

RESEARCH ARTICLE

Open Access



Temperature expression patterns of genes and their coexpression with LncRNAs revealed by RNA-Seq in non-heading Chinese cabbage

Xiaoming Song^{1,2}, Gaofeng Liu¹, Zhinan Huang¹, Weike Duan¹, Huawei Tan¹, Ying Li¹ and Xilin Hou^{1*}

Abstract

Background: Non-heading Chinese cabbage (NHCC, *Brassica rapa* ssp. *chinensis*) is an important leaf vegetable grown worldwide. However, little is known about the molecular mechanisms underlying tolerance for extreme temperature in NHCC. The limited availability of NHCC genomic information has greatly hindered functional analysis and molecular breeding.

Results: Here, we conduct comprehensive analyses of cold and heat treatments in NHCC using RNA-seq. Approximately 790 million paired-end reads representing 136,189 unigenes with N50 length of 1705 bp were obtained. Totally, 14,329 differentially expressed genes (DEGs) were detected. Among which, 10 DEGs were detected in all treatments, including 7 up-regulated and 3 down-regulated. The enrichment analyses showed 25 and 33 genes were enriched under cold and heat treatments, respectively. Additionally, 10,001 LncRNAs were identified, and 9,687 belonged to novel LncRNAs. The expression of miRNAs were more than that of pri-miRNAs and LncRNAs. Furthermore, we constructed a coexpression network for LncRNAs and miRNAs. It showed 67 and 192 genes were regulated by LncRNAs under cold and heat treatments, respectively. We constructed the flowchart for identifying LncRNAs of NHCC using transcriptome. Except conducting the *de novo* transcriptome analyses, we also compared these unigenes with the Chinese cabbage proteins. We identified several most important genes, and discussed their regulatory networks and crosstalk in cold and heat stresses.

Conclusions: We presented the first comprehensive characterization for NHCC crops and constructed the flowchart for identifying LncRNAs using transcriptome. Therefore, this study represents a fully characterized NHCC transcriptome, and provides a valuable resource for genetic and genomic studies under abiotic stress.

Keywords: Cold and heat stresses, RNA-Seq, Expression pattern, LncRNA, Coexpression network, Non-heading Chinese cabbage

Background

Nowadays, frequent occurrences of abnormal weather events have been observed all over the world, such as drought and extreme temperature. These stresses seriously impact plant growth and crops production [1, 2]. Recently, several progresses have been made about the

identification of stress-related genes, which potentially are able to increase the plant tolerance [3–5]. Understanding the molecular mechanism of the abiotic stresses response is important to improve tolerance using molecular techniques.

Generally, these stress signals are converted into cellular responses through two ways, including ABA-dependent and ABA-independent signaling pathways [6, 7]. For the former, ABA is accumulated under osmotic stress caused by drought. It regulates the expression of gene under osmotic stress conditions [6, 8]. The ABA-responsive

* Correspondence: hxl@njau.edu.cn

¹State Key Laboratory of Crop Genetics and Germplasm Enhancement/Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in East China, Ministry of Agriculture, Nanjing Agricultural University, Nanjing 210095, China

Full list of author information is available at the end of the article



element (ABRE) is the major cis-element for ABA-responsive gene expression. ABRE-binding protein (AREB) and ABRE-binding factor (ABF) control gene expression in ABA-dependent manner [1, 9–11]. The molecular studies have revealed that ABA-independent pathway is also important for stress tolerance in plants. Dehydration-responsive element binding protein 1(DREB1)/C-repeat binding factor (CBF) and DREB2 are mainly involved in cold and heat stresses, respectively [12–15]. The DREB/CBF transcription factors (TFs) could specific bind to DRE/CRT cis-elements in promoter of target genes [14, 16, 17]. Several proteins, such as ICE1, ZAT12, CAMTA3, and MYB15, have been identified as regulators of DREB1/CBF genes [12, 18, 19]. In addition, NAC and MYB/MYC also regulate abiotic stress-responsive genes expression [20, 21]. The studies have demonstrated that there are interactions between ABA signaling pathway and other signaling factors in stress responses [1, 22, 23].

Until now, a large number of transcriptome sequencing projects have been conducted in many species. Genome-wide analyses have dramatically improved the efficiency of gene identification [16]. In *Arabidopsis*, about 30 % of the transcripts were related with abiotic stresses, and 2,409 genes played important roles in cold, salt, and drought stresses [24]. In *chrysanthemum*, 8,558 dehydration-responsive transcripts were detected using RNA-seq [25]. In wheat, about 2 % of the wheat genes were related with the cold stress [26]. In *Populus* and switchgrass, heat responsive genes were also identified by transcriptome sequencing [27, 28]. In *A. mongolicus*, 9,309 up-regulated and 23,419 down-regulated genes were identified under cold stress [29].

Brassica rapa contains several subspecies, such as Chinese cabbage (*B. rapa* ssp. *pekinensis*), NHCC, and turnip (*B. rapa* ssp. *rapa*) [30, 31]. The genome of Chinese cabbage had been sequenced, however, there is little information about the NHCC genome and gene dataset. Therefore, we conducted *de novo* assembly and gene annotation without prior genome information in this study. NHCC is one of the most important vegetables in China, and now is cultivated extensively worldwide. It is inevitable injured by low or heat stresses, which can directly lead to the production decrease and affect edible quality. The heat stress can affect the photosynthesis, and even induce the occurrence of several diseases, such as downy mildew, soft rot and virus diseases. The physiological change of temperature response mediated by several genes has been reported in model plants [32, 33]. However, little is known about the temperature-regulated genes and the related pathways in NHCC.

In this study, we conducted the comprehensive characterization for NHCC using RNA-seq, and explored the effect of low and heat temperature on global

change. We identified several most important genes in temperature response, and discussed their regulatory networks and crosstalk in cold and heat stresses. Using Illumina sequencing technology, we generated over 85 billion base of high quality sequence, and identified a larger number of differentially and specifically expressed transcripts. Furthermore, we also identified lots of LncRNAs, and constructed the coexpression network of LncRNAs and protein encoding genes using this transcriptome dataset.

Results and discussion

RNA sequencing and *de novo* assembly of NHCC transcriptome

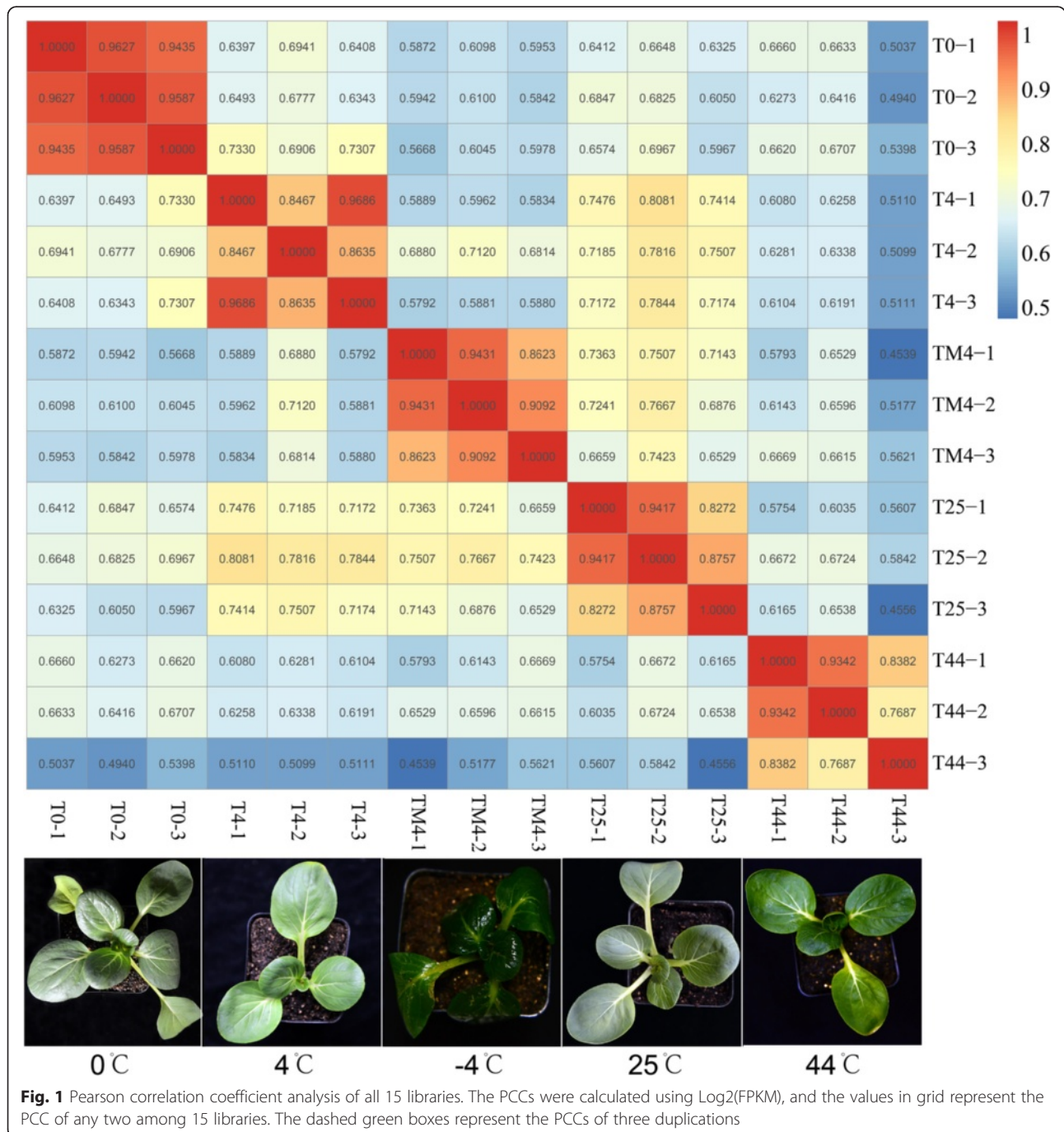
To obtain a global overview of NHCC transcriptome under different temperature treatments, we constructed and sequenced 15 RNA-Seq libraries, including cold treatments (4, 0 and -4 °C), heat treatment (44 °C), and normal condition (25 °C). For each temperature, three samples as the biological replications were sequenced using Illumina HiSeq™ 2000. The base quality of reads was checked using FastQC (Additional file 1: Figure S1). We used relatively stringent criteria for quality control by removing the reads with adaptors and the low quality. Finally, 790,269,418 clean pair-end (PE) reads consisting of 71.12 billion nucleotides (nt) were obtained with an average GC content of 47.30 % (Table 1, Additional file 2: Table S1). After the first assembly, 1,596,012 contigs

Table 1 The summary of the sequencing and assembly

Samples	NHCC
Total raw reads	857,423,614
Total clean reads	790,269,418
Total clean nucleotides (nt)	71,124,247,620
Q20 percentage	98.05 %
N percentage	0.00 %
GC percentage	47.30 %
Contig	
Total number	1,596,012
Total length (nt)	542,865,388
Mean length (nt)	343
N50	593
Unigene	
Total number	136,189
Total length (nt)	153,124,745
Mean length (nt)	1124
N50	1705
Total consensus sequences	136,189
Distinct clusters	73,514
Distinct singletons	62,675

were obtained for all libraries, and the total length over 542.8 Mb (Table 1). The contigs were further joined into 136,189 unigenes using paired-end information and gap filling process. The total length of all unigenes was 153.1 Mb, and the mean length of unigene was 1124 bp (Table 1, Additional file 2: Table S2). The PE sequencing not only increases the depth, but also improves *de novo* assembly efficiency. The N50 achieved 1705 bp, which was larger than most plants *de novo* assembled by RNA-

Seq, such as radish (1095 bp), wax gourd (1132 bp), and celery (1088 bp) [34–36]. This phenomenon indicated that the high quality and accuracy of our assembled transcripts. Based on FRKM, we measured the correlation of three repeats for each temperature. The results showed that there was a good correlation among three repeats. The pearson's correlations of almost all comparisons were larger than 85 % (Fig. 1, Additional file 1: Figure S2).



Functional annotation and classification of the assembled unigenes

Among all 136,189 unigenes, 121,744 (89.39 %) unigenes significantly matched a sequence in at least one of the public databases, including NCBI non-redundant protein (Nr), Gene Ontology (GO), Clusters of Orthologous Group (COG), Swiss-Prot and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Additional file 2: Table S3). The size distribution of BLAST-aligned coding sequence (89.31 %) and predicted proteins are analyzed (Additional file 1: Figure S3a,b). The remaining unigenes that did not match these databases were analyzed by three programs to predict coding regions. Finally, 2793, 2491, and 3119 coding sequences were predicted by ESTScan, CPC, and CNCI programs, respectively (Additional file 1: Figure S3c, Additional file 2: Table S4). The venn diagram showed that there were 684 coding sequences predicted by these three programs, so these genes were relatively reliable as coding genes (Additional file 1: Figure S3d). A total of 105,217 coding transcripts were predicted in our study. Then we aligned these unigenes with the proteins of Chinese cabbage (E-value $<10^{-10}$, identity $>70\%$). The results showed that 93,046 unigenes could align to the 3,2640 Chinese cabbage proteins (Fig. 2a). In addition, we found that over 70 % NHCC transcripts could match with more than 1 Chinese cabbage genes (Fig. 2b). This phenomenon might be caused by the genome duplication of *B. rapa*. The sequences without a homologous hit might represent novel genes in the genome, and some of them might be the specifically expressed in NHCC for temperature treatments. In addition, they also might be the non-coding, alternative transcription, lineage-specific or high allelic variant unigenes.

For Nr annotations, 104,363 unigenes matched in this database (Additional file 2: Table S3). The result indicated that 89.20 % of the top hits showed strong homology with the E-value $<1E-15$ (Additional file 1: Figure S4a). The distribution pattern showed that 87.90 % of unigenes had a similarity higher than 60 % (Additional file 1: Figure S4b). The majority annotated unigenes

were corresponded to the known plant genes, with 41.50 and 39.60 % matching with *A. thaliana* and *A. lyrata*, respectively (Additional file 1: Figure S4c). A total of 96,314 unigenes were assigned at least one GO term, and all GO terms were classified into three groups and further divided into 55 functional subgroups (Additional file 1: Figure S5). Overall, only 45,750 unigenes were assigned to COG classification (Additional file 2: Table S3). Among 25 COG categories, the cluster for 'general functions prediction only' (36.84 %) represented the largest group, followed by 'Transcription', and 'Replication, recombination and repair' (Additional file 1: Figure S6). To identify the biological pathways activated in NHCC, the assembled unigenes were annotated with KEGG. A total of 66,419 unigenes were significantly matched in this database, and were assigned to 128 KEGG pathways (Additional file 2: Table S3,5). The result showed that three largest pathway groups were metabolic pathways (ko01100, 21.12 %), biosynthesis of secondary metabolites (ko01110, 9.72 %), and plant-pathogen interactions (ko04626, 7.27 %). Following these three groups, the plant hormone signal transduction (ko04075) was about 6.60 % of all annotated genes. The level 1 of this pathway was 'Environmental Information Processing', and the level 2 of it was 'Signal transduction' in the KEGG database.

Temperature-dependent gene expression patterns identified by RNA-Seq in NHCC

To view the gene expression, all genes were divided into three categories, including highly (FPKM >50), medium ($5 < \text{FPKM} \leq 50$), and lowly (FPKM ≤ 5) expressed in each library. The results showed that most genes belonged to lowly expressed, followed by medium, and highly expressed (Additional file 1: Figure S7). The purple line shows the cumulative expressed gene number as the library number increased, and 134,980 genes were detected by all libraries.

To evaluate the temperature decrease course and temperature-dependent transcriptomic activities during cold-resistance process in NHCC, we performed a

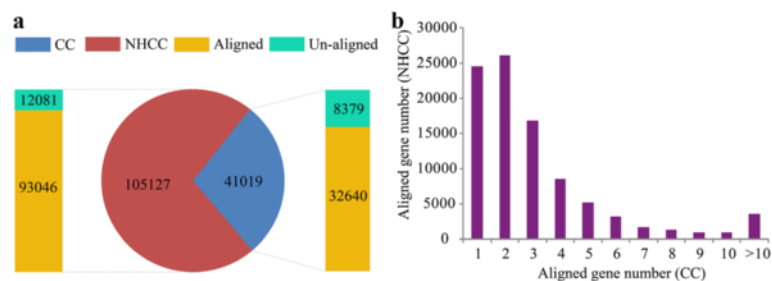


Fig. 2 The alignment analysis for NHCC RNA-Seq transcripts and Chinese cabbage (CC) proteins. **a** The summary of the aligned and un-aligned transcripts. **b** The distribution of the aligned transcripts of the NHCC

temperature decrease course differential gene expression analysis by comparing any two adjacent cold treatments, using the higher temperature treatment as the denominator. There were 27 (3³) possible patterns, including those that increased across all treatment boundaries, termed ‘up-up-up’ (UUU); those that were similar across all boundaries, termed ‘maintain-maintain-maintain’ (MMM); and those that decreased across all boundaries, termed ‘decrease-decrease-decrease’ (DDD). The results showed that genes were non-randomly represented across all patterns, and the overall temperature-dependent patterns are analyzed (Fig. 3, Additional file 2: Table S6). Only few genes continuously decreased (DDD) or increased (UUU) in expression accompanying temperature decrease, and the number were 3 (0.002 %) and 11 (0.008 %), respectively. However, majority genes (79,852, 58.63 %) belonged to the MMM pattern, and the expression almost unchanged over the temperature decrease process. In addition to MMM, the MMU (15,675) and DMM (13,999) were also contained more genes than other expression patterns.

Differential expressed genes detection and compare them among each treatment

To identify the temperature respond genes, 14,329 DEGs were detected between each temperature-treated and control library (FC >2 and q-value >0.8) (Fig. 4). All DEGs were used for clustering analysis, and obtained a

well cluster results (Fig. 5a). Three repeats for each temperature got together, and 0, 4 and 25 °C formed one group, while -4 and 44 °C formed another group. On the whole, all DEGs were divided into three groups, and defined as I, II, and III. Most DEGs in group I belonged to up-regulated genes under -4 °C treatment, while similar in other four temperatures. In group II, the expression of most DEGs under 25 and -4 °C was lower than that under other three temperatures. In group III, most DEGs had relatively low expression, except a few genes under 44 °C.

To further survey the interaction of these treatments for DEGs, we constructed venn diagram using DEGs of each treatment. There were 5445, 7430, 8056, and 488 DEGs under 4, 0, -4, and 44 °C, respectively. Among of them, 1075, 2865, 5254, and 286 belonged to each treatment-specific DEGs (Fig. 5b). Interestingly, we found 10 DEGs were detected in all treatments, including 7 up-regulated and 3 down-regulated genes (Additional file 1: Figure S8). The functional annotation showed that most of them belonged to the stresses related protein, such as LEA14 and KIN2 (Table 2). In addition, we also conducted qRT-PCR experiment to verify the accuracy of the RNA-Seq. The results showed that the expression trends of all genes were consistent with the RNA-Seq, and most genes were also significant differently expressed (*p*-value <0.01) (Additional file 2: Table S7, Additional file 1: Figure S9). Among all DEGs,

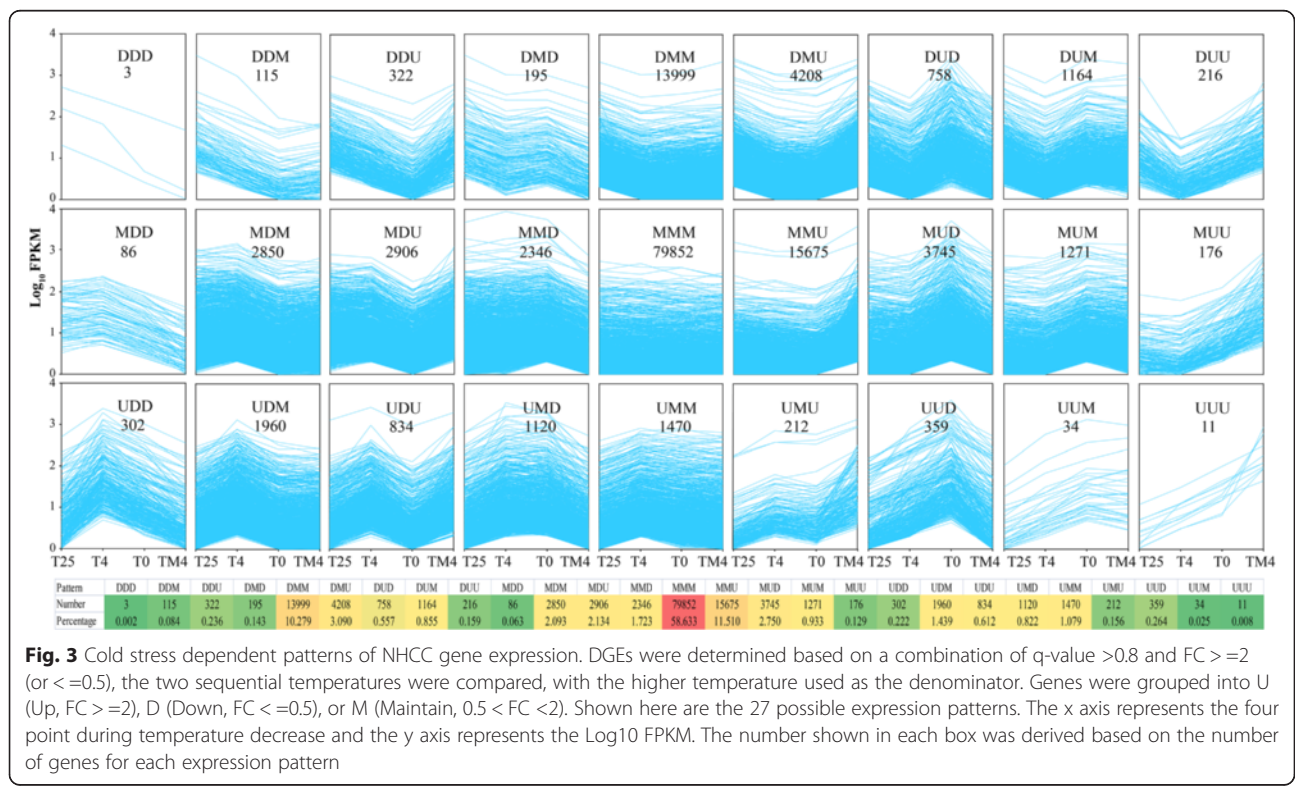
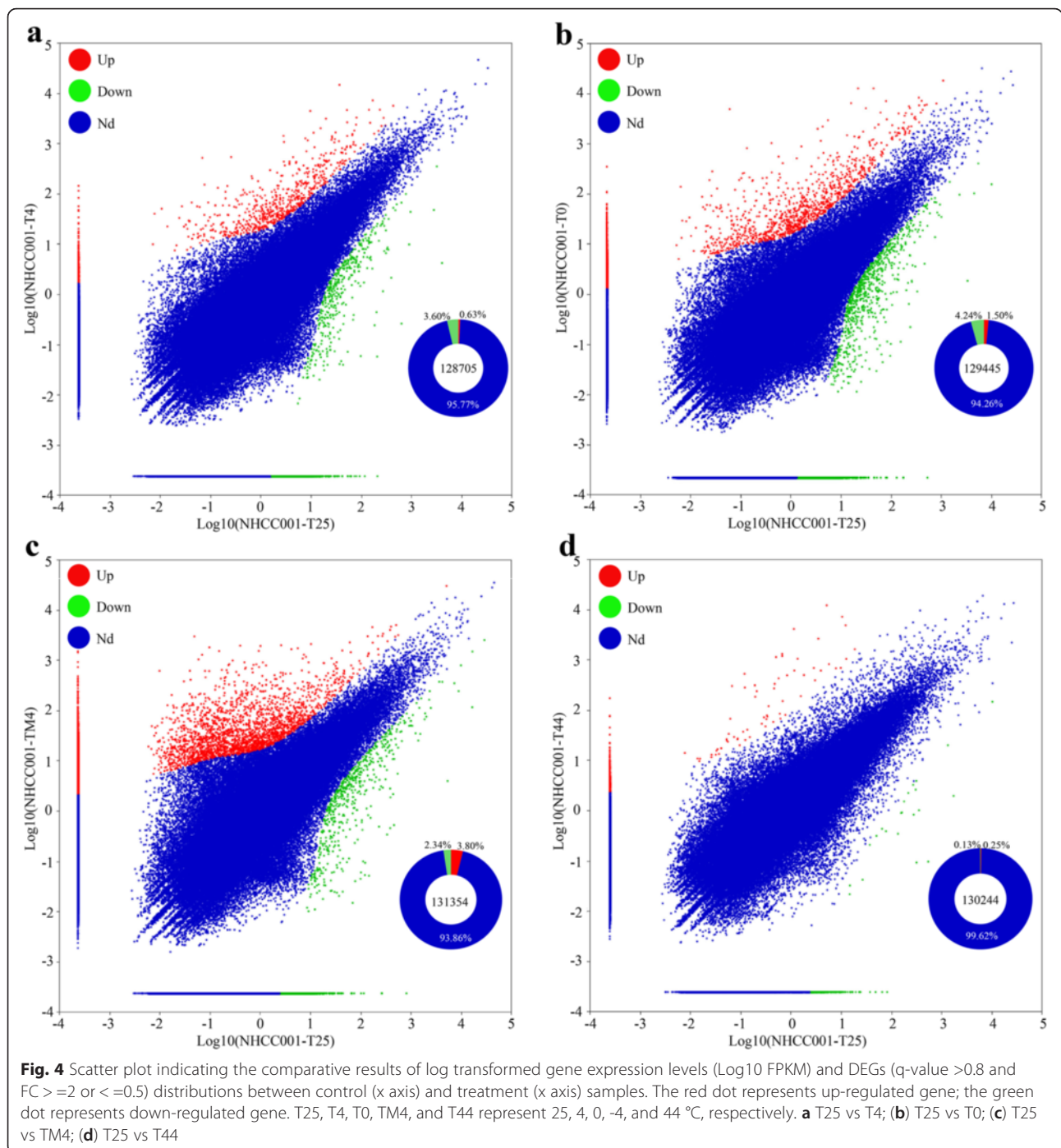


Fig. 3 Cold stress dependent patterns of NHCC gene expression. DEGs were determined based on a combination of q-value >0.8 and FC >=2 (or <=0.5), the two sequential temperatures were compared, with the higher temperature used as the denominator. Genes were grouped into U (Up, FC >=2), D (Down, FC <=0.5), or M (Maintain, 0.5 < FC <2). Shown here are the 27 possible expression patterns. The x axis represents the four point during temperature decrease and the y axis represents the Log10 FPKM. The number shown in each box was derived based on the number of genes for each expression pattern

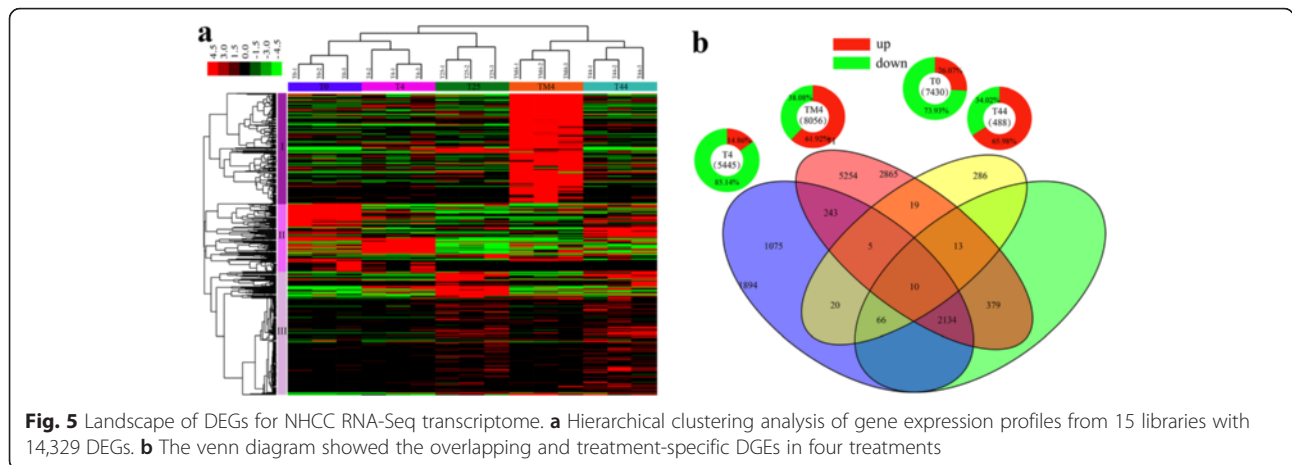


809 were up-regulated, and 1,743 were down-regulated under 4 °C; 1,937 were up-regulated, 5,493 were down-regulated under 0 °C; 4,988 were up-regulated, 3,068 were down-regulated under -4 °C. However, there were fewer DEGs under 44 °C than cold treatment, with only 322 were up-regulated and 166 were down-regulated genes (Additional file 1: Figure S10). Among these DEGs, the most treatment-specific up-regulated genes (4804) were detected under -4 °C, and the most

treatment-specific down-regulated genes (1377) were detected under 0 °C (Additional file 1: Figure S8).

The enrichment analyses revealed most DEGs related with cold and heat stresses

To understand the function of DEGs, we have conducted the GO enrichment analyses using all unigenes as background (Additional file 2: Table S8). Under 4 °C, several cold related GO categories were significantly



enriched, such as response to desiccation, response to cold, response to temperature stimulus, and cold acclimation. The photosynthesis, light harvesting, and translation categories were the top enrichment under 0 °C. Of course, the cold related categories were also enriched at the top 10 categories. Under -4 °C, the mainly enrichment categories were secondary metabolite biosynthetic process, S-glycoside metabolic process, glycosinolate, and glucosinolate metabolic process. However, we did not detect directly cold related categories as 4 and 0 °C at the top 10 categories. This phenomenon indicated the regulatory mechanism might exit several differences between chilling (<7 °C) and freezing (<0 °C) temperatures, which was also consistent with the previous report [37]. Under 44 °C, the mainly enrichment categories were response to high light intensity, heat acclimation, photosynthesis, and response to heat. In addition, the photosynthesis was also enriched under heat stress, which indicated that all temperature stresses could affect plant photosynthesis. We also analyzed genes belonged to the GO enrichment categories. Among all treatments, the most specific genes (2713) were found under -4 °C,

followed by 0 °C (1115), 4 °C (528), and 44 °C (79). In addition, 90 genes were identified under all three cold treatments, and 5 genes were detected by all the cold and heat treatments (Additional file 1: Figure S11a).

In addition, we mapped DEGs to terms in KEGG database to identify significantly enriched pathways. Among the mapped pathways, 20, 11, 22, and 5 pathways were significantly enriched (Qvalue < 0.01) under 4, 0, -4, and 44 °C treatments, respectively (Additional file 1: Figure S12, Table S9). Notably, common enrichments were observed in photosynthesis pathway, metabolic pathway, and photosynthesis-antenna proteins pathway in all treatments. This results indicated that the cold and heat stresses affected the expression of genes involved in these pathways. Most enriched pathways were also detected by the previous reports, which partly reflected the accuracy of our results [24, 27, 29]. Interestingly, we found transcripts involved in protein processing in endoplasmic reticulum pathway were significantly enriched under 44 °C treatment, while it did not enrich in cold treatment. This phenomenon indicated that this pathway might only play roles in heat resistance. We also

Table 2 The expression and functional annotation of 10 DEGs identified by all the cold and heat treatments. The up/down-regulated genes were identified by comparing the treatment (T4, T0, TM4, T44) and control (T25)

GeneID	T25	T4	T0	TM4	T44	Regulation	Annotation
CL4489.Contig2	0.02	24.17	22.42	26.90	11.41	Up	Unknown protein
CL10212.Contig2	0.00	6.52	4.30	7.03	4.30	Up	Glycosyl transferase family 1 protein
CL3727.Contig8	153.31	6.14	26.06	2.08	2.71	Down	Hypothetical protein ARALYDRAFT_910104
CL11270.Contig1	0.78	21.25	81.51	13.92	43.86	Up	At1g01470,a late embryogenesis abundant protein LEA14
Unigene519	0.12	209.31	1750.16	67.05	157.31	Up	BN28b, stress-induced protein KIN2 mRNA
CL8814.Contig1	139.12	12.75	2.07	1.84	3.27	Down	BN28a gene
Unigene16735	0.00	5.36	2.14	8.70	3.29	Up	Unknown protein
CL536.Contig12	0.00	3.12	7.09	10.66	8.15	Up	Wound-induced protein 1
Unigene50726	11.94	0.00	0.00	0.12	0.00	Down	ATP synthase CF1 epsilon chain,DM1-3-516-R44 chloroplast
CL2980.Contig1	8.18	203.56	272.71	191.73	317.44	Up	BN28a, stress-induced protein KIN2,Rapeseed KIN1 protein

analyzed the genes belonged to the KEGG enrichment categories. Among all treatments, the most specific genes (1052) were found under -4 °C, followed by 0 °C (461), 4 °C (160), and 44 °C (44). Forty-five genes were identified under all three cold treatments, and 2 genes were detected by all the cold and heat treatments (Additional file 1: Figure S11b). Furthermore, we also surveyed the enrichment genes identified by combing the GO and KEGG databases. The results showed that 25 and 33 genes were enriched in the two databases under cold and heat treatments, respectively (Additional file 1: Figure S11c,d). These enriched genes will greatly enhance the potential utilization in cold and heat stresses of NHCC.

Identification of abiotic stresses related transcription factors from DEGs

Given that TFs have a major effect on the network of temperature-responsive genes, we also identified the temperature-inducible TFs. Overall, the number of Dehydrin, Chloroa_b-bind, p450, AP2, PSI_PsaH, and EF-hand was more than other TFs in three cold treatments (Additional file 1: Figure S13a,b,c,d,e,f). However, many GST were identified under -4 °C, while they were absent under 0 and 4 °C. This phenomenon indicated that GST might play roles in the cold resistance below 0 °C. Interestingly, HSP70 and HSP20 were identified under 0 and 4 °C, indicating that there was a certain inherent association between cold and heat regulation. Under 44 °C, HSP20 was significant enriched, which

revealed that it played important roles in heat-resistance regulation (Additional file 1: Figure S13g,h). In all treatments, P450 and Chloroa_b-bind TFs were enriched, which indicated that cold and heat stresses had great impact on plant photosynthesis, and thus might affect crop yields. This revealed that they were related with plant photosynthesis, which was also consistent with previous reports [38, 39].

Among all treatments, the most specific TFs were identified under -4 °C (1781), followed by 0 °C (775), and 4 °C (313) in GO enrichment (Additional file 1: Figure S14a). In addition, 51 TFs were identified under all three cold treatments, and 2 genes were detected by all cold and heat treatments. Among all KEGG categories, the most specific TFs (819) were found under -4 °C, followed by 0 °C (368), and 4 °C (128) (Additional file 1: Figure S14b). Twenty-nine genes were identified under all three cold treatments, and 1 genes were detected by all cold and heat treatments. Combing GO and KEGG enrichment analyses, 17 and 31 TFs were enriched only under cold and heat treatments, respectively (Additional file 1: Figure S14c,d, Table 3, Additional file 2: Table S10).

AP2/ERF TFs mainly contained two subgroups, including CBF and DREB2. They interacted with DRE/CRT cis-element and regulated ABA-independent gene expression. The CBF controlled many gene expression under several stresses, such as drought, salinity and freezing stresses. The DREB2 mainly affected gene expression under osmotic and heat stresses, while it

Table 3 The intersection of differentially expressed transcription factors under cold stress in GO and KEGG enrichment categories

Unigene ID	Pfam ID	TF family	E-value for Pfam	T4 vs T25		T0 vs T25		TM4 vs T25	
				log2 ratio	Qvalue	log2 ratio	Qvalue	log2 ratio	Qvalue
CL10543.Contig2	PF00067.17	p450	8.90E-32	6.74	0.8445	6.32	0.8308	6.34	0.838
CL11270.Contig2	PF03168.8	LEA_2	2.10E-18	5.78	0.836	6.05	0.8397	3.24	0.8154
CL11755.Contig1	PF00295.12	Glyco_hydro_28	2.60E-09	-11.3	0.8584	-11.3	0.8635	-11.3	0.8281
CL13372.Contig2	PF00657.17	Lipase_GDSL	5.20E-28	-3.71	0.8189	-3.64	0.818	-2.64	0.806
CL3153.Contig1	PF00201.13	UDPGT	3.90E-26	3.32	0.82	3.33	0.8193	2.79	0.8159
CL6375.Contig2	PF00067.17	p450	1.90E-22	-3.79	0.8254	-2.98	0.8155	-4.06	0.8365
Unigene12001	PF00764.14	Arginosuc_synth	6.50E-52	-11.6	0.8841	-11.6	0.8882	-11.6	0.8574
Unigene16263	PF00206.15	Lyase_1	1.40E-29	-11.32	0.8599	-11.32	0.865	-11.32	0.8298
Unigene20049	PF00314.12	Thaumatococin	1.50E-16	-4.95	0.8083	-3.92	0.8048	-3.35	0.8017
Unigene20728	PF01676.13	Metalloenzyme	6.40E-21	-11.15	0.8443	-11.15	0.8502	-11.15	0.8126
Unigene22237	PF00504.16	Chloroa_b-bind	2.00E-48	2.49	0.8025	3.69	0.8241	3.23	0.8285
Unigene2424	PF00067.17	p450	6.80E-07	6.18	0.806	6.67	0.8301	7.13	0.8527
Unigene2446	PF00206.15	Lyase_1	3.00E-26	-11.94	0.9086	-11.94	0.9119	-11.94	0.8861
Unigene24606	PF02800.15	Gp_dh_C	4.60E-51	-12.68	0.9492	-12.68	0.9515	-12.68	0.9354
Unigene26264	PF02775.16	TPP_enzyme_C	1.10E-10	-11.31	0.8594	-11.31	0.8645	-11.31	0.8292
Unigene42146	PF00006.20	ATP-synt_ab	2.30E-33	-11.43	0.8698	-11.43	0.8745	-11.43	0.841
Unigene50595	PF00764.14	Arginosuc_synth	2.60E-26	-11.15	0.8443	-11.15	0.8502	-11.15	0.8126

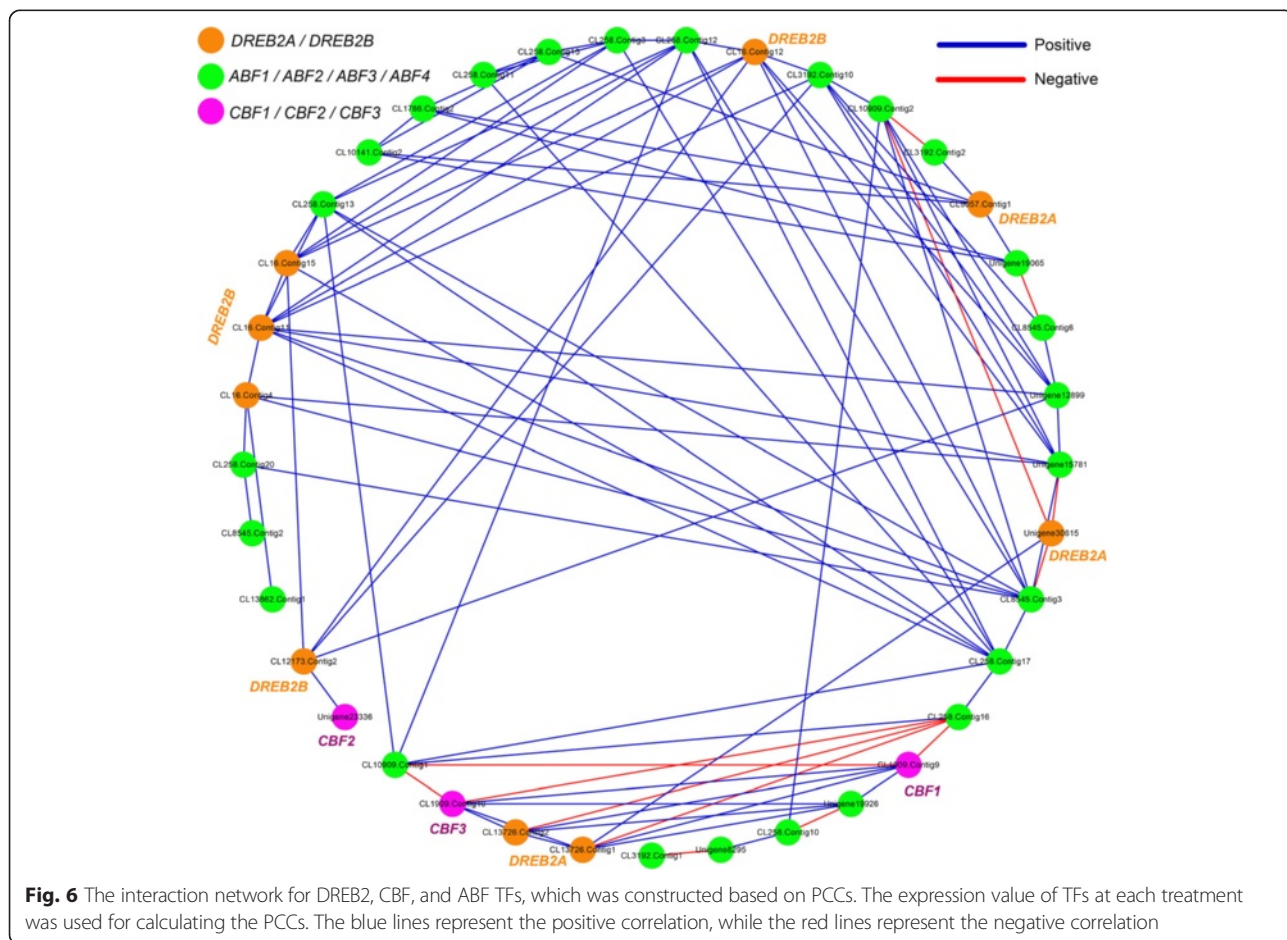
slightly played role in cold stress [1]. Although the functions of these TFs were well explained, the interaction among them was rarely reported, especially in *B. rapa*. Therefore, we conducted correlation analysis for these TFs, and constructed the interaction network of them using expression values. A total of 38 unigenes were detected using BLAST alignment with ABFs, CBFs, and DREB2 of *Arabidopsis* (Additional file 2: Table S11). The Pearson correlation coefficient (PCC) between two of these TFs were calculated using the expression value. Then the interaction network was constructed using partial connections with the PCC larger than 80%. Finally, this network contained 95 connections, including 82 positive and 13 negative connections (Fig. 6). This phenomenon revealed that most connections belonged to positive relationship among of these TFs. However, we noted that CL258.Contig16, a ABF TF, had negative connections with two DREB2A (CL13726.Contig1, CL13726.Contig2), CBF1 (CL1909.Contig9), and CBF3 (CL1909.Contig10).

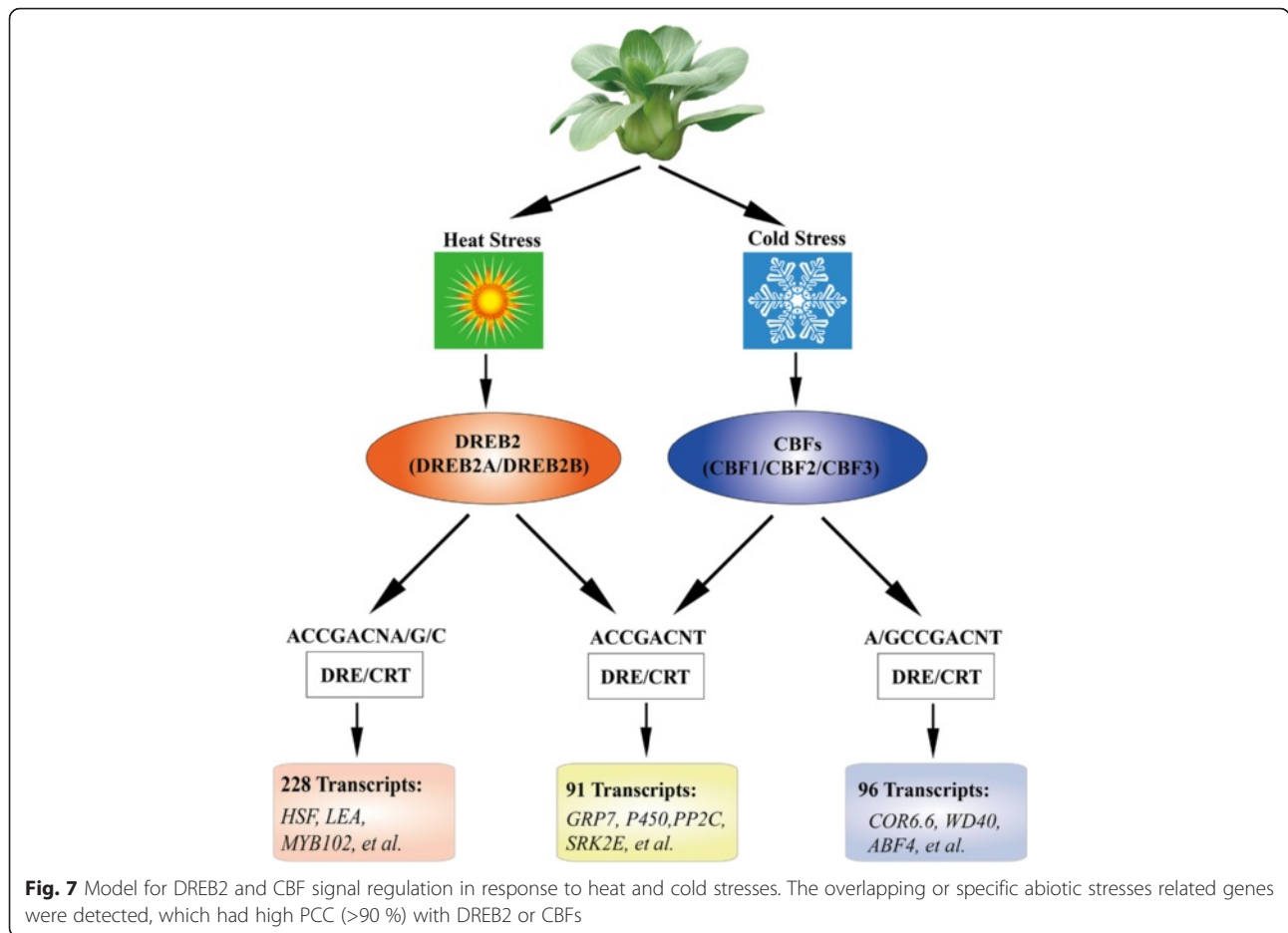
To analyze the correlation of abiotic stresses and these TFs, we collected mainly cold and heat stresses related genes according to previous reports [1, 12]. Then the candidate heat and cold related genes in NHCC were

identified using BLAST alignment with the collected genes. We calculated PCC of these candidate genes and CBF or DREB2. The PCC values larger than 0.8 were selected to construct interaction network (Additional file 1: Figure S15). This network showed that there were more negative connections in DREB2 than that in CBF, which only contained four negative connections. Among all connections, we identified 228 transcripts, which had high PCC (>90%) with DREB2, such as HSF, LEA, and MYB102. By CBFs, 96 transcripts were also identified, such as COR6.6, WD40, and ABF4. Moreover, 91 transcripts were detected by both DREB2 and CBF, including GRP7, P450, PP2C, and SRK2E (Fig. 7, Additional file 2: Table S12).

Identification and characterisation of NHCC LncRNA using RNA-seq

To identify potential LncRNA in NHCC, all sequences from NHCC transcriptome dataset were used. Based on previous reports [40–42], we designed the pipeline for LncRNA analyses. Finally, 10,001 LncRNAs were identified after a series filtering, including transcript length, coding potential, ORF size, and the exclusion of other





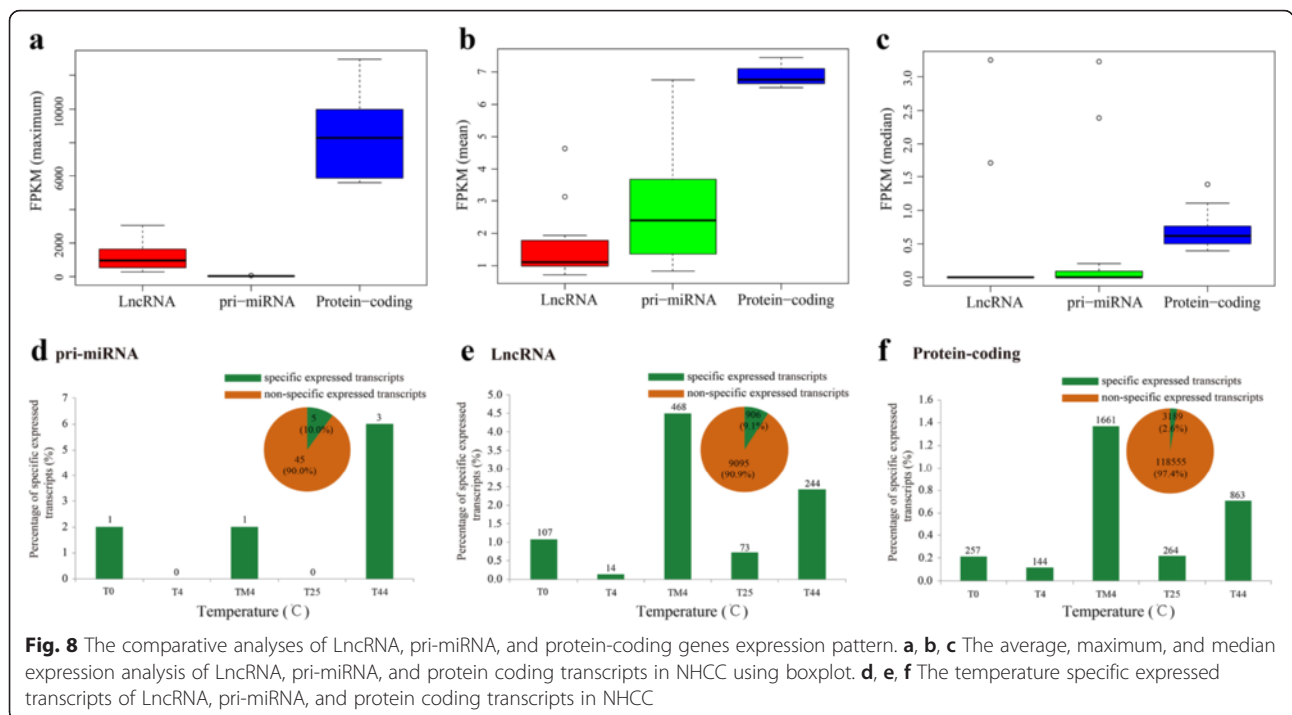
ncRNA. Among these LncRNAs, 9,687 belonged to the novel LncRNAs. In addition, we also identified 50 pri-miRNA through comparing with Rfam and miRBase databases for comparative analyses.

LncRNA gene expression analyses and differently expressed LncRNAs identification

LncRNA is a group of endogenous RNAs that function as regulators of gene expression, which are involved in developmental and physiological processes [43]. They are longer than 200 bp, and several of them can also act as primary transcripts for the production of short RNAs [44]. We assessed the expression pattern under different temperatures using the expressed LncRNA (FPKM >0). A total of 2,236 LncRNAs were expressed in all the treatments. We observed that 73, 14, 107, 468, and 244 LncRNAs belonged to the temperature specificity for T25, T4, T0, TM4, and T44, respectively (Additional file 1: Figure S16). However, a hierarchical clustering of samples showed most LncRNA was low expression in each temperature (Additional file 1: Figure S17). In this study, 50 pri-miRNA, 10,001 LncRNAs, and 121,744 protein coding transcripts were identified. The average, maximum,

and median expression values of these three type transcripts were calculated for comparative analyses. The results showed that the expression of protein coding transcripts were more than that of pri-miRNA and LncRNA (Fig. 8a,b,c, Additional file 2: Table S13). Most pri-miRNA and LncRNA had relatively low expression, which was consistent with the previous reports [45–47], indicating it was a common property of LncRNA. We also investigated the temperature specific expressed transcripts (SETs) among these three types transcripts. The results showed that 10.0 % pri-miRNA, 9.1 % LncRNA, and 2.6 % protein coding transcripts were detected as SETs. For pri-miRNA, 3 SETs were detected in 44 °C, while no SET was found in 4 and 25 °C (Fig. 8d). For LncRNA, the most SETs were identified in -4 °C (468), followed by 44 °C (244), and 0 °C (107) (Fig. 8e). The similar trends were also found in the protein-coding transcripts, except the inverse of 0 and 25 °C (Fig. 8f).

To further investigate the share or specific of these treatments for differently expressed LncRNAs (DELs), we conducted the venn diagram analyses. The results showed that 91, 418, 441, and 34 specifically expressed LncRNAs were identified under 4, 0, -4, and 44 °C,



respectively (Additional file 1: Figure S18a). However, we not detected the share LncRNAs among all these treatments. Among of these DELs, most of them were down-regulated, and there were the most down-regulated LncRNAs under 0 °C (1103) than other three treatments (9 ~ 934) (Additional file 1: Figure S18b). Interestingly, the DEL number under three cold stresses (964 ~ 1327) was more than that of 44 °C, which only contained 44 DELs. However, most LncRNAs (79.55 %) were up-regulated under 44 °C, while most LncRNAs belonged to the down-regulated under other three cold treatments (56.21 % ~ 96.89 %).

Construct the coexpression network between LncRNAs and protein-coding genes

We constructed a coexpression network for LncRNAs and protein-coding genes according to the previously proposed method [40]. For each treatments and for each pair of genes (LncRNA or protein-coding), we computed the PCCs of expression patterns using the expression values. We found that approximately 57.59 % were positive connections, and 42.40 % were negative connections (Additional file 1: Figure S19a). Among all connections, the PCCs of 29.27 % were between -0.4 to -0.2, and followed by 26.29 % connections were between 0.8 to 1. Furthermore, we identified 65,568,352 connections between protein-coding and protein-coding genes, among which, 42.11 % were positive connections, and 57.89 % were negative connections (Additional file 1: Figure S19b). However, the opposite result was found between

LncRNA and LncRNA connections, and the values were 50.12 and 49.88 % for positive and negative connections, respectively.

To be more accurate and intuitive showed the relationship between the LncRNAs and protein-coding genes, we selected the connections with the high correlation ($|PCC| > 0.95$) to construct the interaction network. Overall, the whole network constituted by these connections was divided into 8 clusters, including 3 large networks and 5 relative small networks (Additional file 1: Figure S20a). In the cluster 1, about 67 % connections with the PCC > 0.95, and ~33 % with the PCC = 1. The PCCs of all connections in the cluster 2 were larger than 0.95, but less than 1. Only three connections with PCC less than -0.95 were located in the cluster1 and cluster3. Most connections (362,213) in the networks belonged to positive, and only 636 connections were negative correlation (Additional file 1: Figure S20b). This phenomenon was also found in LncRNA vs LncRNA and protein-coding vs protein-coding genes. Among all of these three types, 99.72 % connections were positive, and 0.28 % were negative connections.

To further analyze the correlation between LncRNA and protein-coding genes under temperature stresses, we annotated the function of the target genes. A total of 67 target genes were regulated by LncRNAs under all three cold treatments, and comparing them with *Arabidopsis* (Additional file 2: Table S14). The annotation showed that most of them belonged to the cold respond proteins, such as CBF1, COR6.6, and LEA14. Similar, we

identified 192 target genes of lncRNAs under 44 °C treatment, and most of them belonged to the heat response genes, such as HSP, LTP, and CBF4 (Additional file 2: Table S15). Furthermore, we also found several target genes of lncRNAs under cold and heat stresses, such as KIN2. This phenomenon indicated that they might play important roles in the crosstalk between cold and heat stresses responses.

Conclusion

In this study, approximately 790 million paired-end reads representing 136,189 unigenes with a total length of 153.1 Mb were obtained. Only few genes were DDD or UUU expression patterns, and majority genes belonged to MMM pattern during the temperature decrease process. 14,329 DEGs were detected between at least one treatment and control library. Among which, 10 DEGs were identified in all treatments, including 7 up-regulated and 3 down-regulated genes. The enrichment analyses demonstrated that most temperature related categories were discovered under cold and heat treatments. Among the enrichment categories, 25 and 33 genes were identified in both CO and KEGG databases under cold and heat treatments, respectively.

Totally, 10,001 lncRNAs were identified from NHCC transcriptome dataset, and 9,687 belonged to novel lncRNAs. The analyses indicated the expression of protein coding transcripts were higher than that of pri-miRNA and lncRNA. We constructed a coexpression network for lncRNAs and protein-coding genes. A total of 67 and 192 target genes were regulated by lncRNAs under three cold and heat treatments, respectively. Furthermore, we also identified several shared target genes of lncRNAs under cold and heat treatments, which indicated that they might play important roles in the crosstalk between cold and heat stresses.

In conclusion, we conduct comprehensive analyses for cold and heat stresses in NHCC using RNA-seq, and identified numerous differentially and specifically expressed transcripts. Many important genes and TFs response to treatment stress were detected, and their crosstalk between cold and heat stress responses was discovered. In addition, we also identified large number of lncRNAs, and constructed the coexpression network of lncRNAs and protein encoding genes. This study provides a platform for elucidating physiologic responses to low and high temperature in *B. rapa*.

Methods

Plant materials, growth conditions, and treatments

The NHCC advanced inbred line, 'Suzhouqing', was used in this study. The surface-sterilized seeds were grown in pots containing a soil: vermiculite mixture (3:1) in a controlled-environment growth chamber programmed

for 16/8 h at 25/18 °C for day/night. Seedlings at the five-leaf stage were transferred to growth chambers set at -4, 0, 4 °C as cold treatments, 25 °C as control, and 44 °C as heat treatment under 4 h for RNA-Seq and qRT-PCR. Three samples of each treatment were generated from different batches of plants for three biological replicates. All leaf samples collected from control and treated plants were washed with distilled water, immediately frozen in liquid nitrogen, and stored at -80 °C for RNA extraction.

RNA extraction for transcriptome sequencing and RT-PCR validation

The RNA was isolated from leaves using RNA kit (Tiangen, Beijing, China) according to manufacturer's instructions. RNA samples were treated with RNase free DNase I to avoid DNA contamination. The RNA was reverse transcribed into cDNA using Prime Script RT reagent Kit (TaKaRa, Kyoto, Japan). The cDNA libraries were constructed using an mRNA-seq assay with a fragment length range of 200 bp (± 25 bp). Finally, the library was sequenced for paired-end reads of 90 bp using Illumina HiSeq™ 2000 platform, which was performed by the Beijing Genomics Institute (BGI) (<http://www.genomics.cn/index>). For qRT-PCR, the *actin* gene (*AF111812*) was used as an internal control to normalize the expression level of the target gene. Primer 5.0 designed the specific primers according to gene sequences. The qRT-PCR assays were performed with three biological and technical replicates. Each reaction was performed in 20- μ L reaction mixtures containing a diluted cDNA sample as template, SYBR Premix Ex Taq (2 \times) (TaKaRa, Kyoto, Japan) and gene-specific primers. qRT-PCR was performed according to our previous report [48]. The comparative Ct value method was adopted to analyze the relative gene expression. RNA expression levels relative to *actin* gene were calculated as $2^{-\Delta\Delta CT}$ according to a previous analysis [48, 49].

Data filtering and de novo assembly

Raw reads generated by Illumina HiSeq™ 2000 were initially processed to get clean reads through the following three steps. i) Remove reads with adaptors contamination; ii) Discard reads with ambiguous sequences "N" larger than 5 %; iii) Remove low quality reads, which contained more than 20 % Q < 20 bases [50]. In addition, we used FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) to check and visualize the quality of RNA-seq reads (Additional file 1: Figure S1). After filtering, all clean reads were assembled using a de novo assembly software Trinity [51]. Firstly, clean reads with a certain length of overlap were combined to generate contigs. Then, the paired-end reads were realigned to contigs to obtain unigene, which could identify different

contigs in the same transcript and ensure the interval among these contigs. The contigs in one transcript were assembled by Trinity and gained the sequence not being extended on either end, which defined as unigene [34]. Then, the TGICL program was used to delete redundant unigene and further assembled all unigenes to form a single set of non-redundant unigenes [52].

Gene expression quantification and differential expression analyses

RNA-Seq reads were aligned to the assembled transcripts using TopHat pipeline with the built-in Bowtie mapping program [53]. The expression of all unigenes was estimated by calculating read density as ‘fragments per kilobase of exon per million mapped reads’ (FPKM) [54]. The DEGs (FC >2, q-value >0.8) between normal and stress-treated conditions were identified using NOISeq (<http://www.bioconductor.org/>) [55]. GO enrichment analyses were performed using Blast2GO [56]. The temperature-dependent gene expression patterns were analyzed according to the previous report [57]. The comparisons were made between two adjacent temperatures—that is, 25 °C vs 4 °C, 4 °C vs 0 °C, and 0 °C vs -4 °C. A gene with FC >=2 was grouped into ‘up’ pattern, a gene with FC <=0.5 was grouped into ‘decrease’, and the remaining genes were grouped into ‘maintain’. Therefore, a gene was grouped to 1 out of 27 patterns, ranging from up-up-up (UUU), maintain-maintain-maintain (MMM), to decrease-decrease-decrease (DDD).

Functional annotation and classification of the transcripts

All assembled transcripts were annotated with the publicly available protein databases, including Nr (<http://www.ncbi.nlm.nih.gov/>), GO (<http://www.geneontology.org/>), COG (<http://www.ncbi.nlm.nih.gov/COG/>), Swiss-Prot protein (<http://www.expasy.ch/sprot/>), and KEGG (<http://www.genome.jp/kegg/>) databases using BLAST (E-value <10⁻⁵). Then, the best alignments were used to decide sequence direction and to predict coding regions of the unigenes. ESTScan software was used to decide sequence direction and coding regions when a unigene unaligned to none of the above databases [58]. WEGO software was used to conduct GO classification for understanding the distribution of gene function [59]. The unigenes were also aligned to COG database to predict and classify possible functions. In addition, KEGG was used to annotate the pathway of the unigenes.

LncRNA detection

To *de novo* detect LncRNAs using RNA-seq, we developed a flowchart according to previous reports with slightly modification (Additional file 1: Figure S21) [40, 41, 60, 61]. We applied several filters to ensure reliability

of LncRNAs. Firstly, all unigenes (136,189) were annotated using BLAST (E-value <10⁻⁵) alignment with NR,NT,Swiss-Prot,KEGG,COG, and GO databases. There were 14,445 unigenes un-annotated by any protein databases above mentioned. Among which, 14,300 unigenes selected for transcripts greater than or equal to 200 bp. Secondly, Coding-Non-Coding Index (CNCI, <http://www.bioinfo.org/software/cnci>) was applied on all candidate unigenes in order to distinguish protein-coding and non-coding sequences [62]. The unigenes with score <0 were defined as non-coding. In addition, Coding Potential Calculator (CPC, <http://cpc.cbi.pku.edu.cn>) was also used for identifying all candidate transcript models in order to assess their coding potential by a second independent method [63]. In order to extract potential non-coding transcripts with a high reliability from our dataset, all transcripts with a score (CPC <-1) were retained as potential non-coding. By combining these two methods, 10,930 unigenes were identified as potential non-coding RNAs. Thirdly, we discard transcripts with an ORF greater than 100 amino acids by ORF-Predictor (E-value <10⁻⁵), and 813 unigenes were discard.

Finally, 10,117 candidate transcripts were identified and compared against several non-coding RNA databases, including Rfam, miRBase, NONCODE with designated threshold value (E-value <10⁻⁵, identity >90 %) by BLAST [64–66]. Candidate transcript models with known protein motifs were discarded. We obtained 50 pri-miRNA sequences by comparing with the miRBase and Rfam databases. To identify LncRNA, we filtered other non-coding RNAs through comparing with Rfam and NONCODE databases. Finally, 10,001 unigenes were identified as LncRNA based on a series of analyses above mentioned. Furthermore, among which, 9,687 belonged to the novel novel LncRNAs by comparing with NONCODE databases.

Statistical analysis

The differential expression levels of genes under the different temperatures were clustered using Cluster program (<http://bonsai.hgc.jp/~mdehoon/software/cluster>), and visualized using Tree View software (<http://jtreeview.sourceforge.net/>) [67]. Using in-house Perl script, PCCs were calculated for correlation studies, including the three repeats correlation, LncRNA and target mRNA correlation, and key candidate mRNA-mRNA correlation. The coexpression interaction networks were constructed using Cytoscape (<http://www.cytoscape.org/>) according to PCC [68]. The numbers of specific and common genes were plotted using Venn diagram in R package [69].

Ethics and consent to participate

This article does not contain any human or animals data performed by any of the other committee.

Consent to publish

Not applicable.

Availability of data and materials

The RNA sequence dataset supporting the results of this article is available in available on NHCC Data Center under Project P002 (<http://nhccdata.njau.edu.cn/>).

Additional files

Additional file 1: Figures S1-S21. (PDF 2762 kb)

Additional file 2: Tables S1-S15. (XLS 9790 kb)

Abbreviations

DEGs: differentially expressed genes; DELs: differently expressed lncRNAs; FPKM: fragments per kilobase of exon per million mapped reads; NHCC: Non-heading Chinese cabbage; PCC: Pearson correlation coefficient; SETs: specific expressed transcripts.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The study was conceived by XH and XS. XS, WD and HT contributed to data collection and bioinformatics analysis. ZH and GL performed the experiments. XS and YL participated in preparing and writing the manuscript. All authors contributed to revising the manuscript. All authors had read and approved the final manuscript.

Acknowledgements

We thank Beijing Genomics Institute (BGI) for their help with the RNA-Seq and bioinformatics analyses.

Funding

This work was supported by the National Natural Science Foundation of China (Key Program, No. 31330067), the Science and Technology Support Program of Tangshan City (grant no. 15120204a), and the Shanghai '2011' program (no. ZF12051301).

Author details

¹State Key Laboratory of Crop Genetics and Germplasm Enhancement/Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in East China, Ministry of Agriculture, Nanjing Agricultural University, Nanjing 210095, China. ²Center of Genomics and Computational Biology, College of Life Sciences, North China University of Science and Technology, Tangshan, Hebei 063000, China.

Received: 22 September 2015 Accepted: 16 April 2016

Published online: 22 April 2016

References

- Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold and heat. *Front Plant Sci.* 2014;5:170.
- Glazebrook J. Genes controlling expression of defense responses in Arabidopsis—2001 status. *Curr Opin Plant Biol.* 2001;4(4):301–8.
- Rout G, Senapati S. *Stress Tolerance in Plants: A Proteomics Approach.* Springer India. 2013;359–386.
- Prasch CM, Sonnewald U. Simultaneous application of heat, drought, and virus to Arabidopsis plants reveals significant shifts in signaling networks. *Plant Physiol.* 2013;162(4):1849–66.
- Krasensky J, Jonak C. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J Exp Bot.* 2012;63(4):1593–608.
- Yoshida T, Mogami J, Yamaguchi-Shinozaki K. ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Curr Opin Plant Biol.* 2014;21C:133–9.
- Tuteja N. Abscisic Acid and abiotic stress signaling. *Plant Signal Behav.* 2007;2(3):135–8.
- Raghavendra AS, Gonugunta VK, Christmann A, Grill E. ABA perception and signalling. *Trends Plant Sci.* 2010;15(7):395–401.
- Fujita Y, Yoshida T, Yamaguchi-Shinozaki K. Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. *Physiol Plant.* 2013;147(1):15–27.
- Fujita Y, Fujita M, Shinozaki K, Yamaguchi-Shinozaki K. ABA-mediated transcriptional regulation in response to osmotic stress in plants. *J Plant Res.* 2011;124(4):509–25.
- Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *Plant J.* 2010;61(4):672–85.
- Chinnusamy V, Zhu JK, Sunkar R. Gene regulation during cold stress acclimation in plants. *Methods Mol Biol.* 2010;639:39–55.
- Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K, Yamaguchi-Shinozaki K. Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc Natl Acad Sci U S A.* 2006;103(49):18822–7.
- Maruyama K, Takeda M, Kidokoro S, Yamada K, Sakuma Y, Urano K, Fujita M, Yoshiwara K, Matsukura S, Morishita Y, et al. Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. *Plant Physiol.* 2009;150(4):1972–80.
- Song X, Li Y, Hou X. Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa ssp. pekinensis*). *BMC Genomics.* 2013;14:573.
- Tian DQ, Pan XY, Yu YM, Wang WY, Zhang F, Ge YY, Shen XL, Shen FQ, Liu XJ. De novo characterization of the Anthurium transcriptome and analysis of its digital gene expression under cold stress. *BMC Genomics.* 2013;14:827.
- Shinozaki K, Yamaguchi-Shinozaki K. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Curr Opin Plant Biol.* 2000;3(3):217–23.
- Doherty CJ, Van Buskirk HA, Myers SJ, Thomashow MF. Roles for Arabidopsis CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. *Plant Cell.* 2009;21(3):972–84.
- Kim Y, Park S, Gilmour SJ, Thomashow MF. Roles of CAMTA transcription factors and salicylic acid in configuring the low-temperature transcriptome and freezing tolerance of Arabidopsis. *Plant J.* 2013;75(3):364–76.
- Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. *Plant Physiol.* 2009;149(1):88–95.
- Saibo NJ, Lourenco T, Oliveira MM. Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. *Ann Bot.* 2009;103(4):609–23.
- Roychoudhury A, Paul S, Basu S. Cross-talk between abscisic acid-dependent and abscisic acid-independent pathways during abiotic stress. *Plant Cell Rep.* 2013;32(7):985–1006.
- Narusaka Y, Nakashima K, Shinwari ZK, Sakuma Y, Furihata T, Abe H, Narusaka M, Shinozaki K, Yamaguchi-Shinozaki K. Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis *rd29A* gene in response to dehydration and high-salinity stresses. *Plant J.* 2003;34(2):137–48.
- Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF. Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiol.* 2002;130(4):2129–41.
- Xu Y, Gao S, Yang Y, Huang M, Cheng L, Wei Q, Fei Z, Gao J, Hong B. Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. *BMC Genomics.* 2013;14:662.
- Winfield MO, Lu C, Wilson ID, Coghill JA, Edwards KJ. Plant responses to cold: Transcriptome analysis of wheat. *Plant Biotechnol J.* 2010;8(7):749–71.
- Chen J, Yin W, Xia X. Transcriptome Profiles of Populus euphratica upon Heat Shock stress. *J Appl Stat.* 2014;15(5):326–40.
- Li YF, Wang Y, Tang Y, Kakani VG, Mahalingam R. Transcriptome analysis of heat stress response in switchgrass (*Panicum virgatum* L.). *BMC Plant Biol.* 2013;13:153.

29. Pang T, Ye CY, Xia X, Yin W. De novo sequencing and transcriptome analysis of the desert shrub, *Ammopiptanthus mongolicus*, during cold acclimation using Illumina/Solexa. *BMC Genomics*. 2013;14:488.
30. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*. 2011;43(10):1035–9.
31. Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biol*. 2011;11:136.
32. Robinson SJ, Parkin IA. Differential SAGE analysis in Arabidopsis uncovers increased transcriptome complexity in response to low temperature. *BMC Genomics*. 2008;9:434.
33. Glaubitz U, Li X, Köhl K, van Dongen JT, Hinch DK, Zuther E. Differential physiological responses of different rice (*Oryza sativa*) cultivars to elevated night temperature during vegetative growth. *Funct Plant Biol*. 2014;41(4):437–48.
34. Wang Y, Pan Y, Liu Z, Zhu X, Zhai L, Xu L, Yu R, Gong Y, Liu L. De novo transcriptome sequencing of radish (*Raphanus sativus* L.) and analysis of major genes involved in glucosinolate metabolism. *BMC Genomics*. 2013;14:836.
35. Jiang B, Xie D, Liu W, Peng Q, He X. De novo assembly and characterization of the transcriptome, and development of SSR markers in wax gourd (*Benicasa hispida*). *PLoS One*. 2013;8(8):e71054.
36. Li M-Y, Wang F, Jiang Q, Ma J, Xiong A-S. Identification of SSRs and differentially expressed genes in two cultivars of celery (*Apium graveolens* L.) by deep transcriptome sequencing. *Horticulture Res*. 2014;1:10.
37. Chinnusamy V, Zhu J, Zhu J-K. Cold stress regulation of gene expression in plants. *Trends Plant Sci*. 2007;12(10):444–51.
38. Pan Y, Michael TP, Hudson ME, Kay SA, Chory J, Schuler MA. Cytochrome P450 Monooxygenases as Reporters for Circadian-Regulated Pathways. *Plant Physiol*. 2009;150(2):858–78.
39. Li XP, Gilmore AM, Caffarri S, Bassi R, Golan T, Kramer D, Niyogi KK. Regulation of photosynthetic light harvesting involves intrathylakoid lumen pH sensing by the PsbS protein. *J Biol Chem*. 2004;279(22):22866–74.
40. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*. 2014;505(7485):635–40.
41. Boerner S, McGinnis KM. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS One*. 2012;7(8):e43047.
42. Illott NE, Ponting CP. Predicting long non-coding RNAs using RNA sequencing. *Methods*. 2013;63(1):50–9.
43. Ounzain S, Micheletti R, Beckmann T, Schroen B, Alexanian M, Pezzuto I, Crippa S, Nemir M, Sarre A, Johnson R, et al. Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs. *Eur Heart J*. 2014;36(6):353–68a.
44. Kaikkonen MU, Lam MTY, Glass CK. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res*. 2011;90(3):430–40.
45. Reis EM, Verjovski-Almeida S. Perspectives of Long Non-Coding RNAs in Cancer Diagnostics. *Front Genet*. 2012;3:32.
46. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25(18):1915–27.
47. Lv J, Cui W, Liu H, He H, Xiu Y, Guo J, Liu H, Liu Q, Zeng T, Chen Y, et al. Identification and characterization of long non-coding RNAs related to mouse embryonic brain development from available transcriptomic data. *PLoS One*. 2013;8(8):e71152.
48. Kanchiswamy CN, Mohanta TK, Capuzzo A, Occhipinti A, Verrillo F, Maffei ME, Malnoy M. Differential expression of CPKs and cytosolic Ca²⁺ variation in resistant and susceptible apple cultivars (*Malus x domestica*) in response to the pathogen *Erwinia amylovora* and mechanical wounding. *BMC Genomics*. 2013;14:760.
49. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29(9):e45.
50. Liu T, Zhu S, Tang Q, Chen P, Yu Y, Tang S. De novo assembly and characterization of transcriptome using Illumina paired-end sequencing and identification of *CesA* gene in ramie (*Boehmeria nivea* L. *Gaud.*). *BMC Genomics*. 2013;14:125.
51. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.
52. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, et al. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 2003;19(5):651–2.
53. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105–11.
54. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
55. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. *Genome Res*. 2011;21(12):2213–23.
56. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36(10):3420–35.
57. Yu Y, Fuscoe JC, Zhao C, Guo C, Jia M, Qing T, Bannon DJ, Lancashire L, Bao W, Du T, et al. A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat Commun*. 2014;5:3230.
58. Iseli C, Jongeneel CV, Bucher P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol* 1999:138–148.
59. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. 2006;34:W293–7.
60. Weikard R, Hadlich F, Kuehn C. Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. *BMC Genomics*. 2013;14:789.
61. Billerey C, Boussaha M, Esquerre D, Rebours E, Djari A, Meersseman C, Klopp C, Gautheret D, Rocha D. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics*. 2014;15:499.
62. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41(17):e166.
63. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35:W345–9.
64. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res*. 2013;41:D226–32.
65. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 2011;39:D152–7.
66. Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, Zhu W, Wu W, Chen R, Zhao Y. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res*. 2014;42:D98–103.
67. Saldanha AJ. Java Treeview—extensible visualization of microarray data. *Bioinformatics*. 2004;20(17):3246–8.
68. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campillo I, Creech M, Gross B, et al. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc*. 2007;2(10):2366–82.
69. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. 2011;12:35.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

