0.0.1 Comparison with other studies for discovery of GSL and FB pathway genes in literature The GSL and flavonoid pathways have been subjected gene co-expression analysis by other groups (Gachon *et al*., 2005; Hirai *et al*., 2007; Saito *et al*., 2008; Yonekura-Sakakibara *et al*., 2008). Using pearson correlation coefficients between combinations of the 22,263 genes, from 1,388 public microarray datasets, Hirai *et al*. (2007) constructed co-expression relationships (correlation coefficient $>$0.65) of *Myb28* and *Myb29* with other genes. In our work these two genes were used in the *seed-gene-set II* to study the GSL pathway (Supplemental Figure 2). Comparing Hirai *et al*.'s. (2007) results (Figure 1 in Hirai *et al*. (2007)) with our work in Supplemental Figure 2, we find 11 overlapped pathway genes. Furthermore, we discovered more pathway genes (e.g., *PMSR2*, *GSTF11* and *GS-OX1-5*, Figure 3) not listed in their Figure 1.

In another study using 54 pre-selected guide genes (encoding 13 transcription factors and 41 enzymes involved in flavonoid and phenylpropanoid pathways), Saito *et al*. (2008) constructed the co-expression networks of a general phenylpropanoid pathway using the analytical tool Correlated genes, resulting in the formation of four modules, representing flavonoid, anthocyanidin, proanthocyanidin and lignin respectively (their Figure 3b). The flavonoid module consisted of sixteen genes (their Figure 3b). Comparing our findings from top 20 genes in Supplemental Table 4 with the 16 genes in their flavonoid module, seven pathway (including *FLS*, which is one of the four seed-genes in *seed-gene-set IV*) genes are commonly detected. As an aside, the three seed genes in our work encoding for TFs (*TT8*, *TT16* and *TTG2*) are not present in their flavonoid module, but instead are found in their proanthocyanidin module (their Figure 3b), which is downstream of flavonoid pathway (Figure 3). However, under stress conditions, from which our datasets are derived, their results suggested that these three TFs could network with the flavonoid network module (their Figure 3 (b), and the broken lines), validated our findings.

0.0.2 Data preprocessing using the RMA normalization We pre-processed the array data from different experiments using the RMA (Robust Multiarray Average) normalization method (Irizarry *et al*., 2003a; Bolstad *et al*., 2003; Irizarry *et al*., 2003b) which is available from Bioconductor website. This widely used normalization method consists of three steps: background adjustment, quantile normalization, and summarization of the probe sets. At the third step, each probe set (consisting of probes that represent the same gene) is assigned a single expression value. That is, the expressions of gene replicates within a probe set are summarized into a single measure.

After RMA normalization, majority genes will be associated with single expression values. For example, in the microarray data with oxidative stress (the data we used to demonstrate our method in the paper), 99.2 % (20832 out of 21009 genes) of genes have single expression measures after RMA normalization. All our seed genes (*seed-gene-set I*, *II*, *III* and *IV*) are associated with single measures.

0.0.3 Robustness of our method with respect to the number of known pathway genes To check the sensitivity of our method with respect to the number of known pathway genes, we studied the performance of our method by applying it to the shoot tissue dataset subjected to oxidative stress from Section 3.2.1 using every possible subset of size 2 and 3 of the genes from seed-gene-set II as seed genes. The seed-gene-set II is composed of 4 regulatory genes (ATR1, MYB28, MYB29, AKN2: these are denoted by a, b, c and d in Supplemental Table 9 respectively). We summarized the application results in Supplemental Table 9.

We first want to point out the importance of gene AKN2 (indexed by d in Supplemental Table 9). With gene AKN2 used as one of the seed genes, the performance of our method looks robust no matter which subset we use (see all gray-shaded columns in Supplemental Table 9). Actually, the results in all gray-shaded columns (even when there are only two seed genes) are very close to what we found when the whole seed-gene-set II was used. However in the absence of gene AKN2 in the seed-gene set, the performance of our method got worse and fluctuated dramatically. These results show that when the critical seed genes are included, our method would perform in a robust way even the size of seed-gene set changes. The results also suggest that to ensure a robust discovery, the selection of seed genes is better guided by appropriate prior biological knowledge. We also checked the results when using only one seed gene: we calculated pair-wise correlations between the single seed gene and all other candidate genes and then ranked the candidate genes accordingly. The results are summarized in Supplemental Table 10. We see that even though gene AKN2 (gene$d$ in the table) performs better than other genes as expected, the results are much worse than the ones using ≥ 2 seed genes (gray-shaded columns in Supplemental Table 9).

On the other hand, however, we may not know which seed genes are critical beforehand, when we apply the method. What can we do in this situation? We notice a large overlapping of identified genes in the gray-shaded columns in Supplemental Table 9. This motivates a practical idea of using different sets/subsets of seed genes (and different expression datasets if available) to find a set of frequently identified pathway genes when the critical seed genes are unknown for a more robust discovery. For example, for the shoot-tissue data presented above, for each gene, we can first construct a list consisting of the ranks (as a pathway candidate) of this gene obtained when different sets of seed genes are used. Next we can compute a new rank score for that gene as the average of top 50% values in the list. Finally we re-rank all the genes based on the new score. The results coming from this procedure are expected to be robust even with the knowledge on *critical* seed genes unknown. In the main text, we also studied the adverse effects of having nosiy genes (i.e., non-pathway genes) in the seed-gene set. Using this idea of searching *frequently identified pathway ways* under different sets of seed genes, the results would also be robust against noisy seed genes. When there are questions on which known pathway genes to include as seed genes or on which expression datasets to use, we suugest implementing this additional step of repeatedly running *pwsrc.knorm* to derive a set of *frequently identified pathway genes* under different datasets with different sets of seed genes for a robust discovery.

0.0.4 Calculating p-values for our test statistics using chi-square approximation To control the discovery error, we assign p-values to each candidate gene. We know that under certain conditions, likelihood ratio test statistics are asymptotically Chi-square distributed with degree freedom of k, where k is the number of seed genes. Following this, we then calculated Chi-square p-values for all candidate genes. We further applied the Bonferroni correction to address the multiple testing issue. We set $\alpha = 0.05$ as

our threshold of corrected p-values. The results are summarized in Supplemental Table 11 for the GSL and FB pathways. For example, the dataset from shoot tissues subjected to oxidative stress with GSL pathway, we found top 98 and 152 genes are significant (whose *p*-values are less than 0.05) with *seed-gene-set I* and *II*, respectively.

We note that the above calculation only provides approximate p-values, especially considering that the bootstrapped expression data matrices are not really independent of each other. An alternative way is to use a permutation test. We can randomly permute the experimental conditions of the candidate genes (keeping the order of the experimental conditions of seed genes unchanged) and apply our method to the permuted data. Repeating this, say 10000 times, we will obtain an empirical distribution for each gene and then can assign p-values accordingly. But this method is quite expensive computationally.

REFERENCES

Gachon,C.M.M., Langlois-Meurinne,M., Henry,Y. and Saindrenan,P. (2005) Transcriptional co-regulation of secondary metabolism enzymes in Arabidopsis: functional and evolutionary implications. *Plant Mol. Biol.*, **58**, 229-245.

Hirai,M.Y., Sugiyama,K., Sawada,Y., Tohge,T., Obayashi,T., Suzuki,A., Araki,R., Sakurai,N., Suzuki,H., Aoki,K., Goda,H., Nishizawa,O.I., Shibata,D. and Saito,K. (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA*, **104**, 6478-6483.

Saito,K., Hirai,M.Y. and Yonekura-Sakakibara,K. (2008) Decoding genes with coexpression networks and metabolomics 'majority report by precogs'. *Trends Plant Sci.*, **13**, 36-43.

Yonekura-Sakakibara,K., Tohge,T., Matsuda,F., Nakabayashi,R., Takayama,H., Niida,R., Watanabe-Takahashi,A., Inoue,E. and Saito,K. (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding genemetabolite correlations in arabidopsis. *Plant Cell*, **20**, 2160-2176.

Irizarry, R.A, Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. and Speed, T.P. (2003a) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249-264.

Bolstad, B.M., Irizarry R. A., Astrand, M., and Speed, T.P. (2003) A Comparison of normalization methods for high density oligonucleotide array data based on bias and variance. *Bioinformatics*, **19**, 185-193.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. and Speed, T.P. (2003b) Summaries of affymetrix GeneChip probe level data, *Nucleic Acids Research*, 31:e15.

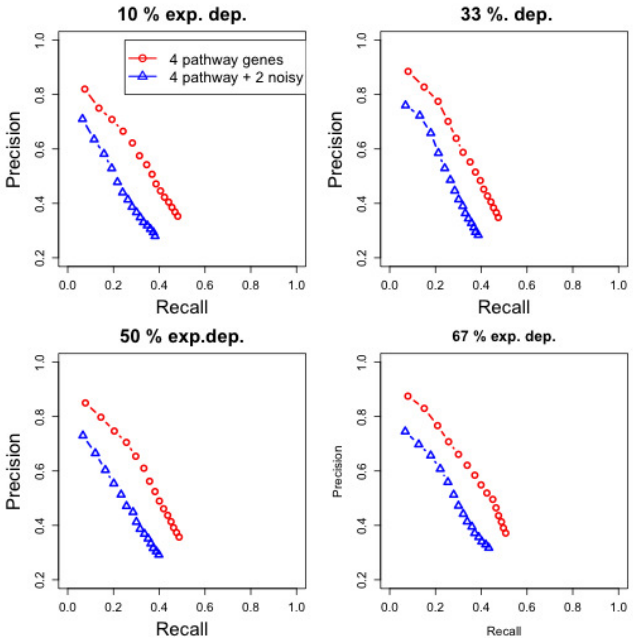

Fig. 1. (**Supplemental Figure 1**) Graphical summary of the simulation study. Simulation datasets are generated with different experiment dependencies (a) 10%, (b) 33%, (c) 50% and (d) 67%. For each plot, *precision* and *recall* are calculated from the top n (n = 1, , 15) genes in the list obtained by *pwsrc.knorm* with seed gene set composed of four pathway genes (red dots) or four pathway genes and 2 non-pathway (noisy) genes.

Table 1. Supplemental Table 1. Description of the experiment conditions used for the *A. thaliana* microarray dataset with different types of stress.

(a) Oxidative stress

Experiment number (shoot tissue)	Treatment	Time points	Experiment number (root tissue)
1	Control	0 h	14
2	Control	0.5 h	15
3	Control	1h	16
4	Control	3 h	17
5	Control	6 h	18
6	Control	12 h	19
7	Control	24 h	20
8	MV, $10\mu M$	0.5 h	21
9	MV, $10\mu M$	1 h	22
10	MV, $10\mu M$	3 h	23
11	MV, $10\mu M$	6 h	24
12	MV, $10\mu M$	12 h	25
13	MV, $10\mu M$	24 h	26

(b) Wouding, (c) UV-B light and (d) Drought stresses

Experiment number (shoot tissue)	Treatment	Time points	Experiment number (root tissue)
1	Control	0 h	16
2	Control	0.25 h	17
3	Control	0.5 h	18
4	Control	1h	19
5	Control	3 h	20
6	Control	6 h	21
7	Control	12 h	22
8	Control	24 h	23
9	Stress	0.25 h	24
10	Stress	0.5 h	25
11	Stress	1 h	26
12	Stress	3 h	27
13	Stress	6 h	28
14	Stress	12 h	29
15	Stress	24 h	30

Table 2. Supplemental Table 2. List of 64 Glucosinolate (GSL) metabolism pathway genes.

Group	AGI code	Gene name
Regulator genes	AT1G18570	MYB51
	AT1G66340	ETR1
	AT3G54640	TRP3/TSA1
	AT4G12030	BAT5
	AT5G03280	EIN2
	AT5G07690	MYB29
	AT5G07700	MYB76
	AT5G46330	FLS2
	AT5G60890	ATR1/Myb34
	AT5G61420	MYB28
	AT1G07640	OBP2
	AT3G09710	IQD1
GSL biosynthesis pathway (verified by experiment)	AT1G12140	GS-OX5
	AT1G16400	CYP79F2
	AT1G16410	CYP79F1
	AT1G18590	SOT17
	AT1G24100	UGT74B1
	AT1G62540	GS-OX2
	AT1G62560	GS-OX3
	AT1G62570	GS-OX4
	AT1G65860	GS-OX1
	AT1G74090	SOT18
	AT1G74100	SOT16
	AT2G20610	SUR1
	AT2G22330	CYP79B3
	AT2G25450	GS-OH
	AT2G43100	IPMI2
	AT3G19710	BCAT4
	AT3G49680	BCAT3
	AT3G58990	IPMI1/AtLeuD2
	AT4G03050	AOP3
	AT4G03060	AOP2
	AT4G13430	AtLeuC1
	AT4G13770	CYP83A1
	AT4G31500	CYP83B1/SUR2/ATR4
	AT4G39950	CYP79B2
	AT5G05260	CYP79A2
	AT5G14200	AtIMD1
	AT5G23010	MAM1
	AT5G23020	MAM3
	AT5G57220	CYP81F2
	AT1G31180	IPMDH1
	AT4G30530	GGP1
GSL biosynthesis pathway (predicted)	AT1G78370	GSTF20
	AT5G07460	PMSR2
	AT5G36160	
	AT2G30860	GSTF9
	AT2G30870	GSTF10
	AT2G31790	UGT74C1
	AT3G03190	GSTF11
GSL biosynthesis pathway (co-substrate pathway)	AT2G14750	AKN1
	AT4G39940	AKN2
	AT4G23100	PAD2
	AT1G65880	BZO1
	AT5G65940	CHY1
	AT1G04580	AAO4
	AT5G63980	FIERY1/SAL1
GSL-catabolic pathway	AT5G44070	PCS1
	AT1G47600	TGG4
	AT1G54030	MVP1
	AT1G59870	PEN3
	AT2G44490	PEN2
	AT1G54040	ESP
	AT3G14210	ESM1

Table 3. Supplemental Table 3. The number of identified GSL pathway genes in the *A. thaliana* microarray dataset from tissues subjected to UV-B light stress using (a) shoot tissue only, *seed-gene-set I*, (b) shoot tissue only, *seed-gene-set II*, (c) shoot and root tissues, *seed-gene-set I*, (d) shoot and root tissues, *seed-gene-set II*.

	Top	*pwsrc.knorm*	*pwsrc.null*	*pearson.mean*	*pearson.max*	*GLM*
(a)	10	3	2	1	0	1
	20	5	2	1	1	2
	30	5	2	2	1	3
	50	5	2	3	1	3
	100	8	3	4	3	5
(b)	10	6	1	3	2	3
	20	9	2	3	4	4
	30	9	2	3	5	5
	50	10	2	4	6	6
	100	13	3	5	8	8
(c)	10	7	1	1	0	6
	20	10	1	2	0	7
	30	10	1	3	0	7
	50	10	1	3	0	7
	100	11	1	3	1	7
(d)	10	8	0	1	1	6
	20	11	0	1	2	9
	30	12	0	1	2	10
	50	16	0	1	2	11
	100	18	1	4	2	11

Table 4. Supplemental Table 4. The number of identified GSL pathway genes in the *A. thaliana* microarray dataset from tissues subjected to drought stress using (a) shoot tissue only, *seed-gene-set I*, (b) shoot tissue only, *seed-gene-set II*, (c) shoot and root tissues, *seed-gene-set I*, (d) shoot and root tissues, *seed-gene-set II*.

	Top	*pwsrc.knorm*	*pwsrc.null*	*pearson.mean*	*pearson.max*	*GLM*
(a)	10	3	2	1	3	0
	20	4	2	2	3	1
	30	5	2	3	6	1
	50	6	3	6	6	2
	100	12	3	7	8	2
(b)	10	5	2	4	5	3
	20	8	3	4	8	3
	30	8	3	6	9	4
	50	8	4	8	12	6
	100	12	9	15	16	6
(c)	10	9	0	3	2	5
	20	10	0	3	3	5
	30	12	0	3	3	9
	50	15	0	3	3	12
	100	17	0	3	5	14
(d)	10	5	0	4	1	5
	20	8	1	4	1	5
	30	9	1	4	1	6
	50	10	2	6	1	7
	100	12	2	7	2	9

Table 5. Supplemental Table 5. List of flavonoid biosynthesis (FB) and phenylpropanoid biosynthesis pathway genes related to our study.

Pathways	Flavonoid biosynthesis (FB)	Phenylpropanoid biosynthesis
AGI code (gene name)	AT4G09820 (*TT8*) [a,b]	AT2G37040 (*PAL1*)
	AT5G23260 (*TT16*)[a,b]	AT3G53260 (*PAL2*)
	AT5G24520 (*TTG1*)[a]	AT5G04230 (*PAL3*)
	AT2G37260 (*TTG2*)[b]	AT3G10340 (*PAL4*)
	AT5G08640 (*FLS*)[a,b]	AT2G30490 (*C4H*)
	AT5G13930 (*CHS*)	AT1G51680 (*4CL1*)
	AT3G55120 (*CHI*)	AT3G21240 (*4CL2*)
	AT3G51240 (*F3H*)	AT1G65060 (*4CL3*)
	AT5G07990 (*F3′H*)	AT3G21230 (*4CL5*)
	AT5G42800 (*DFR*)	AT1G15950 (*CCR1*)
	AT1G61720 (*BAN*)	AT2G23910 (*CCR6*)
	AT5G17220 (*GST*)	
	AT3G59030 (*TT12*)	
	AT5G35550 (*TT2*)	
	AT1G06000 (*UGT89C1*)	
	AT5G17050 (*UGT78D2*)	
	AT1G78570 (*RHM1*)	
	AT4G14090 (*UGT75C1*)	
	AT1G30530 (*UGT78D1*)	
	AT3G29590 (*A5G6999MaT*)	
	AT5G54160 (*OMT1*)	
	AT3G62610 (*MYB11*)	
	AT2G47460 (*MYB12*)	
	AT5G49330 (*MYB111*)	
	AT1G56650 (*PAP1*)	
	AT1G66390 (*PAP2*)	

[a] *seed-gene-set III*
[b] *seed-gene-set IV*

Table 6. Supplemental Table 6. The number of pathway genes identified from flavonoid biosynthesis (FB) and phenylpropanoid biosynthesis pathways in the *A. thaliana* microarray dataset from shoot and root tissues subjected to (a) oxidation, (b) wounding, (c) UV-B light and (d) drought stresses. The number of identified genes from phenylpropanoid pathways is designated in the parenthesis adjacent to the total number of identified genes.

	seed-gene-set	Top rank	*pwsrc.knorm*	*pwsrc.null*	*pearson.mean*	*pearson.max*	*GLM*
(a)	*III*	10	7 (2)	1	0	4	5 (1)
		20	8 (3)	2	0	4	6 (1)
		30	10 (4)	2	0	4	7 (2)
		50	11 (4)	2	0	5 (1)	8 (3)
	IV	10	7 (2)	2	3	5 (1)	4 (1)
		20	9 (3)	3	3	6 (1)	5 (1)
		30	10 (4)	3	3	6 (1)	6 (1)
		50	11 (4)	4	4	6 (1)	7 (2)
(b)	*III*	10	6 (1)	3	0	4	5 (1)
		20	9 (3)	3	0	4	5 (1)
		30	9 (3)	4	0	4	5 (1)
		50	11 (4)	4	0	4	7 (2)
	IV	10	6 (1)	4	0	4	4 (1)
		20	9 (2)	4	1 (1)	4	5 (1)
		30	10 (3)	4	2 (2)	4	6 (1)
		50	11 (4)	4	2 (2)	5 (1)	7 (2)
(c)	*III*	10	6 (1)	1	0	2	4 (1)
		20	8 (3)	1	0	3	4 (1)
		30	11 (4)	1	0	3	4 (1)
		50	11 (4)	1	0	4 (1)	5 (2)
	IV	10	7 (2)	1	0	5 (1)	4 (1)
		20	8 (3)	1	0	6 (1)	4 (1)
		30	10 (3)	1	0	6 (1)	4 (1)
		50	11 (4)	1	1 (1)	6 (1)	4 (1)
(d)	*III*	10	6 (2)	0	0	2	4 (1)
		20	9 (3)	1	0	3	5 (1)
		30	10 (3)	1	0	4	6 (2)
		50	11 (3)	1	0	4	6 (2)
	IV	10	6 (2)	2	1	4	5 (1)
		20	9 (3)	2	2	4	5 (1)
		30	10 (3)	2	2	4	6 (2)
		50	13 (4)	2	2	4	6 (2)

Table 7. Supplemental Table 7. List of identified genes from top 20 list in Table 4. Each gene is designated by the original pathway to which it belongs.

Method	*seed-gene-set III*	*seed-gene-set IV*
pwsrc.knorm	AT1G65060 (*4CL3*)[b]	AT1G65060 (*4CL3*)[b]
	AT1G78570 (*RHM1*)[a]	AT3G51240 (*F3H*)[a]
	AT3G51240 (*F3H*)[a]	AT1G78570 (*RHM1*)[a]
	AT3G55120 (*CHI*)[a]	AT3G55120 (*CHI*)[a]
	AT5G13930 (*CHS*)[a]	AT5G13930 (*CHS*)[a]
	AT2G23910 (*CCR6*)[b]	AT2G23910 (*CCR6*)[b]
	AT5G17050 (*UGT78D2*)[a]	AT5G07990 (*F3′H*)[a]
	AT5G07990 (*F3′H*)[a]	AT5G17050 (*UGT78D2*)[a]
	AT2G37040 (*PAL1*)[b]	AT2G37040 (*PAL1*)[b]
pwsrc.null	AT1G78570 (*RHM1*)[a]	AT1G78570 (*RHM1*)[a]
		AT5G13930 (*CHS*)[a]
pearson.mean		AT4G14090 (*UGT75C1*)[a]
		AT5G42800 (*DFR*)[a]
pearson.max	AT1G78570 (*RHM1*)[a]	AT1G78570 (*RHM1*)[a]
	AT5G13930 (*CHS*)[a]	AT5G13930 (*CHS*)[a]
	AT3G55120 (*CHI*)[a]	AT3G55120 (*CHI*)[a]
		AT3G51240 (*F3H*)[a]
GLM	AT1G65060 (*4CL3*)[b]	AT1G65060 (*4CL3*)[b]
	AT1G78570 (*RHM1*)[a]	AT3G51240 (*F3H*)[a]
	AT3G51240 (*F3H*)[a]	AT3G55120 (*CHI*)[a]
	AT3G55120 (*CHI*)[a]	AT1G78570 (*RHM1*)[a]
	AT5G13930 (*CHS*)[a]	AT5G13930 (*CHS*)[a]

[a] Flavonoid biosynthesis (FB) pathway genes
[b] Phenylpropanoid biosynthesis pathway genes

Table 8. Supplemental Table 8. The number of pathway genes identified from flavonoid biosynthesis (FB) and phenylpropanoid biosynthesis pathways in the *A. thaliana* microarray dataset from shoot and root tissues subjected to drought stresses. For comparison, different seed genes sets are used: (a) *seed-gene-set III*, (b) *seed-gene-set III*, *ATR1*, *AKN2*, (c) *seed-gene-set III*, *MYB28*, *AKN2*, (d) *seed-gene-set III*, *ATR1*, *MYB28*, *AKN2*.

Top rank	(a)	(b)	(c)	(d)
10	6	3	3	3
20	9	5	5	4
30	10	6	7	5

Table 9. Supplemental Table 9. The number of pathway genes identified using the subsets of *seed-gene-set II*.

Top rank	*a,b*	*a,c*	*a,d*	*b,c*	*b,d*	*c,d*	*a,b,c*	*a,b,d*	*a,c,d*	*b,c,d*	*a,b,c,d*
10	3	0	6	3	7	8	1	7	5	7	6
20	4	0	11	3	13	14	2	10	10	12	11
30	4	0	14	3	15	14	3	13	12	13	14
50	6	0	15	5	17	17	4	15	14	15	14

Table 10. Supplemental Table 10. The number of pathway genes identified using single seed genes.

Top rank	*a*	*b*	*c*	*d*
10	0	1	0	3
20	0	2	0	4
30	0	3	0	4
50	0	5	0	6

Table 11. Supplemental Table 11. The number of top candidate genes whose p-values are less than or equal to the threshold $\alpha = 0.05$ for the GSL and FB pathways.

Data	GSL pathway *seed-gene-set I* (shoot)	GSL pathway *seed-gene-set II* (shoot)	GSL pathway *seed-gene-set I* (shoot and root)	GSL pathway *seed-gene-set II* (shoot and root)	FB pathway *seed-gene-set III* (shoot and root)	FB pathway *seed-gene-set IV* (shoot and root)
Oxidation	98	152	88	123	30	28
Wounding	50	104	45	82	25	20
UV-B light	63	262	43	62	32	29
Drought	203	336	73	75	40	25

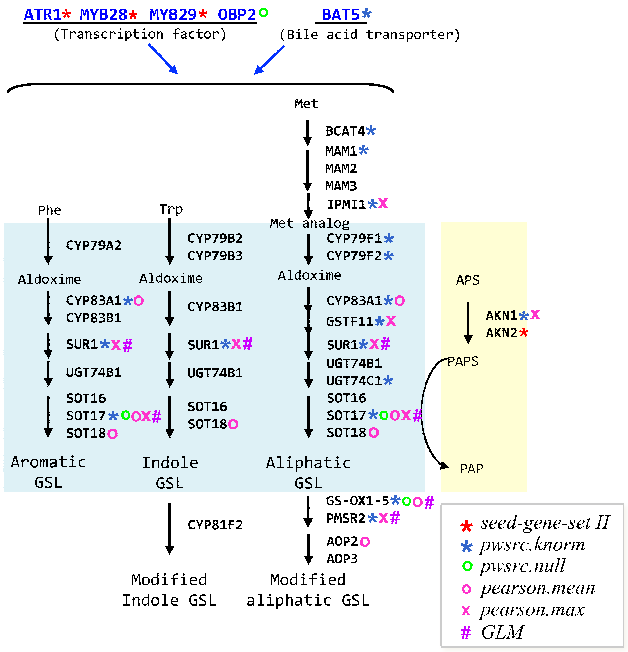

Fig. 2. (**Supplemental Figure 2**) Simplified schematic representation of glucosinolate (GSL) metabolic pathway. Enzymes and regulators are indicated by bold, capital letters. The GSL pathway genes from the top 30 lists identified by different methods are designated by different markers. *A. thaliana* dataset from shoot tissues subjected to oxidative stress and *seed-gene-set II* are used. Compared to other methods, our method uniquely finds six genes, *BAT5*, *BCAT4*, *MAM1*, *CYP79F1*, *CYP79F2* and *UGT74C1*, and misses three genes, *OBP2*, *SOT18* and *AOP2*.

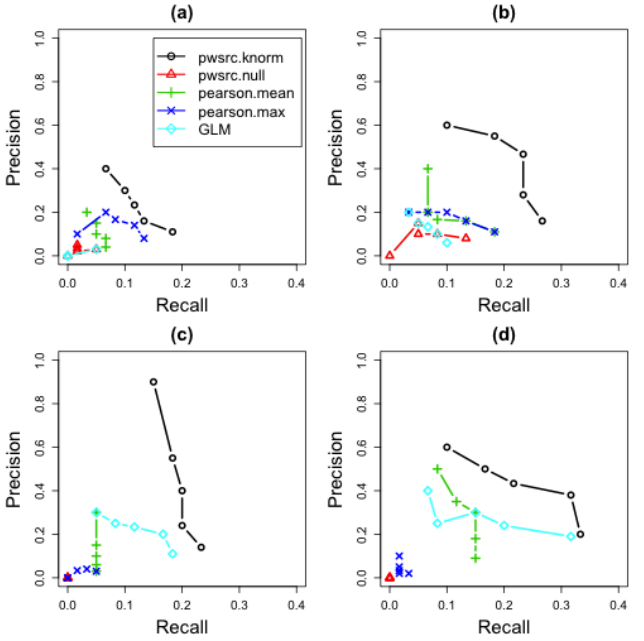

Fig. 3. (**Supplemental Figure 3**)Graphical summary of the *A. Thaliana* microarray dataset subjected to oxidative stress; (a) shoot tissue, *seed-gene-set I*, (b) shoot tissue, *seed-gene-set II*, (c) shoot and root tissues, *seed-gene-set I*, (d) shoot and root tissues, *seed-gene-set II*. *Precision* and *recall* are calculated from the top 10, 20, 30, 50 and 100 genes in the list obtained by different methods.

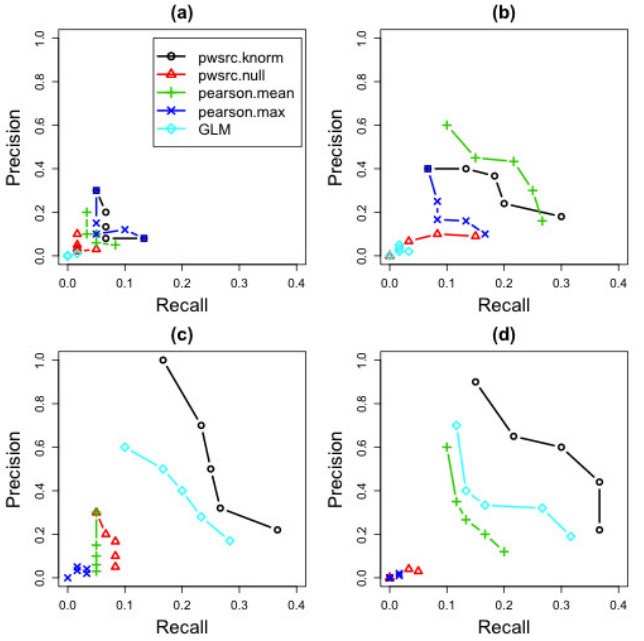

Fig. 4. (**Supplemental Figure 4**) Graphical summary of the *A. Thaliana* microarray dataset subjected wounding stress; (a) shoot tissue, *seed-gene-set I*, (b) shoot tissue, *seed-gene-set II*, (c) shoot and root tissues, *seed-gene-set I*, (d) shoot and root tissues, *seed-gene-set II*. *Precision* and *recall* are calculated from the top 10, 20, 30, 50 and 100 genes in the list obtained by different methods.

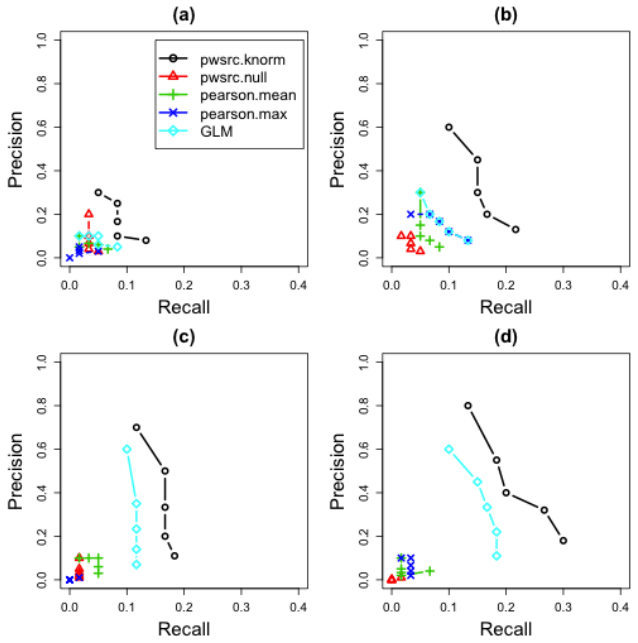

Fig. 5. (**Supplemental Figure 5**) Graphical summary of the *A. Thaliana* microarray dataset subjected to UV-B light stress; (a) shoot tissue, *seed-gene-set I*, (b) shoot tissue, *seed-gene-set II*, (c) shoot and root tissues, *seed-gene-set I*, (d) shoot and root tissues, *seed-gene-set II*. *Precision* and *recall* are calculated from the top 10, 20, 30, 50 and 100 genes in the list obtained by different methods.

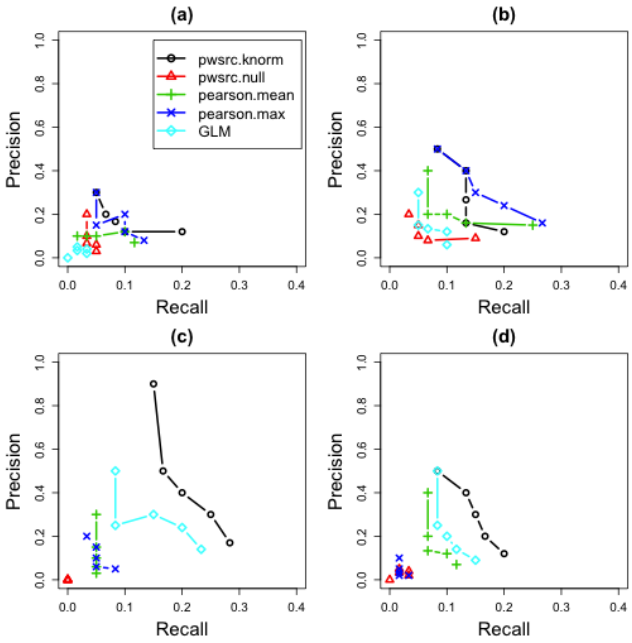

Fig. 6. (**Supplemental Figure 6**) Graphical summary of the *A. Thaliana* microarray dataset subjected to drought stress; (a) shoot tissue, *seed-gene-set I*, (b) shoot tissue, *seed-gene-set II*, (c) shoot and root tissues, *seed-gene-set I*, (d) shoot and root tissues, *seed-gene-set II*. *Precision* and *recall* are calculated from the top 10, 20, 30, 50 and 100 genes in the list obtained by different methods.

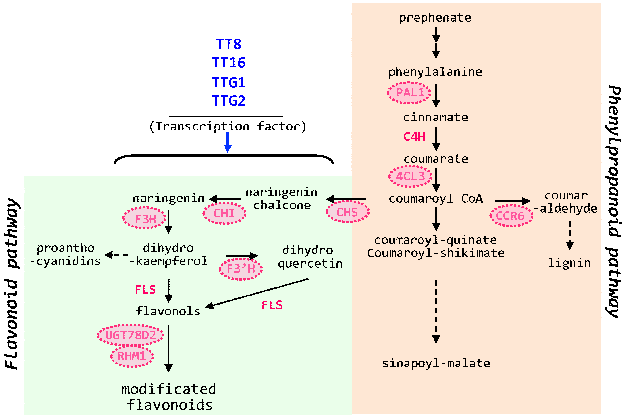

Fig. 7. (**Supplemental Figure 7**) Simplified schematic representation of flavonoid and phenylpropanoid biosynthesis pathways. Enzymes and regulator are indicated by bold, capital letters. Pathway genes identified by *pwsrc.knorm* from top 20 list in Table 4 are marked by dotted circles.