

## ARTICLE OPEN

Hierarchical graphical model reveals HFR1 bridging circadian rhythm and flower development in *Arabidopsis thaliana*Zhana Duren<sup>1,2</sup>, Yaling Wang<sup>3</sup>, Jiguang Wang<sup>4</sup>, Xing-Ming Zhao<sup>5,6</sup>, Le Lv<sup>7</sup>, Xiaobo Li<sup>8</sup>, Jingdong Liu<sup>7</sup>, Xin-Guang Zhu<sup>3</sup>, Luonan Chen<sup>9,10,11,12</sup> and Yong Wang<sup>1,2,10</sup>

To study systems-level properties of the cell, it is necessary to go beyond individual regulators and target genes to study the regulatory network among transcription factors (TFs). However, it is difficult to directly dissect the TFs mediated genome-wide gene regulatory network (GRN) by experiment. Here, we proposed a hierarchical graphical model to estimate TF activity from mRNA expression by building TF complexes with protein cofactors and inferring TF's downstream regulatory network simultaneously. Then we applied our model on flower development and circadian rhythm processes in *Arabidopsis thaliana*. The computational results show that the sequence specific bHLH family TF HFR1 recruits the chromatin regulator HAC1 to flower development master regulator TF AG and further activates AG's expression by histone acetylation. Both independent data and experimental results supported this discovery. We also found a flower tissue specific H3K27ac ChIP-seq peak at AG gene body and a HFR1 motif in the center of this H3K27ac peak. Furthermore, we verified that HFR1 physically interacts with HAC1 by yeast two-hybrid experiment. This HFR1–HAC1–AG triplet relationship may imply that flower development and circadian rhythm are bridged by epigenetic regulation and enrich the classical ABC model in flower development. In addition, our TF activity network can serve as a general method to elucidate molecular mechanisms on other complex biological regulatory processes.

npj Systems Biology and Applications (2019)5:28; <https://doi.org/10.1038/s41540-019-0106-3>

## INTRODUCTION

To study coordination among complex regulatory processes, it is necessary to go beyond individual regulators, *cis*-elements, and target genes (TGs) to provide a global connectome. Network biology holds great potential in unveiling regulatory mechanisms as previously demonstrated.<sup>1–3</sup> A wide variety of approaches can take advantage of the increasingly available gene expression data to infer gene regulatory network (GRN), including ordinary and stochastic differential equations (ODEs/SDEs)-based methods,<sup>4–8</sup> Bayesian and dynamic Bayesian networks (BNs/DBNs),<sup>9–12</sup> and information theoretic or correlation-based methods.<sup>13–15</sup> One fundamental mechanism of gene regulation is that transcription factors (TFs) serve as the key regulators for many important biological processes. TFs bind to specific DNA sequences and thereby regulate the transcription of their TGs. In addition, TFs are known to facilitate biological functions based on its protein/complex activity or interacting network rather than simply on its mRNA concentration, i.e., the mRNA expression of a TF does not represent the activity of this TF on TGs. This fact invalidates many pure gene expression based TF network reconstruction methods. The physical reality is that TF's activity is related to its many protein cofactors during posttranscriptional and posttranslational

modifications.<sup>16</sup> Here we define TF activity (TFA) as DNA-binding fraction (the active part and the TF in regulating status) of each TF. This concept widely considers both the mRNA level and posttranslational regulation, such as differential splicing, post-translational modification, and even the methylation states of the TFs.

Cases have been reported that extensive posttranslational modification on major factors or regulators frequently happens in plant and their expression as well as activity do not agree with each other well.<sup>17</sup> This will affect the transcriptional networks mediated by TFs in the developmental processes of multicellular organisms by specifying when and where genes are expressed. For example, flower development is one of the most important plant developmental process. During flower development, gene expression is precisely regulated both in time and space, which then leads to the formation of flower organ. Flower development has long been known to be triggered by circadian rhythm.<sup>18</sup> In *Arabidopsis thaliana*, a circadian clock, controlled by LATE ELONGATED HYPOCOTYL (LHY) and CIRCADIAN CLOCK ASSOCIATED 1, was reported to promote flowering specifically under long days by controlling the genes GIGANTEA (GI), CONSTANS (CO), and FLOWERING LOCUS T (FT).<sup>19,20</sup> Later on, the GRN related

<sup>1</sup>CEMS, NCMIS, MDIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; <sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China; <sup>3</sup>State Key Laboratory of Molecular Plant Sciences and Center of Excellence for Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China; <sup>4</sup>Division of Life Science, Department of Chemical and Biological Engineering, Center of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Hong Kong, China; <sup>5</sup>Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China; <sup>6</sup>Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Ministry of Education, Shanghai, China; <sup>7</sup>Bayer U.S. – Crop Science, Monsanto Legal Entity, St. Louis, MO 63156, USA; <sup>8</sup>Department of Plant Biology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA; <sup>9</sup>Key Laboratory of Systems Biology, Center for Excellence in Molecular Cell Science, Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China; <sup>10</sup>Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China; <sup>11</sup>School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China and <sup>12</sup>Research Center for Brain Science and Brain-Inspired Intelligence, 201210 Shanghai, China

Correspondence: Jingdong Liu ([jingdong.liu@monsanto.com](mailto:jingdong.liu@monsanto.com)) or Xin-Guang Zhu ([zhuxinguang@picb.ac.cn](mailto:zhuxinguang@picb.ac.cn)) or Luonan Chen ([lnchen@sibs.ac.cn](mailto:lnchen@sibs.ac.cn)) or Yong Wang ([ywang@amss.ac.cn](mailto:ywang@amss.ac.cn))

Received: 26 October 2018 Accepted: 23 July 2019

Published online: 12 August 2019

to flower development has been constructed<sup>21,22</sup> and many regulators were identified known to function also in circadian rhythm,<sup>23,24</sup> which sheds the light on the potential connection between flower development and the circadian rhythm. However, the molecular mechanism of how circadian rhythm links flower development remains largely unexplored.

In this paper, we aim to reveal the associations between flower development and circadian rhythm at the network level, which not only identifies their new regulators or interactions but also opens a new way to study such complex processes. To dissect the molecular mechanism of TFA, we propose a model-based computational method to simultaneously infer the TFA as a hidden variable and reconstruct the TF regulatory network. Our model fully utilizes the upstream and downstream regulation information around TF and integrates the functional linkage data and gene expression data. The upstream information of a TF is its protein (cofactor or kinase)-TF interactions. The downstream information of a TF is its target gene regulations. By systematically mining the consistency of regulatory relationships with functional linkage and gene expression data, our model constructs the TF network by estimating TFA with high accuracy, which can reveal those key TFs even without differential expressions (or “dark” TF).

We reconstructed GRM connecting Flower Development and Circadian Rhythm processes (FDCRNet), and analyzed its topological structure and revealed two hubs in the network bridging these two important processes. We successfully detected the well-known factor LHY. In addition, we found an important factor, HFR1, modulating flower development. Our TFA network analysis revealed that HFR1 modulates interactions among the ABC model genes. Experimental and computational results together show that HFR1 regulates flowering, specifically the differentiation of stamens and carpels, by modulating the regulatory interaction between HAC1 and AG. Furthermore, we hypothesize that HFR1 may play an important role in the interplay between light and temperature.

## RESULTS

### Constructing gene regulatory network with TF activities

We propose a probability framework known as a hierarchical model to construct the network. The primary goal is to predict transcriptional regulations between TFs and TGs and further use network to confer cellular states and modulate the biological activities within a cell. Since the mRNA expression levels for both TFs and TGs can be measured simultaneously by microarray or RNA-seq, pooling large amount of data and investigating TF-TG's correlation is a straightforward way to infer possible regulatory interactions. However, the mRNA levels of the TFs cannot reflect the activity of TFs.<sup>25,26</sup> For example, posttranslational modifications may alter TFs activity by affecting the secondary structure, altering affinity, increasing or decreasing protein degradation, nuclear occupancy, and DNA binding.<sup>27</sup> For example, it is known that HFR1 is ubiquitinated by COP1 E3 ligase and marked for posttranslational degradation during photomorphogenesis.<sup>28</sup> In addition, TF-TF heterodimer and TF-protein cofactors would affect the ability of TF binding to the DNA. Here, we computationally model TFA and interpret gene expression process in a more precise way at the genomic scale.<sup>29</sup> Specifically, our hierarchical mixture model treats TFA as a hidden variable and estimates it by observing regulatory network structure and gene expression profiles, i.e., the functional concentration of each TF, termed TFA, is inferred from its gene expression data, cofactors' expression data, and TGs' expression data. Then correlating the TFA with expression level of TGs will inform us which TG is more likely to be regulated by a TF.

We take an innovative approach for easier inference of an accurate network structure. We first compile a draft network by inferring pairwise and triplet regulatory relationships from correlation and functional linkage data (Step 1 and Step 2 in Fig. 1). Then this draft/candidate network is treated as prior information and updated based on the observed experimental data. TFA is introduced as a hidden variable and determined during the refinement of network structure (Step 3 in Fig. 1).

### Step 1: Pairwise transcriptional regulations by mutual information.

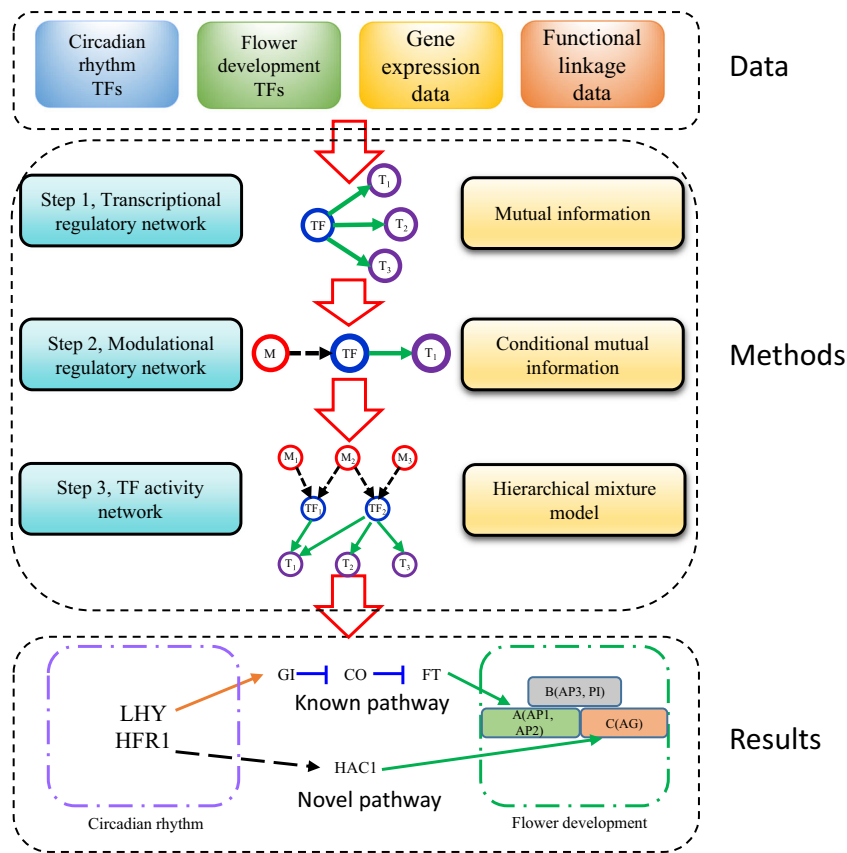
To identify all possible TGs for TFs and further construct TFA network, we constructed the transcriptional regulatory network based on mutual information for all possible TF-TG pairs (Materials and Methods). We fit a t-distribution using mutual information on all possible TF-TG pairs (mean  $\mu$ ), and only select the significant TF-TG pairs ( $H_0: MI(TF, TG) = \mu$ ,  $H_1: MI(TF, TG) > \mu$ ,  $p$ -value  $< 10E-4$ ). Totally, we inferred 4234 candidate TGs for flower development (associated with 97 flowering related TFs) and 1936 candidate TGs for circadian rhythm (associated with 44 circadian-related TFs) (Supplementary Table 3 and 4). The results show that the frost tolerance gene CBF3, histone acetylation-associated gene HAC1, and the flower development-associated gene AG constitute a clique, which means each two genes are connecting. Histone acetyltransferases are known to play an important role both in flowering<sup>30</sup> and cold-induced gene regulation.<sup>31</sup> Interactions among these three genes reveal that histone acetylation may be bridging temperature and flowering.

### Step 2: Modulatory triplet by conditional mutual information.

We next considered the modulator, TF, TG triple regulatory relationships, and constructed the modulation regulatory network by conditional mutual information (CMI). The aim is to identify all the possible modulators of the TF-TG pairs for further constructing TFA network. Given a TF and its target gene pairs, we restricted their candidate modulators in the cofactors gene set which interact with TF in functional linkage network from STRING database. We computed the CMI for all possible cofactors-TF-TG triplets to identify potential modulators for each TF. Given a TF-TG pair, we generate the distribution of CMI of a random cofactor. Specifically, we randomly permute the expression data of the candidate cofactors 1,000 times, and compute CMIs on those permuted data. We test whether or not the candidate cofactor is random one by comparing the CMI with the empirical distribution of the random cofactor's CMI ( $p$ -value cutoff is 0.001). Given a TF (TF- $k$  is noted as  $TF_k$ , and its  $m$  TGs are noted as  $TG_{k1}, TG_{k2}, \dots, TG_{km}$ ), the  $i$ -th cofactor of this TF (noted as  $CF_{ki}$ ) is identified as a modulator if  $CMI(CF_{ki}, TF_k, TG_{kj})$  is significant for at least 50% of the  $TG_{kj}$  ( $j = 1, 2, \dots, m$ ). As a result, we inferred 81,253 modulator-TF-TG triplets for flower development and 27,150 triplets for circadian rhythm process (Supplementary Tables 5 and 6). For example, we found that HFR1 is a modulator between HAC1 and AG. When the expression level of HFR1 is low, the mutual information of HAC1 and AG is 0.34. However, when HFR1 has a high expression level, the mutual information of HAC1 and AG is 0.72 (Fig. 2). The significant increase of mutual information correlates well with the expression change of HFR1.

### Step 3: Integrating upstream and downstream regulations in TF activity network.

We deduced the TFA network by integrating the upstream and downstream regulatory information. We treat the draft network structure obtained by Step 1 and Step 2 as input and estimate the TFA and the regulatory strength parameters (see Methods for details). If the estimated TFA is reliable, it should better correlate with the expression of TGs. To verify the inferred TFA, we collected 26 known TF-TG interactions in flower development process<sup>21,32</sup> (Fig. 3a). On these 26 pairs, we compared two mutual information scores, one is based on mRNA expression data and the other one is based on estimated TFA.



**Fig. 1** TFA network (FDCRNet) construction workflow for Flower Development and Circadian Rhythm processes. Key molecules (TFs) for circadian rhythm and flower development are curated from GO annotation database. Functional linkage and gene expression data are integrated step-by-step to construct the transcriptional regulation network, modulation regulatory network, and TFA network for those key molecules

Mutual information computed by TFA is higher than that computed by mRNA expression on 92.31% (24 out of 26) of the known interactions (Fig. 3c). For example, the mutual information of GI and CO mRNA expression is 0.10, but the mutual information of GI mRNA expression and CO TFA is 0.23. To evaluate the significance of the mutual information score improvement, we generate the null distribution of the mutual information by resampling the gene expression data and TFA data 1000 times by bootstrap method (Fig. 3b). For 92.31% (24 out of 26) of the known interactions, the mutual information computed by the TFA is significantly higher than that computed by the mRNA expression. This result suggests that the TFA generates biologically plausible correlations and serves as a powerful platform for the follow-up hypothesis generation.

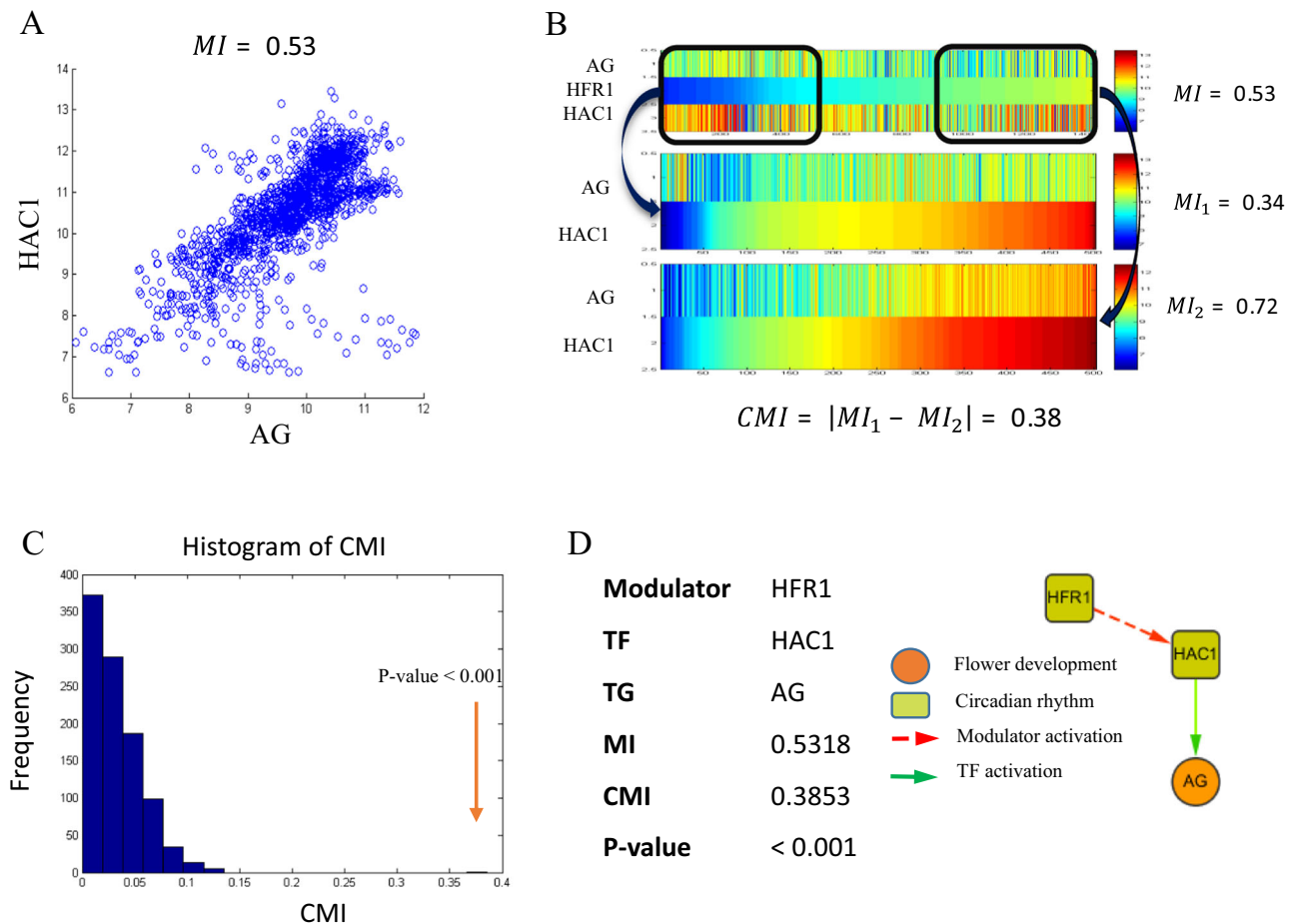
Overall 35,692 out of 81,253 modulator–TF–TG triplets (43.93%) were selected by TFA network (triplets with non-zero  $\alpha$  and  $\beta$ ). To see whether the TFA network provides additional information on CMI or not, we compared the distribution of CMI on the triplets selected by TFA model with those not selected based on TF's mRNA expression (Fig. 3c). We find that 52.27% of the high CMI (>0.2) triplets were not selected by TFA network. The result shows that our TFA network provides additional information on MI and CMI computation, which can further filter the spurious interactions.

#### Genome-wide TF activity network for flower development and circadian rhythm

We applied the TFA network reconstruction method in the 1436 samples from the TAIR database and reconstructed the genome-wide TFA network for 97 flower development-associated

TFs and 44 circadian rhythm-associated TFs (Supplementary Tables 1 and 2). There are 4,556 nodes (including 664 modulators, 116 TFs, and 3,939 TGs) and 8838 edges (including 897 modulation edges and 5941 are transcriptional regulations edges) in TFA network (Supplementary Fig. 2; Supplementary Table 7).

We perform the topological analysis on the inferred TFA network. About 75% (497 out of 664) of the modulators have only one downstream TF. The distribution of the rest 25% modulator out-degree is shown in Supplementary Fig. 2b. MGP, a nuclear-localized putative TFs known to be expressed in flowering stage,<sup>33</sup> modulates 64 TFs including 48 flower development TFs and 16 circadian rhythm TFs. Interestingly, both kinase WEE1 and the kinase activity-associated factor FUS5 are hub modulators. This shows that a kinase may modulate the transcriptional regulation by affecting the TF's activity or its ability to bind the TGs. RING1B, a member of the polycomb repressive core complex 1, modulates many regulations in both FDCRNet. This chromatin remodeling enzymes may modulate the transcriptional regulations by affecting the chromatin state of TF or TGs.<sup>34</sup> Some flower development or circadian rhythm-associated genes, such as HFR1 and AGL28, also have many downstream TFs. A total of 121 TFs are contained in this three-layer network, which consists of 87 flower development TFs and 42 circadian rhythm TFs. In which, 8 TFs are overlapped with flower development TFs. As a median layer, TFs are modulated by modulator and regulate TGs. The in-degrees of some important TFs are high, such as the circadian clock core TF, LHY, have 468 modulators (Supplementary Fig. 2c). The flower development TFs, AP3, and JAG, also have more than 50 modulators. HY5, associated with both FDCRNet, has a high in-degree. The circadian rhythm-associated factor COL9, ERP1, and



**Fig. 2** Modulators affect the pattern of the transcriptional regulation. **a** Expression pattern of HAC1 and AG. Each dot corresponds to the gene expression of HAC1 and AG in one sample. MI represents the mutual information score. **b** Top, expression pattern of a modulator–TF–target triplet: HFR1, HAC1, and AG. Each column represents a sample, which is sorted according to the expression level of the modulator (HFR1). The mutual information (MI) score of the AG and HAC1 on all the samples is 0.53. Middle, expression pattern of TF–target (HAC1 and AG) in HFR1 lowly expressed samples (bottom 35% samples). MI score is 0.34. Bottom, expression pattern of TF–target (HAC1 and AG) in HFR1 highly expressed samples (top 35% samples). TF and target gene are strongly correlated across those samples and the MI score is 0.72. The conditional mutual information (CMI) score is 0.38. **c** Significance test of the conditional mutual information of the HFR1–HAC1–AG. We randomly chose two groups of samples (each group contains 35% samples) and computed the mutual information between TF and target (HAC1–AG) in each group. We repeated this procedure 1000 times to generate the null distribution. Given HFR1’s expression level, the mutual information of HAC1 and AG is significantly changed ( $p$ -value < 0.001). **d** A summary for the example of the HFR1–HAC1–AG

PRR3 have a high out-degree. The key regulators of flower development, AP1, AP2, AP3, SEP2, and LFY have high out-degrees and regulate many TGs (Supplementary Fig. 2).

#### Hub TFs connect flower development and circadian rhythm

Flower development and circadian rhythm are known to be closely connected,<sup>20</sup> and this is demonstrated by the closeness of flower development and circadian rhythm in our global network. We then ask which regulators are central to connect the two biological processes at molecular level. To do this, we analyze a core TFA network in which all the nodes were restricted to TFs (Fig. 4). We calculated the node betweenness in this core network, which is defined as the number of the shortest paths from all nodes to all others that pass through that node. We found two nodes with the largest betweenness scores in our network: LHY and HFR1. In our network, LHY modulates 5 TFs and is modulated by 467 proteins, including 11 flower development and circadian rhythm-associated TFs. Most of the TFs modulating LHY are circadian rhythm associated, and all the TFs modulated by LHY are flower development associated. This network structure reveals that LHY is an important regulator in the process of circadian

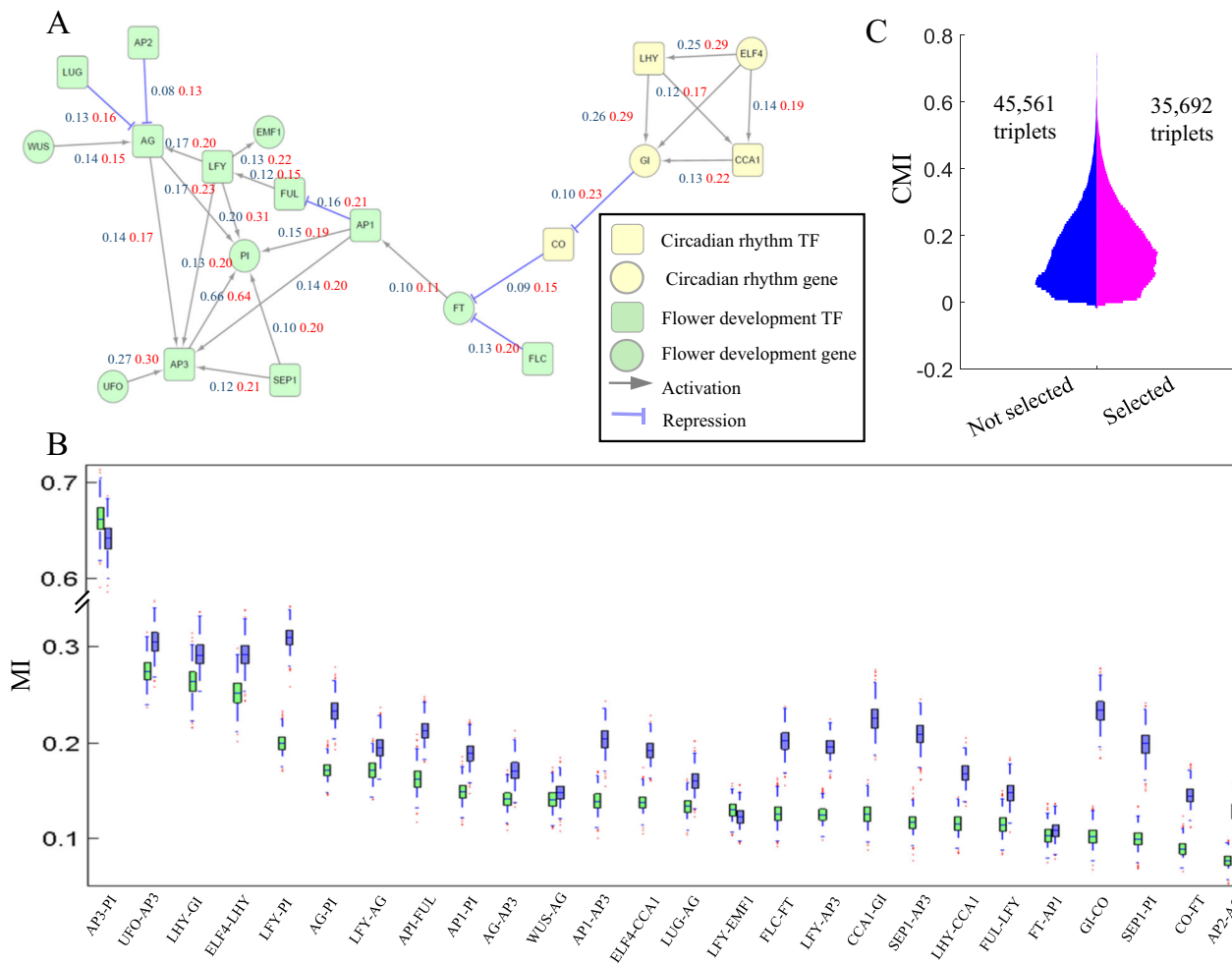
rhythm affecting flower development which is consistent with the previous study.<sup>19,20</sup>

The other central TF with high betweenness score is HFR1. We found that HFR1 modulates many flower development-associated TFs (Fig. 4). HFR1 is known to be involved in circadian rhythm.<sup>35</sup> Interestingly, HFR1 modulates 39 flower development-associated TFs in our network. About 40% (39 out of 97) of the flower development TFs are modulated by HFR1. It seems that HFR1 bridges circadian rhythm and flower development. We gather several evidences to show the possible mechanism of HFR1’s bridging role between circadian rhythm and flower development in the following sections.

HFR1 extends the flower development ABC model in a context specific way

Next, we examine the specific role of HFR1 in flower development. The flower morphogenesis has been modeled with an ABC model, with A specifying sepals, A&B petals, B&C stamens, and C carpels. Figure 5a shows the seven known interactions among the five genes in ABC model (A (AP1, AP2), B (AP3, PI), C (AG)). If HFR1 is an important regulator during flower development, it may have





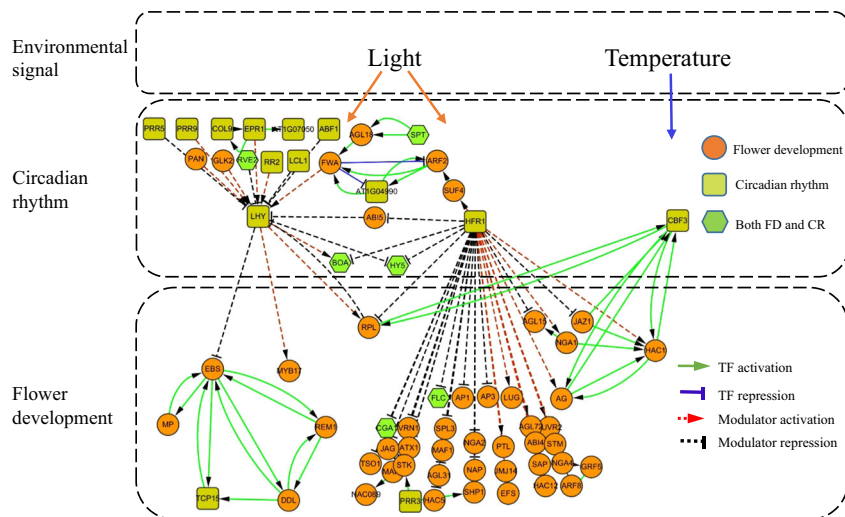
**Fig. 3** TF activity is more accurate than the TF's mRNA expression. **a** A gene regulatory network for flower development and circadian rhythm by correlating TF and target genes. The blue score and red score on edges represent the mutual information computed by TF's mRNA expression and TF activity, respectively. **b** Comparison of the distributions of the mutual information computed by mRNA expression and TF activity on each edge of the standard network. The green box and blue box represent the distributions of the mutual information computed by mRNA expression and TF activity, respectively. **c** Comparison of conditional mutual information on two groups of Modulator–TF–TG triplets, selected by TF activity or by TF's mRNA expression. Note that all triplets used in comparison have significant conditional mutual information

implications in the ABC model. To detect the role of HFR1 in the ABC model, we binned the samples into ten groups based on HFR1's expression level and computed the Pearson correlation coefficient (PCC) on the seven known pairs in each group (Fig. 5d). Given two genes, we binned the PCCs in ten groups into two categories, low and high, by k-means clustering. Then we tested whether the categories are significantly correlated with HFR1 expression or not. We calculated the CMI among HFR1 and each pair of the seven known interactions. Then we permuted HFR1 expression data 10,000 times and calculated CMI to generate background distribution and test the significance. As a result, five out of seven pairs are significant with  $p$ -value  $< 0.0001$ . The regulatory interactions among these five genes are known<sup>21</sup> (Fig. 5a). However, we can find that some interactions dynamically changed with the expression of HFR1 (Fig. 5b, c). In another words, the network rewires among the five genes with different HFR1 expressions (low or high). Interestingly, AP1 regulates AP3 and PI when HFR1 is lowly expressed (Fig. 5b). AG regulates AP1 and PI when HFR1 is highly expressed (Fig. 5c). The ABC model describes that A specify sepals, A&B specify petals, B&C specify stamens, and C specify carpels.<sup>36</sup> Our results show that B connects to A when HFR1 is expressed at low level and connects to C when HFR1 is expressed at high level. These lines of evidence explain that high expression of HFR1 seems to be required in the differentiation of

stamens and carpels by activating C to connect with B. Furthermore, these evidence together reveals that HFR1 seems to play an important modulation role during flowering in *Arabidopsis* as a regulator.

HFR1 modulates flower development by histone acetylation of AG. Next, we explored HFR1's function in flower development. In our TFA network HFR1 modulates many flower development-associated TFs on their TGs. However, the main target of HFR1's regulation to flowering remains unknown. We speculate that it may be AG based on several reasons: (1) it is clear in our TFA network HFR1 regulate flowering gene AG via HAC1 (Fig. 4). AG is one of the most important flowering-associated TFs; (2) high-level expression of HFR1 activates the connection between C (AG) and B (PI) to promote the differentiation of stamens and carpels; (3) AG is the receptor of two out of three main flowering signal pathways. These three reasons encourage us to investigate the regulation of AG by HFR1 in the flower.

We compiled an independent gene expression dataset on floral organs to test if HFR1-HAC1 and AG regulatory triplet is context specific in flower. We computed the PCC on high HFR1 samples (top 35%) and low HFR1 samples (bottom 35%) (Fig. 5e, f). We find that HAC1 and AG are highly correlated when HFR1 is highly



**Fig. 4** The core subnetwork among TFs in FDCRNNet

expressed in the flowing stage. This result in an independent data supports that HFR1 indeed regulates flowering via AG.

The next question is that how HFR1 regulates AG. One hypothesis is that HFR1 regulates AG through histone acetylation since HAC1 is a histone acetyltransferase to activate gene expression.<sup>37</sup> The evidence is as follows: firstly, HFR1 is the modulator of HAC1, HAC5, and HAC12 in our network. These three genes are key players of histone acetylation. Secondly, HFR1 belongs to the bHLH TFs family. Some of the bHLH family members are reported to regulate HAC1 activity.<sup>38</sup> These lines of evidence imply the relationship between HFR1 and HAC1.

To validate the relationship between HAC1 and AG on flower tissue, we compared the histone acetylation mark H3K27ac signal of AG on flower tissue and leaf tissue (Fig. 5g). Encouragingly, there is a H3K27ac peak in AG gene body region on flower tissue and no signal on leaf tissue,<sup>39</sup> which shows that AG may be activated by histone acetyltransferase HAC1 on flower.

Furthermore, we performed motif scanning in the H3K27ac peak of AG on flower tissue. We found a bHLH family TF motif (Jaspar ID: MF0007.1) which matches the center of histone acetylation peak (chr4:10,384,160-10,384,167) (Fig. 5g). This implies a possible binding footprint of HFR1 in acetylated region and the cooperation between HFR1 and HAC1 in regulating flower gene AG.

#### Experimental validation of the HFR1-HAC1 physical interaction

One simple explanation for HFR1's modulation on HAC1's regulation to AG is a physical interaction between HFR1 and HAC1. To test the interaction between HFR1 and HAC1, we carried out a yeast two-hybrid experiment (Fig. 6a–d). Neither the yeast strain with empty bait and prey vectors nor the strain expressing GAL4 binding domain (BD) and HFR1-AD (activation domain) fusion protein was able to grow on a synthetic dropout media (SD) plate without Trp, Leu, His, and Ade supplements, but both the strain with empty AD and HAC1-BD constructs and the strain with HFR-AD and HAC1-BD constructs could grow on the quadruple dropouts selective media, suggesting that HAC1 itself is able to activate expression of the reporter genes. To control the self-activation activity of HAC1, we added 5 mM 3-amino-1,2,4-triazole (3-AT) to the quadruple dropouts selective media. Under this condition, the strain expressing HFR-AD and HAC1-BD fusion proteins was still able to form colonies while the growth of control strain expression activation domain and HAC1-BD chimeric protein was completely inhibited. This result established that

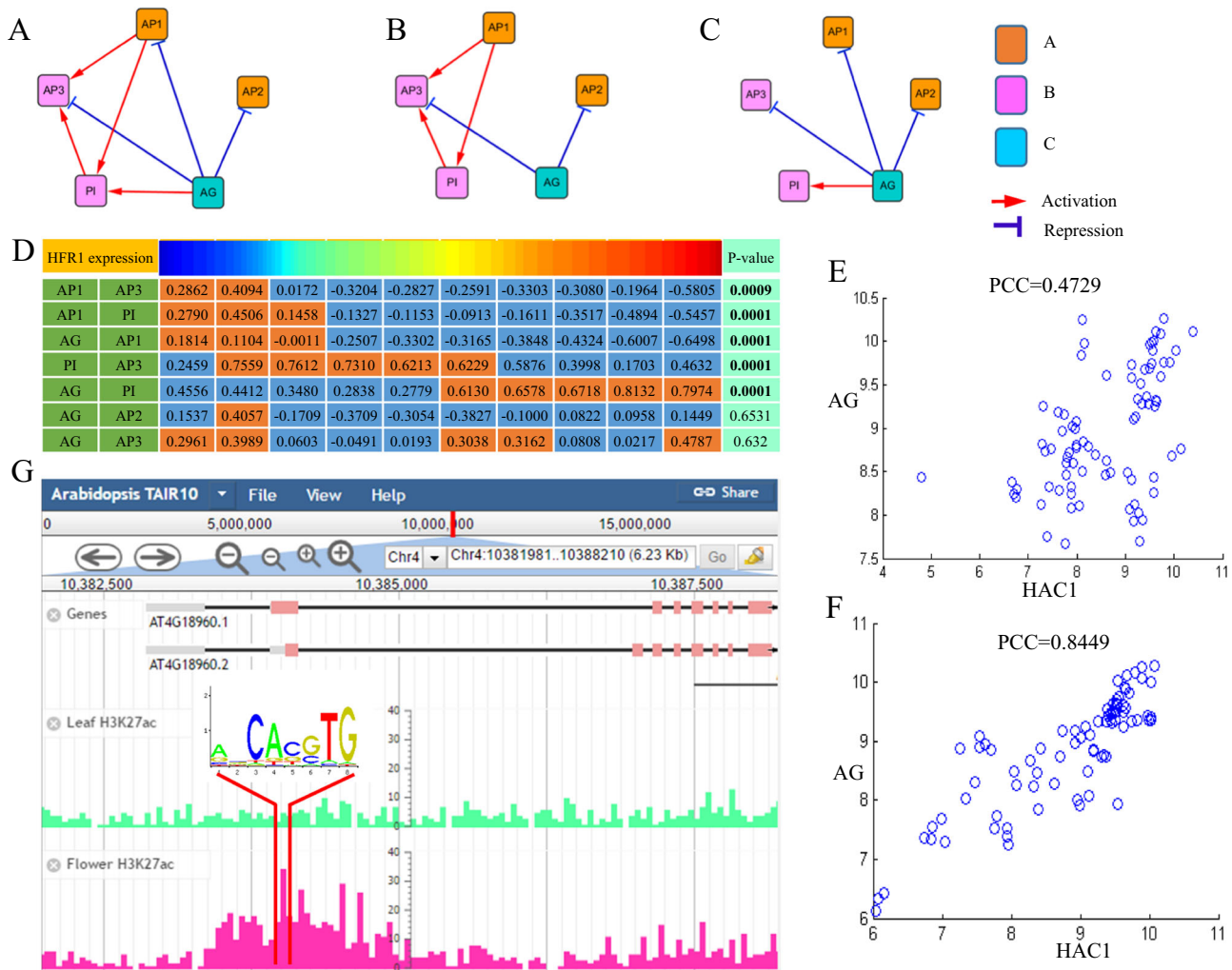
HFR1 and HAC1 are truly able to physically interact with each other.

## DISCUSSIONS

We proposed a new method to infer the GRM by systematically integrating prior knowledge, i.e., TF upstream regulation and downstream regulation. There are three levels of regulatory networks reported in this study: transcription regulatory network, modulation regulatory network, and TFA network. We inferred the direct regulation (transcription regulation) and indirect regulation (modulation regulation), and at last integrated these regulations in a graphical model to get a consistent network by introducing the TFA concept.

We applied our method in two important biological processes: flower development and circadian rhythm in *Arabidopsis thaliana*. As a result, we found that two hub TFs, LHY and HFR1, bridge these two biological processes. We hypothesize that HFR1 is an important regulator of flower development in *Arabidopsis thaliana*. HFR1 modulates some important flower development regulations, such as the interactions among the ABC model genes. TFA network indicates that regulation of HAC1 on AG is modulated by HFR1. Then, we found that AG is acetylated on flower tissue and but not in leaf. Interestingly, a bHLH family TF motif (Jaspar ID: MF0007.1) occurs in H3K27ac peak near AG, which indicates the footprint of HFR1 together with histone acetylation signal. At last, we validated that HFR1 is physically interacting with HAC1 by a yeast two-hybrid experiment. These results together suggest a model that HFR1 and HAC1 physically interact and bind to AG to activate its expression via histone acetylation. More specifically, transcriptional factor HFR1 sequence-specifically binds to AG and recruits the chromatin regulator HAC1 to acetylate AG, which activates AG's expression on flower tissue (Fig. 6e). Both experimental and public data validation support our results. These phenomena revealed that HFR1 regulates flowering, specifically regulates the differentiation of stamens and carpels by recruiting HAC1 to bind to AG, which further activates it by histone acetylation. This triplet relationship HFR1-(HAC1-AG) controlling flowering differentiation also enriches the classical ABC model in flower development.

Our study indicates that HFR1 plays a crucial role during the interplay between light and temperature. In general, the environmental signals, such as light and temperature, promote flowering by controlling the circadian rhythm.<sup>40,41</sup> Interestingly, HFR1 encodes a light-inducible protein and is involved in phytochrome signaling.<sup>42</sup> We notice in our network that there are three major



**Fig. 5** HFR1 is an important regulator for flower development and extends ABC model. **a** The known regulatory interactions within the ABC model (AP1, AP2, AP3, PI, and AG). **b, c** The predicted edges among the five genes by FDCRNet under the condition that HFR1 is lowly expressed and highly expressed respectively. **d** The correlations among the five genes in ABC model are dynamically changed along with the expression of HFR1. We binned the samples into ten groups by HFR1's expression level. We then computed the Pearson correlation coefficient (PCC) among the five genes in each group. Given two genes, we binned the PCCs in ten groups into two categories, low and high, by k-means clustering. The high PCC categories are colored in orange and low PCC categories are colored in blue. The *p*-values of conditional mutual information are computed in each category by the permutation test. **e, f** Validation of the modulation relationship (HFR1-HAC1-AG) in an independent floral organ data. Consistent with the FDCRNet results, the correlation between HAC1 and AG is strengthened by HFR1's high expression. **g** Comparison of H3K27ac signal of AG on flower and leaf tissues

pathways which link environmental signals to flower differentiation (Fig. 7). The first one is a well-known pathway: light-LHY circadian clock-GI-CO-FT pathway.<sup>32</sup> The second pathway is the light-HFR1 pathway to associate with some important flower development TFs. The third one is the temperature-CBF3-HAC1-AG pathway. The first pathway was reported in many studies<sup>19,20</sup> and we can correctly reconstruct it by our method. The first two pathways are regulated by light<sup>20,42,43</sup> and the third pathway is regulated by temperature.<sup>20,42,43</sup> TF HFR1 serves a critical role in the second and third pathways. In the second pathway, HFR1 is a modulator of 39 flower development-associated TFs, including AP1, AP3, AG, FLC, and LUG. In the third pathway, HFR1 serves as a modulator between HAC1 and AG (Figs. 4–6). These evidences indicate that HFR1 plays a key role in the transducing signals from light and temperature to influence circadian and hence flowering development.

We discuss some limitations of our analysis. In our model, we did not use time ordering information, which is important in

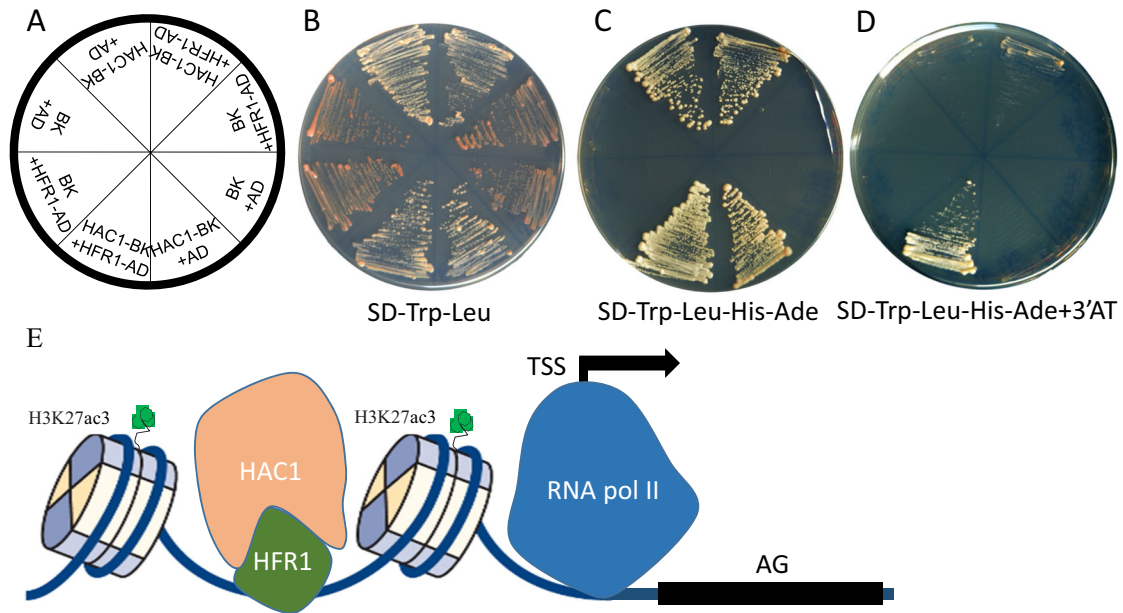
circadian rhythm, due to the following two reasons: (1) our model needs a large dataset as input data, but most of the time-course data have small number of samples; (2) we select the circadian rhythm genes from gene ontology, which have included most of the genes related to circadian rhythm. In the future, we will extend this model to consider the dynamical information<sup>23,24</sup> from time-course data.

## METHODS

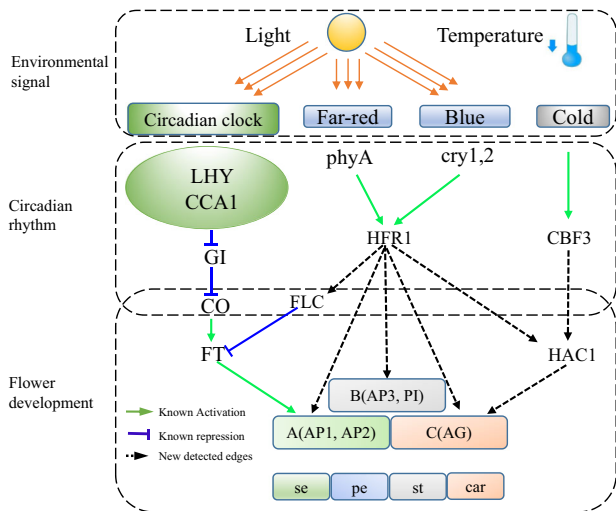
### Sources of datasets

We sourced the key players in flower development and circadian rhythm by searching the GO database. Finally we collected 97 flower development-associated TFs and 44 circadian rhythm-associated TFs (Supplementary Table 1 and 2). We used a preprocessed gene expression data from the TAIR database in our study.<sup>44</sup> The dataset contains 1436 hybridization experiments using the Affymetrix *A. thaliana* ATH1 (25 K) array, which contains 22,810 probe sets. These expression data are from multiple tissues (including leaf, root, flower, and seed) on multiple





**Fig. 6** HFR1 physically interacts with HAC1 and regulates flowering by activating AG. **a** A sketch map of yeast strains on the selective plates. BD + AD, yeast strain AH109 transformed with empty vectors pGBK-T7 and pGAD-T7; HAC1-BD + AD, AH109 transformed with pGBK-T7 background vector expressing HAC1-BD and empty vector pGAD-T7; BD + HFR1-AD, AH109 transformed with pGBK-T7 empty vector and pGAD-T7 background vector expressing HFR1-AD; HAC1-BD + HFR1-AD, AH109 expressing HAC1-BD and HFR1-AD in pGBK-T7 and pGAD-T7 vectors respectively. **b–d** Growth of the strains streaked on selective media plates with SD/-Trp/-Leu (**b**), SD/-Trp/-Leu/-His/-Ade (**c**) and SD/-Trp/-Leu/-His/-Ade/+3'AT (**d**). **e** A mechanistic model for HFR1 and HAC1's physical interaction to activate AG by histone acetylation



**Fig. 7** Three major pathways by which the circadian rhythm regulates flowering by environmental signature. HFR1 senses the light signal and regulates the flowering process

experimental conditions including biotic treatment, development, abiotic treatment, nutrient treatment, hormone treatment, chemical treatment, cell cycle, and genotype comparison. The probe IDs were converted into gene IDs. Probes that match more than one gene were removed. For those multiple probes that matched a single gene, the maximum expression value was assigned to that gene. The functional linkage data from STRING database are used to screen the candidate cofactors for the TFs.

In order to validate the relationships between HFR1, HAC1, and AG in an independent dataset, we collected public gene expression data of floral organ.<sup>44</sup> This data are from 11 series (GSE5632, GSE9605, GSE9721, GSE9799, GSE15689, GSE16056, GSE2848, GSE2847, GSE27281, GSE576 and GSE607) and contain 232 samples. We used the robust multi-array average method to normalize the data. ChIP-seq data of H3K27ac on flower tissue and leaf tissue are from PlantDHS database.<sup>39</sup>

#### Overview of methods

We introduce TFA  $P$  as a random variable.  $D$  represents the observed experimental data and  $G$  is the prior information from the draft network structure of the TF–TG regulations. Then the joint probability for the gene expression ( $D$ ), TFA ( $P$ ), and regulatory structure ( $G$ ) can be expressed as:

$$Pr(D, P, G) = Pr(G)Pr(D, P|G)$$

Regulatory structure  $G$  is derived from functional linkage data and the mRNA expression correlation among TG, TF, and other cofactors. The cofactors are defined as modulator and have influences on the regulation of a TF–TG pair.<sup>45</sup> Specifically, we compile a draft regulatory structure  $G$  by applying mutual information to detect TF–TG pairs and CMI to reveal the modulator–TF–TG triplets (Methods). Then we apply a Gaussian graphical model to estimate parameters and hidden variables by maximizing the conditional probability  $Pr(D, P|G)$ .

We decomposed gene expression ( $D$ ) into two parts, modulator expression data  $Q$  and target gene expression data  $E$ . We represent concentrations of modulators, TF activities, and expression levels of targets with random variables  $Q = (Q_1, \dots, Q_L, \dots, Q_K)$ ,  $P = (P_1, \dots, P_L)$ , and  $E = (E_1, \dots, E_N)$ , respectively (Supplementary Fig. 1). Here,  $Q_1, \dots, Q_L$  are the mRNA concentrations of TFs.  $Q_{L+1}, \dots, Q_K$  are the concentrations of cofactors or modulators,  $P$  are the activities and  $E$  are the expressions of TGs.  $Q$ ,  $P$ , and  $E$  are matrices of  $K \times M$ ,  $L \times M$ , and  $N \times M$ , respectively, where  $M$  is the number of samples. According to the dependent and conditional independent relationship in the network, we can get the likelihood as follows,

$$Pr(Q, E) = \prod_{k=1}^K Pr(Q_k) \prod_{l=1}^L Pr(P_l | pa(P_l)) \prod_{n=1}^N Pr(E_n | pa(E_n)), \quad (1)$$

where  $pa(X)$  represents all parents of node  $X$  in the network. TFA  $P$  is a linear combination of modulator  $Q$ , and the regulatory strength of the  $k$ -th modulator on the  $l$ -th TFA (TFA) is represented as  $\beta_{lk}$ . Target gene expression  $E$  is fitted by linear combination of TFA  $P$ , and the regulatory strength of the  $l$ -th TF on the  $n$ -th target gene is represented as  $\alpha_{nl}$ . The matrix  $B = (\beta_{lk})_{L \times K}$  is the effect strength of modulator on TFA and the matrix  $A = (\alpha_{nl})_{N \times L}$  is the regulatory strength of TF on target gene. Thus the input of TFA network inference is expression data of target gene ( $E$ ) and modulator ( $Q$ ). The parameters to be inferred are the regulatory strength matrix of modulator on TFA (matrix  $B$ ), the regulatory strength matrix of TF on TGs (matrix  $A$ ), and the hidden variables TFA ( $P$ ).



In summary, we use a three-step procedure to reconstruct the TFA network. At first, we use the existing data to mine the possible regulations to derive the draft network structure. Step 1 and Step 2 identify the potential target and modulator for the TFs and provide the candidate connections in the network. Then in Step 3, we consider the derived network structure into Eq. (1) and estimate the parameters and hidden variables by maximum likelihood estimation (MLE) using an Expectation Maximization (EM) algorithm.

### Reconstruction of the TF activity network

**Step 1, Infer transcription regulatory network.** To identify the potential TGs for each TF, we detect the possible interactions by assuming that functionally related two genes are always coexpressed on the gene expression data.<sup>46</sup> In this study, we use the mutual information, which measures the degree of statistical dependency between two variables to infer the transcriptional regulations. The mutual information of TF ( $X$ ) and a potential target gene ( $Y$ ) is defined as follows:

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left( \frac{p(x, y)}{p_1(x)p_2(y)} \right). \quad (2)$$

More specifically, we apply the Gaussian-kernel estimator to estimate the mutual information. We use the optimal Gaussian bandwidth<sup>46,47</sup> in Gaussian-kernel estimator.

**Step 2, Infer modulation network.** A modulator is a gene that influences the regulation of a TF-target gene pair.<sup>45</sup> For example, TF  $A$  regulate target gene  $B$  on the condition of gene  $C$  is high expressed (or low expressed) but not regulate target gene  $B$  on the condition of gene  $C$  is low expressed (or high expressed). For this example, gene  $C$  is a modulator between TF  $A$  and target gene  $B$ . The interaction between a modulator and a TF-target regulation is defined as modulation. There are two types of interactions in a modulation network. One is TF-target regulation and the other is modulation.

With the inferred transcriptional regulations in Step 1, we then apply the CMI to detect the modulation network. MINDy, an information-theoretic algorithm, is used to estimate the CMI.<sup>45</sup> The approximation of CMI of a TF ( $X$ ) and target gene ( $Y$ ) given a modulator ( $M$ ) is as follows:

$$CMI(X, Y|M) = MI(X, Y|M \in L_m^+) - MI(X, Y|M \in L_m^-), \quad (3)$$

where  $L_m^+$  is a set of samples on which gene  $M$  is highly expressed. Similarly, the  $L_m^-$  represents the set of samples on which the gene  $M$  is lowly expressed.

**Step 3, Infer TFA network.** In Steps 1 and Step 2, we inferred the potential structure of the network. Starting from those candidates structure, we further used the observed data to estimate the TFA and then the regulatory strengths of regulations between TFs and TGs. In this step, we applied Gaussian distribution to model the conditional distribution. We assumed that random variables  $Q$  (modulator expression),  $P$  (TFA), and  $E$  (Target gene expression) obey Gaussian distribution with mean linear combination of the parent nodes. In this linear Gaussian model, we have the activities of TFs depending on mRNA concentrations of TFs and corresponding modulators, and thus we have

$$Pr[p_l | pa(p_l)] \propto N \left( \sum_{k \in pa(p_l)} \beta_{lk} q_k, \sigma_{p_l|pa}^2 \right), \quad (4)$$

where  $l = 1, 2, \dots, L$ ;  $p_l$  is the activity of  $l$ -th TF (hidden variables),  $q_k$  is the observed value of  $Q_k$ ,  $pa(p_l)$  represent the parent nodes of  $p_l$ ,  $\beta_{lk}$  are the parameters to be estimated. Similarly, we have

$$Pr[E_n | pa(E_n)] \propto N \left( \sum_{l \in pa(E_n)} \alpha_{nl} p_l, \sigma_{E_n|pa}^2 \right), \quad (5)$$

where  $n = 1, 2, \dots, N$ ;  $E_n$  is expression of  $n$ -th TG,  $\alpha_{nl}$  are the parameters to be estimated.

According to the formulae (1), (4), and (5), we have

$$Pr(Q, E) \propto \sum_{k=1}^K N(\mu_{Q_k}, \sigma_{Q_k}^2) \sum_{l=1}^L N \left( \sum_k \beta_{lk} q_k, \sigma_{p_l|pa}^2 \right) \sum_{n=1}^N N \left( \sum_l \alpha_{nl} p_l, \sigma_{E_n|pa}^2 \right) \quad (6)$$

We need to estimate the parameter  $\Theta = (A, B, \sigma_{p_l|pa}^2, \sigma_{E_n|pa}^2)$  and the hidden variable  $P$ . To infer the optimal parameters and regulatory strengths of the graphical model, we follow the MLE to maximize the probability of observed data.<sup>48</sup> We maximize the logarithm likelihood

function

$$l(Q, E|\Theta) = \log(L(Q, E|\Theta)) = \sum_{i=1}^M \sum_{l=1}^L \log(\Pr(Q_i, E_i|P_i, \Theta)) \quad (7)$$

$$\max_{\Theta} \log(L(Q, E|\Theta)).$$

We should find a "good"  $\Theta$  to maximize the objective function in formula (7). Clearly, this problem belongs to the parameter estimation problem with missing data. The parameters are the regulatory strengths in the network structure and the missing data is the TFA.

### EM algorithm to estimate the parameters

We used the EM strategy to solve this problem. The EM strategy is an alternative iteration strategy. In E-step, we estimate the missing data  $P$  given the parameter  $\Theta$ , and in M-step we estimate the parameter  $\Theta$  given missing data  $P$ . Alternating these two steps, we get a local maximum of the logarithm likelihood function.

Given initial parameter  $\Theta$ , we use hard-assignment version EM strategy. In E-step, given the parameter (candidate network structure)  $\Theta^{(i-1)}$  we computed the missing data (TFA)  $P^{(i-1)}$  as follows:

$$P^{(i-1)} = \arg \max_P \log \left( L \left( P | \Theta^{(i-1)}, Q, E \right) \right) \quad (8)$$

In M-step, we computed the parameter (network structure)  $\Theta^{(i)}$  given the missing data (TFA)  $P^{(i-1)}$  as follows:

$$\Theta^{(i)} = \arg \max_{\Theta} \log \left( L \left( \Theta | P^{(i-1)}, Q, E \right) \right). \quad (9)$$

In formulae (8) and (9), the objective functions are all differentiable concave functions. Hence, the point with zero gradient is the global maximum. The solution of (8) is obtained by solving linear equations  $CP^{(i-1)} = D$ , and we have  $P^{(i-1)} = C^{-1}D$ , where

$$C = \left( \sum_{n=1}^N \frac{\alpha_{nl} \alpha_{nl}}{\sigma_{E_n|pa}^2} \right)_{L \times L} + \text{diag} \left( \frac{1}{\sigma_{p_l|pa}^2} \right)_{L}$$

$$D = \left( \sum_{k=1}^K \beta_{lk} q_k + \sum_{n=1}^N \frac{\alpha_{nl} e_n}{\sigma_{E_n|pa}^2} \right)_{L \times 1}$$

The solution of (9) is obtained by linear equations  $E = A^{(i)} P^{(i-1)}$ ,  $P^{(i-1)} = B^{(i)} Q$ . If  $P^{(i-1)}$  and  $Q$  have full rank, then  $A^{(i)} = EP^{(i-1)+}$ ,  $B^{(i)} = P^{(i-1)}Q^+$ , where  $X^+$  represents the pseudo-inverse matrix of the matrix. The condition that  $P^{(i-1)}$  and  $Q$  have full rank cannot be easily satisfied. In such a case, we add a condition that the network structure should be sparse. Therefore, we have a model as follows:

$$\min_{A, B} \|E - A^{(i)} P^{(i-1)}\|_2^2 + \|P^{(i-1)} - B^{(i)} Q\|_2^2 + \lambda \|A_1 + B_1\| \quad (10)$$

We use ADMM<sup>49</sup> to solve this L1 regularization optimization problem. The estimation of the variance  $\sigma_{p_l|pa}^2, \sigma_{E_n|pa}^2$  is as follows:

$$\sigma_{p_l|pa}^{2(i)} = \frac{\sum_{m=1}^M \left( P_{lm}^{(i-1)} - \sum_{k=1}^K \beta_{lk} q_{km} \right)^2}{M} \quad (11)$$

$$\sigma_{E_n|pa}^{2(i)} = \frac{\sum_{m=1}^M \left( e_{nm}^{(i-1)} - \sum_{l=1}^L \alpha_{nl} p_{lm} \right)^2}{M} \quad (12)$$

### Yeast two-hybrid experiment to validate HFR1-HAC1 interaction

The yeast two-hybrid experiment was performed by using Matchmaker™ GAL4 two-hybrid system 3 (Clontech Laboratories, Inc. USA). To make the expression vectors, the coding sequence of *HAC1* was de novo-synthesized and in frame inserted into the bait vector pGBK-T7 to form the express vector pGBK-HAC1. Meanwhile, the coding fragment of *HFR1* was amplified with primers 5'-CATATGATGGGGAGAGCTCCGTGC-3' (introduced with a *NdeI* restriction site) and 5'-GTCGACAATTCATCCCAAGCTTCCCGG-3' (introduced with a *Sall* restriction site) from Col-0 cDNA library. The fragment was then in frame recombined into the *NdeI* and *XhoI* site of bait vector pGBK-T7 by enzyme restriction to generate the expression vector pGBK-HFR1. The yeast strain AH109 was then transformed with these vectors according to the manual of Matchmaker™ GAL4 Two-Hybrid System 3. Positive transformants were screened on selective SD/-Trp/-Leu media. Medium-stringency examination of interaction was performed on SD/-Trp/-Leu/-His/-Ade media. To minimize the leaky expression of *HIS3*

that can be caused by self-activation of HAC1, 3-AT was added into the SD/-Trp/-Leu/-His/-Ade media at a final concentration of 5 mM.

### Method statement

Our method was performed in compliance with relevant guidelines and regulations and was approved by Academy of Mathematics and Systems Science, Chinese Academy of Sciences.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

Data that support the findings of this study is from the TAIR database (<https://www.arabidopsis.org/>).

### CODE AVAILABILITY

Code for TF activity network is in github (<https://github.com/durenzn/TF-activity-network>).

### ACKNOWLEDGEMENTS

The authors thank Prof. Xiaofeng Cao for helpful discussions. This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB13000000), the National Key R&D Program of China (No. 2017YFA0505500, 2018YFC09105), and the Monsanto Company. The authors are also supported by The National Natural Science Foundation of China (NSFC) under Grants (Nos. 11871463, 61671444, 61621003, 31771476, 61772368, 61572363), Natural Science Foundation of Shanghai (17ZR1445600), Shanghai Municipal Science and Technology Major Project (2018SHZDZX01), ZHANGJIANG LAB and CAS strategic leading project (No. XDA08020301).

### AUTHOR CONTRIBUTIONS

J.L., X.G.Z., L.C., and Y.W. conceived the research. Z.D., J.W., L.C., and Y.W. designed the model and the computational framework. Y.W. and X.G.Z. carried out the experiments. X.M.Z., L.L., and X.L. contributed to the results interpretation. Z.D., J.L., X.G.Z., L.C., and Y.W. wrote the paper.

### ADDITIONAL INFORMATION

**Supplementary information** accompanies the paper on the *npj Systems Biology and Applications* website (<https://doi.org/10.1038/s41540-019-0106-3>).

**Competing interests:** The authors declare no competing interests.

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

### REFERENCES

- Wang, J., Zhang, S., Wang, Y., Chen, L. & Zhang, X.-S. Disease-aging network reveals significant roles of aging genes in connecting genetic diseases. *PLoS Comput. Biol.* **5**, e1000521 (2009).
- Teichmann, S. A. & Babu, M. M. Gene regulatory network growth by duplication. *Nat. Genet.* **36**, 492–496 (2004).
- Ma, S., Jiang, T. & Jiang, R. Differential regulation enrichment analysis via the integration of transcriptional regulatory network and gene expression data. *Bioinformatics* **31**, 563–571 (2014).
- Greenfield, A., Madar, A., Ostrer, H. & Bonneau, R. DREAM4: combining genetic and dynamic information to identify biological networks and dynamical models. *PLoS ONE* **5**, e13397 (2010).
- Cantone, I. et al. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell* **137**, 172–181 (2009).
- Penfold, C. A. & Wild, D. L. How to infer gene networks from expression profiles, revisited. *Interface Focus* **1**, 857–870 (2011).
- Wang, Y., Joshi, T., Zhang, X.-S., Xu, D. & Chen, L. Inferring gene regulatory networks from multiple microarray datasets. *Bioinformatics* **22**, 2413–2420 (2006).
- Zhana, D., Yong, W., Shigeru, S. & Katsuhisa, H. *Control Conference (CCC), 31st Chinese*. 7437–7442 (IEEE, 2012).
- Beal, M. J., Falciani, F., Ghahramani, Z., Rangel, C. & Wild, D. L. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics* **21**, 349–356 (2005).
- Werhli, A. V., Grzegorzczak, M. & Husmeier, D. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. *Bioinformatics* **22**, 2523–2531 (2006).
- Zou, C. & Feng, J. Granger causality vs. dynamic Bayesian network inference: a comparative study. *BMC Bioinform.* **10**, 122 (2009).
- Duren, Z. et al. Modeling gene regulation from paired expression and chromatin accessibility data. *Proc. Natl Acad. Sci.* **114**, E4914–E4923 (2017).
- Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7**, S7 (2006).
- Zhang, X. et al. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics* **28**, 98–104 (2012).
- Zhao, J., Zhou, Y., Zhang, X. & Chen, L. Part mutual information for quantifying direct associations in networks. *Proc. Natl Acad. Sci.* **113**, 5130–5135 (2016).
- Duren, Z. & Wang, Y. A systematic method to identify modulation of transcriptional regulation via chromatin activity reveals transcription factor network during mESC differentiation. *Sci. Rep.* **6**, 22656, <https://doi.org/10.1038/srep22656> (2016).
- Guerra, D. et al. Post-transcriptional and post-translational regulations of drought and heat response in plants: a spider's web of mechanisms. *Front. Plant Sci.* **6**, 57 (2015).
- Schaffer, R. et al. The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* **93**, 1219–1229 (1998).
- Mizoguchi, T. et al. Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in Arabidopsis. *Plant Cell Online* **17**, 2255–2270 (2005).
- Park, D. H. et al. Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. *Science* **285**, 1579–1582 (1999).
- Espinosa-Soto, C., Padilla-Longoria, P. & Alvarez-Buylla, E. R. A gene regulatory network model for cell-fate determination during Arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *Plant Cell Online* **16**, 2923–2939 (2004).
- Mendoza, L. & Alvarez-Buylla, E. R. Dynamics of the genetic regulatory network for Arabidopsis thaliana flower morphogenesis. *J. Theor. Biol.* **193**, 307–319 (1998).
- Pruneda-Paz, J. L., Breton, G., Para, A. & Kay, S. A. A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. *Science* **323**, 1481–1485 (2009).
- Locke, J., Millar, A. & Turner, M. Modelling genetic networks with noisy and varied experimental data: the circadian clock in Arabidopsis thaliana. *J. Theor. Biol.* **234**, 383–393 (2005).
- Liao, J. C. et al. Network component analysis: reconstruction of regulatory signals in biological systems. *Proc. Natl Acad. Sci.* **100**, 15522–15527 (2003).
- Li, Z., Shaw, S. M., Yedwabnick, M. J. & Chan, C. Using a state-space model with hidden variables to infer transcription factor activities. *Bioinformatics* **22**, 747–754 (2006).
- Mehra, A., Baker, C. L., Loros, J. J. & Dunlap, J. C. Post-translational modifications in circadian rhythms. *Trends Biochem. Sci.* **34**, 483–490 (2009).
- Jang, I.-C., Yang, J.-Y., Seo, H. S. & Chua, N.-H. HFR1 is targeted by COP1 E3 ligase for post-translational proteolysis during phytochrome A signaling. *Genes Dev.* **19**, 593–602 (2005).
- Tran, L. M., Brynildsen, M. P., Kao, K. C., Suen, J. K. & Liao, J. C. gNCA: a framework for determining transcription factor activity based on transcriptome: identifiability and numerical implementation. *Metab. Eng.* **7**, 128–141 (2005).
- He, Y., Michaels, S. D. & Amasino, R. M. Regulation of flowering time by histone acetylation in Arabidopsis. *Science* **302**, 1751–1754 (2003).
- Pavangadkar, K., Thomashow, M. F. & Triezenberg, S. J. Histone dynamics and roles of histone acetyltransferases during cold-induced gene regulation in Arabidopsis. *Plant Mol. Biol.* **74**, 183–200 (2010).
- Fowler, S. et al. GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *EMBO J.* **18**, 4679–4688 (1999).
- Colasanti, J. et al. The maize INDETERMINATE1 flowering time regulator defines a highly conserved zinc finger protein family in higher plants. *BMC Genom.* **7**, 1 (2006).
- Wang, Y., Jiang, R. & Wong, W. H. Modeling the causal regulatory network by integrating chromatin accessibility and transcriptome data. *Natl Sci. Rev.* **3**, 240–251 (2016).
- Yamashino, T. et al. A link between circadian-controlled bHLH factors and the APRR1/TOC1 quintet in Arabidopsis thaliana. *Plant Cell Physiol.* **44**, 619–629 (2003).

36. Barrio, R. A., Hernandez-Machado, A., Varea, C., Romero-Arias, J. R. & Alvarez-Buylla, E. Flower development as an interplay between dynamical physical fields and genetic networks. *PLoS ONE* **5**, e13523 (2010).
37. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).
38. Hamamori, Y. et al. Regulation of histone acetyltransferases p300 and PCAF by the bHLH protein twist and adenoviral oncoprotein E1A. *Cell* **96**, 405–413 (1999).
39. Zhang, T., Marand, A. P. & Jiang, J. PlantDHS: a database for DNase I hypersensitive sites in plants. *Nucleic Acids Res.* **44**, D1148–D1153 (2015).
40. Min, N. Integration of light signaling with photoperiodic flowering and circadian rhythm. *Cell Res.* **15**, 559–566 (2005).
41. Chao, L.-M. et al. Arabidopsis transcription factors SPL1 and SPL12 confer plant thermotolerance at reproductive stage. *Mol. Plant* **10**, 735–748 (2017).
42. Yang, J. et al. Light regulates COP1-mediated degradation of HFR1, a transcription factor essential for light signaling in Arabidopsis. *Plant Cell* **17**, 804–821 (2005).
43. Seo, E. et al. Crosstalk between cold response and flowering in Arabidopsis is mediated through the flowering-time gene SOC1 and its upstream negative regulator FLC. *Plant Cell* **21**, 3185–3197 (2009).
44. Huala, E. et al. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **29**, 102–105 (2001).
45. Wang, K. et al. Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat. Biotechnol.* **27**, 829–837 (2009).
46. Steuer, R., Kurths, J., Daub, C. O., Weise, J. & Selbig, J. The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics* **18**, S231–S240 (2002).
47. Silverman, B. W. Density estimation for statistics and data analysis. (Chapman and Hall, London, UK, 1986).
48. Ye, C., Galbraith, S. J., Liao, J. C. & Eskin, E. Using network component analysis to dissect regulatory networks mediated by transcription factors in yeast. *PLoS Comput. Biol.* **5**, e1000311 (2009).
49. Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Vol. 3, 1–122 (Foundations and Trends® in Machine Learning, 2011).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019