African Journal of Pharmacy and Pharmacology Vol. 6(22), pp. 1625-1630, 15 June, 2012
Available online at http://www.academicjournals.org/AJPP
DOI: 10.5897/AJPP12.494
ISSN 1996-0816 ©2012 Academic Journals

Full Length Research Paper

Exploration of the biological basis of coronary heart disease angina pectoris with Qi deficiency and Qi stagnation based on GenCLiP gene mining software

Xueling Ma[#], Xing Zhai[#], Yubo Li, Yulin-OuYang, Qi Shi, Hua Xie, Huihui Zhao, Shuzhen Guo, Yong Wang, Jianxin Chen* and Wei Wang*

Beijing University of Chinese Medicine, 11 Bei San Huan Dong Lu, ChaoYang District, Beijing 100029, Peoples Republic of China.

Accepted 3 May, 2012

The aim of this study is taking coronary heart disease (CHD) angina pectoris (AP) as an example to approach the manipulation and application of GenCLip, gene mining software in searching disease-syndrome related genes. According to results from CIPHER, a prediction software, 100 genes were predicted based on the similarity of the characterization of CHD. GenCLiP gene mining software was applied to find the disease-syndrome related genes of CHD AP with qi deficiency and qi stagnation. There are 9 genes, ANG, APOA1, MEF2A, PPP1R12C, SREBF1, TCAP, TNNI3, TNNT2, and TPM1 related to both qi deficiency and qi stagnation syndromes of CHD angina pectoris. 7 genes (AST, ELN, FCN1, MYLK, MYLK2, MYOG, and PRTN3) are specifically related to qi deficiency syndrome, while 4 (IL32, PAM, TNFSF8, and TNNC1) are specifically related to qi stagnation syndrome. The study concluded that application of GenCLiP gene-mining software to explore the disease-syndrome related genes is an effective, rapid and feasible method, which has certain reference value in the study of the biological basis of traditional Chinese medicine (TCM) syndrome and can be extended to the identification of the syndrome related genes in other diseases.

Key words: Coronary heart disease angina pectoris, Qi deficiency, Qi stagnation, disease-syndrome related genes, GenCLiP.

INTRODUCTION

GenCLiP is a literature mining tool that can cluster a list of genes with keywords that are auto-extracted from their up-to-date related literature and then manually created by the user. GenCLiP can display literature mentioning specified genes and keywords for manual verification of their associations (Huang et al., 2008). By using GenCLiP, we may interpret disease pathogenesis and find novel genes or pathways for further research. There are many difficulties in researching the biological basic of syndromes as well as in plant biology research, though lots of softwares had been developed and in use (Loraine, 2009; Usadel et al., 2009). The result of differentially expressed gene number by using gene chip analysis is quite large. If to verify the genes one by one not only costs a lot, but also takes a long time. In this study, we take coronary heart disease (CHD) angina pectoris (AP) as an example, using GenCLip software for CHD abnormal expression of genes to conduct a preliminary analysis, exploring the biological basis of qi deficiency and qi stagnation syndrome in the level of gene similarities and differences.

*Corresponding author. E-mail: wangwei@bucm.edu.cn: cjx@bucm.edu.cn. Tel: +86 10 64286283. Fax: +86 10 64286283.

# These authors contributed equally to this work.

MATERIALS AND METHODS

1) Genes of CHD CIPHER platform was applied to predict CHD associated genes (Wu et al., 2008). Top 100 of genes were collected and used as probe to search CHD related literatures in MEDLINE database. The CIPHER was a novel algorithm that uses phenotype similarity to predict gene clusters for CHD. The predicted

Table 1. Syndrome diagnostics standard of Qi deficiency and stagnation.

Syndrome	Symptoms
Qi deficiency	Secret anguish, dyspnea, heart-throb, spiritlessness, hypodynamia, disinclination to say, light tongue, weak pulse
Qi stagnation	Chest tightness, sigh, depression, dysphoria, string pulse

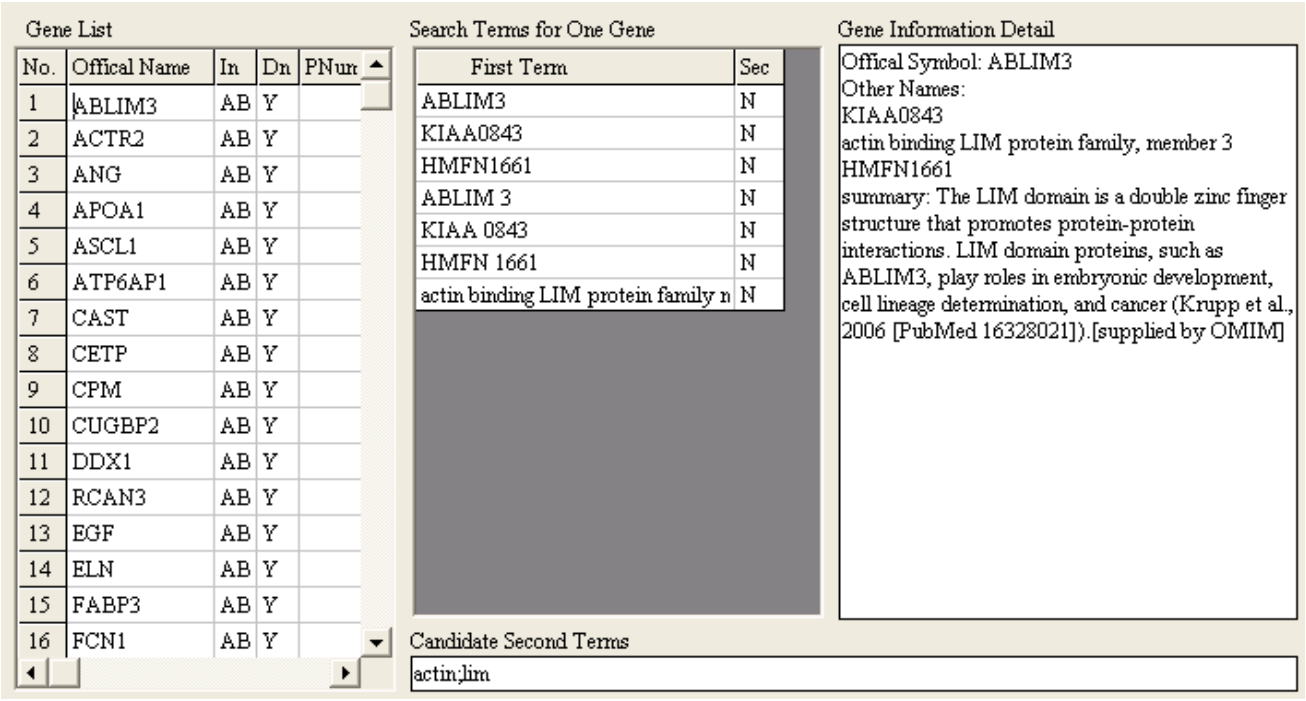

Figure 1. Literature retrieval of the 67 differentially expressed genes of CHD.

results majorly accord with the clinics.
2) Syndrome diagnostic criteria of CHD AP with qi deficiency, qi stagnation and blood stasis (Jia et al., 2008) was shown in Table 1.
3) Application and manipulation of GenCLiP software.

a) Download the up-to-date related literature for each gene from PUBMED, and automatically extract keywords from the literature, and manually curated the keywords, remove unrelated keywords, add keywords that are closely related to CHD, and set the weight for certain keywords that are perceived to be more important than others. The literature retrieval of the 67 differentially expressed genes of CHD is shown in Figure 1 and the manually curated 47 keywords of the genes are shown in Table 2.
b) Based on the manually curated keywords above, cluster analysis of genes with these keywords was conducted then, in order to infer the current study the pathogenesis of the disease and syndrome-disease related genes. The result of keyword-based cluster analysis of these genes is shown in Figure 2.
c) Select specific key words of CHD AP with qi deficiency and qi stagnation to search and conduct the gene-network of genes related to the key words. Set positive and negative genes through stochastic simulation test to infer whether or not the genes are specifically related to the specified key words. Finally, use the software to retrieve the artificial genes from the literature of the specified 67 genes and use the multiple keywords to further verify their relationship.

RESULTS AND DISCUSSION

The key word-based genes cluster analysis results of qi deficiency and qi stagnation showed that there are 9 genes related to both qi deficiency and qi stagnation syndromes of CHD angina pectoris, ANG, APOA1, MEF2A, PPP1R12C, SREBF1, TCAP, TNNI3, TNNT2 and TPM1. Seven (7) genes specifically related to qi deficiency syndrome, CAST, ELN, FCN1, MYLK, MYLK2, MYOG and PRTN3 (Figure 3) while 4 genes specifically related to qi stagnation syndrome, IL32, PAM, TNFSF8 and TNNC1 (Figure 4).

Select fatigue as the specific keywords of CHD AP with qi deficiency and select depressive as the specific keywords of qi stagnation to search and conduct the gene-networks. There are total 47 related genes, while 23 genes form 33 related gene pairs. The results of the gene-network of CHD AP with qi deficiency and qi stagnation were shown in Figures 5 and 6.

Table 2. The 47 keywords of CHD angina pectoris with Qi deficiency, Qi stagnation and blood stasis after manually curated.

Keywords	Plural flag	Weight	Synonym flag
CHD angina pectoris	0	10000	1
CHD	0	10000	1
Angina pectoris	0	10000	1
Vascular	0	100	2
Ventricular	0	100	2
Actin	0	11	3
Angiogenesis	0	10	4
Angiogenin	0	10	4
Angiotensin	0	100	5
Apolipoprotein	0	100	6
Apoptosis	0	100	7
CA2	0	10	8
Calcineurin	0	10	8
Calcium	0	10	8
Calmodulin	0	10	8
Calsarcin	1	10	8
CAM	0	10	8
Cardiomyopathy	0	100	2
Cholesterol	0	11	9
Cholesteryl	0	11	9
CK	0	11	10
Collagen	0	11	11
Creatine	0	11	12
CTNI	0	1000	13
Cytokine	1	10	14
Cytoskeleton	0	10	14
Dynamics	0	1	15
Elastic	0	1	16
Elastin	0	10	16
Fatty	0	10	17
HDL	0	30	17
Hypertrophic	0	2	18
IL	0	10	19
Immune	0	2	20
Immunoglobulin	0	2	20
Infarction	0	10	21
Inflammatory	0	100	22
LDL	0	10	17
Lipid	1	10	23
Lipoprotein	1	10	23
MB	0	10	24
MLCK	0	10	25
Myosin	0	10	25
PRE	0	10	26
Protease	0	10	27
Ribonuclease	0	10	28
RNASE	0	10	28
Spectrin	0	10	29
Syntrophin	0	10	30
Titin	0	10	31
Triglyceride	0	10	32
Troponin	0	100	13
Qi deficiency	0	10000	33
Secret anguish	0	3000	34
Cardiothoracic vague pain	0	3000	34
Dull pain	0	3000	34
Heart-throb	0	100	35
Palpitation	0	100	35
Palmus	0	100	35
Cardiopalmus	0	100	35
Heart palpitation	0	100	35
Spiritlessness	0	3000	36
Loss of vitality	0	3000	36
Hypodynamia	0	3000	37
Acratia	0	3000	37
Lassitude	0	3000	37
Fatigue	0	3000	37
Disinclination to say	0	300	38
No desire to speak	0	300	38
Dyspnea	0	3000	39
Shortness of breath	0	3000	39
Palpitation and short breath	0	3000	39
Light tongue	0	100	40
Weak pulse	0	100	41
Qi stagnation	0	10000	42
Chest tightness	0	3000	43
Chest distress	0	3000	43
Chest stuffiness	0	3000	43
Sigh	0	3000	44
Sighing	0	3000	44
Preference for sighing	0	3000	44
Depression	0	3000	45
Depressive	0	3000	45
Dysphoria	0	300	46
Irritability	0	300	46
Fidgeting	0	300	46
Restlessness	0	300	46
String pulse	0	300	47

Conclusion

Applying GenCLip gene mining software is easy to differentiate gene-specific functional clustering and conduct specific functional gene network, meanwhile the software size is moderate which can be operated by ordinary computer. The most convenient part is that the user can add and combine the keywords according to their own research directions and then conduct gene mining to analyze the relationship between the related genes, finally establish a gene-network of the related genes based on the specific keywords of specific traditional Chinese medicine (TCM) syndrome.

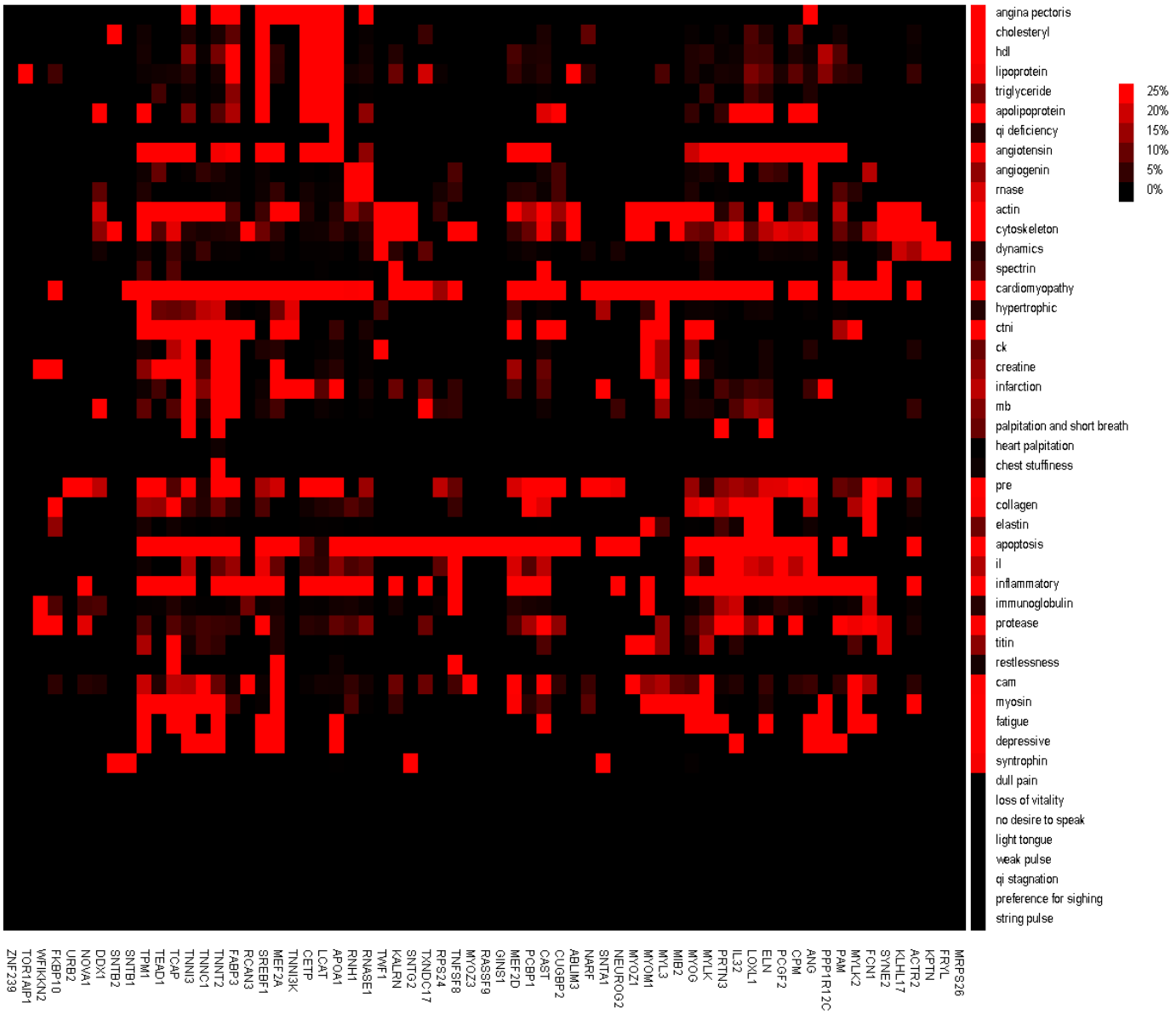

Figure 2. Keyword-based cluster analysis of genes related to CHD.

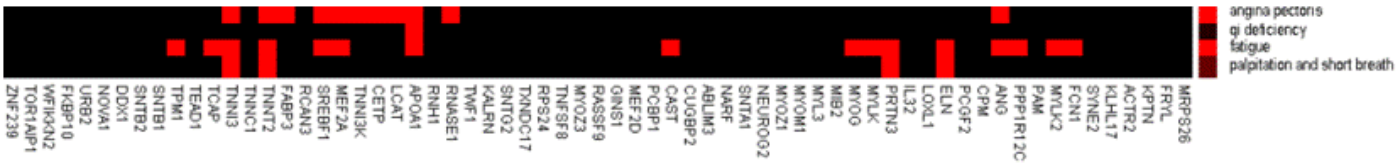

Figure 3. Cluster analysis results of qi deficiency keywords with related genes of AP.

Figure 4. Cluster analysis results of qi stagnation keywords with related genes of AP.

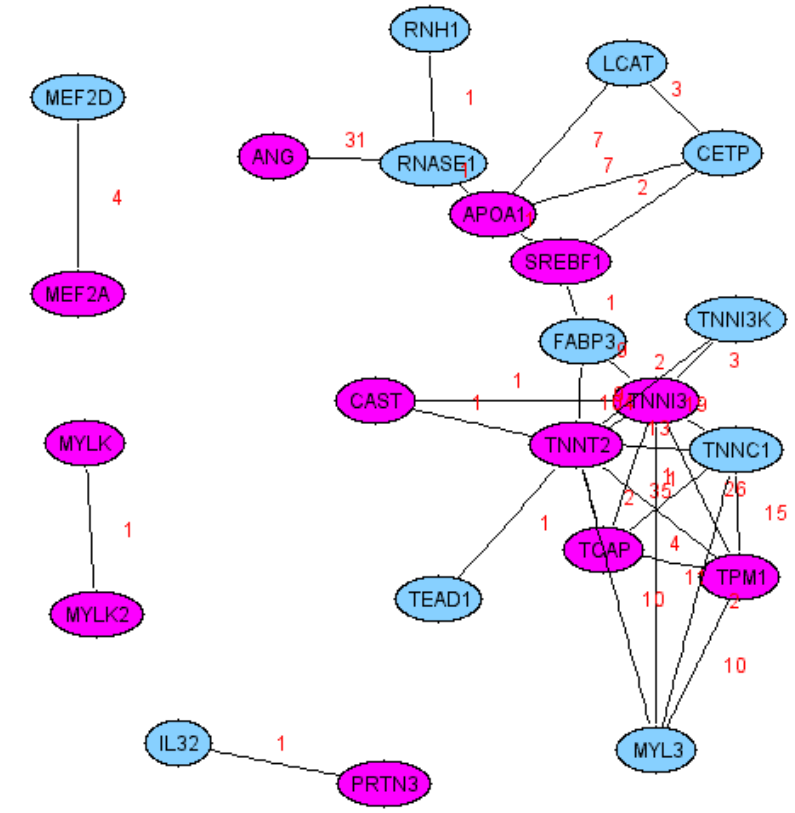

Figure 5. Gene-network of CHD angina pectoris with qi deficiency. The genes in magenta color are related to qi deficiency syndrome.

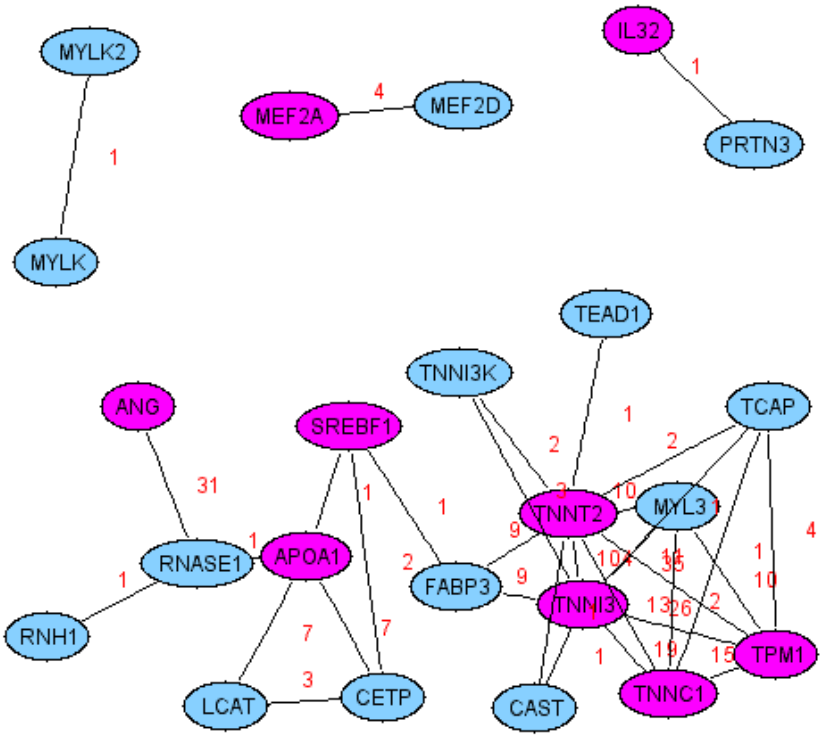

Figure 6. Gene-network of CHD angina pectoris with qi stagnation. The genes in magenta color are related to qi stagnation syndrome.

In this study, we take CHD AP as an example, using GenCLip software for CHD abnormal expression of genes to conduct a preliminary analysis, exploring the biological basis of qi deficiency and qi stagnation syndrome in the level of gene similarities and differences. The results showed that there are 9 genes related to both qi deficiency and qi stagnation syndromes of CHD angina pectoris, ANG, APOA1, MEF2A, PPP1R12C, SREBF1, TCAP, TNNI3, TNNT2 and TPM1. 7 genes specifically related to qi deficiency syndrome, CAST, ELN, FCN1, MYLK, MYLK2, MYOG and PRTN3 while 4 genes specifically related to qi stagnation syndrome, IL32, PAM, TNFSF8 and TNNC1. This study suggests that the application of GenCLiP gene-mining software in exploring the disease-syndrome related genes is an effective, rapid and feasible method, which has certain reference value in the study of the biological basis of TCM syndrome and can be extended to the identification of syndrome related genes in other diseases.

ACKNOWLEDGMENTS

This work was supported by the National Basic Research Program of China (973 Program) under grant No. 2011CB505106, The National Department Public Benefit Research Foundation (200807007), The International Science and Technology Cooperation of China (2008DFA30610), and National Science Foundation of China (30902020 and 81173463).

REFERENCES

Huang ZX, Tian HY, Hu ZF, Zhou YB, Zhao J, Yao KT. (2008). GenCLiP: a software program for clustering gene lists by literature profiling and constructing gene co-occurrence networks related to custom keywords. BMC Bioinformatics, 13(9): 308.

Jia Z, Wang Y, Wu Y, Li Y, Gao H, Xi G, Chen J, Gu C,Yuan G, Wu X, Wei C (2007).Entropy-based complex system partition method in extracting CHD angina pectoris syndrome elements and their distribution. Acad. J. Second Mil. Med. Univ., 28(7): 775-777.

Loraine A (2009). Co-expression analysis of metabolic pathways in plants. Methods Mol. Biol., 553: 247-264.

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ (2009). Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats. Plant Cell Environ., 32(12): 1633-1651.

Wu X, Jiang R, Zhang MQ, Li S (2008). Network-based global inference of human disease genes. Mol. Syst. Biol., 4: 189, pp.1-11.