

# SCIENTIFIC REPORTS



OPEN

## Differential Coexpression Analysis Reveals Extensive Rewiring of *Arabidopsis* Gene Coexpression in Response to *Pseudomonas syringae* Infection

Received: 26 April 2016  
Accepted: 23 September 2016  
Published: 10 October 2016

Zhenhong Jiang<sup>1</sup>, Xiaobao Dong<sup>1</sup>, Zhi-Gang Li<sup>1</sup>, Fei He<sup>1,2</sup> & Ziding Zhang<sup>1</sup>

Plant defense responses to pathogens involve massive transcriptional reprogramming. Recently, differential coexpression analysis has been developed to study the rewiring of gene networks through microarray data, which is becoming an important complement to traditional differential expression analysis. Using time-series microarray data of *Arabidopsis thaliana* infected with *Pseudomonas syringae*, we analyzed *Arabidopsis* defense responses to *P. syringae* through differential coexpression analysis. Overall, we found that differential coexpression was a common phenomenon of plant immunity. Genes that were frequently involved in differential coexpression tend to be related to plant immune responses. Importantly, many of those genes have similar average expression levels between normal plant growth and pathogen infection but have different coexpression partners. By integrating the *Arabidopsis* regulatory network into our analysis, we identified several transcription factors that may be regulators of differential coexpression during plant immune responses. We also observed extensive differential coexpression between genes within the same metabolic pathways. Several metabolic pathways, such as photosynthesis light reactions, exhibited significant changes in expression correlation between normal growth and pathogen infection. Taken together, differential coexpression analysis provides a new strategy for analyzing transcriptional data related to plant defense responses and new insights into the understanding of plant-pathogen interactions.

Plants as sessile organisms are subject to numerous attacks from microbes during their lifetime. As a result, plants have evolved a sophisticated immune system that enables each cell to monitor every invasion by microbes and to mount an appropriate defense response when necessary. Typical immune responses include the generation of reactive oxygen species, activation of the MAPK pathway, deposition of callose and the production of phytohormones, involving complicated transcriptional reprogramming<sup>1–3</sup>. These immune responses are interconnected and collaborative for resisting pathogens.

Microarray technology has provided a powerful approach for analyzing genome-wide gene expression profiling during plant immune responses. Typically, differentially expressed genes (DEGs) during plant immune responses are identified as potential plant defense-related gene candidates. Differential expression analysis considers each gene individually, while their potential interactions are ignored. However, genes or their protein products do not act in isolation; instead, they are interrelated with each other and act in close coordination<sup>4–6</sup>. Taking interactions between genes into account, various approaches based on gene sets<sup>7</sup>, coexpression<sup>8–10</sup>, machine learning<sup>11</sup> and biological networks<sup>12</sup> have been adopted to analyze transcriptional datasets involved in plant immunity<sup>13–15</sup>. For example, a large-scale immune coexpression network was constructed to identify immune-related functional modules involved in plant defense<sup>10</sup>. Previously, we employed an advanced machine learning method to integrate *Arabidopsis* gene networks with transcriptome data<sup>11</sup>. Through comprehensive network analysis, we revealed

<sup>1</sup>State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China. <sup>2</sup>Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA. Correspondence and requests for materials should be addressed to F.H. (email: plane83@gmail.com) or Z.Z. (email: zidingzhang@cau.edu.cn)

shared and distinct network organizations between pattern-triggered immunity and effector-triggered immunity in *Arabidopsis*.

Recently, differential coexpression analysis is emerging as an important complement to the traditional differential expression analysis<sup>16–18</sup>. Instead of focusing on individual genes, the goal of differential coexpression analysis is to identify gene pairs with different expression correlation between two conditions (such as healthy and diseased samples)<sup>16</sup>. One of the possible mechanisms behind differential coexpression is that the regulation of a pair of genes may be disrupted/activated under a certain condition<sup>17,19</sup>. A number of approaches have been developed to detect differential coexpression. Some methods focus on the detection of differential coexpression gene sets<sup>20,21</sup>, whereas others aim to identify differential coexpression gene pairs<sup>22,23</sup>. In 2013, Amar *et al.* developed a method called DICER to detect differentially coexpressed gene sets using a probabilistic model<sup>20</sup>. DICER not only identifies a gene set with correlation significantly altered between two conditions but also provides pairs of gene sets with changed correlational relationships. Based on Fisher's z-test, Fukushima *et al.* explored tomato gene function via differential coexpression analysis<sup>24</sup>. Later, they published an R package called DiffCorr to identify differential coexpression gene pairs<sup>22</sup>. Yang *et al.* implemented several methods in the R package DCGL to identify differential coexpression gene pairs, differentially coexpressed genes (DCGs) and transcription factors (TFs) that regulate differential coexpression<sup>23</sup>.

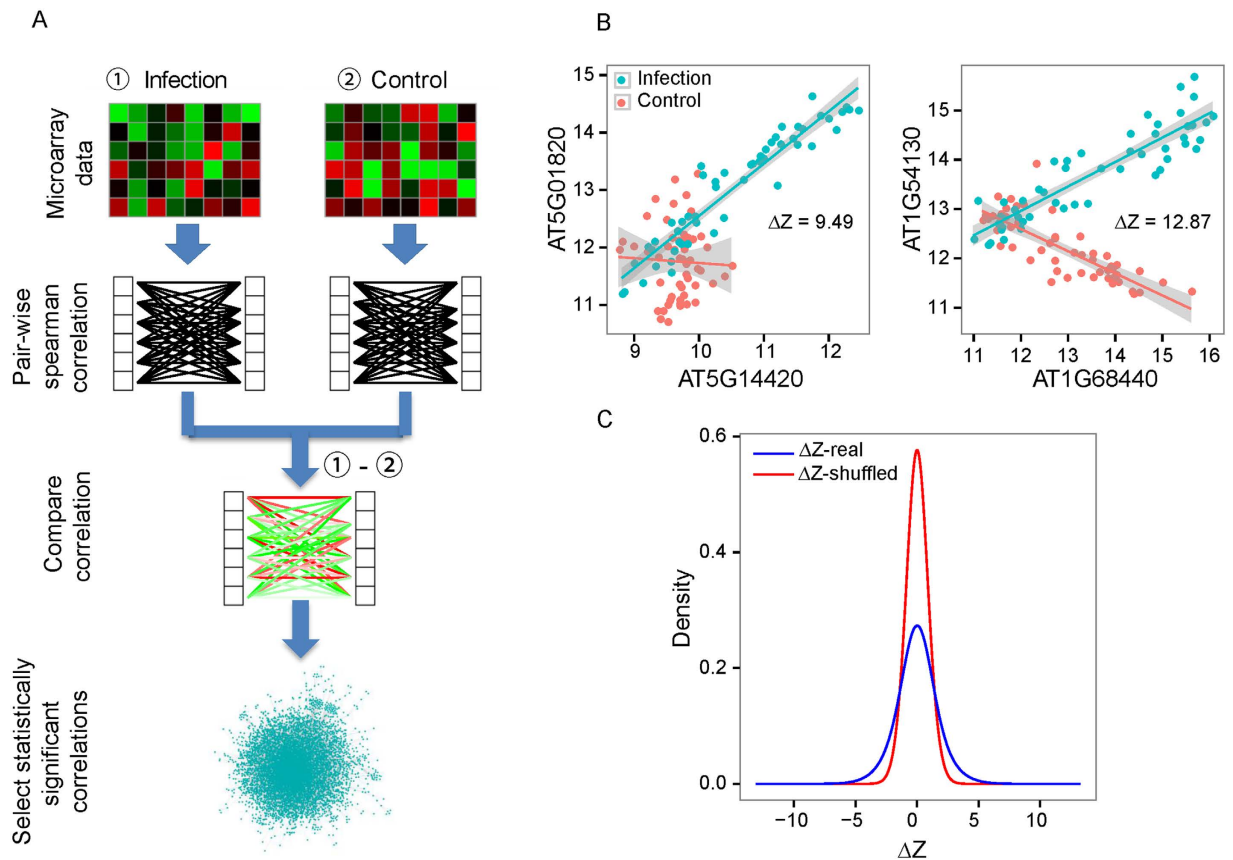
Differential coexpression analysis has been widely used to analyze human disease-related transcriptional data<sup>16,17</sup>. Choi *et al.* applied differential coexpression analysis to detect major cellular changes in tumor cells and found gene groups whose co-regulation may contribute to malignant transformation<sup>25</sup>. Hudson *et al.* successfully inferred myostatin as the gene containing the causal mutation of their interested phenotype through differential coexpression analysis<sup>26</sup>. Recently, Reznik and Sander studied differential coexpression across metabolic pathways in both breast and kidney tumors, and they integrated regulatory information to study the drivers (*i.e.*, TF) of these changes<sup>27</sup>.

Although a plethora of transcriptional datasets measuring *Arabidopsis* gene expression during plant immunity are available<sup>28</sup>, to the best of our knowledge no study has been conducted to analyze plant immune responses using differential coexpression analysis. *Pseudomonas syringae* is a Gram-negative bacterial pathogen that causes diseases in a wide range of plant species. The *Arabidopsis*-*P. syringae* interaction is recognized as one of the most important model systems for understanding plant-pathogen interactions<sup>29</sup>. Transcriptomics studies focusing on this model system have already broadened our understanding of plant-pathogen interactions. For example, de Torres-Zabala *et al.* generated a high-resolution time series expression profile [Gene Expression Omnibus (GEO) accession number: GSE56094] measuring *Arabidopsis* gene expression following either mock, *P. syringae* pv. *tomato* DC3000 or *P. syringae* pv. *tomato* DC3000 *hrpA* challenges. They analyzed the expression changes of nuclear-encoded chloroplast-targeted *Arabidopsis* genes and showed that chloroplast was a key component of early immune responses<sup>30</sup>. In combination with hormone profiling, reverse genetics and RNA-seq analyses, they also explored the dynamics, interaction and contribution of jasmonic acid, coronatine (COR) and jasmonate ZIM-domain (JAZ) proteins to *P. syringae* disease progression<sup>31</sup>. Moreover, Lewis *et al.* conducted a comparative analysis on the same transcriptional data for exploring the transcriptional dynamics during microbial-associated molecular pattern-triggered immunity induced by *P. syringae* pv. *tomato* DC3000 *hrpA* treatment and effector-triggered susceptibility caused by *P. syringae* pv. *tomato* DC3000 challenge<sup>32</sup>.

To date, the GEO database<sup>33</sup> contains over 400 samples related to *Arabidopsis* immune responses to the infection of *P. syringae*, which provides a great opportunity to revisit the transcriptomics data through some new computational methods and thus convert these data into new biological discoveries. In this work, we focused on two sub-series from GSE56094. We first showed that differential coexpression was a common phenomenon during plant immune responses. Then, we identified 1,315 genes (*i.e.*, DCGs) that frequently changed their coexpression partners. To explore the biological significance of differential coexpression, we identified some potential TFs regulating differential coexpression in plant immune responses with the assistance of known *Arabidopsis* gene regulatory networks. Furthermore, we investigated differential coexpression in the context of metabolic pathways. These results further indicated that the *Arabidopsis* gene network has been extensively rewired in response to infections by plant pathogens.

## Results and Discussion

**Differential coexpression is extensive during plant immune responses.** The microarray data GSE56094 is composed of 156 distinct samples from 13 time points in three conditions: mock treatments or infections by either virulent *P. syringae* pv. *tomato* DC3000 or the corresponding nonpathogenic *hrpA* mutant, with four replicates for each condition<sup>30–32</sup>. In our work, we focused on mock and virulent *P. syringae* pv. *tomato* DC3000 treatments. Therefore, 104 samples from GSE56094 were used in this work, including 52 samples from mock-treated control and 52 samples infected by bacteria (Fig. 1A). We only used 6,775 genes with an expression variance larger than 0.2 in either of the two conditions. The R package DiffCorr<sup>22</sup> was used to identify differentially coexpressed gene pairs. Briefly, Spearman correlation coefficients (SCCs) and associated *p*-values for all possible gene pairs among these 6,775 genes were calculated for each condition. Then, the difference in correlation between two conditions for every gene pair was further evaluated using  $\Delta Z$ . A pair of genes is defined as differentially coexpressed if the expression correlations of the gene pair under two conditions (pathogen infection and mock treatment) are significantly different (see the Materials and Methods section for details). In total, we obtained 124,115 differential coexpression gene pairs among 4,748 genes (The complete list of these differential coexpression gene pairs is available at: <http://systbio.cau.edu.cn/SI/index.php>). Figure 1B illustrated two examples of differentially coexpressed gene pairs from our work. In the first example, the strength of coexpression between AT5G14420 and AT5G01820 increased from  $-0.03$  to  $0.96$  after infection. With respect to the AT1G68440 and AT1G54130 gene pair in the second example, the strength of coexpression was reversed between control ( $-0.83$ ) and infection ( $0.89$ ).



**Figure 1. Overview of the differential coexpression analysis.** (A) The computational framework used to identify differential coexpression gene pairs. (B) Two real cases of differentially coexpressed gene pairs. The  $x$ -axis and  $y$ -axis reflect gene expression level of the differentially coexpressed gene pairs. The solid lines reflect the linear correlation between two genes. (C) The distribution of transformed difference in correlation ( $\Delta Z$ ) in real (blue) and shuffled (red) data. Statistical analysis shows that the absolute values of  $\Delta Z$  in real data are significantly larger than values in shuffled data (Student's  $t$ -test,  $p$ -value  $< 2.2 \times 10^{-16}$ ).

To investigate whether these drastically changed coexpression patterns are biologically significant, we followed the strategy of Amar *et al.*<sup>20</sup> to shuffle data and then assessed the distribution of coexpression differences for all possible gene pairs on the microarray. The coexpression difference for a gene pair is measured using  $\Delta Z$  (Eq. 1, see the Materials and Methods section for details). A larger absolute value of  $\Delta Z$  indicates a larger coexpression difference. If the differential coexpression is common, the real biological data will have larger coexpression differences than the shuffled data. We permuted the 104 samples by shuffling sample labels (*i.e.*, infection or control) and calculated  $\Delta Z$  based on the shuffled data. As expected, the absolute values of  $\Delta Z$  are larger in real data compared with shuffled data (Student's  $t$ -test,  $p$ -value  $< 2.2 \times 10^{-16}$ , Fig. 1C). To test whether this phenomenon was prevalent during plant immune responses or it is specific to the microarray dataset we used, we repeated the analysis on a combined gene expression dataset generated by several different studies (GEO accession number: GSE5685, GSE5615, GSE21920, GSE18978 and GSE17500<sup>34</sup>) and found that the absolute values of  $\Delta Z$  are larger in real data compared with shuffled data (Student's  $t$ -test,  $p$ -value  $< 2.2 \times 10^{-16}$ , Figure S1A and Table S1). Furthermore, the same pattern was observed not only for gene expression under the infection of *P. syringae* but also for other pathogens, such as *Golovinomyces orontii* (GEO accession number: GSE5686, Student's  $t$ -test,  $p$ -value  $< 2.2 \times 10^{-16}$ , Figure S1B, Table S1) and *Botrytis cinerea* (GEO accession number: GSE29642<sup>35</sup>, Student's  $t$ -test,  $p$ -value  $< 2.2 \times 10^{-16}$ , Figure S1C). Based on the above analyses, we concluded that differential coexpression was a common phenomenon during plant immune responses.

**DCGs are heavily involved in plant immune responses.** In our analysis, we defined genes that were frequently differentially coexpressed with other genes as DCGs. To identify DCGs, we tested the enrichment of differentially coexpressed gene pairs for each gene. For the gene  $i$  that participated in  $k_i$  differential coexpression gene pairs, the  $p$ -value was calculated based on a binomial probability model (see Eq. 2 in the Materials and Methods section). If the corrected  $p$ -value of gene  $i$  was smaller than a given cutoff of 0.05, gene  $i$  was assigned as a DCG. According to the criterion, we totally obtained 1,315 DCGs (Table S2).

Gene Ontology (GO)<sup>36</sup> enrichment analysis of 1,315 DCGs revealed that the most significantly enriched GO term was “response to stimulus” (hypergeometric test,  $p$ -value =  $9.80 \times 10^{-20}$ , Table S3A), indicating the extensive involvement of DCGs in plant responses to stimuli. Other biological processes related to plant defense responses

have also been enriched in DCGs, such as “defense response” and “cell wall modification”. In addition to GO enrichment analysis, we compared 1,315 DCGs with 1,393 curated defense-related genes (see the Materials and Methods section for details) and found that DCGs were significantly enriched in defense-related genes (hypergeometric test,  $p$ -value =  $2.73 \times 10^{-09}$ ). This observation highlighted the important functional roles of DCGs in plant defense responses.

For 129 plant defense-related DCGs, some have already been proved to play vital roles in plant immunity to *P. syringae*. For example, *FLS2*, which encodes an LRR receptor-like kinase involved in the perception of flagellin in *Arabidopsis*<sup>37</sup>, was differentially coexpressed with 229 genes. The observation reflected that the coexpression between *FLS2* and these 229 genes were extensively altered after *P. syringae* infection. Other important defense-related genes included the R gene *ADRI*, a JA synthetase *JARI*, and so on. Notably, the DCG with the largest number of differential coexpression events was AT3G03440, which was involved in 1,001 differential coexpression gene pairs. AT3G03440 encodes a chloroplast-located armadillo (ARM) repeat family protein, but its physiological function remains elusive. Even so, some studies have suggested their potential functional importance in many biological processes, including disease resistance<sup>38,39</sup>. Through analyzing its 1,001 differentially coexpressed partners, “defense response” was found to be over-represented (hypergeometric test,  $p$ -value =  $4.39 \times 10^{-3}$ , Table S3B). Based on these observations, AT3G03440 may be a good candidate to study plant immune responses to *P. syringae*.

**Differential coexpression analysis provides additional information that is complementary to differential expression analysis.** Differential expression analysis is a routine method to conduct microarray data analysis. Generally, differential expression analysis treats each gene separately and does not consider its relationship with potential interaction partners. In this context, it is interesting to decipher the relationship between differential coexpression and differential expression based on the transcriptional dataset used in this work.

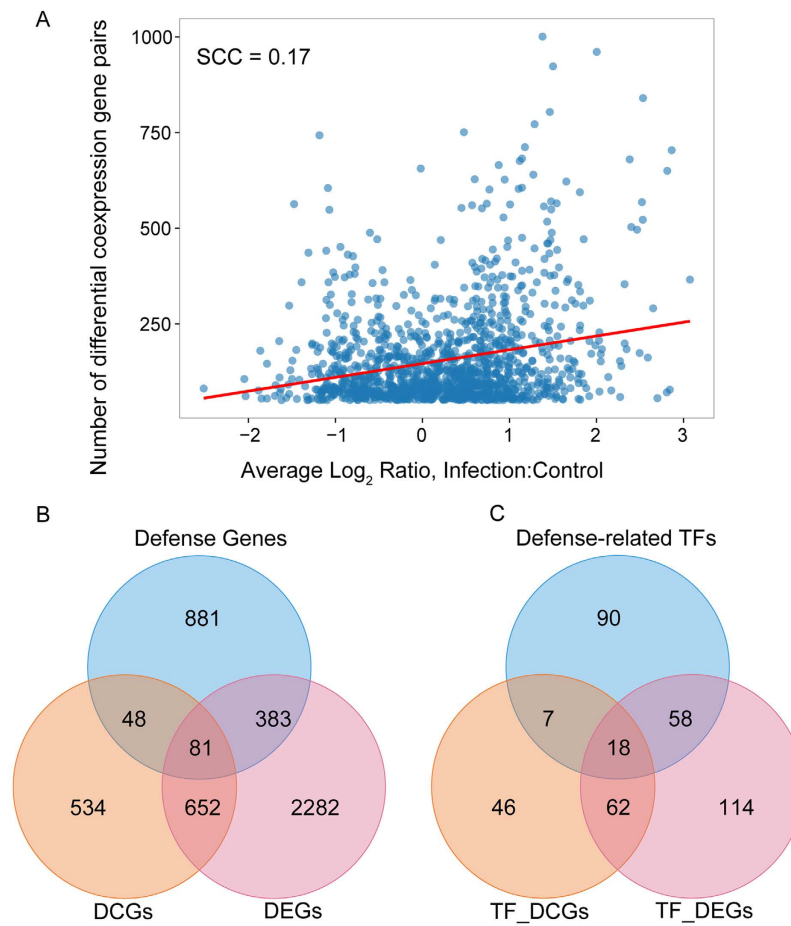
We first calculated the magnitude of differential expression (measured by average  $\log_2$ -fold change) for each DCG (Table S2). Then, we compared the magnitude of differential expression with the number of differentially coexpressed gene pairs on the identified 1,315 DCGs. As shown in Fig. 2A, the correlation between these two measures was generally weak (SCC = 0.17). The result was consistent with the previous observation that differential expression analysis sheds little light on differential coexpression analysis<sup>27</sup>. We also compared 1,315 DCGs with 3,398 DEGs (Table S4) detected using the R package maSigPro<sup>40</sup> (see the Materials and Methods section). The overlap between DCGs and DEGs was generally large, but there were 582 DCGs that were not detected as differentially expressed by maSigPro (Fig. 2B, Table S2). Moreover, we also calculated the overlap between DEGs and DCGs from another two datasets, *i.e.*, the dataset of *G. orontii* infection (GSE5686) and the dataset of *B. cinerea* infection (GSE29642). Using the same analysis procedures performed on GSE56094, we obtained 154 DEGs and 404 DCGs from GSE5686, and 4,817 DEGs and 1,676 DCGs from GSE29642. DCGs from GSE5686 and GSE29642 were both enriched for defense-related genes (hypergeometric test,  $p$ -value <  $2.26 \times 10^{-16}$  for DCGs from GSE5686 and  $p$ -value =  $2.27 \times 10^{-14}$  for DCGs from GSE29642). Despite the high overlap between DCGs and DEGs, we also observed that many DCGs could not be detected by differential expression analysis (Figure S2). These analyses further revealed the general importance of DCGs in plant defenses.

TFs are major players in regulating transcriptional reprogramming during plant immunity<sup>41</sup>. Therefore, we analyzed the relationship among DCGs, DEGs and plant defense-related genes in the context of TFs. In total, we obtained 1,996 TFs from the *Arabidopsis* Gene Regulatory Information Server (AGRIS)<sup>42</sup> and the Plant Transcription Factor Database (PlantTFDB)<sup>43</sup>. As shown in Fig. 2C, DCGs and DEGs were both significantly enriched in defense-related TFs (hypergeometric test,  $p$ -value =  $1.28 \times 10^{-4}$  for TFs from DCGs and  $p$ -value <  $2.20 \times 10^{-16}$  for TFs from DEGs), indicating that TFs from DCGs or DEGs were important in plant immune responses. In total, 53 (40%) DCG-unique TFs were not defined as DEG, 7 of which participate in plant immune responses (Table S2).

The above comparative analyses quantified the relationship between DCGs and DEGs and highlighted their roles in plant immunity. Consistent with a previous study<sup>27</sup>, it is reasonable that some genes not only tend to have different coexpression partners under different conditions but are also differentially expressed. It is worth mentioning that the purpose of such comparative analysis is not to argue which measure is better. Rather, we conclude that differential coexpression analysis is complementary to differential expression analysis, which can provide additional candidates for the study of plant immunity.

**Potential TFs regulate the observed differential coexpression.** Direct identification of TFs involved in plant immunity through differential expression analysis is insufficient, given that TFs also tend to be stably expressed and regulated at the protein level<sup>26,44</sup>. The disruption of TF binding sites can affect the expression correlation between its targets, which is recognized as a potential regulatory mechanism of differential coexpression<sup>16,17,45</sup>. To account for this phenomenon, we explored the underlying regulatory mechanisms of differential coexpression in plant immune responses to *P. syringae* by integrating the known regulatory network data available in public databases.

After removing redundant interactions, we collected 50,824 TF-target binary interactions between 529 TFs and 22,793 target genes from the *Arabidopsis* Transcriptional Regulatory Map (ATRM)<sup>46</sup>, AGRIS<sup>42</sup> and AthaMap<sup>47</sup> (Table S5). In total, 13,879 interactions between 333 TFs and 5,371 targets were retained for further analysis by filtering out stably expressed genes under each condition (variance < 0.2 in both conditions). We investigated the relationship between TFs and their targets to identify TFs that potentially regulate differential coexpression. For this assessment, we implemented a metric for prioritizing TFs that were putative regulators of differential coexpression during plant immune responses. It was based on the hypothesis that a TF of greater importance in regulating differential coexpression should have a greater number of its targets forming differentially coexpressed



**Figure 2. The relationship between differential expression and differential coexpression.** (A) Differential expression and differential coexpression are weakly correlated. Differential expression is measured by the average  $\log_2$ -fold change across different time points. Each node indicates a DCG, and there are 1,315 DCGs in total. The red line shows linear fit, and the reported SCC shows the correlation between the two measures. (B) Overlap among DCGs, DEGs and plant defense-related genes. (C) Overlap among DCGs, DEGs and plant defense-related genes in the context of TFs. TF\_DCGs represent differentially coexpressed TFs, whereas TF\_DEGs denote differentially expressed TFs.

gene pairs<sup>23</sup>. We applied a binomial test (see the Materials and Methods section) to determine whether the number of differentially coexpressed gene pairs between targets of a TF was higher than expected.

Among 333 TFs in the regulatory network we examined, 30 TFs were identified with a significantly greater number of differentially coexpressed gene pairs formed between targets of each TF (Table 1). In fact, two of them were not designed on the microarray platform we analyzed, and 22 of them cannot be identified by traditional differential expression analysis (Figure S3), further demonstrating the power of our method. Moreover, 10 of 30 TFs have been reported to play important roles in plant immune system (Marked with asterisk in Table 1). For example, *ANAC019*, *ANAC055*, and *ANAC072* participate in COR-triggered systemic susceptibility to *P. syringae* pv. *maculicola* ES4326, and triple-mutant plants exhibited enhanced resistance to *P. syringae* ES4326<sup>48</sup>. *MYB51* regulates the biosynthesis of phytoalexin, which is a critical defense metabolite in *Arabidopsis*<sup>49</sup>. In addition, *DREB1A*, *DREB1B* and *DREB1C* (also referred to as *CBF1*, *CBF2* and *CBF3*) from the DREB subfamily A-1 of ERF/AP2 TF family were all detected as TFs regulating differential coexpression. All of these three TFs positively regulate plant tolerance to cold stress<sup>50</sup>, but their roles in plant immunity have not been explored. Considering the potential crosstalk between plant responses to biotic and abiotic stresses<sup>51</sup>, these genes can serve as important candidates for further experimental validation.

For these 20 TFs with unknown roles in plant immunity, we assessed whether any indirect evidence is available indicating that they may be involved in plant immunity. By analyzing their targets, we found that the targets of 9 TFs are significantly enriched with defense-related GO terms (Table S6), suggesting a potential role for those TFs in regulating defense-related gene network. For example, the MADS TF *AGL15* was not differentially expressed between control and infection conditions (Figure S4A). However, the expression correlations between one of its targets (*HYS1*) and the other 87 targets were significantly increased after infection by *P. syringae* (Figure S4B). Interestingly, those targets regulated by *AGL15* are enriched in defense-related functions (Table S3C). The role of *HYS1* in plant immunity has been previously studied<sup>52</sup>, but how it is regulated is not clear. Based on the fact that

Gene <sup>a</sup>	Symbol	Target number	Differential coexpression gene pairs	p-value <sup>b</sup>
AT5G13790	<i>AGL15</i>	1,520	9,150	$<2.20 \times 10^{-16}$
AT5G61850	<i>LFY3</i>	1,499	7,353	$<2.20 \times 10^{-16}$
AT4G16110*	<i>ARR2</i>	370	585	$<2.20 \times 10^{-16}$
AT2G43010	<i>PIF4</i>	726	1,995	$<2.20 \times 10^{-16}$
AT1G24260	<i>SEP3</i>	1,689	11,245	$<2.20 \times 10^{-16}$
AT5G07100	<i>WRKY26</i>	150	133	$1.51 \times 10^{-13}$
AT3G16857	<i>ARR1</i>	222	238	$2.10 \times 10^{-13}$
AT2G46130	<i>WRKY43</i>	132	109	$6.62 \times 10^{-13}$
AT5G22570*	<i>WRKY38</i>	133	109	$1.49 \times 10^{-12}$
AT5G11260	<i>HYS</i>	372	511	$6.38 \times 10^{-8}$
<b>AT4G27410*</b>	<i>ANAC072</i>	8	5	$7.98 \times 10^{-8}$
AT4G25490	<i>DREB1B</i>	16	8	$2.39 \times 10^{-7}$
AT4G25480	<i>DREB1A</i>	40	19	$2.89 \times 10^{-7}$
AT4G36990*	<i>TBF1</i>	1,090	3,705	$4.00 \times 10^{-7}$
AT3G14230*	<i>RAP2.2</i>	278	294	$4.40 \times 10^{-6}$
AT1G13450	<i>GT-1</i>	115	68	$1.12 \times 10^{-5}$
AT2G20180	<i>PIF1</i>	356	434	$4.02 \times 10^{-4}$
AT4G27330	<i>SPL</i>	3	1	$4.02 \times 10^{-4}$
<b>AT1G27730*</b>	<i>ZAT10</i>	5	1	$5.57 \times 10^{-3}$
AT4G25470	<i>DREB1C</i>	18	4	$7.37 \times 10^{-3}$
<b>AT5G05410</b>	<i>DREB2A</i>	32	8	$9.05 \times 10^{-3}$
<b>AT1G18570*</b>	<i>MYB51</i>	6	1	$1.05 \times 10^{-2}$
<b>AT3G15500*</b>	<i>ANAC055</i>	6	1	$1.05 \times 10^{-2}$
AT5G11590	<i>TINY2</i>	20	4	$1.55 \times 10^{-2}$
<b>AT1G52890*</b>	<i>NAC019</i>	7	1	$1.89 \times 10^{-2}$
AT1G69490*	<i>ANAC029</i>	25	5	$2.32 \times 10^{-2}$
AT2G16770	<i>bZIP23</i>	13	2	$2.84 \times 10^{-2}$
AT5G56110	<i>MYB80</i>	129	62	$3.95 \times 10^{-2}$
AT4G35040	<i>bZIP19</i>	14	2	$3.95 \times 10^{-2}$
AT3G24050	<i>GATA1</i>	307	304	$3.95 \times 10^{-2}$

**Table 1. Transcription factors with targets that significantly form differential coexpression gene pairs.**

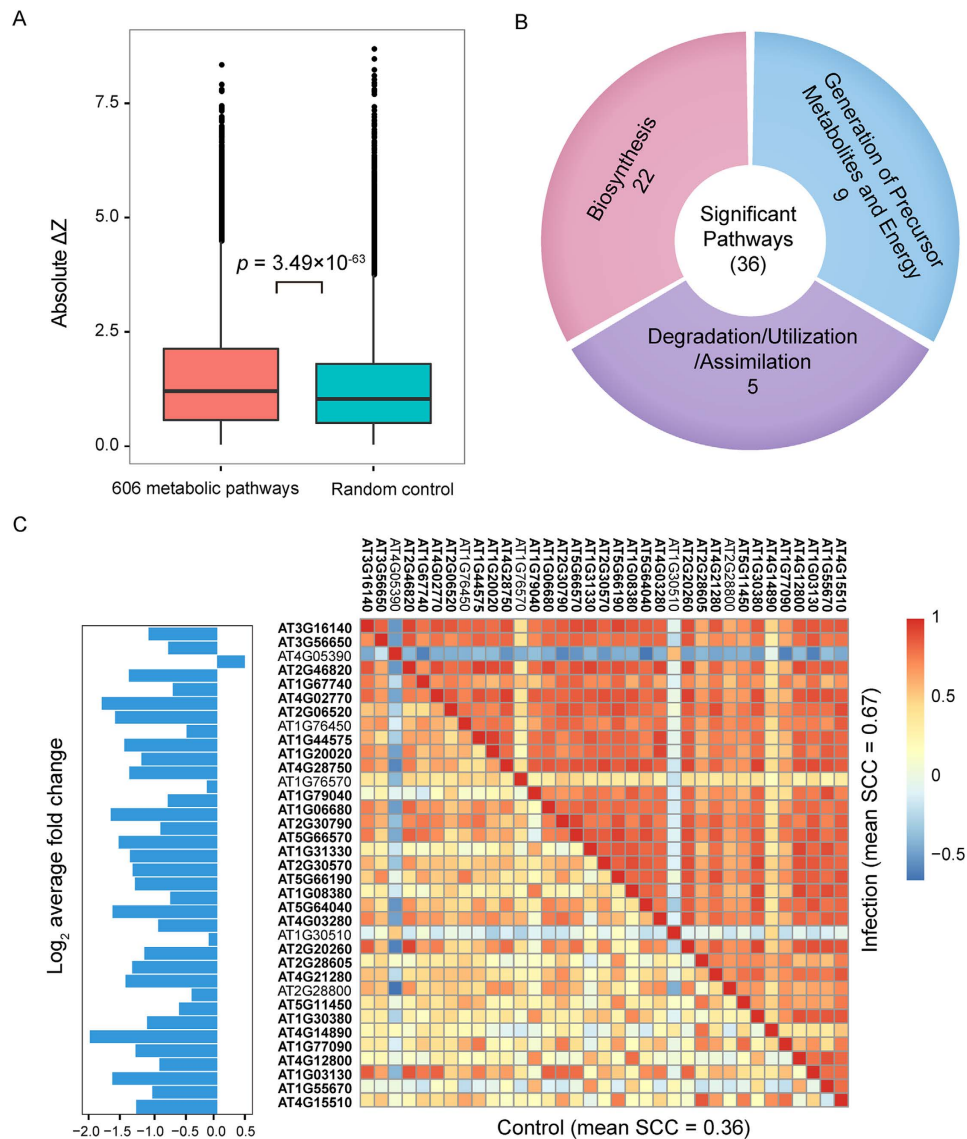
<sup>a</sup>DEGs are in bold. TFs that are vital for *Arabidopsis* defense are marked with an asterisk. <sup>b</sup>p-values are corrected using the Benjamini-Hochberg correction.

the promoter of *HYS1* can bind with *AGL15*<sup>42</sup> and the increased coexpression between *HYS1* and other targets of *AGL15*, it is reasonable to assume *AGL15* may play a role when *Arabidopsis* is infected by *P. syringae*.

### Extensive gene coexpression rewiring in metabolic pathways during plant immune responses.

From the GO annotation of DCGs, we noticed that metabolic processes were significantly overrepresented (Table S3A). Plant metabolism constitutes integral part of the plant immune system. The primary metabolic pathways, such as photosynthesis, photorespiration and TCA cycle, provide the energy required for the defense responses<sup>53,54</sup>. The indispensable contribution of secondary metabolites to plant immunity, such as phytoalexin and various phenolic compounds, has also been well established<sup>55-57</sup>. Considering the important role of metabolites in plant immunity, we wonder how the coexpression pattern changes within metabolic pathways during plant immune responses to *P. syringae*. For this purpose, we first collected 606 metabolic pathways from the AraCyc database (V13.0) deposited in the Plant Metabolic Network (PMN) project<sup>58</sup>. Then, we analyzed the difference in correlation between infection and control for all possible gene pairs from a common pathway. We found that the absolute differences in correlation of metabolic gene pairs were larger for a pathway compared with a group of randomly selected genes (Fig. 3A, Wilcoxon test, p-value =  $3.49 \times 10^{-63}$ ). Based on this finding, we hypothesized that the coordination of metabolic genes in the same metabolic pathway was widely influenced during plant immune responses to *P. syringae*.

Then, we were interested in which metabolic pathway significantly changes coexpression patterns between its genes during plant immune responses to *P. syringae*. For this purpose, we assigned a pathway score to each pathway by measuring the difference in coexpression between control and infection conditions (see the Materials and Methods section). A larger pathway score indicated a larger difference in coexpression for the pathway. We used the concept of dysregulated pathways<sup>59</sup> to represent those pathways. In total, 36 dysregulated pathways were identified (Table 2) and divided into three categories according their super pathways in PMN, including biosynthesis, degradation/utilization/assimilation, and generation of precursor metabolites and energy (Fig. 3B). As a comparison, the Gene Set Enrichment Analysis (GSEA) method<sup>60</sup> was used to identify metabolic pathways with



**Figure 3. Differential coexpression in metabolic pathways.** (A) The distribution of difference in correlation of 58,158 intra-pathway gene pairs (red) and 58,158 random gene pairs (green). (B) The categories of dysregulated pathways with significantly changed correlations between control and infection groups. (C) Photosynthesis light reactions pathway (PWY-101). The mean correlation of gene pairs in photosynthesis light reactions was 0.36 in the control group and increased to 0.67 after pathogen infection. Average  $\log_2$ -fold change values of the 35 metabolic genes are presented in the bar plot on the left. The 30 DEGs detected using the R package maSigPro are indicated in bold.

significantly changed expression levels, and 80 metabolic pathways were identified (nominal GSEA  $p$ -value  $< 0.05$ , Table S7). GSEA is a statistical method used to determine whether predefined gene sets (*i.e.*, metabolic pathway in our work) are differentially expressed between two conditions. Compared with GSEA, our method aims to detect dysregulated pathways with changed expression correlation rather than changed expression levels. For 36 dysregulated metabolic pathways detected using our method, only 16 pathways were revealed by GSEA. These results suggest that differential coexpression can provide new predictions for the identification of important pathways involved in plant immunity compared with the method only considering differential expression.

We further explored the relationship between dysregulated pathways and differentially expressed pathways. Regarding the 16 pathways detected as significant by both of our method and GSEA, it is possible that the coordinated changes in expression of genes in the pathway resulted in the altered gene coexpression. For example, the metabolic pathway “photosynthesis light reactions” (Pathway ID: PWY-101) was identified as differentially expressed as well as dysregulated (*i.e.*, differentially coexpressed). The average correlation among genes within this pathway increased from 0.36 to 0.67 after pathogen infection, whereas most genes in this pathway (30 out of 35) were significantly down-regulated (Fig. 3C). Increasing evidences showed that photosynthesis is inhibited during immune responses<sup>61,62</sup>. Not surprisingly, the original study reported the down-regulation of genes

Pathway ID <sup>a</sup>	Pathway Name	#Gene	#DEG <sup>b</sup>	P_score <sup>c</sup>	P-value <sup>d</sup>
SUCSYN-PWY*	Sucrose biosynthesis I	45	20	1.79	$<1.00 \times 10^{-03}$
PWY-5046	2-Oxoisovalerate decarboxylation to isobutanoyl-CoA	10	2	2.34	$<1.00 \times 10^{-03}$
CITRULBIO-PWY	Citrulline biosynthesis	26	7	2.02	$<1.00 \times 10^{-03}$
PWY-101*	Photosynthesis light reactions	35	30	2.60	$<1.00 \times 10^{-03}$
THRESYN-PWY	Threonine biosynthesis	19	4	2.00	$<1.00 \times 10^{-03}$
PWY66-21	Ethanol degradation II	22	9	1.95	$<1.00 \times 10^{-03}$
GLYCLEAV-PWY*	Glycine cleavage	10	5	2.36	$<1.00 \times 10^{-03}$
PWY-7328	Super pathway of UDP-glucose-derived O-antigen building blocks biosynthesis	28	5	1.99	$<1.00 \times 10^{-03}$
PWY-5464*	Super pathway of cytosolic glycolysis, pyruvate dehydrogenase and TCA cycle	110	31	1.56	$<1.00 \times 10^{-03}$
PWY-622*	Starch biosynthesis	27	12	1.90	$<1.00 \times 10^{-03}$
CALVIN-PWY*	Calvin-Benson-Bassham cycle	37	25	1.87	$<1.00 \times 10^{-03}$
COLANSYN-PWY	Colanic acid building blocks biosynthesis	38	7	1.93	$<1.00 \times 10^{-03}$
PHOTOALL-PWY*	Oxygenic photosynthesis	72	55	2.14	$<1.00 \times 10^{-03}$
PWY-5004	Super pathway of citrulline metabolism	31	7	1.98	$<1.00 \times 10^{-03}$
PWYQT-4470	$\gamma$ -Glutamyl cycle	3	0	4.84	$<1.00 \times 10^{-03}$
NONOXIPENT-PWY	Pentose phosphate pathway	12	6	2.43	$<1.00 \times 10^{-03}$
GLUTATHIONESYN-PWY	Glutathione biosynthesis	2	0	8.30	$<1.00 \times 10^{-03}$
PWY-5723*	Rubisco shunt	35	17	1.80	$<1.00 \times 10^{-03}$
CAROTENOID-PWY*	Super pathway of carotenoid biosynthesis	20	11	2.03	$2.04 \times 10^{-02}$
PWY-5805	Nonaprenyl diphosphate biosynthesis I	2	2	6.63	$2.04 \times 10^{-02}$
PWY-4302	Aerobic respiration III	56	4	1.63	$2.04 \times 10^{-02}$
PWY-6415	L-ascorbate biosynthesis V	23	5	1.99	$2.04 \times 10^{-02}$
PWY-2501	Fatty acid $\alpha$ -oxidation I	10	6	2.45	$2.04 \times 10^{-02}$
HOMOSERSYN-PWY	Homoserine biosynthesis	7	1	2.72	$2.04 \times 10^{-02}$
TYRFUMCAT-PWY*	Tyrosine degradation I	8	2	2.66	$2.04 \times 10^{-02}$
HEXPPSYN-PWY	Hexaprenyl diphosphate biosynthesis	2	2	6.63	$2.04 \times 10^{-02}$
COMPLETE-ARO-ARA-PWY*	Super pathway of phenylalanine, tyrosine and tryptophan biosynthesis	26	10	1.94	$2.04 \times 10^{-02}$
PWY-5114*	UDP-sugars interconversion	29	5	1.78	$3.24 \times 10^{-02}$
PWY-1422	Vitamin E biosynthesis	6	3	2.72	$3.24 \times 10^{-02}$
PWY-6444	Benzoate biosynthesis II	12	7	2.10	$3.24 \times 10^{-02}$
PWY-601*	Glucosinolate biosynthesis from tryptophan	25	7	1.78	$3.24 \times 10^{-02}$
PWYQT-4450*	Aliphatic glucosinolate biosynthesis, side chain elongation cycle	7	4	2.53	$3.24 \times 10^{-02}$
PWY-5783	Octaprenyl diphosphate biosynthesis	2	2	6.63	$3.24 \times 10^{-02}$
GLUTAMINDEG-PWY	Glutamine degradation I	18	5	1.99	$3.24 \times 10^{-02}$
TRPSYN-PWY*	Tryptophan biosynthesis	14	7	2.06	$4.59 \times 10^{-02}$
PWY-181*	Photorespiration	26	12	1.72	$4.59 \times 10^{-02}$

**Table 2. Dysregulated pathways with significantly changed expression correlations.** <sup>a</sup>The pathways detected as significant through the GSEA analysis are marked with asterisk. <sup>b</sup>DEGs were detected using the R package maSigPro. <sup>c</sup>P\_scores are calculated using Eq. 4. <sup>d</sup>p-values were corrected using the Benjamini-Hochberg correction.

located in chloroplasts during plant responses to *P. syringae*<sup>30</sup>. Therefore, genes in PWY-101 were coordinately down-regulated during plant immune responses (Figure S5).

For the 20 pathways specifically identified by our method (not differentially expressed between two conditions), we also observed some plant defense-related pathways. For example, the glutathione biosynthesis pathway (Pathway ID: GLUTATHIONESYN-PWY) was detected as a dysregulated pathway by our method. There were two enzymes in GLUTATHIONESYN-PWY, and the coexpression between them was increased from  $-0.41$  (control) to  $0.85$  (infection) (Figure S6). Glutathione participates in detoxification and signaling in plant defense against biotic and abiotic stresses<sup>63</sup>. The abundance of glutathione increases in the plants infected with avirulent pathogens or insect *Mayetiola destructor*<sup>64</sup>. The increased correlation of GLUTATHIONESYN-PWY under pathogen infection may favor the production of glutathione, which contributes to plant immune responses to *P. syringae*.

Pathways with roles in plant immunity that have not been reported may serve as good candidates for further analysis. Our expression correlation-based method found that aerobic respiration III (Pathway ID: PWY-4302) was dysregulated during plant immune responses to *P. syringae*, but the GSEA analysis recognized it as



non-differentially expressed. Within this pathway, a majority of genes (93%) displayed a stable expression pattern in control and infection conditions. Regarding the expression correlation between genes in this pathway, we observed large changes in correlation between a subset of genes in this pathway (Figure S7). Although genes in the subset exhibited minimal changes in gene expression, the expression correlation between these genes significantly changed. In general, the generation of energy and release of ROS during aerobic respiration may indicate the potential role of this pathway in plant immunity<sup>65</sup>. Further validation is needed to decipher how these coexpression changes influence plant immune responses to *P. syringae*.

**Comparison of current work with previous studies.** Our current work is primarily based on the dataset of GSE56094, and the resultant new findings have been clearly described in the aforementioned sections. Here, we focus on discussing the technical differences between our work and the original studies based on the same microarray data<sup>30–32</sup>. In these original studies, de Torres Zabala *et al.* focused on the expression of nuclear-encoded chloroplast-targeted genes<sup>30</sup>. The group further analyzed the expression pattern of genes related to jasmonate signaling<sup>31</sup>. Lewis *et al.* aimed to determine the difference between transcriptional reprogramming associated with microbial-associated molecular pattern-triggered immunity and effector-triggered susceptibility<sup>32</sup>. In contrast to these original studies, in this work GSE56094 was used to conduct differential coexpression analysis. A series of computational analyses were performed in our work, including DCG analysis, metabolic pathway analysis and transcriptional regulation analysis.

As reported in Lewis *et al.*, three techniques were utilized to identify DEGs<sup>32</sup>. We also compared the list of 9,782 DEGs between mock-infiltrated leaves and DC3000-infiltrated leaves identified by the original study of Lewis *et al.*<sup>32</sup> with 1,315 DCGs detected in this work. We found that 238 out of 1,315 DCGs were not included in the DEG list of Lewis *et al.* (2015) (Table S2), further indicating the complementarity between DEGs and DCGs. Regarding the module/pathway analysis, Lewis *et al.* used coexpression-based clustering method to identify gene clusters coexpressed within each treatment (mock treatment, infection by virulent *P. syringae* or avirulent *P. syringae* with *hrpA* mutant)<sup>32</sup>. In our analysis, we first demonstrated extensive gene coexpression rewiring in metabolic pathways during plant immune responses and then identified 36 dysregulated metabolic pathways with significant changes in expression correlation between normal growth and pathogen infection (Table 2). This finding differs from the coexpression-based clustering analysis performed by Lewis *et al.* With respect to the coregulation analysis, Lewis *et al.* aimed to predict the specific regulation of pathogen-response genes, whereas we sought to identify potential TFs regulating the observed differential expression during plant immune responses. In summary, our current analyses exhibit technical complementarity to the three original studies based on the same dataset GSE56094 and thus provide new insights into our understanding of plant immunity.

## Conclusions

Plants are challenged by various pathogens during all phases of their development. To effectively defend against these infections, plants have evolved a powerful immune system, and extensive transcriptional reprogramming exists during this process. In this work, we initiated a study to investigate *Arabidopsis* immune responses to the bacterium *P. syringae* through differential coexpression analysis. We concluded that differential coexpression was a common phenomenon during plant immunity. Moreover, we found that DCGs, which are complementary to DEGs, also played important roles in plant immune responses. By integrating regulatory networks into our analysis, we identified several TFs that may regulate this differential coexpression during plant immune responses. We noticed extensive differential coexpression of metabolic genes during plant immunity and identified 36 dysregulated metabolic pathways. In the future, the transcriptional data measuring plant responses to pathogen infection will be increasingly available, allowing us to perform the current analysis using sizable expression data and achieve more robust results. In the meantime, the flourish of transcriptional data will also allow us to answer some specific biological questions in the context of differential coexpression. For instance, comparative analysis of differential coexpression during plant immune responses to different pathogens should be an important topic. It is our expectation that the differential coexpression analysis can boost the study of plant immune response-related transcriptomics and provide new insights into deciphering the molecular mechanisms of plant-pathogen interactions.

## Materials and Methods

**Data collection.** Normalized expression dataset (GEO accession number: GSE56094) were directly downloaded from the GEO database<sup>33</sup>. GSE56094 was designed to measure *Arabidopsis* gene expression at a high resolution under three conditions: mock treatment, or infections with either *P. syringae* pv. *tomato* DC3000 or the corresponding nonpathogenic *hrpA* mutant<sup>30–32</sup>. The detailed experimental procedures of the pathogen treatments and microarray data processing were documented by Lewis *et al.*<sup>32</sup>. Probe sets were mapped to their corresponding AGI (*Arabidopsis* Genome Initiative) gene identifiers according to the annotation file from GEO, and replicated probes of the same gene were averaged. After data processing, we obtained 23,974 unique genes. *Arabidopsis* metabolic pathways were obtained from the AraCyc database<sup>58</sup> at PMN (<http://www.plantcyc.org/>). Genes in each pathway were filtered using expression data, and only pathways with genes detected on the microarray were retained for analysis. *Arabidopsis* transcriptional regulatory networks were collected from ATRM<sup>46</sup>, AGRIS<sup>42</sup> and AthaMap<sup>47</sup>, and TFs were collected from PlantTFDB<sup>43</sup> and AGRIS. For predicted regulatory networks from AthaMap, only TFs with binding sites determined by pattern-based screenings were considered in our work.

Plant defense-related genes were collected through analyzing annotation information from the TAIR GO.Slim files and the literature<sup>41</sup>. We first downloaded the GO.Slim files from the FTP site of TAIR (<ftp://ftp.arabidopsis.org/home/tair/Ontologies/>). Then, for each record in the annotation file, if the description of a gene contained

at least one of the following biological process keywords, *i.e.*, “systemic acquired resistance”, “disease resistance”, and “immune” and “defense”, we selected the gene as a plant defense-related gene. Moreover, Tsuda and Somssich recently provided a list of TFs associated with plant immunity in their review article<sup>41</sup>. These TFs were also assigned as defense-related genes. Furthermore, collected defense-related genes were filtered using expression data, and only genes detected on the microarray were reserved.

**Detection of differentially coexpressed gene pairs.** Differentially coexpressed gene pairs were detected using the R package DiffCorr<sup>22</sup>, and the workflow is summarized as follows (Fig. 1A). First, the obtained microarray data were divided into infection and control groups. Prior to the calculation of the coexpression correlation coefficient, we removed genes with expression variance less than 0.2 across both infection samples and control samples. Then, for each group, SCCs and Benjamini-Hochberg<sup>66</sup> adjusted *p*-values for all possible gene pairs were calculated. Here, a gene pair was defined as significantly coexpressed if its corrected *p*-value was less than a given cutoff of  $10^{-8}$ . Next, the difference in correlation of a gene pair (*i, j*) between infection and control conditions was evaluated by the following equation:

$$\Delta Z_{i,j} = \frac{Z_I - Z_C}{\sqrt{\frac{1}{N_I - 3} + \frac{1}{N_C - 3}}} \quad (1)$$

where  $N_I$  and  $N_C$  was the sample number in the infection and control groups, respectively;  $Z_I$  and  $Z_C$  were Fisher's transformation of SCC between the gene pair (*i, j*) under infection and control conditions, respectively, and had been used in previous differential coexpression analysis<sup>22,27</sup>. The  $\Delta Z$  followed a Gaussian distribution<sup>22</sup>, and the associated *p*-value was evaluated based on the distribution and corrected using the Benjamini-Hochberg correction. If the corrected *p*-value is less than  $10^{-8}$ , the difference in correlation was regarded as significant. Finally, we defined a pair of genes as differentially coexpressed if they exhibited both a significant difference in correlation and a significant correlation under at least one of the two conditions.

**Identification of DCGs.** In our work, DCGs were defined as genes whose expression correlations with many other genes were significantly changed between infection and control conditions. The binomial probability model was used to identify DCGs<sup>67</sup>. Given that the probability of observing a “differential coexpression gene pair” was *p* ( $p = 5.41 \times 10^{-3}$  in this work), the *p*-value of differential coexpression for the gene *i* with  $k_i$  differentially coexpressed gene pairs was evaluated as follows:

$$p_{DCG}^i = \sum_{x=k_i}^n C_n^x (p)^x (1-p)^{n-x} \quad (2)$$

where *n* was the number of genes used to identify differentially coexpressed gene pairs. Then,  $p_{DCG}^i$  was corrected using the Benjamini-Hochberg correction for multiple testing.

**Identification of DEGs.** The Bioconductor package maSigPro was used to identify DEGs<sup>40</sup>. The obtained 104 samples of 52 infection samples and 52 control samples were used as input. The maSigPro method follows a two-step regression strategy to identify DEGs from time-course microarray data. Regarding the parameters of maSigPro, Q value and “alfa” were set as 0.01, and a default  $R^2$  threshold of 0.6 was adopted. After MaSigPro analysis, a cutoff of  $\text{abs}(\log_2\text{-fold change}) \geq 0.58$  (*i.e.*, at least 1.5-fold change) was applied to designate genes as differentially expressed.

**GO enrichment analysis.** GO enrichment analysis was performed using BiNGO 3.02 in Cytoscape with the “GO Biological Process” category<sup>68</sup>. Using the whole annotation as the reference set, over-represented terms were selected with a significance level of 0.05 (hypergeometric test) after Benjamini-Hochberg correction.

**Statistical test for TFs regulating differential coexpression.** Our analyses were based on the assumption that if the targets of a TF tend to form differentially coexpressed gene pairs, the TF must be more important in regulating differential coexpression<sup>23</sup>. Given that the probability of observing a “differential coexpression gene pair” was *p* ( $p = 5.41 \times 10^{-3}$  in this work), the probability of forming at least *e* differential coexpression gene pairs between TF targets by chance was:

$$P = \sum_{k=e}^N C_N^k p^k (1-p)^{N-k} \quad (3)$$

If a TF had *n* targets, *N* was equal to  $n \times (n-1)/2$ . A correction for multiple testing using Benjamini-Hochberg method was then applied to the vector of *P*-values for all TFs.

**Calculation of pathway score for metabolic pathways.** To identify dysregulated pathway, we first calculated average difference in expression correlation for all possible gene pairs within the given pathway. The pathways associated with only one gene were discarded from this analysis. For a pathway *i* with *m* genes, there were a total  $m \times (m-1)/2$  possible gene pairs in this pathway, and we defined this gene pair set as  $G_i$ . The pathway score was calculated as follows:

$$P\_score_i = \frac{2}{m \times (m - 1)} \sum_{(k,l) \in G_i} |\Delta Z_{k,l}| \quad (4)$$

where  $\Delta Z_{k,l}$  was calculated using Eq. 1 representing difference in correlation for the gene pair  $(k, l)$  from  $G_i$ .

Then, we generated a random pathway score distribution for each pathway to evaluate the significance of the pathway score. For a pathway with  $m$  genes, we randomly selected  $m$  genes from the microarray and calculated a pathway score using Eq. 4. We repeated the process 1,000 times to form a random pathway score distribution. An empirical  $p$ -value for the pathway was calculated as the fraction of permutation values that were greater than or equal to the observed value and corrected for multiple testing using the Benjamini-Hochberg correction. Finally, a pathway was defined as a dysregulated pathway in plant immune responses to *P. syringae* if its corrected  $p$ -value was less than a given cutoff.

GSEA was also used to identify metabolic pathways with significantly changed expression levels and was performed using the GSEA software<sup>60</sup>. The minimum and maximum gene set sizes were set to 5 and 1,000, respectively. The metric for ranking genes was set to “Diff\_of\_Classes”. 1000 permutations were used to calculate  $p$ -value and the permutation type was set to “Gene\_set”. All other parameters were set as default. A cutoff of 0.05 was used to define statistical significance.

## References

- Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
- Tsuda, K. & Katagiri, F. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr. Opin. Plant Biol.* **13**, 459–465 (2010).
- Moore, J. W., Loake, G. J. & Spoel, S. H. Transcription dynamics in plant immunity. *Plant Cell* **23**, 2809–2820 (2011).
- Rachlin, J., Cohen, D. D., Cantor, C. & Kasif, S. Biological context networks: a mosaic view of the interactome. *Mol. Syst. Biol.* **2**, 66–66 (2006).
- He, F., Zhou, Y. & Zhang, Z. Deciphering the *Arabidopsis* floral transition process by integrating a protein-protein interaction network and gene expression data. *Plant Physiol.* **153**, 1492–1505 (2010).
- He, F., Karve, A. A., Maslov, S. & Babst, B. A. Large-scale public transcriptomic data mining reveals a tight connection between the transport of nitrogen and other transport processes in *Arabidopsis*. *Front Plant Sci* **7**, 1207 (2016).
- Barah, P., Winge, P., Kusnierczyk, A., Tran, D. H. & Bones, A. M. Molecular signatures in *Arabidopsis thaliana* in response to insect attack and bacterial infection. *PLoS One* **8**, e58987 (2013).
- Amrine, K. C. H., Blanco-Ulate, B. & Cantu, D. Discovery of core biotic stress responsive genes in *Arabidopsis* by weighted gene co-expression network analysis. *PLoS One* **10**, e0118731 (2015).
- Leal, L. G., López, C. & López-Kleine, L. Construction and comparison of gene co-expression networks shows complex plant immune responses. *PeerJ* **2**, e610 (2014).
- Tully, J. *et al.* Expression-based network biology identifies immune-related functional modules involved in plant defense. *BMC Genomics* **15**, 421 (2014).
- Dong, X., Jiang, Z., Peng, Y.-L. & Zhang, Z. Revealing shared and distinct gene network organization in *Arabidopsis* immune responses by integrative analysis. *Plant Physiol.* **167**, 1158–1185 (2015).
- Jiang, Z., Dong, X. & Zhang, Z. Network-based comparative analysis of *Arabidopsis* immune responses to *Golovinomyces orontii* and *Botrytis cinerea* infections. *Scientific Reports* **6**, 19149 (2016).
- Windram, O., Penfold, C. A. & Denby, K. J. Network modeling to understand plant immunity. *Annu. Rev. Phytopathol.* **52**, 93–111 (2014).
- Mine, A., Sato, M. & Tsuda, K. Towards a systems understanding of plant-microbe interactions. *Front Plant Sci* **5**, 423 (2014).
- Li, H. & Zhang, Z. Systems understanding of plant-pathogen interactions through genome-wide protein-protein interaction networks. *Front. Agr. Sci. Eng.* **3**, 102–112 (2016).
- de la Fuente, A. From ‘differential expression’ to ‘differential networking’ – identification of dysfunctional regulatory networks in diseases. *Trends Genet.* **26**, 326–333 (2010).
- Gaiteri, C., Ding, Y., French, B., Tseng, G. C. & Sibille, E. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes, Brain and Behav.* **13**, 13–24 (2014).
- Hsu, C.-L., Juan, H.-F. & Huang, H.-C. Functional analysis and characterization of differential coexpression networks. *Scientific Reports* **5**, 13295 (2015).
- Li, K.-C. Genome-wide coexpression dynamics: Theory and application. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16875–16880 (2002).
- Amar, D., Safer, H. & Shamir, R. Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput Biol* **9**, e1002955 (2013).
- Tesson, B. M., Breitling, R. & Jansen, R. C. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics* **11**, 497–497 (2010).
- Fukushima, A. DiffCorr: An R package to analyze and visualize differential correlations in biological networks. *Gene* **518**, 209–214 (2013).
- Yang, J. *et al.* DCGL v2.0: An R package for unveiling differential regulation from differential co-expression. *PLoS One* **8**, e79729 (2013).
- Fukushima, A. *et al.* Exploring tomato gene functions based on coexpression modules using graph clustering and differential coexpression approaches. *Plant Physiol.* **158**, 1487–1502 (2012).
- Choi, J. K., Yu, U., Yoo, O. J. & Kim, S. Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics* **21**, 4348–4355 (2005).
- Hudson, N. J., Reverter, A. & Dalrymple, B. P. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput Biol* **5**, e1000382 (2009).
- Reznik, E. & Sander, C. Extensive decoupling of metabolic genes in cancer. *PLoS Comput Biol* **11**, e1004176 (2015).
- He, F. *et al.* Large-scale atlas of microarray data reveals the distinct expression landscape of different tissues in *Arabidopsis*. *Plant J* **86**, 472–480 (2016).
- Katagiri, F., Thilmony, R. & He, S. Y. The *Arabidopsis Thaliana*-*Pseudomonas Syringae* Interaction. *The Arabidopsis Book/American Society of Plant Biologists* **1**, e0039 (2002).
- de Torres Zabala, M. *et al.* Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. *Nature Plants* **1**, 15074 (2015).
- de Torres Zabala, M. *et al.* Novel JAZ co-operativity and unexpected JA dynamics underpin *Arabidopsis* defence responses to *Pseudomonas syringae* infection. *New Phytol.* **209**, 1120–1134 (2016).

32. Lewis, L. A. *et al.* Transcriptional dynamics driving MAMP-Triggered immunity and pathogen effector-mediated immunosuppression in *Arabidopsis* leaves following infection with *Pseudomonas syringae* pv *tomato* DC3000. *Plant Cell* **27**, 3038–3064 (2015).
33. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
34. Robert-Seilaniantz, A. *et al.* The microRNA miR393 re-directs secondary metabolite biosynthesis away from camalexin and towards glucosinolates. *Plant J* **67**, 218–231 (2011).
35. Windram, O. *et al.* *Arabidopsis* defense against *Botrytis cinerea*: chronology and regulation deciphered by high-resolution temporal transcriptomic analysis. *Plant Cell* **24**, 3530–3557 (2012).
36. Consortium, G. O. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).
37. Gómez-Gómez, L. & Boller, T. FLS2: An LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Mol. Cell* **5**, 1003–1011 (2000).
38. González-Lamothe, R. *et al.* The U-Box protein CMPG1 Is required for efficient activation of defense mechanisms triggered by multiple resistance genes in Tobacco and Tomato. *Plant Cell* **18**, 1067–1083 (2006).
39. Samuel, M. A., Salt, J. N., Shiu, S. H. & Goring, D. R. Multifunctional ARM repeat domains in plants. *Int. Rev. Cytol.* **253**, 1–26 (2006).
40. Conesa, A., Nueda, M. J., Ferrer, A. & Talón, M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* **22**, 1096–1102 (2006).
41. Tsuda, K. & Somssich, I. E. Transcriptional networks in plant immunity. *New Phytol.* **206**, 932–947 (2015).
42. Yilmaz, A. *et al.* AGRIS: the *Arabidopsis* gene regulatory information server, an update. *Nucleic Acids Res.* **39**, D1118–D1122 (2011).
43. Jin, J., Zhang, H., Kong, L., Gao, G. & Luo, J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* **42**, D1182–D1187 (2014).
44. Zampieri, M., Soranzo, N. & Altafini, C. Discerning static and causal interactions in genome-wide reverse engineering problems. *Bioinformatics* **24**, 1510–1515 (2008).
45. Xu, F. *et al.* Differential co-expression and regulation analyses reveal different mechanisms underlying major depressive disorder and syndromal symptomatic depression. *BMC Bioinformatics* **16**, 112 (2015).
46. Jin, J. *et al.* An *Arabidopsis* transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. *Mol. Biol. Evol.* **32**, 1767–1773 (2015).
47. Bülow, L., Brill, Y. & Hehl, R. AthaMap-assisted transcription factor target gene identification in *Arabidopsis thaliana*. *Database* **2010**, baq034 (2010).
48. Zheng, X.-y. *et al.* Coronatine promotes *Pseudomonas syringae* virulence in plants by activating a signaling cascade that inhibits salicylic acid accumulation. *Cell Host & Microbe* **11**, 587–596 (2012).
49. Frerigmann, H., Glawischnig, E. & Gígolashvili, T. The role of MYB34, MYB51 and MYB122 in the regulation of camalexin biosynthesis in *Arabidopsis thaliana*. *Front Plant Sci* **6**, 654 (2015).
50. Chinnusamy, V., Zhu, J.-K. & Sunkar, R. Gene regulation during cold stress acclimation in plants. *Methods in Molecular Biology* **639**, 39–55 (2010).
51. Fujita, M. *et al.* Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr. Opin. Plant Biol.* **9**, 436–442 (2006).
52. Yoshida, S., Ito, M., Nishida, I. & Watanabe, A. Identification of a novel gene HYS1/CPR5 that has a repressive role in the induction of leaf senescence and pathogen-defence responses in *Arabidopsis thaliana*. *Plant J* **29**, 427–437 (2002).
53. Kangasjärvi, S., Neukermans, J., Li, S., Aro, E.-M. & Noctor, G. Photosynthesis, photorespiration, and light signalling in defence responses. *J. Exp. Bot.* **63**, 1619–1636 (2012).
54. Bolton, M. D. Primary metabolism and plant defense—fuel for the fire. *Mol. Plant-Microbe Interact.* **22**, 487–497 (2009).
55. Ahuja, I., Kissen, R. & Bones, A. M. Phytoalexins in defense against pathogens. *Trends Plant Sci.* **17**, 73–90 (2012).
56. Lv, Q., Cheng, R. & Shi, T. Regulatory network rewiring for secondary metabolism in *Arabidopsis thaliana* under various conditions. *BMC Plant Biol.* **14**, 180 (2014).
57. Piasecka, A., Jedrzejczak-Rey, N. & Bednarek, P. Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytol.* **206**, 948–964 (2015).
58. Mueller, L. A., Zhang, P. & Rhee, S. Y. AraCyc: A biochemical pathway database for *Arabidopsis*. *Plant Physiol.* **132**, 453–460 (2003).
59. Tegge, A. N., Caldwell, C. W. & Xu, D. Pathway correlation profile of gene-gene co-expression for identifying pathway perturbation. *PLoS One* **7**, e52127 (2012).
60. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545–15550 (2005).
61. Bilgin, D. D. *et al.* Biotic stress globally downregulates photosynthesis genes. *Plant, Cell Environ.* **33**, 1597–1613 (2010).
62. Kangasjärvi, S., Tikkanen, M., Durian, G. & Aro, E.-M. Photosynthetic light reactions – An adjustable hub in basic production and plant immunity signaling. *Plant Physiol. Biochem.* **81**, 128–134 (2014).
63. Ghanta, S. & Chattopadhyay, S. Glutathione as a signaling molecule: Another challenge to pathogens. *Plant Signaling & Behavior* **6**, 783–788 (2011).
64. Liu, X., Zhang, S., Whitworth, R. J., Stuart, J. J. & Chen, M.-S. Unbalanced activation of glutathione metabolic pathways suggests potential involvement in plant defense against the gall midge *Mayetiola destructor* in wheat. *Scientific Reports* **5**, 8092 (2015).
65. Tripathy, B. C. & Oelmüller, R. Reactive oxygen species generation and signaling in plants. *Plant Signaling & Behavior* **7**, 1621–1633 (2012).
66. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).
67. Yu, H. *et al.* Link-based quantitative methods to identify differentially coexpressed genes and gene Pairs. *BMC Bioinformatics* **12**, 315 (2011).
68. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).

## Acknowledgements

We thank Drs. Saijun Tang, Xiangfeng Wang, and Yuan Zhou at China Agricultural University for their helpful discussions on this topic. This work was supported by grants from the National Natural Science Foundation of China (31271414, 31471249).

## Author Contributions

Z.J. and Z.Z. conceived and designed the study. Z.J. collected data and performed the research. Z.J., Z.-G.L. and X.D. analyzed the data. Z.J., F.H. and Z.Z. wrote the manuscript. All authors approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

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

**How to cite this article:** Jiang, Z. *et al.* Differential Coexpression Analysis Reveals Extensive Rewiring of *Arabidopsis* Gene Coexpression in Response to *Pseudomonas syringae* Infection. *Sci. Rep.* **6**, 35064; doi: 10.1038/srep35064 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016