Plant Biotechnology 25, 583–587 (2008)

Gene Expression Note

Over-expression of transcription associated factor genes co-expressed with genes of the mevalonate pathway, upstream of isoprenoid biosynthesis, in *Arabidopsis* cultured cells

Ryosuke Sano[1], Yoshiyuki Ogata[1], Hideyuki Suzuki[1], Yoichi Ogawa[1], Tomoko Dansako[1,a], Nozomu Sakurai[1], Koei Okazaki[1], Koh Aoki[1], Kazuki Saito[2], Daisuke Shibata[1,*]

[1] Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan; [2] Graduate School of Pharmaceutical Sciences, Chiba University, Inage, Chiba 263-8522, Japan

* E-mail: shibata@kazusa.or.jp Tel: +81-438-52-3947 Fax: +81-438-52-3948

Received September 10, 2008; accepted October 31, 2008 (Edited by M. Umeda)

Abstract From a list of 3087 *Arabidopsis* putative transcription associated factor (TAF) genes including those possibly involved in transcription processes via their protein interaction, five TAF genes were related to genes upstream of isopentenyl diphosphate biosynthesis, i.e., the mevalonate pathway, by analyzing their network relationships with gene-to-gene Pearson's correlation coefficients calculated from 1388 DNA microarray results. We up-regulated the five candidate genes under the control of a cauliflower mosaic virus 35S promoter in *Arabidopsis* suspension-cultured T87 cells, and evaluated gene expression in the transgenic cells by DNA microarray analysis. All five of the genes were substantially up-regulated in each transgenic cell line, whereas none of the enzyme genes for the mevalonate pathway exhibited any considerable up-regulation. The inconsistency between the co-expression relationship and lack of consequent up-regulation suggests that it is difficult to find genes controlling isoprenoid upstream metabolites through co-expression analysis. Nevertheless, the results of this study are a basis for further study of isoprenoid synthesis.

Key words: *Arabidopsis* T87 cell line, co-expression network analysis, isoprenoid biosynthesis, microarray.

In the last decade, a large number of microarray experiments have been carried out for various organisms (Edgar et al. 2002; Parkinson et al. 2007) and the results from these experiments have been accumulated in public databases. In *Arabidopsis*, over 3000 datasets of microarray experiments produced with the Affymetrix ATH1 array which contains 22746 probe sets are now available from the AtGenExpress consortium (AtGenExpress JPN: http://pfg.psc.riken.jp/AtGenExpress/index.html; Schmid et al. 2005). With the wealth of microarray experimental data, co-expression network analysis is now a common method for relating transcription factor genes with a set of genes of interest, which provides cues for further study of the transcriptional regulatory mechanisms among those genes (for a review, see Aoki et al. 2007). Correlation coefficients (higher than or equal to 0.5) among *Arabidopsis* genes based on 1388 datasets from AtGenExpress are also available from the ATTED-II database (Obayashi et al. 2007). Recently, co-expression analysis has been successfully applied to find transcription factor genes that control glucosinolate biosynthesis in *Arabidopsis* (Hirai et al. 2007). A large set of *Arabidopsis* full-length cDNA clones are now available (Seki et al. 2002). Ogawa et al. (2008) recently developed a protocol for efficient and high-throughput vector construction using full-length cDNA clones and *Agrobacterium*-mediated transformation of *Arabidopsis* cultured cells, which will facilitate functional genomics with gain-of-function technology utilizing full-length cDNA libraries such as the RAFL clones provided from the RIKEN BioResource Center. Thus, the resources and technologies for *Arabidopsis* functional genomics are now well-developed.

In plants, two isoprenoid biosynthetic pathways, the mevalonate pathway in cytosol and the mevalonate-independent pathway in plastids, consequently produce a large variety of metabolites, including useful compounds such as natural rubber, carotenoid pigments, and tocopherol (Lange and Ghassemian 2003). The mevalonate and mevalonate-independent pathways as the 'upstream' reactions might be crucial to control the

Abbreviations: Transcription associated factor, TAF

[a] Present address: Core Laboratory, Nara Prefectural small and medium-sized enterprises Support Corporation, 88 Shijo, Kashihara, Nara 634-0813, Japan

This article can be found at http://www.jspcmb.jp/

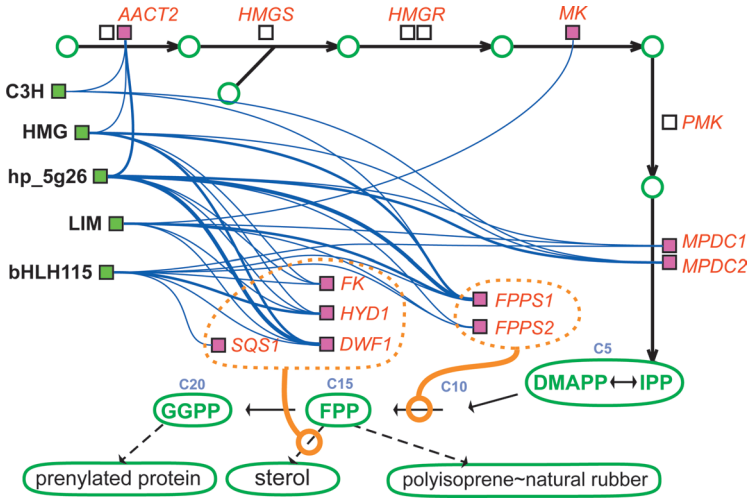

Figure 1. Mevalonate pathway and derived isoprenoid biosynthesis (cytosolic) with five candidate TAF genes detected by co-expression analysis. Circles and ovals (letters inside) denote metabolites, and squares represent genes. Green squares represent candidate TAF genes, and red squares represent enzyme genes having co-expression relationships with the candidates. Enzyme genes for the mevalonate pathway and some derivative pathways are shown. Blue lines connecting genes designate a co-expression relationship; the width of a line corresponds to the value of the correlation coefficient (thin≥0.5; medium≥0.6; thick≥0.7). C5, C10, C15 and C20 indicate the carbon number of the condensing metabolite, respectively. The enzyme gene names are abbreviated as follows. *AACT1,2*: acetyl-CoA C-acyltransferase, *HMGS*: hydroxymethylglutaryl-CoA synthase, *HMGR*: 3-hydroxy-3-methylglutaryl CoA reductase, *MK*: mevalonate kinase, *PMK*: phosphomevalonate kinase, *MPDC1,2*: diphosphomevalonate decarboxylase, *FPPS1,2*: farnesyl diphosphate synthase, *SQS1*: squalene synthase, *FK*: sterol C14 reductase, *HYD1*: C-8,7 sterol isomerase, *DWF1*: sterol C24-reductase. The metabolite names are also abbreviated; IPP: isopentenyl diphosphate, DMAPP: dimethylallyl diphosphate, FPP: farnesyl diphosphate, GGPP: geranylgeranyl diphosphate. A full description of the abbreviated TAF gene names is shown in Table 1.

production of 'downstream' derivative metabolites. In this study, we searched for and examined transcription associated factor (TAF) genes co-expressed with genes of the mevalonate pathway in *Arabidopsis* with the goal of eventually improving the production of isoprenoid metabolites.

We obtained a large gene-to-gene expression correlation dataset from the ATTED-II database (http://www.atted.bio.titech.ac.jp; Obayashi et al. 2007). A gene list of 3087 *Arabidopsis* putative TAFs (available on request) was based on transcription factor genes collected from 3 databases (AGRIS: http://arabidopsis.med.ohio-state.edu; DATF: http://attfdb.cbi.pku.edu.cn; RARTF: http://rarge.gsc.riken.jp/rartf) (Davuluri et al. 2003; Guo et al. 2005; Iida et al. 2005) as described in Hirai et al. (2007). Additional genes possibly involved in transcription processes via their protein interactions, etc., were also listed by filtering the "molecular function aspect" of Gene Ontology term arranged in the KAGIANA tool (Ogata et al. 2008) with the words 'nucleotide binding' and 'protein binding'. Meanwhile, a total of 58 enzyme genes were listed as genes assigned to cytosolic isoprenoid biosynthesis, the mevalonate pathway, isoprene condensation and sterol biosynthesis (Figures 1, 2) in reference to the list in Lange and Ghassemian (2003), pathway maps in the KaPPA-View (http://kpv.kazusa.or.jp/kappa-view; Tokimatsu et al. 2005), and some information from recent reports (e.g., Morikawa et al. 2006). With the selected TAF genes as the query and isoprenoid enzyme genes as the objective, we extracted groups of query genes that are highly correlated to the objective genes, using the "co-expression network analysis" function of the KAGIANA tool (Ogata et al. 2008). Obtained co-expression relationships were depicted using the Pajek software (Batagelj and Mrvar 1998) to facilitate a grasp of the network structure. Five TAF genes, bHLH115 (*At1g51070*), LIM (*At2g39900*), C3H (*At4g22250*), HMG (*At4g23800*), and hp_5g26 (*At5g26850*) had at least two co-expression links to mevalonate pathway genes with >0.5 Pearson coexpression coefficients (Figure 1; Table 1). The multiple co-expression correlation links might imply some involvement of these TAF genes in the mevalonate pathway and other pathways of cytosolic isoprenoid biosynthesis.

To clarify the function of these TAF genes in cytosolic isoprenoid biosynthesis, we up-regulated them in *Arabidopsis* cultured cells. Five full-length cDNA clones (RAFL clones) for the candidate TAF genes were obtained from the RIKEN BioResource Center (Tsukuba, Japan), and introduced into the Gateway-based binary vector pGWB2 (having a CaMV 35S promoter to drive the target). *Arabidopsis* T87 cells were used for *Agrobacterium*-mediated transformation as described previously (Ogawa et al. 2008). Twenty independent transgenic calli were isolated and subjected to northern analysis to confirm the up-regulation of the introduced gene. Preparation of T87 suspension cells,

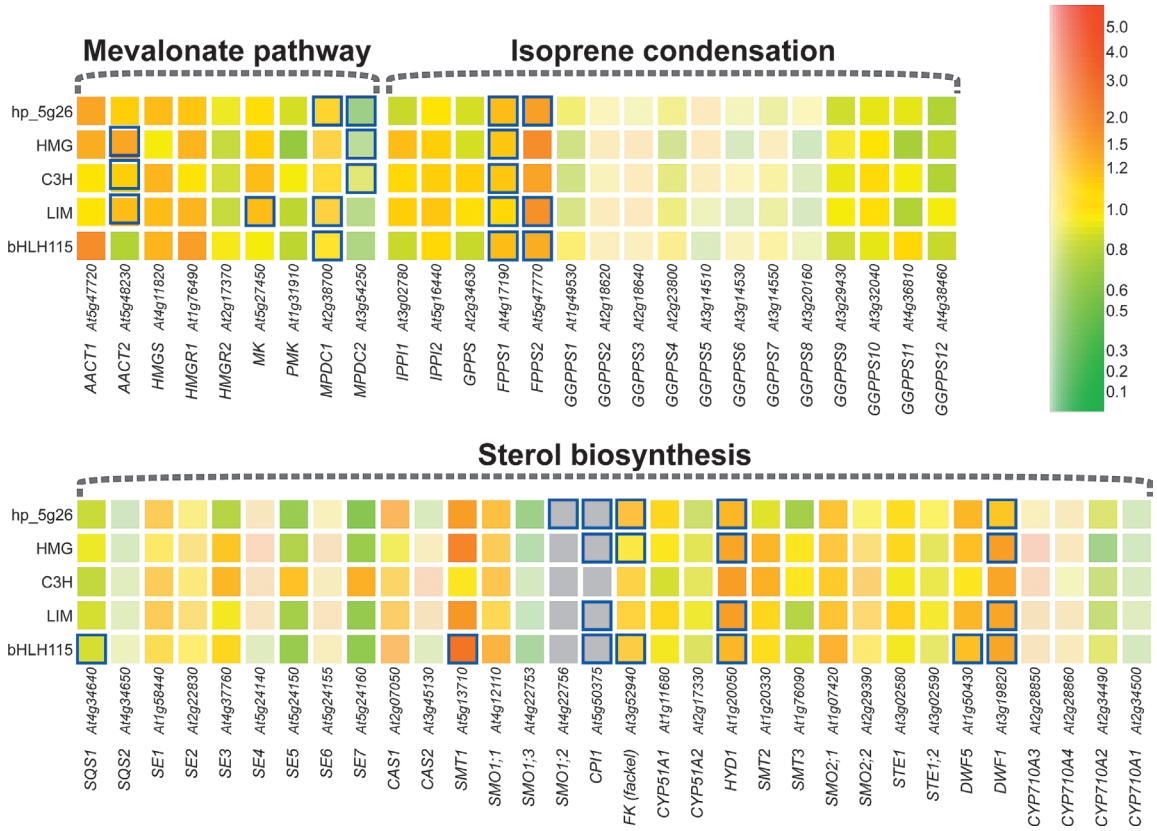

Figure 2. Gene expression of cytosolic isoprenoid biosynthetic genes in over-expressed cell lines for five candidate TAF genes. Coloring of tiles denotes the relative expression level of each enzyme gene designated below; each row is the result from the candidate TAF gene designated left. Color scaling was done with GeneSpring version 7.3.1; relative expression levels are expressed in green-yellow-red color spectra (down-regulation, no change and up-regulation, respectively, at fold change designated below) and the lightness represents reliability (lighter means less reliable). Gray colors for *SMO1;2* and *CPI1* indicate no signal information for these genes because of no probe sets on the Agilent Arabidopsis2 microarray. Blue frames represent a co-expression relationship (correlation coefficient ≥0.5) between enzyme gene (below) and TAF gene (left). The enzyme gene names are abbreviated as follows. *IPPI1,2*: isopentenyl diphosphate isomerase, *GPPS*: geranyl diphosphate synthase, *FPPS1,2*: farnesyl diphosphate synthase, *GGPPS1-12*: geranylgeranyl diphosphate synthase, *SQS1,2*: squalene synthase, *SE1-7*: squalene monooxygenase, *CAS1,2*: cycloartenol synthase, *SMT1*: cycloartenol C24 methyltransferase, *SMO1;1-3*: 24-methylenecycloartenol C-4 methyl oxidase, *CPI1*: cycloeucalenol cycloisomerase, FK (*fackel*): sterol C14 reductase, *CYP51A1,2*: obtusifoliol C14 methyl oxidase, *HYD1*: C-8,7 sterol isomerase, *SMT2,3*: sterol C-methyltransferase, *SMO2;1,2*: sterol C-4 methyl oxidase, *STE1*: sterol-C5-desaturase, *DWF5*: sterol Δ7 reductase, *DWF1*: sterol C24-reductase, *CYP710A1-4*: sterol C 22-desaturase. Full descriptions of abbreviated mevalonate pathway enzyme names are shown in Figure 1.

extraction of total RNA, and northern blot analyses were performed essentially as reported in Ogawa et al. (2008). Ten-day-old cell cultures following two weeks of preculture (in 100-ml flasks containing 30 ml of liquid medium with hygromycin and meropenem) were used as samples for RNA extraction. RNA samples from transgenic T87 cells were extracted using an RNeasy Plant Mini kit (Qiagen, Hilden, Germany). From each set of the five TAF transgenic lines, three independent clones with high transgene expression were selected for microarray experiments. Three independent hybridizations with the transgenic cell lines were performed using the Agilent Arabidopsis2 Oligo DNA Microarray (contains 21500 probe sets; Agilent Technologies, Palo Alto, USA) with the two-color procedure. For the two-color procedure, total RNAs from each TAF transgenic clone and a mixture of three independent control vector lines (with equal amounts) were labeled with Cy3 ('green') and Cy5 ('red') dyes, respectively. Probe preparation and hybridization of the microarray was carried out according to the manufacturer's instructions (Gene Expression Hybridization Kit; Agilent Technologies). The whole dataset was analyzed by the GeneSpring software (version 7.3.1; Agilent Technologies). Relative signal strengths (Cy3 / Cy5) from the three hybridizations were averaged to each probe and used for the latter analysis. To evaluate reproducibility among hybridizations, a calculation of chip-to-chip expression correlation was performed between all pairs of the control vector data (net 'red' signal strength) using Pearson's correlation coefficient. The result was that the coefficients were all higher than 0.98, indicating the high reproducibility of the present dataset in gene expression.

As shown in Table 1, all of the five TAF candidate genes were considerably up-regulated in each over-

Table 1. Description of the five candidate TAF genes related to the mevalonate pathway in co-expression analysis.

Gene name abbreviated	AGI code[1]	Description[1]	Net signal value for control vector line[2]	Fold change in over-expressed transgenic[3]
bHLH115	*At1g51070*	basic helix-loop-helix (bHLH) family protein; similar to ILR3	11844 ± 1578	5.6 ± 0.73
LIM	*At2g39900*	LIM type Zinc finger domain-containing protein	1747 ± 77	29.41 ± 10.84
C3H	*At4g22250*	zinc finger (C3HC4-type RING finger) family protein	2309 ± 195	66.69 ± 10.3
HMG	*At4g23800*	high mobility group (HMG1/2) family protein	2625 ± 119	49.16 ± 10.03
hp_5g26	*At5g26850*	similar to hypothetical protein OsJ_010154; contains fold Armadillo-type	177 ± 39	4.7 ± 1.01

[1] Based on the latest version of the *Arabidopsis* genome annotation (TAIR8)

[2] Net signal value for control vector pGWB2 lines (rProcessedSignal); average of three independent lines (±SD) is shown to represent a basal expression level of each candidate gene.

[3] Actual up-regulation level (fold change) of each candidate gene in the over-expressing T87 cell line; average of three independent lines (±SD) is shown.

expressed cell line. On the other hand, almost all of the enzyme genes for the mevalonate pathway, which have co-expression relationships to the candidate TAF genes (blue frames in Figure 2), exhibited no significant ($>2^1$ or $<2^{-1}$) up- or down-regulation in the transgenic T87 cells (Figure 2). The only exception was *MPDC2* in hp_5g26 over-expressing lines, which revealed more than two-fold ($2^{-1.18}$) down-regulation. The genes involved in cytosolic isoprenoid pathways, isoprene condensation, and sterol biosynthesis had almost the same tendency as those from the mevalonate pathway genes. Only *SMT1* in bHLH115 over-expressing lines was found to be $2^{1.27}$-fold up-regulated. It should be noted, however, that there are two ambiguities in expression levels of sterol biosynthesis genes (*SMO1;2*, *CPI1*), which have co-expression relationships to the candidate TAF genes (blue frames in Figure 2), due to the lack of probe sets on the Arabidopsis2 microarray corresponding to the two genes.

There are several reasons why over-expression of the candidate TAF genes did not directly lead to up-regulation of the cytosolic isoprenoid biosynthesis genes with co-expression relationships. First, the correlation coefficient values obtained from the 1388 microarray datasets might involve some biological bias. It may be that the datasets emphasize some expression tendency, possibly brought about by differences in expression profiles between shoot and root. In fact, some co-expression relationships are considerately altered when correlation coefficients are calculated from more recent datasets with over 3000 microarrays instead of the 1388 we used (Ogata Y, unpublished observation). In addition to the usage of adequate array datasets for calculating correlation coefficients, the application of more sophisticated algorithms for co-expression analysis could help find other candidate TAF genes correlated with metabolic pathways, including the mevalonate pathway. Second, the co-expression and over-expression analyses were conducted with different source materials; the correlation coefficients were derived mainly from plant tissues, whereas the over-expression experiments were carried out in T87 cultured cells. Thus, we cannot rule out the possibility that T87 cells lack expression of some essential genes required for full regulatory function of the candidate TAF genes examined in this study. Alternately, it is possible that some regulatory factor which is comparatively abundant in T87 cultured cells hinders the responsiveness to the over-expression of candidate TAF genes. Third, it should be noted that successful metabolic pathway activation (Hirai et al., 2007) in *Arabidopsis* cultured cells was demonstrated with a secondary metabolic pathway, which could be regulated through a simple regulatory mechanism operating under some restricted conditions. On the other hand, the mevalonate pathway, followed by the sterol biosynthetic route, is one of the arterial pathways to maintain the plant life cycle, and such important pathways should be regulated through complex mechanisms. In fact, one of the mevalonate pathway enzyme genes, *HMGR*, which has no co-expression relationships with the five candidate TAF genes (Figure 1), is a target for positive feedback regulation by downstream metabolites both in animals (Goldstein and Brown 1990) and plants (Wentzinger et al. 2002; Hemmerlin et al. 2003).

Aside from the assumption that co-expression relationships imply that over-expression of candidate TAF genes should result in the up-regulation of isoprenoid enzyme genes, it is likely that these co-expression relationships hint at some biological relatedness. We could assume that an unidentified hypothetical common regulator leads to the co-expression of candidate TAF genes and isoprenoid biosynthetic genes. If this is the case, then these TAF genes may have some regulatory function in other metabolic pathways which require some coordinated regulation with the mevalonate pathway. In fact, some enzyme genes for other pathways 'downstream' of isoprenoid biosynthesis (cytosolic and plastidal) without any co-expression relationship were up-regulated in our TAF over-expressed transgenic lines (data not shown). Further analyses, including those of loss-of-function

mutants or transgenic cultured cells with RNAi constructs, will be needed to decipher the co-expression relatedness of the candidate TAF genes into biological relationships between these genes and metabolic pathways, including isoprenoid biosynthesis.

Acknowledgements

The authors thank F. Fujii, T. Matsuura, M. Hasegawa, K. Mori, K. Moriya, Y. Asami, and M. Iwata for their technical assistance. We thank the RIKEN BioResource Center (Tsukuba, Japan) for RAFL cDNA clones and *Arabidopsis* T87 cells, and Dr. T. Nakagawa (Shimane University) for his gift of pGWB2. This work was supported by a grant from the New Energy and Industrial Technology Development Organization (NEDO) as part of the project "Development of Fundamental Technologies for Controlling the Material Production Process of Plants".

References

Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol* 48: 381–390

Batagelj V, Mrvar A (1998) Pajek—program for large network analysis. *Connections* 21: 47–57

Davuluri R, Sun H, Palaniswamy S, Matthews N, Molina C, Kurtz M, Grotewold E (2003) AGRIS: Arabidopsis Gene Regulatory Information Server, an information resource of Arabidopsis *cis*-regulatory elements and transcription factors. *BMC Bioinformatics* 4: 25

Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucl Acids Res* 30: 207–210

Goldstein JL, Brown MS (1990) Regulation of the mevalonate pathway. *Nature* 343: 425–430

Guo A, He K, Liu D, Bai S, Gu X, Wei L, Luo J (2005) DATF: a database of Arabidopsis transcription factors. *Bioinformatics* 21: 2568–2569

Hemmerlin A, Hoeffler J-F, Meyer O, Tritsch D, Kagan IA, Grosdemange-Billiard C, Rohmer M, Bach TJ (2003) Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco Bright Yellow-2 cells. *J Biol Chem* 278: 26666–26676

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

Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K (2005) RARTF: Database and tools for complete sets of Arabidopsis transcription factors. *DNA Res* 12: 247–256

Lange BM, Ghassemian M (2003) Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant Mol Biol* 51: 925–948

Morikawa T, Mizutani M, Aoki N, Watanabe B, Saga H, Saito S, Oikawa A, Suzuki H, Sakurai N, Shibata D, Wadano A, Sakata K, Ohta D (2006) Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in *Arabidopsis* and Tomato. *Plant Cell* 18: 1008–1022

Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, Shibata D, Saito K, Ohta H (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in *Arabidopsis*. *Nucl Acids Res* 35: D863–869

Ogata Y, Sakurai N, Provart NJ, Steinhauser D, Krall L (2008) Bioinformatics tools to discover co-expressed genes in plants. In: Kahl G, Meksem K (eds) *The Handbook of Plant Functional Genomics Concepts and Protocols*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, pp 309–335

Ogawa Y, Dansako T, Yano K, Sakurai N, Suzuki H, Aoki K, Noji M, Saito K, Shibata D (2008) Efficient and high-throughput vector construction and *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* suspension-cultured cells for functional genomics. *Plant Cell Physiol* 49: 242–250

Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R, Rayner T, Sharma A, William E, Sarkans U, Brazma A (2007) ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucl Acids Res* 35: D747–750

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37: 501–506

Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K, Shinagawa A, Shinozaki K (2002) Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science* 296: 141–145

Tokimatsu T, Sakurai N, Suzuki H, Ohta H, Nishitani K, Koyama T, Umezawa T, Misawa N, Saito K, Shibata D (2005) KaPPA-View. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol* 138: 1289–1300

Wentzinger LF, Bach TJ, Hartmann M-A (2002) Inhibition of squalene synthase and squalene epoxidase in tobacco cells triggers an up-regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Plant Physiol* 130: 334–346