencing a given gene, the value of

97   each parameter β in its equation was modified by multiplying it with a factor *a* ranging from 0.05

98   to 10 in steps of 0.05 and subsequently from 10 to 100 in steps of 1. The resulting flowering time

99   after simulating the modified model was obtained, as well as the expression value of the gene

100   itself at day 10 (this timepoint was used because it matches closely with the timepoint used in

101   much of the experimental datasets that we used). Out of the resulting expression values, a range

102   of ten-fold expression change was chosen around the unperturbed expression level at day 10. For

103   *SVP* and *FLC* there is no ODE because these genes are present as external inputs in the model.

104   For these, variation in upstream signalling pathways was simulated by simply setting the level of

105   the gene to different fixed levels. For *SVP* this again involved a range of ten-fold expression

106   change; for *FLC* this range was arbitrarily made larger because of the small effect of ten-fold

107   expression change.

108

109   *Experimental data*

110   We use data from a randomly chosen subset of genes for which mutations are described as

111   impacting flowering time [22]. Our dataset has at least several examples per floral pathway

112   integrator gene. Data was extracted from figures or tables in papers describing the effect of

113   mutations of particular genes on flowering time, and presenting the expression level of genes

114   involved in signal integration. Expression measurements in different experiments are made at

115   different days and/or different tissues, but such differences are not taken into account. Also, in

116   particular for *FT*, often values are provided for several timepoints during one day (to capture the

117   circadian rhythm). Although in such case in principle it would be best to record the total area

118   under the curve (sum of expression), for simplicity, the highest observed value was used as

119   approximation in this case.

120   To analyse the data, a straight line was fitted through each of the datasets: *T=Sensitivity*x+ T_0,*

121   where *T* is flowering time and *x* is expression level; *Sensitivity* and T_0 are parameters for which

122   values are obtained in the fit. The R-function *lm* was used for the linear fit, and *cor.test* to test the

123   statistical significance.

124   One important point in our data analysis is that various datasets were obtained using different

125   ways of normalizing the expression values. Multiplicative normalization should effect *Sensitivity*

126   in a multiplicative way: if $T=S*x+T_0$, then for x'=a*x, $T=(S/a)*x'+T_0$, i.e. S'=S/a. Hence, we

127   can compare the value of *Sensitivity* for different genes only when the same reference gene is

128   used for normalization, and no additional relative normalization is used. The parameter T_0 should

129   be independent of the normalization that is used for expression data. It would only depend on the

130   unit of flowering time. This unit was either total leaf number or rosette leaf number; we did not

131   observe a systematic difference for data reported in either unit and hence did not discriminate

132   between these cases in presenting our results.

133 In addition to separately fitting the various datasets available for a given floral pathway

134 integrator gene, we also obtained one model for each floral pathway integrator gene in which the

135 various datasets were fitted simultaneously. This was performed using the R-function *nls*. In

136 these models, each dataset obtained its own value of *Sensitivity*, but only one global value of T_0

137 was used for each floral pathway integrator gene.

138 **Results**

139 We aim to obtain a comprehensive picture of how variation in signalling pathways influences

140 flowering time via affecting floral pathway integrator genes. To do so, we first analysed our

141 recently published mechanistic model for the floral pathway integrator gene network [18]. This

142 model describes regulatory interactions between the various integrator genes and is able to

143 predict the effect of a specific perturbation in one of the genes, on all the other genes in the

144 network. By assessing how this finally influences *AP1* expression, the model predicts flowering

145 time: flowering is predicted to start when *AP1* expression passes a certain threshold. This model

146 was developed using expression data and flowering time of wild-type *Arabidopsis thaliana*, as

147 well as mutants of floral pathway integrator genes. In our current work, we focus on genetic

148 variation in upstream signalling pathways, which were not used previously for modelling. To

149 simulate variation in these upstream signalling pathways, parameters describing input to the

150 floral pathway integrator genes were modified in the model (see Methods). This allowed to

151 observe the dependency of predicted flowering time on expression levels of floral pathway

152 integrator genes (Fig. 1). These plots indicate that for each gene, in an expression range of five-

153 to tenfold around its nominal expression, the response of flowering time to expression change is

154 approximately linear. To further analyse the response curves obtained from our model (Fig. 1) a

155 linear model was fitted. The p-value associated with the linear fit is significant ($<10^{-15}$) for all the

156 genes over the full range of expression displayed in Fig. 1. The obtained Pearson R^2 values for

157 the linear fits are all above 0.75.

158 Hence, analysis of our floral pathway integrator gene regulatory network model predicts a

159 gradual and rather linear dependence of flowering time response on changes in input to the floral

160 regulatory network. To assess the validity of this prediction, we chose to analyze large amounts

161   of datasets available in literature. Numerous studies present measurements of flowering times in

162   various conditions and for various genetic backgrounds. Since one often knows which floral

163   pathway integrator gene is relevant for the specific signalling pathway involved, the expression

164   levels of the specific gene thought to be responsible for mitigating the input from the signal

165   transduction pathway are measured as well. Although one has to extract most of this data

166   manually from tables or figures in relevant publications, it is an advantage that large amounts of

167   data can be analysed in this way. Even though some of the individual datasets are small, in its

168   totality the data consists of over 200 pairs of measurements of expression level and flowering

169   time. This data has so far been scattered throughout literature and we demonstrate that it can be

170   integrated. We use this data as a means to describe in a quantitative way the effect of changes in

171   genetic background in signalling pathway components on flowering time. We start with a

172   specific example regarding the floral pathway integrator gene *SOC1*.

173

174   *Introductory example for SOC1*

175   *SOC1* expression measurements (qPCR) were obtained in different genetic backgrounds (*cry2*

176   and *fri*, affecting the photoperiod pathway and the vernalization pathway, respectively) and

177   different conditions [23]. For the same conditions, flowering time was also measured [23]. It is

178   straightforward to combine these two sets of measurements in a quantitative way, although this

179   has not yet been done so far. As shown in Figure 2A, across the different genetic backgrounds,

180   there is a quite strong linear dependency of flowering time on the expression level of *SOC1*

181   (R^2=0.80). It is this dependency that is the focus of investigation of this study, for *SOC1* as well

182   as for floral pathway integrator genes. In our analysis, we focus on the effect of differences in

183   genetic backgrounds on each particular gene in the floral pathway integrator gene network. For

184   that particular gene, expression level measurements might then be explanatory for flowering time

185   changes. By analysing data as shown in Figure 2A from various publications, we are able to get a

186   comprehensive quantitative picture how floral pathway integrator gene expression mediates

187   transmission of environmental and endogenous cues to flowering time.

188   When integrating and comparing data for different experiments or different genes, one particular

189   complication is that reported qPCR gene expression levels are normalized in various ways. In

190   order to be able to combine datasets from different publications, one of the two following

191   conditions should hold: (1) The same reference gene was used for normalization, and we assume

192   that the expression level of the reference gene is constant in the different conditions applied in

193   the various publications. In this scenario, expression levels of different genes in various

194   publications can be quantitatively compared. Alternatively, (2) the reported expression level was

195   scaled using wildtype expression levels of the gene of interest. In this case, in order to compare

196   data from different publications, it is essential that the wildtype expression level that is used is

197   the same. This seems less likely than the assumption that a reference gene such as actin or

198   tubulin has a constant gene expression level. In several cases, the two scenarios are combined, in

199   the sense that qPCR data are first normalized to a reference gene but that the reported expression

200   level is subsequently scaled to a wildtype expression level.

201   For *SOC1*, the data analysed above were reported after scaling the expression level to wildtype

202   *SOC1* expression levels. Two additional examples of data for flowering time and *SOC1*

203   expression were obtained in which expression levels were normalized relative to a reference

204   gene [24, 25] (Figure 2B-C). In one of these (Fig. 2B), there was again a clear linear

205   relationships between the observed *SOC1* expression levels and flowering time in various

206   backgrounds, with Pearson R^2 value of 0.76. In the third one, there was less evidence for a linear

207   relationship, with Pearson R^2 value of 0.46 (p-value 0.3). Remarkably, it can be observed in

208   Figure 2 that one of the two parameters in the linear equation is quite similar for each of the three

209   datasets (78, 98 and 91, respectively). This observation is more generally true, and we will come

210   back to it in the next section. Note that the fits in Fig. 2B and 2C are less robust than the one in

211   Fig. 2A, but we discuss below how we can combine multiple datasets for one gene in a

212   simultaneous fit.

213

214   *Dependency of flowering time on floral pathway integrator gene expression levels*

215   Datasets reporting gene expression levels for various floral pathway integrator genes in different

216   genetic backgrounds, in combination with flowering time values in these genetic backgrounds,

217   were obtained (Figure 3; Table I). We start by fitting multiple models for each gene (one per

218   dataset). Because in some cases, the number of data points in a dataset is rather small, we

219   subsequently fit one model per floral pathway integrator gene (see below).

220   As presented above for *SOC1*, linear relationships were observed between flowering time and

221   gene expression levels (SI Figures 1-5; SI Table I). These can be described by the following

222   equation:

$$\text{T} = \textit{Sensitivity}*\text{ExpressionLevel} + T_0 \qquad (1)$$

224   Here, T is the observed flowering time, and the coefficients *Sensitivity* and T_0 are specific for

225   each floral pathway integrator gene. This equation describes how the measured flowering time T

226   in a given genetic background can be modelled as a linear function of the expression level of a

227   floral pathway integrator gene. The parameter *Sensitivity* describes the slope, in other words, the

228   sensitivity of flowering time to changes in expression of the flowering time integrator network

229   gene. Parameter T_0 describes the intercept with the line where ExpressionLevel equals zero.

230   Because, as explained above, expression data can only be directly compared if the same

231   normalization has been applied, we present values of *Sensitivity* and T_0 for each floral pathway

232   integrator gene separately for each possible type of normalization (Table II; SI Table I). SI

233   Figure 6 presents a histogram of the Pearson R^2 values obtained with the different models,

234   indicating that in the large majority of cases the value of R^2 is higher than 0.75, meaning that

235   more than 75% of the variation is explained by a simple linear model. The majority of the linear

236   models has a significant p-value and this mainly depends on the number of datapoints available;

237   for the cases with more than 5 datapoints, 9 out of 12 have a p-value below 0.05 (SI Table I).

238   In contrast to Sensitivity, T_0 should not depend on normalization applied to the expression data

239   (see Methods for explanation). Hence, T_0 values for the same floral pathway integrator gene

240   obtained from different datasets should be quite similar. This was indeed observed for the *SOC1*

241   datasets presented above. More generally, although there is some variation, the different values

242   of T_0 obtained for a given gene are indeed significantly similar to each other compared to the

243   values for the other genes (SI Text; SI Fig. 7). For the values of Sensitivity this is not the case, in

244   line with our expectation.

245   One concern with respect to the analysis so far could be that for some of the datasets, the number

246   of data points is rather small. We still chose to analyze such datasets initially separately because

247   the combination of perturbations of various input pathways for the same floral pathway

248   integrator gene allowed to demonstrate the similarity of T_0 values. To further deal with the

249   concern that some of the datasets are small, we subsequently fitted one final model per floral

250 pathway integrator gene. This was done by allowing one T_0 value per floral pathway integrator

251 gene, but a different value of *Sensitivity* per dataset. In this setup, the number of data points is for

252 each gene larger than the number of parameters; the number of degrees of freedom ranges from 2

253 for *SVP* to 72 for *FT*, and for all genes except *SVP* and *LFY* it is at least 30. Comparing the linear

254 model predictions with the experimental flowering time values indicates in most cases a clear

255 correspondence (Fig. 4A; SI Fig. 8). Note that *FT* has the most deviating behaviour in the sense

256 that the relationship between experimental and predicted flowering time values is less linear.

257 The values of T_0 are ordered as follows: $T_{0,\mathrm{SVP}} < T_{0,\mathrm{FLC}} \sim T_{0,\mathrm{AGL24}} \sim T_{0,\mathrm{LFY}} < T_{0,\mathrm{FT}} < T_{0,\mathrm{SOC1}}$. T_0

258 indicates the flowering time predicted by the linear relation in case of zero gene expression,

259 which should be later for a flowering activator than for a flowering repressor. Hence, one would

260 expect activators to have higher values than repressors. This is indeed the case. Given that the

261 values of T_0 indicate the expected flowering time when the level of a specific floral pathway

262 integrator gene is set to zero, the values of T_0 can be used to predict the flowering time for

263 knock-out mutants of each of the floral pathway integrator genes. To validate these predictions,

264 we compare them with our recently obtained set of flowering times for knock-out mutants [18]

265 (Fig. 4B). There is a good correspondence between predictions and experimental data, although

266 *FT* deviates from this pattern (Pearson R^2 including all cases is 0.38 between T_0 and flowering

267 time of knock-out mutants; excluding *FT*, the value of R^2 is 0.96 and the p-value ~0.02). Note

268 that *LFY* is not included in this figure because a *lfy* mutant does not flower properly at all [26].

269 The discordant behaviour of *LFY* cannot be predicted by the simple linear analysis presented

270 here. We provide an alternative analysis of our flowering time ODE model for prediction of *LFY*

271 mutant flowering time in SI Fig. 9. *LFY* expression was fixed at given levels and the resulting

272 flowering time predicted by the ODE model was recorded. For values of LFY below ~ 1nM, the

273 model predicts that there is no flowering. This behaviour is in accordance with the known

274 behaviour of the *lfy* null-mutant which was not used for training the model, providing additional

275 independent validation for the model.

276 The value of the slope of the fitted line in Figure 4B is much lower than 1. This line relates the

277 value of T_0, our prediction of flowering time, to the observed flowering time in knock-out

278 mutants. One reason for this small slope could be the fact that knock-out mutants in general will

279 not have exactly zero expression *in planta*, leading to a smaller effect on flowering time than

280 predicted. Nevertheless, the clear relationship between predicted and experimental flowering

281 time provides independent validation of the simple linear model fits from which the value of T_0

282 was obtained. Note that the flowering time and expression data used to obtain these fits are from

283 genetic backgrounds in which upstream signal components have been mutated. Hence, the input

284 data are independent from the floral pathway integrator gene knock-out mutants from which

285 flowering time data is used in Figure 4 for validation.

Discussion

Input from the environment is transduced by signalling pathways and integrated by a small number of floral pathway integrator genes. The complexity of the signalling pathways and their connection with the floral pathway integrator genes is overwhelming. Hence, understanding the effect of genetic variation in signalling pathways on flowering time is a daunting task. Our analysis indicates that in spite of this complexity, the effect of differences in genetic background can be quantitatively understood by focussing on expression level changes of floral pathway integrator genes. Perturbations in upstream signalling pathways effect floral pathway integrator genes mostly in such a way that the effect on flowering time is linear in the change in gene expression level. The fact that a linear response is significant in most cases, and that this response is observed for different floral pathway integrator genes, suggests that it is an import aspect of the way in which plants adapt to their local environment. The measured expression level changes are often up to tenfold or higher (Fig. 2, SI Fig 1-5). Hence, the linearity is observed over a large range of expression values.

Our findings on the role of gene expression variation in transducing the effect of genetic background variation to flowering time can be compared with more general analyses focusing on understanding the effect of variation in genetic background on phenotypes. For example, it was found in *C. elegans* that the effect of genetic background on the severity of RNAi and mutant phenotypes could be predicted from variation in the expression level of the affected gene [27]. Also, it has been observed that genetic variation associated with trait variation is likely to influence expression variation as well [28], suggesting that this expression variation is intermediate in establishing the link between change in genotype and change in phenotype. A recent method estimated genetically regulated gene expression and correlated these estimates

309   with phenotype values to identify genes involved in causing the phenotype [29]. In a broad

310   perspective, our analysis demonstrates the possibility of analysing the dependence of quantitative

311   traits on expression of key genes involved, which could be applied to a variety of plant traits.

312   Our findings are based on literature data that differ in experimental conditions such as the day or

313   the timepoint during the day used for measurement. Although this obviously puts limit on the

314   level of comparability between these data sets, our analysis shows that it is possible to integrate

315   such data. One particular complicating factor is the fact that qPCR data are reported in a variety

316   of ways. For one parameter in our model we overcome this problem by comparing data

317   normalized in the same way. For the other parameter, this is not needed because it is independent

318   of normalization. Nevertheless, the use of multiple qPCR reference genes would be of great

319   value, both for better comparability between studies and also to ensure accuracy of

320   measurements [30].

321   In addition to different ways of reporting expression, also different ways of reporting flowering

322   time are used. The data we used either reported the total number of leaves, or the number of

323   rosette leaves. Days to flowering is not often reported but would be a useful addition, in

324   particular since leaf number and days to flowering are not always congruent [31]. A more

325   systematic storage of qPCR data and of phenotypic measurements [32] such as flowering time

326   would clearly also be helpful to enable large scale comparative analyses such as we present.

327   The linear model appeared to be successful, but less so for *FT* than for other genes: the value of

328   T_0 obtained for *FT* did not correlate well with the experimental flowering time of an *ft* mutant

329   (Fig. 4), and when fitting the various datasets simultaneously for each gene, there was a less

330   clear linear relationship between predicted and observed expression for *FT* compared to the other

331   genes (SI Fig. 8). This might relate to the fact that in particular for *FT*, the mRNA levels

332   measured by qPCR are only a weak proxy for the real amount of active component. This is

333   because FT protein is transported from leaves to meristem before it may exert its effect on *SOC1*

334   and *FT*. Molecular aspects of this transport are not known in much detail yet, but one could

335   imagine that there would be some kind of threshold above which not all FT is transported. If this

336   would be the case, the predicted value of T_0 in our analysis would be too low, as is indeed

337   observed when the predicted values are compared with experimental flowering times for mutants

338   (Fig. 4). A similar threshold behaviour seems to be present in SI Figure 8 for *FT*. A more general

339   scenario in which the response of flowering time to expression level of a particular floral

340   pathway integrator gene would not necessarily be expected to be linear is if multiple floral

341   pathway integrator genes are simultaneously effected by upstream changes. Yet another

342   complicating factor is the fact that various floral pathway integrator genes regulate each other.

343   This could lead to correlations in expression levels of various floral pathway integrator genes,

344   which in turn might influence our analysis. If a gene which is directly influenced by an upstream

345   pathway regulates another floral pathway integrator gene, both might in principle display a clear

346   correlation between flowering time response and expression level.

347   In the literature, the quantitative, continuous nature of flowering time and its gradual response to

348   changing input is often neglected when analysing the effect of variation on flowering time. In

349   many cases, the measured response of flowering time to perturbations is reported just as leading

350   to early or late flowering. Only a few studies analyse quantitative relationships between gene

351   expression levels and flowering time. This includes a study in which *AGL24* is shown to be a

352   dosage-dependent mediator of flowering signals [33]. *FLC* levels in Arabidopsis accessions are

353   correlated to flowering times of these accessions [34]. For rice, there is one example of analysis

354   of quantitative relationship between expression of an FT ortholog and flowering time [35]. Our

355   comprehensive quantitative analysis neatly fits with these previous findings and quantifies the

356   dosage dependence of flowering time for various floral pathway integrator genes. It indicates that

357   the effect size of genetic variation in input pathways on flowering time can be understood via

358   expression changes of floral pathway integrator genes. This proportional response of flowering

359   time to upstream changes enables a gradual adaptation to changing environmental factors such as

360   temperature and light. The continuous nature of flowering time is therefore an essential aspect of

361   the potential of plants to adapt to various environments.

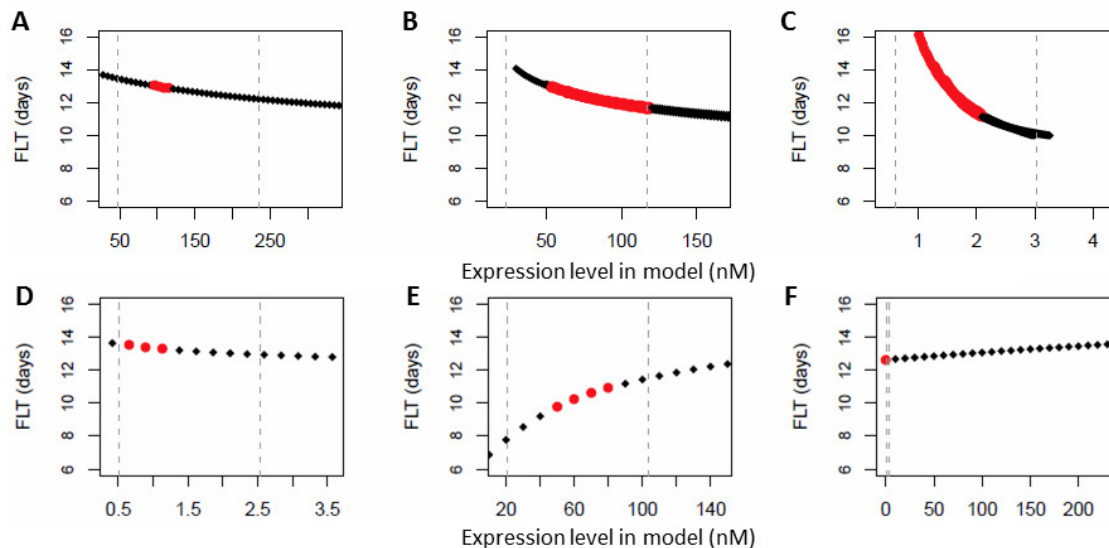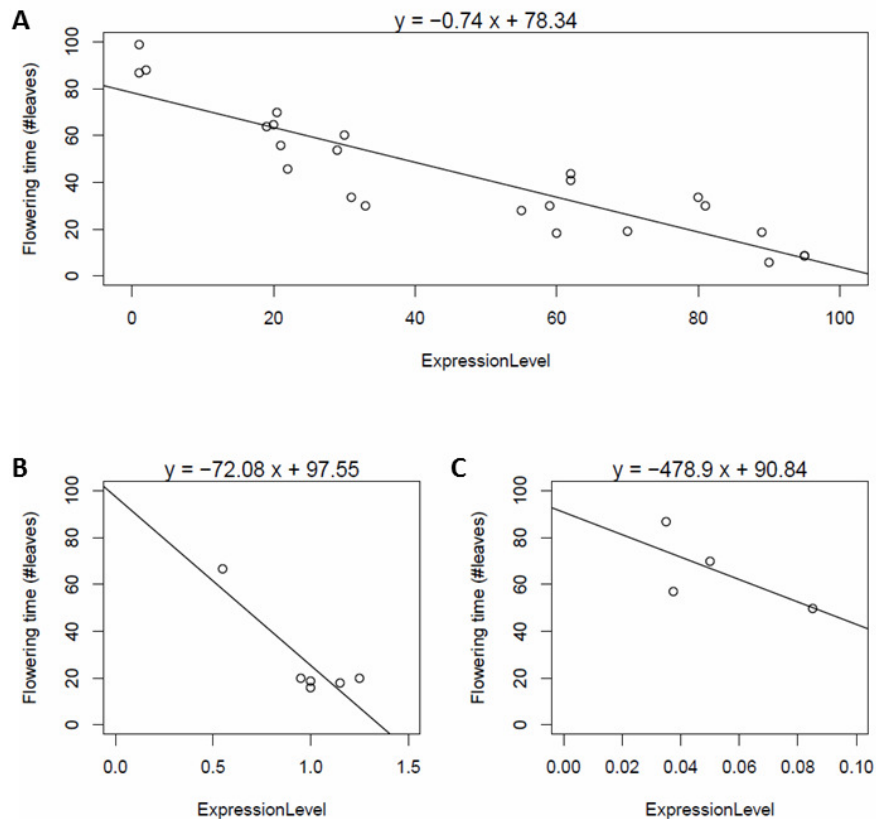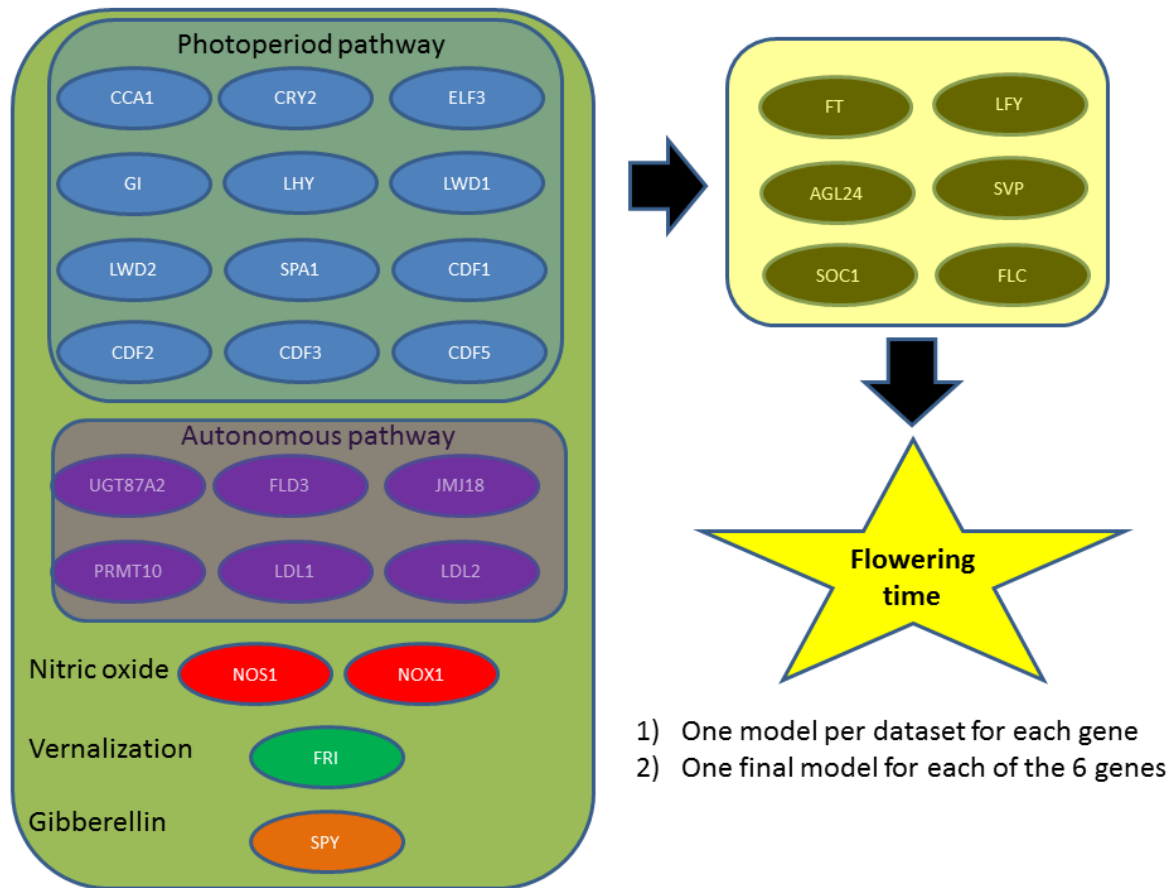362   **Figures**

363

364

365

366   **Figure 1. Dynamic model predicts linear dependency of flowering time in different genetic**

367   **backgrounds on floral pathway integrator gene expression levels.** The dynamic Ordinary

368   Differential Equation (ODE) model for flowering time regulation in ref. [18] was used to

369   simulate how flowering time (FLT) depends on gene expression level measured at day 10 for **(A)**

370   *AGL24* **(B)** *SOC1* **(C)** *LFY* **(D)** *FT* **(E)** *SVP* **(F)** *FLC*. To mimic genetic variation in upstream

371   signalling pathways, parameter values in the ODE model were modified as explained in

372   Methods. Red points indicate the expression level of the gene at day 9-11 in the unperturbed

373   model. Vertical dotted grey lines indicate five-fold expression range around the expression level

374   at day 10. For *FLC*, the five-fold range is small compared to the displayed range and the two

375   vertical lines fall on top of each other.

Figure 2. Dependency of flowering time (vertical axis) on *SOC1* expression levels (horizontal axis) in various genetic backgrounds and various conditions, obtained in three different studies (A-C). Flowering time is reported in number of leaves; expression is normalized by scaling to wildtype expression level (A), normalized to actin (B) or normalized to tubulin (C).

382

383

Figure 3. Overview of data and analysis. Available flowering time measurements and expression levels of floral pathway integrator genes (right hand side) were obtained from literature for various genetic backgrounds. The genes from different upstream signalling pathways which were mutated in these genetic backgrounds are indicated at the left hand side. We analyse the data by modelling how expression level changes in floral pathway integrator genes (caused by genetic variation in the upstream signalling pathways) lead to quantitative changes in flowering time. In a first step, several models were obtained for each of the floral pathway integrator genes. Subsequently, one final model was obtained for each of these genes.
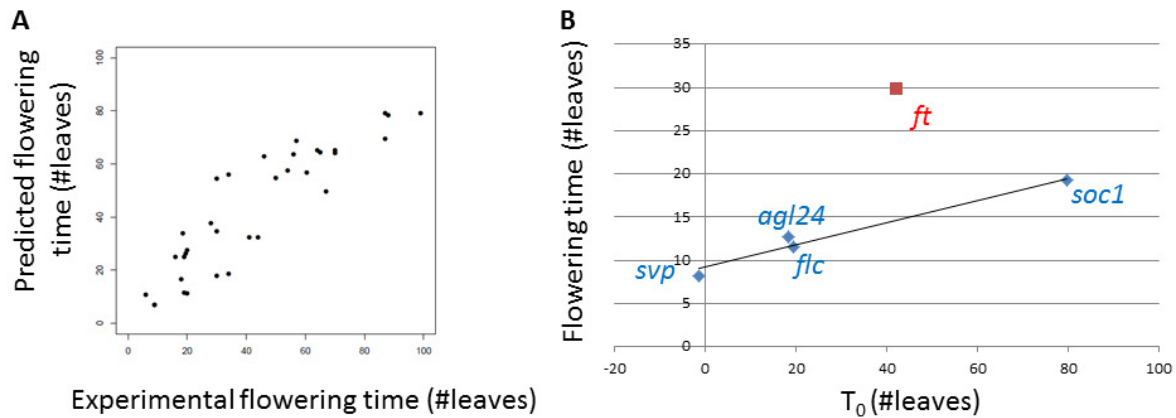
392

393

394

395 **Figure 4. Comparison between predictions and experimental data. (A)** Comparison between

396 predicted and experimental flowering time for single linear model fitted to various *SOC1*

397 datasets. These datasets are the same as the ones used in Fig. 2, but here they are all fitted

398 simultaneously using different values of Sensitivity but one single value of T_0. The number of

399 degrees of freedom in this fit is 30. **(B)** Comparison between T_0 and flowering time of knock-out

400 mutants. Based on fits of quantitative relationships between expression levels and flowering

401 time, T_0 predicts flowering time in knock-out mutants for different floral pathway integrator

402 genes. These predictions show a good relationship with experimentally observed flowering time

403 for these knock-outs. Each point in this plot represents one particular floral pathway integrator

404 gene; red outlier point indicates *ft*.

405

406

407

408 **Table I. Datasets obtained from literature[a]**

Gene / reference	Mutant genotypes	Wildtype genotype	Conditions[b]	Flowering time [c]
SOC1				
[23]	*cry2*, FLC-Sf2, FRI-Sf2, *cry2*;FLC-Sf2, *cry2*;FRI-Sf2, *cry2*;FRI-Sf2;FLC-Sf2	Ler, Cvi	LD,SD; 25C; day 21	TL
[25]	*gi*,35S::GI, 35S::GI;*gi*	Col	LD; 22C; day 15	TL
[24]	35S::AGL24, 35S::SOC1, *agl24*	Col	SD; 22C; day 21	RL
FT				
[36]	*gi*; 35S::GI,*lhy,lhy;cca1, 35S::GI;lhy,* gi;lhy;cca1	Ler	SD; 22C; day 10	TL
[37]	*agl24*, 35S::SVP, *svp*, *soc1*	Col,Ler,C24	LD, SD; GA; 22C; day 11	TL
[5]	*cry2, cyr2;spa1*	Col,RLD	LD; day 14	RL
[38]	*Cdf1,cdf2,cdf3,cdf5*	Col	LD,SD; day 10	RL
[39]	35S:JMJ18, jmj18, tissue specific JMJ18	Col	LD; 22/18C; day 11	TL
[25]	*gi*,35S::GI, 35S::GI;*gi*	Col	LD; 22C; day 15	TL
[40]	*cry2*, tissue specific CRY2	Col	LD; day 9	RL
[41]	*elf3*; *elf3* enhancer and suppressor lines	Col,Ler	LL; day 14	RL
[42]	*gi*; *35S::gi*; tissue specific GI	Col	LD,SD; 23/16C; day 10	TL
[43]	*gi*, *spy*	Col,Ler	LD; 22C; day 14	TL
[44]	*lwd1;lwd2,lwd1;lwd2/LWD1*	Col	LD,SD; day 18	RL
FLC				
[39]	35S:JMJ18, jmj18, tissue specific JMJ18	Col	LD; 22/18C;	TL

			day 11	
[23]	*cry2*, FLC-Sf2, FRI-Sf2, *cry2*;FLC-Sf2, *cry2*;FRI-Sf2, *cry2*;FRI-Sf2;FLC-Sf2	Ler, Cvi	LD,SD; 25C; day 21	TL
[45]	*nox1, nos1, NO-donor treatment*	Col	LD; 22C; day 10	RL
[46]	*prmt10, prmt5*	Col	LD; day 11	TL
[47]	*ldl1,ldl2,ldl1/ldl2*	Col	LD; day 10	TL
[48]	*ugt87a2*	Col	LD; 22C; day 21	RL
SVP				
[41]	*elf3*; *elf3* enhancer and suppressor lines	Col,Ler	LL; day 14	RL
[37]	*ft*	Col	LD; 22C; day 11	TL
LFY				
[45]	*nox1, nos1*	Col	LD; 22C; day 10	RL
[48]	*ugt87a2*	Col	LD; 22C; day 21	RL
AGL24				
[33]	AGL24-RNAi, 35S-AGL24	Col,Ler	LD; 23C; day 5	RL
[37]	agl24-1, 35S::SVP, svp-41, soc1-2	Col,Ler,C24	LD, SD; GA; 22C; day 11	TL

409 a Flowering time and expression data for specific floral pathway integrator genes were obtained from literature.

410 Table includes data for each floral pathway integrator gene in which genetic background and expression data was

411 measured. Values obtained from each dataset are presented in Fig. 2 and SI Figures 1-5. Results of fitting these data

412 using a linear model are shown in Table II and SI Table I.

413 b Experimental conditions: LD indicates long day, SD indicates short day, LL indicates continuous light, GA

414 indicates gibberellin. Day indicates age of plant for which measurements were taken. If reported, temperature is

415 indicated as well.

416   [c] Flowering time measurement: RL indicates number of rosette leaves, TL indicates total number of leaves.

417

Table II. Linear dependencies of flowering time on expression levels[a]

Gene	Normalization (number of datasets)	Sensitivity	T_0
SOC1	scaled (1x)	-0.74	78.3
	actin (1x)	-72	97.5
	tubulin (1x)	-478.9	90.8
FT	scaled (3x)	-0.30 (0.06)	38.5 (6.0)
	actin (2x)	-19.6 (9.95)	45.4 (11.2)
	tubulin (1x)	-11.5	29.9
	IPP2 (4x)	-4.0 (1.1)	53.4 (15)
	UBQ10 (3x)	-363 (451)	45.8 (24.0)
FLC	scaled (7x)	5.8 (7.1)	12.7 (5.1)
	actin (1x)	81.0	8.1
SVP	scaled (1x)	0.29	4
	tubulin (1x)	37.2	-12.5
LFY	scaled (3x)	-5.0 (1.5)	14.6 (2.7)
AGL24	scaled (3x)	-1.7 (1.8)	19.6 (2.1)

[a] Values for parameters in linear fit T = Sensitivity*ExpressionLevel + T_0 for data shown in Fig. 2 and SI Figures 1-5. Normalization method used in the different datasets is indicated (scaled means normalization by scaling with wildtype or maximum expression value). Different normalization renders values of Sensitivity incomparable, but should not affect comparisons between values of T_0. Reported values are average (standard deviation) in case multiple datasets are available for the same normalization. Characteristics of individual datasets are reported in Table I. Values for Sensitivity and T_0 in individual datasets are reported in SI Table I.

426 **References**

427 1. Srikanth, A. and M. Schmid, *Regulation of flowering time: all roads lead to Rome.* Cell Mol Life Sci, 2011.
428   **68**(12): p. 2013-37.
429 2. Andres, F. and G. Coupland, *The genetic basis of flowering responses to seasonal cues.* Nat Rev Genet,
430   2012. **13**(9): p. 627-39.
431 3. Simpson, G.G. and C. Dean, *Arabidopsis, the Rosetta stone of flowering time?* Science, 2002. **296**(5566):
432   p. 285-9.
433 4. Bouche, F., et al., *FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis*
434   *thaliana.* Nucleic Acids Res, 2016. **44**(D1): p. D1167-71.
435 5. Zuo, Z., et al., *Blue light-dependent interaction of CRY2 with SPA1 regulates COP1 activity and floral*
436   *initiation in Arabidopsis.* Curr Biol, 2011. **21**(10): p. 841-7.
437 6. Yoo, S.K., et al., *CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through*
438   *FLOWERING LOCUS T to promote flowering in Arabidopsis.* Plant Physiol, 2005. **139**(2): p. 770-8.
439 7. Abe, M., et al., *FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot*
440   *apex.* Science, 2005. **309**(5737): p. 1052-6.
441 8. Wigge, P.A., et al., *Integration of spatial and temporal information during floral induction in Arabidopsis.*
442   Science, 2005. **309**(5737): p. 1056-9.
443 9. Verhage, L., G.C. Angenent, and R.G. Immink, *Research on floral timing by ambient temperature comes*
444   *into blossom.* Trends Plant Sci, 2014. **19**(9): p. 583-91.
445 10. Capovilla, G., M. Schmid, and D. Pose, *Control of flowering by ambient temperature.* J Exp Bot, 2015.
446   **66**(1): p. 59-69.
447 11. Immink, R.G., et al., *Characterization of SOC1's central role in flowering by the identification of its*
448   *upstream and downstream regulators.* Plant Physiol, 2012. **160**(1): p. 433-49.
449 12. Michaels, S.D., et al., *AGL24 acts as a promoter of flowering in Arabidopsis and is positively regulated by*
450   *vernalization.* Plant J, 2003. **33**(5): p. 867-74.
451 13. Lee, J., et al., *SOC1 translocated to the nucleus by interaction with AGL24 directly regulates leafy.* Plant J,
452   2008. **55**(5): p. 832-43.
453 14. Eriksson, S., et al., *GA4 is the active gibberellin in the regulation of LEAFY transcription and Arabidopsis*
454   *floral initiation.* Plant Cell, 2006. **18**(9): p. 2172-81.
455 15. Wagner, D., R.W. Sablowski, and E.M. Meyerowitz, *Transcriptional activation of APETALA1 by LEAFY.*
456   Science, 1999. **285**(5427): p. 582-4.
457 16. Kaufmann, K., et al., *Orchestration of floral initiation by APETALA1.* Science, 2010. **328**(5974): p. 85-9.
458 17. Jung, C. and A.E. Muller, *Flowering time control and applications in plant breeding.* Trends Plant Sci, 2009.
459   **14**(10): p. 563-73.
460 18. Leal Valentim, F., et al., *A quantitative and dynamic model of the Arabidopsis flowering time gene*
461   *regulatory network.* PLoS One, 2015. **10**(2): p. e0116973.
462 19. Dong, Z., et al., *A gene regulatory network model for floral transition of the shoot apex in maize and its*
463   *dynamic modeling.* PLoS One, 2012. **7**(8): p. e43450.
464 20. Jaeger, K.E., et al., *Interlocking feedback loops govern the dynamic behavior of the floral transition in*
465   *Arabidopsis.* Plant Cell, 2013. **25**(3): p. 820-33.
466 21. Salazar, J.D., et al., *Prediction of photoperiodic regulators from quantitative gene circuit models.* Cell,
467   2009. **139**(6): p. 1170-9.
468 22. Lloyd, J. and D. Meinke, *A comprehensive dataset of genes with a loss-of-function mutant phenotype in*
469   *Arabidopsis.* Plant Physiol, 2012. **158**(3): p. 1115-29.
470 23. El-Din El-Assal, S., et al., *The role of cryptochrome 2 in flowering in Arabidopsis.* Plant Physiol, 2003.
471   **133**(4): p. 1504-16.
472 24. Liu, C., et al., *Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis.*
473   Development, 2008. **135**(8): p. 1481-91.
474 25. Gunl, M., et al., *Analysis of a post-translational steroid induction system for GIGANTEA in Arabidopsis.*
475   BMC Plant Biol, 2009. **9**: p. 141.
476 26. Blazquez, M.A., et al., *LEAFY expression and flower initiation in Arabidopsis.* Development, 1997.
477   **124**(19): p. 3835-44.
478 27. Vu, V., et al., *Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes.* Cell,
479   2015. **162**(2): p. 391-402.
480 28. Nicolae, D.L., et al., *Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery*
481   *from GWAS.* PLoS Genet, 2010. **6**(4): p. e1000888.
482 29. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome*
483   *data.* Nat Genet, 2015.
484 30. Remans, T., et al., *Reliable gene expression analysis by reverse transcription-quantitative PCR: reporting*
485   *and minimizing the uncertainty in data accuracy.* Plant Cell, 2014. **26**(10): p. 3829-37.
486 31. Takahashi, M. and H. Morikawa, *Nitrogen dioxide accelerates flowering without changing the number of*
487   *leaves at flowering in Arabidopsis thaliana.* Plant Signal Behav, 2014. **9**(10): p. e970433.

488 32.   Krajewski, P., et al., *Towards recommendations for metadata and data handling in plant phenotyping.* J
489   Exp Bot, 2015. **66**(18): p. 5417-27.
490 33.   Yu, H., et al., *AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals.* Proc Natl Acad
491   Sci U S A, 2002. **99**(25): p. 16336-41.
492 34.   Lempe, J., et al., *Diversity of flowering responses in wild Arabidopsis thaliana strains.* PLoS Genet, 2005.
493   **1**(1): p. 109-18.
494 35.   Takahashi, Y., et al., *Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to*
495   *diversity of flowering time in cultivated rice.* Proc Natl Acad Sci U S A, 2009. **106**(11): p. 4555-60.
496 36.   Mizoguchi, T., et al., *Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms*
497   *in Arabidopsis.* Plant Cell, 2005. **17**(8): p. 2255-70.
498 37.   Li, D., et al., *A repressor complex governs the integration of flowering signals in Arabidopsis.* Dev Cell,
499   2008. **15**(1): p. 110-20.
500 38.   Fornara, F., et al., *Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression*
501   *and are essential for a photoperiodic flowering response.* Dev Cell, 2009. **17**(1): p. 75-86.
502 39.   Yang, H., et al., *A companion cell-dominant and developmentally regulated H3K4 demethylase controls*
503   *flowering time in Arabidopsis via the repression of FLC expression.* PLoS Genet, 2012. **8**(4): p. e1002664.
504 40.   Endo, M., et al., *CRYPTOCHROME2 in vascular bundles regulates flowering in Arabidopsis.* Plant Cell, 2007.
505   **19**(1): p. 84-93.
506 41.   Nefissi, R., et al., *Double loss-of-function mutation in EARLY FLOWERING 3 and CRYPTOCHROME 2 genes*
507   *delays flowering under continuous light but accelerates it under long days and short days: an important*
508   *role for Arabidopsis CRY2 to accelerate flowering time in continuous light.* J Exp Bot, 2011. **62**(8): p.
509   2731-44.
510 42.   Sawa, M. and S.A. Kay, *GIGANTEA directly activates Flowering Locus T in Arabidopsis thaliana.* Proc Natl
511   Acad Sci U S A, 2011. **108**(28): p. 11698-703.
512 43.   Tseng, T.S., et al., *SPINDLY and GIGANTEA interact and act in Arabidopsis thaliana pathways involved in*
513   *light responses, flowering, and rhythms in cotyledon movements.* Plant Cell, 2004. **16**(6): p. 1550-63.
514 44.   Wu, J.F., Y. Wang, and S.H. Wu, *Two new clock proteins, LWD1 and LWD2, regulate Arabidopsis*
515   *photoperiodic flowering.* Plant Physiol, 2008. **148**(2): p. 948-59.
516 45.   He, Y., et al., *Nitric oxide represses the Arabidopsis floral transition.* Science, 2004. **305**(5692): p. 1968-
517   71.
518 46.   Niu, L., et al., *Regulation of flowering time by the protein arginine methyltransferase AtPRMT10.* EMBO
519   Rep, 2007. **8**(12): p. 1190-5.
520 47.   Jiang, D., et al., *Arabidopsis relatives of the human lysine-specific Demethylase1 repress the expression of*
521   *FWA and FLOWERING LOCUS C and thus promote the floral transition.* Plant Cell, 2007. **19**(10): p. 2975-
522   87.
523 48.   Wang, B., et al., *UGT87A2, an Arabidopsis glycosyltransferase, regulates flowering time via FLOWERING*
524   *LOCUS C.* New Phytol, 2012. **194**(3): p. 666-75.

525