Computers in Biology and Medicine 42 (2012) 582–589

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Extracting plants core genes responding to abiotic stresses by penalized matrix decomposition

Jin-Xing Liu [a,b,*], Chun-Hou Zheng [c], Yong Xu [a]

[a] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China
[b] College of Information and Communication Technology, Qufu Normal University, Rizhao, China
[c] College of Electrical Engineering and Automation, Anhui University, Hefei, Anhui 230039, China

ARTICLE INFO

Article history:
Received 31 May 2011
Accepted 1 February 2012

Keywords:
Genes expression data
Penalized matrix decomposition
Sparse method
Data analysis
Abiotic stresses

ABSTRACT

Sparse methods have a significant advantage to reduce the complexity of genes expression data and to make them more comprehensible and interpretable. In this paper, based on penalized matrix decomposition (PMD), a novel approach is proposed to extract plants core genes, i.e., the characteristic gene set, responding to abiotic stresses. Core genes can capture the changes of the samples. In other words, the features of samples can be caught by the core genes. The experimental results show that the proposed PMD-based method is efficient to extract the core genes closely related to the abiotic stresses.

1. Introduction

Plants, as sessile organisms, have evolved an enormous capacity to realize their genetically predetermined developmental program despite ever-changing environmental conditions. As a result, they are able to cope with environmental conditions, such as cold, drought, heat, osmotic pressure, salt, UV-B light stress, etc [1,2]. The literatures of plant stress response are replete with title of "some stresses", whose tolerance was conferred by stress inducible protein that activates the plant's response to a specific stress of many abiotic stresses [3]. The underlying concept is that there exists a specific set of interacting genes that respond to each abiotic stress.

There are many conventional methods, such as RT-PCR or Northern blotting, to research the expression law of characteristic genes [4]. However these methods have some drawbacks, e.g., only one or a small number of genes can be studied at one time. So, how to obtain the interacting genes responding to abiotic stresses is still a challenge.

The rapid development of DNA microarray technology has made it possible to monitor gene expression levels on a genomic scale [5,6]. The gene expression data captured using the high-throughput technique can potentially provide systematic information regarding the underlying dynamics and mechanisms in biology, which greatly enhances the fundamental understanding of life on the molecular level. However, large numbers of gene expression data pose the problem of finding the knowledge of interest. Analyzing these data needs well founded mathematical tools which are adaptable to the large quantity of information and noise they carried on. Until now, many mathematical methods, such as singular value decomposition (SVD), principal component analysis (PCA) and independent component analysis (ICA), have been demonstrated to be able to analyze these data. For example, Alter et al. proposed to use SVD for processing and modeling the gene expression data [7]. Kumar et al. used SVD to mine health care data [8]. PCA was used to select genes for microarray data analysis by Wang et al. [9]. Ma et al. used PCA to identify differential gene pathways [10]. Mustafa et al. combined NMR with PCA to analyze the effect of salicylic acid on the metabolite profile of *Catharanthus roseus* cell suspension culture [11]. Huang et al. proposed a penalized discriminant method based on ICA for classifying gene expression data [12]. Li et al. used locally linear discriminant embedding to classify the gene expression data [13].

Though the classical methods such as SVD and PCA have been successfully applied in analysis of gene expression data, they still have some drawbacks, e.g., the left and right singular vectors of SVD or principal components (PCs) of PCA are usually dense. These make it difficult to interpret the singular vectors or PCs without subjective judgment. To make the data more comprehensible and interpretable, many mathematical tools are developed for reducing the complexity of the data. Among these approaches, sparse methods have a significant advantage, while giving up little statistical efficiency. Until now, many different

* Corresponding author at: Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China.
Tel.: +86 15065581070.
E-mail addresses: sdcavell@126.com, 313339002@qq.com (J.-X. Liu).

0010-4825/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2012.02.002

sparse methods were introduced, such as sparse PCA using the lasso given by Zou et al. [14], sparse PCA using convex relaxation techniques proposed by Zhang et al. [15], penalized matrix decomposition proposed by Witten et al. [16]. Moreover, many sparse methods have been widely used for gene expression data analysis. A sparse linear discriminant analysis was used to analyze the gene expression data by Wu et al. [17]. Lass et al. used the SPCA for clustering and feature selection [18]. Zheng et al. used the nonnegative matrix factorization for microarray data analysis [19]. In [16], Witten et al. proposed penalized matrix decomposition (PMD), which was used to discover the transcriptional modules by Zheng et al. [20] and Zhang et al. [21].

Though sparse methods are useful, main applications of them are focused on the clustering and classification at present, which may not give the reasonable and intelligible results on the selection of characteristic gene. To solve this problem, in this paper, a novel method, based on Penalized Matrix Decomposition (PMD), is proposed to extract "core genes" from the gene expression data. Here, core genes denote the characteristic gene set, which respond to the abiotic stresses. Owing to the penalty function, the factor matrices generated by PMD are sparse, so PMD can make the results more comprehensible and interpretable. The core genes extracted by PMD can capture the changes of the samples belonging to the same condition, i.e., the features of samples can be captured by the core genes. Moreover, we focus on how to extract the core genes responding to abiotic stresses by our method. The experimental results show that the proposed method can identify the core genes responding to the abiotic stresses. The contribution of this paper lies in the proposition of a PMD-based approach for extracting the core genes responding to the abiotic stresses.

The remainder of the paper is organized as follows. In Section 2, firstly, the PMD method is introduced. Then how to extract core genes using PMD is described. The experimental results are shown in Section 3. Section 4 concludes this paper and outlines the future works.

2. Methods

2.1. Mathematical definition of PMD

This subsection briefly introduces the Penalized Matrix Decomposition (PMD) proposed by Witten et al. [16]. Let $\mathbf{X}$ denote an $m \times n$ matrix of real-valued data, which consists of m genes in n samples, in general, $m \gg n$. In the case of microarray data, x_{ij} is the expression level of the ith gene in the jth sample. The elements of the ith row of $\mathbf{X}$ form the n-dimensional vector $\mathbf{r}_i$, which is referred to as the *transcriptional response* of the ith gene. Correspondingly, the elements of the jth column of $\mathbf{X}$ form the m-dimensional vector $\mathbf{s}_j$, which is referred to as the *expression profile* of the jth sample. Without loss of generality, let the column means of $\mathbf{X}$ be zero, the singular value decomposition (SVD) of matrix $\mathbf{X}$ can be written as follows:

$$\mathbf{X}=\mathbf{UDV}^T, \quad \mathbf{U}^T\mathbf{U}=\mathbf{I}_m, \quad \mathbf{V}^T\mathbf{V}=\mathbf{I}_n . \tag{1}$$

The PMD generalizes this decomposition by additional constraints on $\mathbf{U}$ and $\mathbf{V}$. The rank-1 PMD can be formulated as the following optimization problem:

$$\underset{d,\mathbf{u},\mathbf{v}}{minimize}\ \tfrac{1}{2}\|\mathbf{X}-d\mathbf{u}\mathbf{v}^T\|_F^2$$
$$s.t. \|\mathbf{u}\|_2^2=1,\ \|\mathbf{v}\|_2^2=1,\ P_1(\mathbf{u})\le\alpha_1,\ P_2(\mathbf{v})\le\alpha_2,\ d\ge 0 \tag{2}$$

where $\mathbf{u}$ is a column of $\mathbf{U}$, $\mathbf{v}$ is a column of $\mathbf{V}$, d is a diagonal element of $\mathbf{D}$, $\|\bullet\|_F$ is the Frobenius norm, P_1 and P_2 are convex penalty functions that can take a variety of forms [16].

Let $\mathbf{U}$ and $\mathbf{V}$ be $m \times p$ and $n \times p$ matrices, respectively, and $\mathbf{D}$ be a diagonal matrix with diagonal elements d_k, it can be proved as follows [16]:

$$\frac{1}{2}\|\mathbf{X}-\mathbf{UDV}^T\|_F^2=\frac{1}{2}\|\mathbf{X}\|_F^2-\sum_{k=1}^{p}\mathbf{u}_k^T\mathbf{X}\mathbf{v}_k d_k+\frac{1}{2}\sum_{k=1}^{p}d_k^2 \tag{3}$$

Hence, using the case $p=1$, we can see that $\mathbf{u}$ and $\mathbf{v}$ satisfying Eq. (2) can also satisfy the following problem:

$$\underset{\mathbf{u},\mathbf{v}}{maximize}\ \mathbf{u}^T\mathbf{X}\mathbf{v}$$
$$s.t. \|\mathbf{u}\|_2^2=1,\ \|\mathbf{v}\|_2^2=1,\ P_1(\mathbf{u})\le\alpha_1,\ P_2(\mathbf{v})\le\alpha_2 \tag{4}$$

and the d satisfying Eq. (2) is $d=\mathbf{u}^T\mathbf{X}$. The objective function $\mathbf{u}^T\mathbf{X}$ in Eq. (4) is bilinear in $\mathbf{u}$ and $\mathbf{v}$, that is, with $\mathbf{u}$ fixed, it is linear in $\mathbf{v}$, and vice versa. In fact, with $\mathbf{v}$ fixed, the criterion in Eq. (4) takes the following form:

$$\underset{\mathbf{u}}{maximize}\ \mathbf{u}^T\mathbf{X}\mathbf{v}$$
$$s.t. \|\mathbf{u}\|_2^2=1,\ P_1(\mathbf{u})\le\alpha_1. \tag{5}$$

This criterion is not convex due to L_2-equality penalty on $\mathbf{u}$.

The optimization problem in Eq. (4) can be finessed to the following biconvex optimization [16]:

$$\underset{\mathbf{u},\mathbf{v}}{maximize}\ \mathbf{u}^T\mathbf{X}\mathbf{v}$$
$$s.t. \|\mathbf{u}\|_2^2\le 1,\ \|\mathbf{v}\|_2^2\le 1,\ P_1(\mathbf{u})\le\alpha_1,\ P_2(\mathbf{v})\le\alpha_2. \tag{6}$$

It can be turned out that the solution to Eq. (6) satisfies Eq. (4) provided that α is chosen appropriately [16].

Eq. (6) is called the rank-1 PMD, and the iterative algorithm used to optimize is summarized as Algorithm 1.

Algorithm 1. *The iterative algorithm of the rank-1 PMD*

Step 1: Initialize $\mathbf{v}$ to have unit L_2-norm.

Step 2: Iterate until convergence:

(a) $\mathbf{u}\leftarrow\arg\max_{\mathbf{u}}\mathbf{u}^T\mathbf{X}\mathbf{v}, s.t. \|\mathbf{u}\|_2^2\le 1, P_1(\mathbf{u})\le\alpha_1$

(b) $\mathbf{v}\leftarrow\arg\max_{\mathbf{v}}\mathbf{u}^T\mathbf{X}\mathbf{v}, s.t. \|\mathbf{v}\|_2^2\le 1, P_2(\mathbf{v})\le\alpha_2$

Step 3: $d\leftarrow\mathbf{u}^T\mathbf{X}$

To obtain multiple factors of PMD, we can maximize the criterion in Eq. (6) repeatedly, each time using the residuals obtained by subtracting the product of previous factors $d\mathbf{uv}$ from $\mathbf{X}$, i.e. $\mathbf{X}^{k+1}\leftarrow\mathbf{X}^k-d_k\mathbf{u}_k\mathbf{v}_k^T$. Without the P_1- and P_2-penalty constraints, it can be shown that the K-factor PMD algorithm leads to the rank-K SVD of $\mathbf{X}$. In particular, the successive solutions are orthogonal. This can be seen since the solutions $\mathbf{u}_k$ and $\mathbf{v}_k$ are in the column and row spaces of $\mathbf{X}^k$, which has been orthogonalized with respect to $\mathbf{u}_j,\mathbf{v}_j$ for $j\in 1,\ldots,k-1$. With P_1 and/or P_2 present, the solutions are no longer in the column and/or row spaces, and so the orthogonality does not hold. The detailed algorithm of PMD can be found in [16]. In this paper, characteristic genes are selected according to $\mathbf{u}$, so we only take the penalty on $\mathbf{u}$, i.e. $P_1(\mathbf{u})\le\alpha_1$, and do not take the penalty on $\mathbf{v}$. By choosing appropriately the parameters α_1, PMD can result in sparse factors $\mathbf{u}$. Generally speaking, α_1 should be restricted to the ranges $1\le\alpha_1\le\sqrt{m}$, which can be selected according to the algorithm in [16].

2.2. Extracting core genes by PMD

The PMD algorithm decomposes the matrix $\mathbf{X}$ of gene expression data into two bases matrices $\mathbf{U}$ and $\mathbf{V}$, one defined by the left singular vectors and the other by right singular vectors. Referring to the definition in the Section 2.1, the left singular vectors span the space of the sample expression profiles $\{\mathbf{s}_j\}$ and the right singular vectors span the space of the gene transcriptional

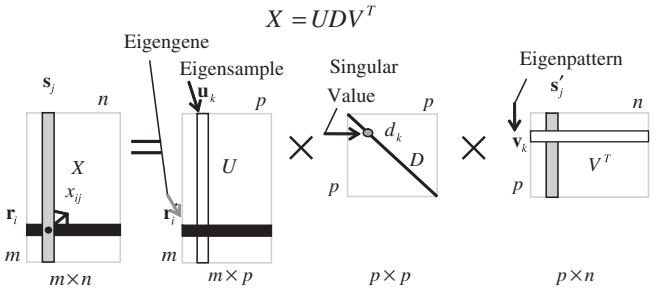

Fig. 1. Graphical depiction of PMD of a matrix **X**, annotated with adopted in the paper.

responses $\{\mathbf{r}_i\}$. Following the convention [22], we refer to the right singular vectors $\{\mathbf{v}_k\}$, i.e., the columns of **V**, as *eigenpatterns*, to the left singular vectors $\{\mathbf{u}_k\}$, i.e., the columns of **U**, as *eigensamples* and to the rows of **U** as *eigengenes*. eigensamples, eigengenes, eigenpattern and other definitions are shown in Fig. 1.

To reduce the dimensionality of the data, a subset of eigensamples are often selected to represent **X**. The underlying rationale is that the eigensamples have extracted the characteristic structure of the data. Corresponding to the eigensamples, the eigengenes are also extracted.

In this paper, our goal is to find the core genes that denote the characteristic gene set responding to the abiotic stresses. Referring to the definitions in Section 2.1, the interest signals in this case are the sample expression profile $\mathbf{s}_j$. By Eq. (1), the PMD equation for $\mathbf{s}_j$ is

$$\mathbf{s}_j = \sum_{k=1}^{p} v_{jk} d_k \mathbf{u}_k, \ j = 1,2,\ldots,n, \tag{7}$$

which is a linear combination of the eigensamples $\{\mathbf{u}_k\}$. The jth column of V^T, $\mathbf{s}'_j$ (see Fig. 1), contains the coordinates of the jth sample in the coordinate system (basis) of the scaled eigensamples, $d_k\mathbf{u}_k$. Using the vector $\mathbf{s}'_j$, the expression profiles of the samples may be captured by $p \le n$ variables, which are always fewer than the m variables in the vector $\mathbf{s}_j$. So, PMD can generally reduce the number of variables used to represent the sample expression profiles.

However, in order to reconstruct the original data, the eigensamples are needed, which are m-dimensional vectors. As mentioned above, the elements of the jth column of **X** form the m-dimensional vector $\mathbf{s}_j$. So, the sample $\mathbf{s}_j$ can be represented by the eigensample $\{\mathbf{u}_k\}$. By choosing appropriate penalty function P_1, the sparse $\{\mathbf{u}_k\}$ can be achieved. That is, responding to abiotic stresses, the characteristic gene set, namely core genes, can be used to represent the samples. When core genes are extracted by PMD-based method, each of them can capture the changes of the samples belonging to the same condition. In other words, the features of samples can be captured by the core genes.

The schemes to obtain the core genes can be summarized as follows:

Firstly, **X** is decomposed into the bases matrices **U** and **V** using PMD.
Secondly, the sparse eigensamples $\{\mathbf{u}_k\}$ are obtained.
Thirdly, the genes corresponding to nonzero entries in the eigensamples $\{\mathbf{u}_k\}$ are selected as the core genes.
Finally, the core genes are checked using Gene Ontology (GO).

The workflow diagram of our method is shown in Fig. 2.

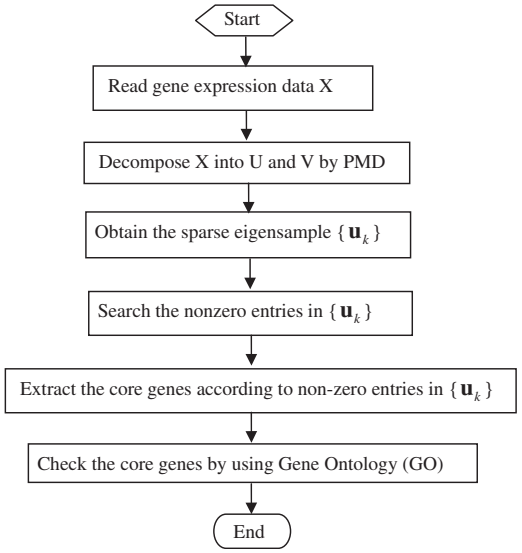

Fig. 2. Workflow diagram of our method.

3. Results and discussion

In this section the proposed method is evaluated by applying it to extract the core genes responding to abiotic stresses. Section 3.1 gives the data source. How to select the parameters is shown in Section 3.2. In Section 3.3, the gene ontology (GO) analysis is executed to evaluate the performance of the proposed method. The verifications of core genes based on literatures are given in Section 3.4.

3.1. Data source

For gene expression analysis, the Affymetrix CEL files are downloaded from NASCArrays [⟨http://affy.arabidopsis.info/⟩, reference numbers are: cold stress, NASCArrays-138; osmotic stress, NASCArrays-139; salt stress, NASCArrays-140; drought stress, NASCArrays-141; UV-B light stress, NASCArrays-144; heat stress, NASCArrays-146] [23]. The arrays are adjusted to avoid the background of optical noise using the GC-RMA software by Wu et al. [24] and normalized using quartile normalization. The results of GC-RMA are gathered in a matrix to be processed by PMD.

3.2. Parameters selection

The raw data include two classes, i.e. shoot and root, under each stress. Because the sparse principal component analysis (SPCA) given by Journee et al. outperforms existing algorithms both in quality of the obtained solution and computational speed [25], we compare PMD with SPCA method for extracting the core genes from these datasets. In this paper, the l_1-norm of $\mathbf{u}$ is taken as the penalty function, i.e. $\|\mathbf{u}\|_1 \le \alpha_1$. By choosing an appropriate α_1, a sparse $\mathbf{u}$ with many entries being zeros can be obtained. Because of $1 \le \alpha_1 \le \sqrt{m}$, let $\alpha_1 = \alpha^*\sqrt{m}$, where $1/\sqrt{m} \le \alpha \le 1$. For simplicity, let $p=1$, that is, only one factor is used. The results given with l_0- and l_1-norm penalty in SPCA are similar, which are also shown in [25]. Since the l_0-norm is faster than l_1-norm, we take l_0-norm penalty. The parameter γ in SPCA is used to regulate the sparse degree of PCs. For fair comparison, 500 genes are roughly selected by these two methods via choosing appropriate parameters. The parameters α and γ of the two methods, i.e., PMD and SPCA, on different data set are listed in Table 1.

3.3. Gene Ontology (GO) analysis

The Gene Ontology (GO) Term Enrichment tool can find significant shared GO terms or parents of those GO terms, used to describe the genes in the query/input set and to help discover what those genes may have in common [26]. In this paper, GOTermFinder is used to investigate the enrichment of functional annotations of genes

Table 1
The values of α and γ on different data sets.

Stress	*Shoot*		*Root*	
	PMD	SPCA	PMD	SPCA
	α	γ	α	γ
Drought	0.0928	0.4224	0.0999	0.4065
Salt	0.0924	0.4920	0.1057	0.5261
UV-B	0.1036	0.4505	0.0966	0.4329
Cold	0.1026	0.4660	0.0983	0.4726
Heat	0.0765	0.3770	0.0931	0.3710
Osmotic	0.1049	0.5139	0.0946	0.5338

Table 2
Response to stimulus (GO:0050896).

Stress		*PMD*		*SPCA*	
		P-value	Sample frequency	*P*-value	Sample frequency
Drought	s	**4.55E-64**	**247/500 (49.4%)**	2.63E-37	205/500 (41.0%)
Drought	r	9.77E-34	199/500 (39.8%)	**1.41E-36**	**205/500 (41.0%)**
Salt	s	**3.01E-29**	**190/500 (38.0%)**	4.36E-28	188/500 (37.6%)
Salt	r	**2.60E-47**	**222/500 (44.4%)**	3.16E-24	180/500 (36.0%)
UV-B	s	**3.48E-62**	**244/500 (48.8%)**	4.85E-61	242/500 (48.4%)
UV-B	r	**1.02E-19**	**170/500 (34.0%)**	3.41E-12	151/500 (30.2%)
Cold	s	**7.85E-47**	**221/500 (44.2%)**	3.59E-35	201/500 (40.2%)
Cold	r	**9.26E-33**	**197/500 (39.4%)**	2.22E-25	183/500 (36.6%)
Heat	s	3.78E-15	159/500 (31.8%)	**1.09E-26**	**185/500 (37.0%)**
Heat	r	**4.08E-11**	**148/500 (29.6%)**	1.85E-10	146/500 (29.2%)
Osmotic	s	**1.70E-39**	**209/500 (41.8%)**	2.92E-23	178/500 (35.6%)
Osmotic	r	1.26E-15	160/500 (32.0%)	**6.91E-22**	**175/500 (35.0%)**

In this table, "s" and "r" denote shoot and root samples, respectively.

identified by our method. The analysis of GOTermFinder provides significant information for the biological interpretation of high-throughput experiments. GOTermFinder is publicly available at ⟨http://go.princeton.edu/cgi-bin/GOTermFinder⟩ [27]. The core genes selected by PMD and SPCA are checked by GO, its threshold parameters are set as follows: maximum p-value=0.01 and minimum number of gene products=2. The results of GO Term Enrichment are given in supplementary file (Sup1).

3.3.1. Response to stimulus

Table 2 lists the response to stimulus (GO:0050896), which is the ancestor of all the abiotic stresses. Background frequency of response to stimulus (GO:0050896) in TAIR set is 4570/29556 (15.3%).

As Table 2 listed, both methods, PMD and SPCA, can extract the significant results with very lower *P*-value and very higher sample frequency. In Table 2, the superior results are shown in bold type. In the twelve items, there are only three of them (drought on root, heat on shoot and osmotic on root) that SPCA outperforms our method. In other nine items, our method is superior to SPCA.

3.3.2. Core genes responding to the stresses

To evaluate the performance of both methods for exacting core genes, the characteristic terms closely related to the stresses are investigated in detail. Table 3 lists the results, with the superior ones shown in bold type.

As Table 3 listed, for the drought stress, PMD method can select the genes responding to water deprivation (47 in shoot and 26 in root), but SPCA method only selects 23 in shoot and 24 in root. For the salt stress, PMD method selects the 41 and 33 genes responding to salt stress in shoot and root, respectively, yet SPCA can select only 28 genes in shoot and 22 genes in root. For the UV-B stress in shoot sample, PMD and SPCA can select 104 and 83 genes defending response, respectively. For the cold stress, the genes responding to cold can be selected by both PMD (44 in shoot and 43 in root) and SPCA (34 in shoot and 33 in root). For the heat stress, the genes responding to heat can be selected by both PMD (45 in shoot and 43 in root) and SPCA (30 in shoot and 28 in root). For the osmotic stress, the genes responding to osmotic stress can be given by both PMD (55 in shoot and 39 in root) and SPCA (29 in shoot and 27 in root).

In summary, on all the characteristic items, our method has superiority over SPCA.

3.4. Verifications based on the literatures

Many literatures about plants responding to abiotic stresses have been published recently. The literatures are searched to verify the core genes to abiotic stresses selected by PMD. In this paper, we make our detailing investigation on the core genes responding to drought-, salt- and cold-stresses.

3.4.1. Core genes responding to draught

For drought-stress, the core genes responding to water deprivation in shoot samples are listed in Table 4. In Table 4, the column of *References* gives the searching results that the authors have affirmed in the literature. The column of *Response to* denotes what the genes respond to.

As Section 3.3.2 mentioned, PMD-based method identified 47 genes responding to water deprivation in shoot samples with drought stress. All these genes can be searched in literatures. And as Table 4 listed, all the core genes are indeed related to drought stress, some of which are also related to cold and/or salt stresses.

3.4.2. Core genes responding to salt

For salt-stress, the core genes responding to salt in root samples are listed in Table 5.

As Section 3.3.2 mentioned, PMD-base method identified 33 genes responding to salt in root samples with salt stress. Table 5 lists the references about core genes responding to salt in root samples extracted by PMD. As Table 5 listed, there are 32 genes indeed related to salt stress, some of which are also related to cold and/or drought stresses. And only one gene (AT5G52190) is not related to responding to salt stress.

3.4.3. Core genes responding to cold

For cold-stress, Table 6 lists the core genes responding to cold in shoot samples.

As Section 3.3.2 mentioned, our method identified 44 genes responding to cold in shoot samples with cold stress. An amazing result can be found that all the genes are responding to cold stress. Some of them are also concerned with salt and/or drought stresses.

What should be pointed out is that, as Tables 4–6 listed, some genes are responding to two or more abiotic stresses. These overlapping cases of genes reveal that they take part in some different gene regulatory networks, and the expression pattern of them under cold, drought and salt stresses are similar [100].

From the verifications, it can be concluded that our method can select the core genes responding to the abiotic stresses.

4. Conclusions

In this paper, a novel method was proposed for extracting core genes based on Penalized Matrix Decomposition (PMD). For processing gene expression data, the method can achieve

Table 3
Characteristic terms selected from GO by algorithms.

Stress		GO Terms	Background frequency	PMD		SPCA	
				P-value	Sample frequency	P-value	Sample frequency
Drought	s	GO:0009414 response to water deprivation	207/29887 (0.7%)	**2.86E-33**	**47/500 (9.4%)**	1.17E-08	23/500 (4.6%)
Drought	r	GO:0009415 response to water deprivation	207/29887 (0.7%)	**2.96E-11**	**26/500 (5.2%)**	1.8E-09	24/500 (4.8%)
Salt	s	GO:0009651 response to salt stress	395/29887 (1.3%)	**3.16E-16**	**41/500 (8.2%)**	1.10E-06	28/500 (5.6%)
Salt	r	GO:0009651 response to salt stress	395/29887 (1.3%)	**4.98E-10**	**33/500 (6.6%)**	3.03E-03	22/500 (4.4%)
UV-B	s	GO:0006952 Defense response	919/29887 (3.1%)	**1.54E-52**	**104/500 (20.8%)**	8.66E-34	83/500 (16.6%)
UV-B	r	GO:0006953 Defense response	919/29887 (3.1%)	**2.97E-05**	**40/500 (8.0%)**	1.67E-03	36/500 (7.2%)
Cold	s	GO:0009409 response to cold	276/29887 (0.9%)	**9.31E-25**	**44/500 (8.8%)**	3.69E-15	34/500 (6.8%)
Cold	r	GO:0009410 response to cold	276/29887 (0.9%)	**9.92E-24**	**43/500 (8.6%)**	2.51E-14	33/500 (6.6%)
Heat	s	GO:0009408 response to heat	140/29887 (0.5%)	**4.61E-40**	**45/500 (9.0%)**	1.82E-20	30/500 (6.0%)
Heat	r	GO:0009409 response to heat	140/29887 (0.5%)	**3.04E-37**	**43/500 (8.6%)**	3.58E-18	28/500 (5.6%)
Osmotic	s	GO:0006970 response to osmotic stress	474/29887 (1.6%)	**6.96E-27**	**55/500 (11.0%)**	1.78E-06	29/500 (5.8%)
Osmotic	r	GO:0006970 response to osmotic stress	474/29887 (1.6%)	**2.16E-13**	**39/500 (7.8%)**	2.73E-05	27/500 (5.4%)

In this table, "s" and "r" denote shoot and root samples, respectively.

Table 4
References about core genes responding to water deprivation in shoot samples.

Gene name	Response to	References
ABCG22	Drought	Benschop et al. (2007) [28]
ABF3	Drought, salt	Abdeen et al. (2010) [29]
AKR4C8	Drought, salt, cold	Simpson et al. (2009) [30]
ANNAT3	Drought, salt, cold	Cantero et al. (2006) [31]
ANNAT4	Drought, salt	Huh et al. (2010) [32]
AOC1	Drought	Peltier et al. (2004) [33]
AT3G02480	Drought	Huang et al. (2008) [34]
AT3G05640	Drought	Huang et al. (2008) [34]
AT5G38710	Drought	Funk et al. (2010) [35]
BGLU18	Drought, salt	Huibers et al. (2009) [36]
COR413IM1	Drought, cold	Okawa et al. (2008) [37]
DR4	Drought	Gosti et al. (1995) [38]
DREB1A	Drought, cold	Seo et al. (2009) [39]
DREB2A	Drought, cold	Seo et al. (2009) [39]
EDL3	Drought, salt	Koops et al. (2011) [40]
ERD10	Drought, salt, cold	Kim et al. (2010) [41]
ESL1	Drought, salt	Yamada et al. (2010) [42]
GolS2	Drought, cold	Maruyama et al. (2009) [43]
GSTF6	Drought, salt	Jiang et al. (2007) [44]
HAI1	Drought	Huang et al. (2008) [34]
HB-12	Drought, salt	Huang et al. (2008) [34]
HB-7	Drought, salt	Huang et al. (2008) [34]
KIN1	Drought, cold	Huang et al. (2008) [34]
LEA14	Drought, salt	Huibers et al. (2009) [36]
LOX2	Drought	Bannenberg et al. (2009) [45]
LTI30	Drought, cold	Chung et al. (2008) [46]
LTI78	Drought, salt, cold	Vergnolle et al. (2005) [47]
LTP3	Drought	Huang et al. (2008) [34]
LTP4	Drought, salt	Huang et al. (2008) [34]
MYB60	Drought, salt	Oh et al. (2011) [48]
MYBR1	Drought, salt	Huang et al. (2008) [34]
MYC2	Drought	Abe et al. (2003) [49]
NAC019	Drought	Tran et al. (2004) [50]
NAC3	Drought, salt	Tran et al. (2004) [50]
NCED3	Drought, salt	Iuchi et al. (2001) [51]
PP2CA	Drought, cold	Lan et al. (2011) [52]
RAB18	Drought, cold	Tanaka et al. (2005) [53]
RAP2.6	Drought, salt, cold	Krishnaswamy et al. (2011) [54]
Rap2.6L	Drought, salt, cold	Krishnaswamy et al. (2011) [54]
RPK1	Drought, salt, cold	Osakabe et al. (2011) [55]
RD20	Drought, salt	Aubert et al. (2010) [56]
RD26	Drought	Kunieda et al. (2008) [57]
RD28	Drought	Alexandersson et al. (2005) [58]
SAG21	Drought, cold	Fowler et al. (2002) [59]
STZ	Drought, salt, cold	Rossel et al. (2007) [60]
WRKY33	Drought, salt, cold	Jiang et al. (2009) [61]
ZF2	Drought, salt	Drechsel et al. (2010) [62]

more comprehensible and interpretable results. It also can extract the core genes which capture the changes of the samples belonging to the same stresses.

Table 5
References about core genes responding to salt in root samples.

Gene name	Response to	References
AT5G52190		
bZIP1	Salt	Dietrich et al. (2011) [63]
CAMBP25	Salt, drought	Perruc et al. (2004) [64]
CBL1	Salt, drought, cold	Du et al. (2004) [64]
DDF1	Salt, drought, cold	Kang et al. (2011) [65]
EDL3	Drought, salt	Koops et al. (2011) [40]
GolS2	Salt, drought, cold	Maruyama et al. (2009) [43]
HB-12	Salt, drought	Huang et al. (2008) [34]
HVA22B	Salt	Chen et al. (2002) [66]
HVA22D	Salt	Chen et al. (2002) [66]
LTI78	Salt, drought, cold	Vergnolle et al. (2005) [47]
MKK9	Salt	Zhou et al. (2002) [67]
MYB108	Salt	Kraepiel et al. (2011) [68]
MYB15	Salt	Chen et al. (2006) [69]
MYB2	Salt, drought	Guo et al. (2011) [70]
MYB49	Salt	Chen et al. (2006) [69]
MYB51	Salt	Chen et al. (2006) [69]
MYB74	Salt	Chen et al. (2006) [69]
MYB96	Salt, drought	Chen et al. (2006) [69]
NAC6	Salt, drought	Tran et al. (2004) [50]
NCED3	Salt, drought	Iuchi et al. (2001) [51]
NSL1	Salt	Noutoshi et al. (2006) [71]
NUDT7	Salt	Jambunathan et al. (2010) [72]
RAP2.6	Salt, drought, cold	Krishnaswamy et al. (2011) [54]
RHL41	Salt, cold	Vogel et al. (2005) [73]
S6K2	Salt, cold	Pislariu et al. (2007) [74]
SOT12	Salt	Baek et al. (2010) [75]
STZ	Salt, drought, cold	Rossel et al. (2007) [60]
TSPO	Salt	Balsemao-Pires et al. (2011) [76]
UGT74E2	Salt, drought	Tognetti et al. (2010) [77]
WRKY25	Salt, drought, cold	Jiang et al. (2009) [61]
WRKY33	Salt, drought, cold	Jiang et al. (2009) [61]
ZF2	Drought, salt	Drechsel et al. (2010) [62]

Furthermore, the genes selected by the methods were analyzed using GO terms enrichment analysis. The results indicated that the proposed PMD-based method has superiority over SPCA on extracting the characteristic terms closely related to the stresses. In the end, the verifications of core genes based on literatures demonstrated that the proposed PMD-based method is potentially effective. In future, we will focus on the biological interpretation of the core genes.

Conflict of interest statement

None declared.

Table 6
References about core genes responding to cold in shoot samples.

Gene name	Response to	References
AT4G30660	Cold, salt	Fabro et al. (2008) [78]
AT4G34150	Cold	Van Leene et al. (2010) [79]
AT4G38840	Cold	Lee et al. (2005) [80]
AT5G54470	Cold	Mikkelsen et al. (2009) [81]
ATPMEPCRB	Cold	Mechli et al. (1998) [82]
BAP1	Cold	Yang et al. (2006) [83]
CBF1	Drought, cold	Vergnolle et al. (2005) [47]
CBF2	Cold	Shinwari et al. (1998) [84]
CBL1	Cold, drought, salt	Cheong et al. (2003) [85]
CCA1	Cold, drought	Lau et al. (2011) [86]
CEJ1	Cold	Tsutsui et al. (2009) [87]
CIPK7	Cold	Huang et al. (2011) [88]
CIPK9	Cold, salt	Pandey et al. (2007) [89]
COR15A	Cold	Thalhammer et al. (2010) [90]
COR15B	Cold	Thalhammer et al. (2010) [90]
COR27	Cold	Mekkelsen et al. (2009) [81]
COR47	Cold, drought	Huibers et al. (2009) [36]
CT-BMY	Cold	Vergnolle et al. (2005) [47]
CZF1	Cold	Vergnolle et al. (2005) [47]
DDF1	Cold, drought	Kang et al. (2011) [50]
DREB1A	Cold, drought	Seki et al. (2001) [91]
DREB26	Cold, drought, salt	Krishnaswamy et al. (2011) [54]
ELIP1	Cold	Vergnolle et al. (2005) [47]
ERD10	Cold, drought	Reyes et al. (2008) [92]
ERD7	Cold, salt, drought	Kimura et al. (2003) [93]
ERF5	Cold	Catala et al. (2003) [94]
GolS1	Cold, salt	Nishizawa et al. (2008) [95]
GolS2	Salt, drought, cold	Maruyama et al. (2009) [43]
GolS3	Salt, drought, cold	Maruyama et al. (2009) [43]
HVA22D	Cold	Chen et al. (2002) [66]
KIN1	Cold, drought	Huang et al. (2008) [34]
LEA4-5	Cold, drought	Reyes et al. (2008) [92]
LHY	Cold, salt	Lau et al. (2011) [86]
LTI30	Cold, drought	Chung et al. (2008) [46]
LTI78	Cold, drought, salt	Vergnolle et al. (2005) [47]
NAC062	Cold	Seo et al. (2011) [96]
RHL41	Cold, salt	Vogel et al. (2005) [73]
SAG21	Cold, drought	Seki et al. (2001) [91]
STZ	Salt, drought, cold	Sakamoto et al. (2004) [97]
SUS1	Cold	Bermejo et al. (2011) [98]
TCH2	Cold	Delk et al. (2005) [99]
TCH4	Cold	Lee et al. (2005) [80]
WRKY25	Cold, salt, drought	Jiang et al. (2009) [61]
WRKY33	Cold, salt, drought	Jiang et al. (2009) [61]

Acknowledgments

This work was supported by Program for the New Century Excellent Talents in University (no. NCET-08–0156), NSFC under Grant no. 61071179, Provincial Natural Science Research Program of Higher Education Institutions of Anhui Province under Grant no. KJ2012A005 and the Foundation of Qufu Normal University under Grant no. XJ200947.

Appendix A. Supplementary materials

Supplementary materials associated with this article can be found in the online version at doi:10.1016/j.compbiomed.2012.02.002.

References

[1] G.J. Allen, S.P. Chu, K. Schumacher, C.T. Shimazaki, D. Vafeados, A. Kemper, S.D. Hawke, G. Tallman, R.Y. Tsien, J.F. Harper, Alteration of stimulus-specific guard cell calcium oscillations and stomatal closing in *Arabidopsis det3* mutant, Science 289 (2000) 2338–2342.
[2] H.S. Ma, D. Liang, P. Shuai, X.L. Xia, W.L. Yin, The salt-and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in *Arabidopsis thaliana*, J. Exp. Bot. 61 (2010) 4011–4019.
[3] N.V. Fedoroff, Systems biology of abiotic stress: the elephant and the blind men, Abiotic Stress Adaptation Plants (2010) 485–502.
[4] N.S. Maan, S. Maan, K. Nomikou, D.J. Johnson, M. El Harrak, H. Madani, H. Yadin, S. Incoglu, K. Yesilbag, A.B. Allison, RT-PCR assays for seven serotypes of epizootic haemorrhagic disease virus and their use to type strains from the Mediterranean Region and North America, Plos. One 5 (2010) e12782.
[5] J. Kilian, D. Whitehead, J. Horak, D. Wanke, S. Weinl, O. Batistic, C. D'Angelo, E. Bornberg-Bauer, J. Kudla, K. Harter, The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses, Plant J. 50 (2007) 347–363.
[6] C.K. Sarmah, S. Samarasinghe, Microarray gene expression: a study of between-platform association of Affymetrix and cDNA arrays, Comput. Biol. Med. 41 (2011) 980–986.
[7] V.N.M. Aradhya, F. Masulli, S. Rovetta, A. Novel, Approach for biclustering gene expression data using modular singular value decomposition, Lect. Notes Bioinformatics 6160 (2010) 254–265.
[8] C.A. Kumar, S. Srinivas, Mining associations in health care data using formal concept analysis and singular value decomposition, J. Biol. Syst. 18 (2010) 787–807.
[9] A.T. Wang, E.A. Gehan, Gene selection for microarray data analysis using principal component analysis, Stat. Med. 24 (2005) 2069–2087.
[10] S. Ma, M.R. Kosorok, Identification of differential gene pathways with principal component analysis, Bioinformatics 25 (2009) 882–889.
[11] N.R. Mustafa, H.K. Kim, Y.H. Choi, R. Verpoorte, Metabolic changes of salicylic acid-elicited *Catharanthus roseus* cell suspension cultures monitored by NMR-based metabolomics, Biotechnol. Lett. 31 (2009) 1967–1974.
[12] D.S. Huang, C.H. Zheng, Independent component analysis-based penalized discriminant method for tumor classification using gene expression data, Bioinformatics 22 (2006) 1855–1862.
[13] B. Li, C.H. Zheng, D.S. Huang, L. Zhang, K. Han, Gene expression data classification using locally linear discriminant embedding, Comput. Biol. Med. 40 (2010) 802–810.
[14] H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis, J. Comput. Graphical Stat. 15 (2006) 265–286.
[15] Y. Zhang, A. d'Aspremont, L.E. Ghaoui, Sparse P.C.A.: Convex relaxations, algorithms and applications, Optim. Control (math.OC), (2010).
[16] D.M. Witten, R. Tibshirani, T. Hastie, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, Biostatistics 10 (2009) 515–534.
[17] M.C. Wu, L.S. Zhang, Z.X. Wang, D.C. Christiani, X.H. Lin, Sparse linear discriminant analysis for simultaneous testing for the significance of a gene set/pathway and gene selection, Bioinformatics 25 (2009) 1145–1151.
[18] R. Luss, A. d'Aspremont, Clustering and feature selection using sparse principal component analysis, Optim. Eng. 11 (2010) 145–157.
[19] C.H. Zheng, V. To-Yee Ng, L. Zhang, C.K. Shiu, H.Q. Wang, Tumor classification based on non-negative matrix factorization using gene expression data, NanoBiosci. IEEE Trans. (2011) 86–93.
[20] C.H. Zheng, L. Zhang, T.Y. Ng, C.K. Shiu, S.L. Wang, Inferring the transcriptional modules using penalized matrix decomposition, Advanced Intelligent Computing Theories and Applications: With Aspects of Artificial Intelligence, vol. 6216, (2010), pp. 35–41.
[21] J. Zhang, C.H. Zheng, J.X. Liu, H.Q. Wang, Discovering the transcriptional modules using microarray data by penalized matrix decomposition, Comput. Biol. Med. 41 (2011) 1041–1050.
[22] F. Liang, Use of SVD-based probit transformation in clustering gene expression profiles, Comput. Stat. Data Anal. 51 (2007) 6355–6366.
[23] D.J. Craigon, N. James, J. Okyere, J. Higgins, J. Jotham, S. May, NASCArrays: a repository for microarray data generated by NASC's transcriptomics service, Nucleic Acids Res. 32 (2004) D575–D577.
[24] Z.J. Wu, R.A. Irizarry, R. Gentleman, F. Martinez-Murillo, F. Spencer, A model-based background adjustment for oligonucleotide expression arrays, J. Am. Stat. Assoc. 99 (2004) 909–917.
[25] M. Journée, Y. Nesterov, P. Richtarik, R. Sepulchre, Generalized power method for sparse principal component analysis, J. Mach. Learn. Res. 11 (2010) 517–553.
[26] T.Z. Berardini, V.K. Khodiyar, R.C. Lovering, P. Talmud, The Gene Ontology in 2010: extensions and refinements, Nucleic Acids Res. 38 (2010) D331–D335.
[27] E.I. Boyle, S.A. Weng, J. Gollub, H. Jin, D. Botstein, J.M. Cherry, G. Sherlock, GO:TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes, Bioinformatics 20 (2004) 3710–3715.
[28] J.J. Benschop, S. Mohammed, M. O'Flaherty, A.J.R. Heck, M. Slijper, F.L.H. Menke, Quantitative phosphoproteomics of early elicitor signaling in *Arabidopsis*, Mol. Cell. Proteomics 6 (2007) 1198–1214.
[29] A. Abdeen, J. Schnell, B. Miki, Transcriptome analysis reveals absence of unintended effects in drought-tolerant transgenic plants overexpressing the transcription factor ABF3, BMC Genomics 11 (2010).
[30] P.J. Simpson, C. Tantitadapitak, A.M. Reed, O.C. Mather, C.M. Bunce, S.A. White, J.P. Ride, characterization of two novel aldo–keto reductases from *Arabidopsis*: expression patterns, broad substrate specificity, and an open active-site structure suggest a role in toxicant metabolism following stress, J. Mol. Biol. 392 (2009) 465–480.
[31] A. Cantero, S. Barthakur, T.J. Bushart, S. Chou, R.O. Morgan, M.P. Fernandez, G.B. Clark, S.J. Roux, Expression profiling of the *Arabidopsis* annexin gene

family during germination, de-etiolation and abiotic stress, Plant Physiol. Biochem. 44 (2006) 13–24.
[32] S.M. Huh, E.K. Noh, H.G. Kim, B.W. Jeon, K. Bae, H.C. Hu, J.M. Kwak, O.K. Park, *Arabidopsis* annexins AnnAt1 and AnnAt4 interact with each other and regulate drought and salt stress responses, Plant Cell Physiol. 51 (2010) 1499–1514.
[33] J. Peltier, A. Ytterberg, Q. Sun, New functions of the thylakoid membrane proteome of *Arabidopsis thaliana* revealed by a simple, fast, and versatile fractionation strategy, J. Biol. Chem. 279 (2004) 49367–49383.
[34] D.Q. Huang, W.R. Wu, S.R. Abrams, A.J. Cutler, The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors, J. Exp. Bot. 59 (2008) 2991–3007.
[35] D. Funck, S. Eckard, G. Muller, Non-redundant functions of two proline dehydrogenase isoforms in *Arabidopsis*, BMC Plant Biol. 10 (2010).
[36] R.P. Huibers, M. de Jong, R.W. Dekter, G. Van den Ackerveken, Disease-specific expression of host genes during downy mildew infection of *Arabidopsis*, Mol. Plant–Microbe Interact. 22 (2009) 1104–1115.
[37] K. Okawa, K. Nakayama, T. Kakizaki, T. Yamashita, T. Inaba, Identification and characterization of Cor413im proteins as novel components of the chloroplast inner envelope, Plant Cell Environ. 31 (2008) 1470–1483.
[38] F. Gosti, N. Bertauche, N. Vartanian, J. Giraudat, Abscisic acid-dependent and -independent regulation of gene-expression by progressive drought in *Arabidopsis thaliana*, Mol. Genet. 246 (1995) 10–18.
[39] E. Seo, H. Lee, J. Jeon, H. Park, J. Kim, Y.S. Noh, I. Lee, 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 (2009) 3185–3197.
[40] P. Koops, S. Pelser, M. Ignatz, C. Klose, K. Marrocco-Selden, T. Kretsch, EDL3 is an F-box protein involved in the regulation of abscisic acid signalling in *Arabidopsis thaliana*, J. Exp. Bot. (2011).
[41] S.Y. Kim, K.H. Nam, Physiological roles of ERD10 in abiotic stresses and seed germination of *Arabidopsis*, Plant Cell Rep. 29 (2010) 203–209.
[42] K. Yamada, Y. Osakabe, J. Mizoi, K. Nakashima, Y. Fujita, K. Shinozaki, K. Yamaguchi-Shinozaki, Functional analysis of an *Arabidopsis thaliana* abiotic stress-inducible facilitated diffusion transporter for monosaccharides, J. Biol. Chem. 285 (2010) 1138–1146.
[43] K. Maruyama, M. Takeda, S. Kidokoro, K. Yamada, Y. Sakuma, K. Urano, M. Fujita, K. Yoshiwara, S. Matsukura, Y. Morishita, R. Sasaki, H. Suzuki, K. Saito, D. Shibata, K. Shinozaki, K. Yamaguchi-Shinozaki, Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A, Plant Physiol. 150 (2009) 1972–1980.
[44] Y. Jiang, B. Yang, N.S. Harris, M.K. Deyholos, Comparative proteomic analysis of NaCl stress-responsive proteins in *Arabidopsis* roots, J. Exp. Bot. 58 (2007) 3591–3607.
[45] G. Bannenberg, M. Martínez, M. Hamberg, C. Castresana, Diversity of the enzymatic activity in the lipoxygenase gene family of *Arabidopsis thaliana*, Lipids 44 (2009) 85–95.
[46] S. Chung, R.W. Parish, Combinatorial interactions of multiple cis-elements regulating the induction of the *Arabidopsis* XERO2 dehydrin gene by abscisic acid and cold, Plant J. 54 (2008) 15–29.
[47] C. Vergnolle, M.N. Vaultier, L. Taconnat, J.P. Renou, J.C. Kader, A. Zachowski, E. Ruelland, The cold-induced early activation of phospholipase C and D pathways determines the response of two distinct clusters of genes in *Arabidopsis* cell suspensions, Plant Physiol. 139 (2005) 1217–1233.
[48] J.E. Oh, Y. Kwon, J.H. Kim, H. Noh, S.W. Hong, H. Lee, A dual role for MYB60 in stomatal regulation and root growth of *Arabidopsis thaliana* under drought stress, Plant Mol. Biol. 77 (2011) 91–103.
[49] H. Abe, T. Urao, T. Ito, M. Seki, K. Shinozaki, K. Yamaguchi-Shinozaki, *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling, Plant Cell 15 (2003) 63–78.
[50] L.S.P. Tran, K. Nakashima, Y. Sakuma, S.D. Simpson, Y. Fujita, K. Maruyama, M. Fujita, M. Seki, K. Shinozaki, K. Yamaguchi-Shinozaki, Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter, Plant Cell 16 (2004) 2481–2498.
[51] S. Iuchi, Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*, Plant J 30 (2002). 611–611 (vol 27, pg 325, 2001).
[52] W.Z. Lan, S.C. Lee, Y.F. Che, Y.Q. Jiang, S. Luan, Mechanistic analysis of AKT1 regulation by the CBL–CIPK–PP2CA interactions, Mol. Plant 4 (2011) 527–536.
[53] Y. Tanaka, T. Sano, M. Tamaoki, N. Nakajima, N. Kondo, S. Hasezawa, Ethylene inhibits abscisic acid-induced stomatal closure in *Arabidopsis*, Plant Physiol. 138 (2005) 2337–2343.
[54] S. Krishnaswamy, S. Verma, M.H. Rahman, N.N.V. Kav, Functional characterization of four APETALA2-family genes (RAP2.6, RAP2.6L, DREB19 and DREB26) in *Arabidopsis*, Plant Mol. Biol. 75 (2011) 107–127.
[55] Y. Osakabe, S. Mizuno, H. Tanaka, K. Maruyama, K. Osakabe, D. Todaka, Y. Fujita, M. Kobayashi, K. Shinozaki, K. Yamaguchi-Shinozaki, Overproduction of the membrane-bound receptor-like protein kinase 1, RPK1, enhances abiotic stress tolerance in *Arabidopsis*, J. Biol. Chem. 285 (2010) 9190–9201.
[56] Y. Aubert, D. Vile, M. Pervent, D. Aldon, B. Ranty, T. Simonneau, A. Vavasseur, J.P. Galaud, RD20, a stress-inducible caleosin, participates in stomatal control, transpiration and drought tolerance in *Arabidopsis thaliana*, Plant Cell Physiol. 51 (2010) 1975–1987.
[57] T. Kunieda, N. Mitsuda, M. Ohme-Takagi, S. Takeda, M. Aida, M. Tasaka, M. Kondo, M. Nishimura, I. Hara-Nishimura, NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in *Arabidopsis*, Plant Cell 20 (2008) 2631–2642.
[58] E. Alexandersson, L. Fraysse, S. Sjovall-Larsen, S. Gustavsson, M. Fellert, M. Karlsson, U. Johanson, P. Kjellbom, Whole gene family expression and drought stress regulation of aquaporins, Plant Mol. Biol. 59 (2005) 469–484.
[59] S. Fowler, M.F. Thomashow, *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway, Plant Cell 14 (2002) 1675–1690.
[60] J.B. Rossel, P.B. Wilson, D. Hussain, N.S. Woo, M.J. Gordon, O.P. Mewett, K.A. Howell, J. Whelan, K. Kazan, B.J. Pogson, Systemic and intracellular responses to photooxidative stress in *Arabidopsis*, Plant Cell 19 (2007) 4091–4110.
[61] Y.Q. Jiang, M.K. Deyholos, Functional characterization of *Arabidopsis* NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses, Plant Mol. Biol. 69 (2009) 91–105.
[62] G. Drechsel, S. Raab, S. Hoth, *Arabidopsis* zinc-finger protein 2 is a negative regulator of ABA signaling during seed germination, J. Plant Physiol. 167 (2010) 1418–1421.
[63] K. Dietrich, F. Weltmeier, A. Ehlert, C. Weiste, M. Stahl, K. Harter, W. Droge-Laser, Heterodimers of the *Arabidopsis* transcription factors bZIP1 and bZIP53 reprogram amino acid metabolism during low energy stress, Plant Cell 23 (2011) 381–395.
[64] E. Perruc, M. Charpenteau, B.C. Ramirez, A. Jauneau, J.P. Galaud, R. Ranjeva, B. Ranty, A novel calmodulin-binding protein functions as a negative regulator of osmotic stress tolerance in *Arabidopsis thaliana* seedlings, Plant J. 38 (2004) 410–420.
[65] H.G. Kang, J. Kim, B. Kim, H. Jeong, S.H. Choi, E.K. Kim, H.Y. Lee, P.O. Lim, Overexpression of FTL1/DDF1, an AP2 transcription factor, enhances tolerance to cold, drought, and heat stresses in *Arabidopsis thaliana*, Plant Sci. 180 (2011) 634–641.
[66] C.N. Chen, C.C. Chu, R. Zentella, S.M. Pan, T.H.D. Ho, AtHVA22 gene family in *Arabidopsis*: phylogenetic relationship, ABA and stress regulation, and tissue-specific expression, Plant Mol. Biol. 49 (2002) 633–644.
[67] C.J. Zhou, Z.H. Cai, Y.F. Guo, S.S. Gan, An *Arabidopsis* mitogen-activated protein kinase cascade, MKK9-MPK6, plays a role in leaf senescence, Plant Physiol. 150 (2009) 167–177.
[68] Y. Kraepiel, J. Pedron, O. Patrit, E. Simond-Cote, V. Hermand, F. Van Gijsegem, Analysis of the plant bos1 mutant highlights necrosis as an efficient defence mechanism during *D. dadantii*/*Arabidospis thaliana* interaction, Plos One 6 (2011).
[69] Y.H. Chen, X.Y. Yang, K. He, M.H. Liu, J.G. Li, Z.F. Gao, Z.Q. Lin, Y.F. Zhang, X.X. Wang, X.M. Qiu, Y.P. Shen, L. Zhang, X.H. Deng, J.C. Luo, X.W. Deng, Z.L. Chen, H.Y. Gu, L.J. Qu, The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family, Plant Mol. Biol. 60 (2006) 107–124.
[70] Y.F. Guo, S.S. Gan, AtMYB2 regulates whole plant senescence by inhibiting cytokinin-mediated branching at late stages of development in *Arabidopsis*, Plant Physiol. 156 (2011) 1612–1619.
[71] Y. Noutoshi, T. Kuromori, T. Wada, T. Hirayama, A. Kamiya, Y. Imura, M. Yasuda, H. Nakashita, K. Shirasu, K. Shinozaki, Loss of necrotic spotted lesions 1 associates with cell death and defense responses in *Arabidopsis thaliana*, Plant Mol. Biol. 62 (2006) 29–42.
[72] N. Jambunathan, A. Penaganti, Y.H. Tang, R. Mahalingam, Modulation of redox homeostasis under suboptimal conditions by *Arabidopsis* nudix hydrolase 7, BMC Plant Biol. 10 (2010).
[73] J.T. Vogel, D.G. Zarka, H.A. Van Buskirk, S.G. Fowler, M.F. Thomashow, Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*, Plant J. 41 (2005) 195–211.
[74] C.I. Pislariu, R. Dickstein, An IRE-like A.G.C. kinase gene, MtIRE, has unique expression in the invasion zone of developing root nodules in *Medicago truncatula*, Plant Physiol. 144 (2007) 682–694.
[75] D. Baek, P. Pathange, J.S. Chung, J.F. Jiang, L.Q. Gao, A. Oikawa, M.Y. Hirai, K. Saito, P.W. Pare, H.Z. Shi, A stress-inducible sulphotransferase sulphonates salicylic acid and confers pathogen resistance in *Arabidopsis*, Plant Cell Environ. 33 (2010) 1383–1392.
[76] E. Balsemao-Pires, Y. Jaillais, B.J.S.C. Olson, L.R. Andrade, J.G. Umen, J. Chory, G. Sachetto-Martins, The *Arabidopsis* translocator protein (AtTSPO) is regulated at multiple levels in response to salt stress and perturbations in tetrapyrrole metabolism, BMC Plant Biol. 11 (2011).
[77] V.B. Tognetti, O. Van Aken, K. Morreel, K. Vandenbroucke, B.V. de Cotte, I. De Clercq, S. Chiwocha, R. Fenske, E. Prinsen, W. Boerjan, B. Genty, K.A. Stubbs, D. Inze, F. Van Breusegem, Perturbation of indole-3-butyric acid homeostasis by the UDP-glucosyltransferase UGT74e2 modulates *Arabidopsis* architecture and water stress tolerance, Plant Cell 22 (2010) 2660–2679.
[78] G. Fabro, J.A. Di Rienzo, C.A. Voigt, T. Savchenko, K. Dehesh, S. Somerville, M.E. Alvarez, Genome-wide expression profiling *Arabidopsis* at the stage of *Golovinomyces cichoracearum* haustorium formation, Plant Physiol. 146 (2008) 1421–1439.
[79] J. Van Leene, J. Hollunder, D. Eeckhout, G. Persiau, E. Van de Slijke, H. Stals, G. Van Isterdael, A. Verkest, S. Neirynck, Y. Buffel, S. De Bodt, S. Maere, K. Laukens, A. Pharazyn, P.C.G. Ferreira, N. Eloy, C. Renne, C. Meyer, J.D. Faure, J. Steinbrenner, J. Beynon, J.C. Larkin, Y. Van de Peer, P. Hilson, M. Kuiper, L. De Veylder, H. Van Onckelen, D. Inze, E. Witters, G. De Jaeger,

Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*, Mol. Syst. Biol. 6 (2010).
[80] B.H. Lee, D.A. Henderson, J.K. Zhu, The *Arabidopsis* cold-responsive transcriptome and its regulation by ICE1, Plant Cell 17 (2005) 3155–3175.
[81] M.D. Mikkelsen, M.F. Thomashow, A role for circadian evening elements in cold-regulated gene expression in *Arabidopsis*, Plant J. 60 (2009) 328–339.
[82] F. Micheli, C. Holliger, R. Goldberg, L. Richard, Characterization of the pectin methylesterase-like gene AtPME3: a new member of a gene family comprising at least 12 genes in *Arabidopsis thaliana*, Gene 220 (1998) 13–20.
[83] H.J. Yang, Y.Q. Li, J. Hua, The C2 domain protein BAP1 negatively regulates defense responses in *Arabidopsis*, Plant J. 48 (2006) 238–248.
[84] Z.K. Shinwari, K. Nakashima, S. Miura, M. Kasuga, M. Seki, K. Yamaguchi-Shinozaki, K. Shinozaki, An *Arabidopsis* gene family encoding DRE/CRT binding proteins involved in low-temperature-responsive gene expression, Biochem. Biophys. Res. Commun. 250 (1998) 161–170.
[85] Y.H. Cheong, K.N. Kim, G.K. Pandey, R. Gupta, J.J. Grant, S. Luan, CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in *Arabidopsis*, Plant Cell 15 (2003) 1833–1845.
[86] O.S. Lau, X. Huang, J.B. Charron, J.H. Lee, G. Li, X.W. Deng, Interaction of *Arabidopsis* DET1 with CCA1 and LHY in mediating transcriptional repression in the plant circadian clock, Mol. Cell 43 (2011) 703–712.
[87] T. Tsutsui, W. Kato, Y. Asada, K. Sako, T. Sato, Y. Sonoda, S. Kidokoro, K. Yamaguchi-Shinozaki, M. Tamaoki, K. Arakawa, T. Ichikawa, M. Nakazawa, M. Seki, K. Shinozaki, M. Matsui, A. Ikeda, J. Yamaguchi, DEAR1, a transcriptional repressor of DREB protein that mediates plant defense and freezing stress responses in *Arabidopsis*, J. Plant Res. 122 (2009) 633–643.
[88] C.L. Huang, S. Ding, H. Zhang, H. Du, L.Z. An, CIPK7 is involved in cold response by interacting with CBL1 in *Arabidopsis thaliana*, Plant Sci. 181 (2011) 57–64.
[89] G.K. Pandey, Y.H. Cheong, B.G. Kim, J.J. Grant, L.G. Li, S. Luan, CIPK9: a calcium sensor-interacting protein kinase required for low-potassium tolerance in *Arabidopsis*, Cell Res. 17 (2007) 411–421.
[90] A. Thalhammer, M. Hundertmark, A.V. Popova, R. Seckler, D.K. Hincha, Interaction of two intrinsically disordered plant stress proteins (COR15A and COR15B) with lipid membranes in the dry state, BBA-Biomembr. 1798 (2010) 1812–1820.
[91] M. Seki, M. Narusaka, H. Abe, M. Kasuga, K. Yamaguchi-Shinozaki, P. Carninci, Y. Hayashizaki, K. Shinozaki, Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses using a full-length cDNA microarray, Plant Cell 13 (2001) 61–72.
[92] J.L. Reyes, F. Campos, H. Wei, R. Arora, Y.I. Yang, D.T. Karlson, A.A. Covarrubias, Functional dissection of hydrophilins during in vitro freeze protection, Plant Cell Environ. 31 (2008) 1781–1790.
[93] M. Kimura, Y. Yamamoto, M. Seki, T. Sakurai, M. Sato, K. Shinozaki, K. Manabe, M. Matsui, Identification of *Arabidopsis* genes regulated by high light stress using cDNA microarray, Plant Cell Physiol. 43 (2002) S159-S159.
[94] R. Catalá, E. Santos, J.M. Alonso, J.R. Ecker, J.M. Martínez-Zapater, J. Salinas, Mutations in the Ca^{2+}/H^{+} transporter CAX1 increase CBF/DREB1 expression and the cold-acclimation response in *Arabidopsis*, Plant Cell Online 15 (2003) 2940–2951.
[95] A. Nishizawa, Y. Yabuta, S. Shigeoka, Galactinol and raffinose constitute a novel function to protect plants from oxidative damage, Plant Physiol. 147 (2008) 1251–1263.
[96] P.J. Seo, M.J. Kim, J.S. Song, Y.S. Kim, H.J. Kim, C.M. Park, Proteolytic processing of an *Arabidopsis* membrane-bound NAC transcription factor is triggered by cold-induced changes in membrane fluidity, Biochem. J. 427 (2010) 359–367.
[97] H. Sakamoto, K. Maruyama, Y. Sakuma, T. Meshi, M. Iwabuchi, K. Shinozaki, K. Yamaguchi-Shinozaki, *Arabidopsis* Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions, Plant Physiol. 136 (2004) 2734–2746.
[98] R. Bermejo, T. Capra, R. Jossen, A. Colosio, C. Frattini, W. Carotenuto, A. Cocito, Y. Doksani, H. Klein, B. Gomez-Gonzalez, A. Aguilera, Y. Katou, K. Shirahige, M. Foiani, The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores, Cell 146 (2011) 233–246.
[99] N.A. Delk, K.A. Johnson, N.I. Chowdhury, J. Braam, CML24, regulated in expression by diverse stimuli, encodes a potential Ca^{2+} sensor that functions in responses to abscisic acid, daylength, and ion stress, Plant Physiol. 139 (2005) 240–253.
[100] M. Mistry, P. Pavlidis, Gene ontology term overlap as a measure of gene functional similarity, BMC Bioinformatics 9 (2008) 327.

Jin-Xing Liu received the B.S. degree in Electronic Engineering in 1997 from the Shandong University, the M.S. degree in Control Theory and Control Engineering in 2003 from the QuFu Normal University and the Ph.D. degree in Pulp and Paper Engineering in 2008, from the South China University of Technology. He is currently a Postdoctoral Fellow in Shenzhen Graduate School, Harbin Institute of Technology, China. His research interests include feature extraction, pattern recognition and bioinformatics.

Chun-Hou Zheng received the B.S. degree in Physics Education in 1995 and the M.S. degree in Control Theory and Control Engineering in 2001 from the QuFu Normal University and the Ph.D. degree in Pattern Recognition and Intelligent System in 2006, from the University of Science and Technology of China. From February 2007 to June 2009 he worked as a Postdoctoral Fellow in the Hefei Institutes of Physical Science, Chinese Academy of Sciences. From July 2009 to July 2010 he worked as a Postdoctoral Fellow in the Department of Computing, The Hong Kong Polytechnic University. He is currently a Professor in the College of Electrical Engineering and Automation, Anhui University, Hefei, Anhui, China. His research interests include Pattern Recognition and Bioinformatics.

Yong Xu was born in Sichuan, China, in 1972. He received his B.S. degree, M.S. degree in 1994 and 1997, respectively. He received the Ph.D. degree in Pattern recognition and Intelligence System at the NUST (China) in 2005. Now he works at the Shenzhen Graduate School, Harbin Institute of Technology. His current interests include feature extraction, biometric, face recognition, machine learning and image processing.