Available online at www.sciencedirect.com

ScienceDirect

Genomics 90 (2007) 620–628

GENOMICS

www.elsevier.com/locate/ygeno

A draft gene regulatory network for cellular totipotency reprogramming during plant somatic embryogenesis

Fanchang Zeng, Xianlong Zhang *, Lei Cheng, Lisong Hu, Longfu Zhu, Jinglin Cao, Xiaoping Guo

National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei 430070, People's Republic of China

Received 2 March 2007; accepted 17 July 2007
Available online 19 September 2007

Abstract

The complexity of the somatic embryogenesis (SE) transcriptome suggests that numerous molecules are involved. To understand better the functional genomics of complex molecular systems during this important reprogramming process, we used bioinformatics and a pathway database to construct a draft network based on transcriptionally regulated SE-related genes, from functional genomics assays readout to high-level biological data interpretation. Here, a complex molecular system was unraveled by this network. This draft network is a potential reservoir for hundreds of testable predictions about cellular processes in early SE. This work could provide a useful test for modeling of a systems network and may have merit as a study presenting an advanced technology application due to its biological and economical importance. The approach presented here is scalable and can be extended to include additional data types. In particular, this effective system approach will be applied to various targeted gene networks in the future.

Keywords: Somatic embryogenesis; Regulatory network; Functional genomics; Systems biology

Somatic embryogenesis (SE) is the capacity of somatic cells to form in culture complete new embryos via a process that resembles zygotic embryogenesis. During SE the development of somatic cells is reprogrammed to the embryogenic pathway, and SE forms the basis of cellular totipotency in higher plants. Steward et al. [1] first described SE, which can serve as a model system for the study of the regulation of gene expression required for the earliest developmental events in the life of higher plants, such as the developmental mechanism of embryogenesis [2]. As the initial basis for cellular and genetic engineering, SE has also played an important role in genetic transformation, somatic hybridization, and somaclonal variation. Differential gene expression in somatic cells confers the ability to manifest embryogenic potential and is involved in developmental program switching. However, few of the molecular events involved in the transition of a somatic cell to an embryogenic-competent cell are known [2,3]. Plant somatic embryonic cultures are not readily achievable, and genetic expression differs between somatic embryonic and nonembryonic cells. Furthermore, the SE-related genes discovered from functional genomic assays in our lab [4] are expected to be relevant to cell totipotency and SE because, unlike embryonic cells, the various types of nonembryonic cells do not exhibit the embryogenic potential and differentiation phenotype.

Owing to a genotype-dependent response, the genetic engineering of cotton requires a highly successful plant regeneration procedure from the somatic cells [5–8]. Our previous work revealed specific transcriptionally regulated SE-related genes, which are true to a high degree of certainty and appear to represent best the genetic expression in somatic embryonic cells of cotton [4]. The discovery of specific transcriptionally regulated genes further facilitates understanding of the functional genomics of complex molecular systems. Experiments demonstrated that the phenotype and transcriptional regulation of SE are generated by a complex combination of processes involving the interaction of multiple molecules [4].

Although numerous fundamental aspects of development have been revealed through the study of individual genes and proteins, system-level models are still lacking for most

* Corresponding author. Fax: +86 27 8728 0016.
E-mail address: xlzhang@mail.hzau.edu.cn (X. Zhang).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.07.007

developmental processes. Although difficult, it is imperative to unravel the mechanisms underlying the totipotency association network. The biological processes of SE constitute an ideal model for a system-level approach. Early SE, including processes such as cell division and establishment of cellular polarity, is readily amenable to large-scale functional analysis.

A first step toward a system-level understanding of SE is to draw up a draft molecular interaction network of the molecular assemblies involved as well as the functional connections between them. Here we show that such putative interaction models can be derived from an integrated gene/protein network generated from expression profiles and a comprehensive pathways database and verified by published experimental evidence. The integrated network suggests that early plant SE is achieved through coordination of a set of molecular machines. We assessed the overall predictive value of this molecular machine network by examining its fit to published SE data based on experiments using cultured plant cells.

Logical networks inferred from the functions of known genes, combined with computational biology, experiment-based molecular biology, and information generated through various genome projects, can be used to predict SE-related genes that have been independently validated in cultured plant cells. Together, these provide a rich framework for understanding the precise mechanisms of SE in higher plants and, ultimately, the detailed steps by which these genes direct a plant's specific SE process.

Results

To model early SE globally, we generated network graphs (Fig. 1). Each node represents an early embryogenesis gene and its product(s), and each control represents a potential functional connection. The network graphs were based on the world's largest pathway database, the *Arabidopsis* database, whose data have been verified by published experimental evidence, and specific transcriptionally regulated SE-related genes discovered in our lab (for definitions of *node* and *control*, see Materials and methods).

The molecular interaction regulation network in Fig. 1 was constructed using the Build Pathway tool based on high-throughput interaction datasets from the model organism *Arabidopsis*. The draft network of various cellular pathways that may be involved, directly or indirectly, in regulating SE contains 1570 total members (nodes and controls). Property values can be distinguished by the shape and color of the node in the pathway network. Here, we show the thickly dense and integrated draft network. Each type of node or control has a unique name and graphic representation. The legible and magnified illustration of the whole draft somatic embryogenesis association networks is shown in Supplementary Material 8. (Zoom in to see the detailed and legible nodes and controls. For detailed information on all control and node types, see Supplementary Materials 10 and 12). By clicking on any of the biological objects within the Build Pathway tool workboard, more detailed information for each object (e.g., annotation characters and cellular localization in the interaction network) can be obtained (example data are shown in Supplementary Material 11 and all the properties of all controls and nodes in the draft SE association networks in Supplementary Material 7).

Pathway controls in the putative SE association molecular network

Fig. 1 illustrates the comprehensive putative molecular interaction pathway network. It covers 850 total controls, including 23 binding, 14 chemical reaction, 31 expression, 23 molecular synthesis, 26 molecular transport, 5 protein modification, and 728 regulation controls. To help readers understand how this model works, details of the specific transcriptionally regulated SE-related genes discovered in our lab and the control elements in the draft network are given in Supplementary Materials 3, 4, and 12. A list of SE association controls of interest is provided in Supplementary Material 1.

Pathway nodes in the putative SE association molecular network

The draft network of various cellular pathways that may be involved, directly or indirectly, in regulating SE contains 720 total nodes, including 3 cell objects, 35 cell processes, 2 glycans, 203 proteins, 364 small molecules, 2 treatments, 9 complexes, and 102 functional classes. A list of SE association nodes of interest is provided in Supplementary Material 1. Supplementary Material 10 provides details of all the nodes.

Two extremely noteworthy cell processes are unraveled in this draft network as the significant SE association biomarkers

The network devised in this study revealed numerous interconnected molecular pathways, with several radialized conglomerated cores like hubs of a wheel. These hubs have high connectivity, as described in Supplementary Material 1 (Table S2), and a large number of individual connectivity elements, as described in Supplementary Material 1 (Table S1). These clustered nodes likely play a key role in embryogenic developmental pathways and thus can be considered as main regulators of the developmental switch during SE induction. An extremely noteworthy feature is that the hubs share two cell processes, cell death and cell proliferation, which sheds light on distinct SE association biomarkers.

Changes in mRNA abundance of genes characteristic of oxidative stress and cell division during SE suggest that the arrangement of the new cells into organized structures might depend on a genetically controlled balance between cell proliferation and cell death [9]. Experiments conducted in our laboratory [4] and the draft network are consistent with these findings. During SE, we observed many genes normally expressed during senescence and death and in response to oxidative bursts, although genes characteristic of cell proliferation were also expressed abundantly. Furthermore, molecules related to cell death and indicative of cell proliferation constituted the largest set. The relative expression levels of

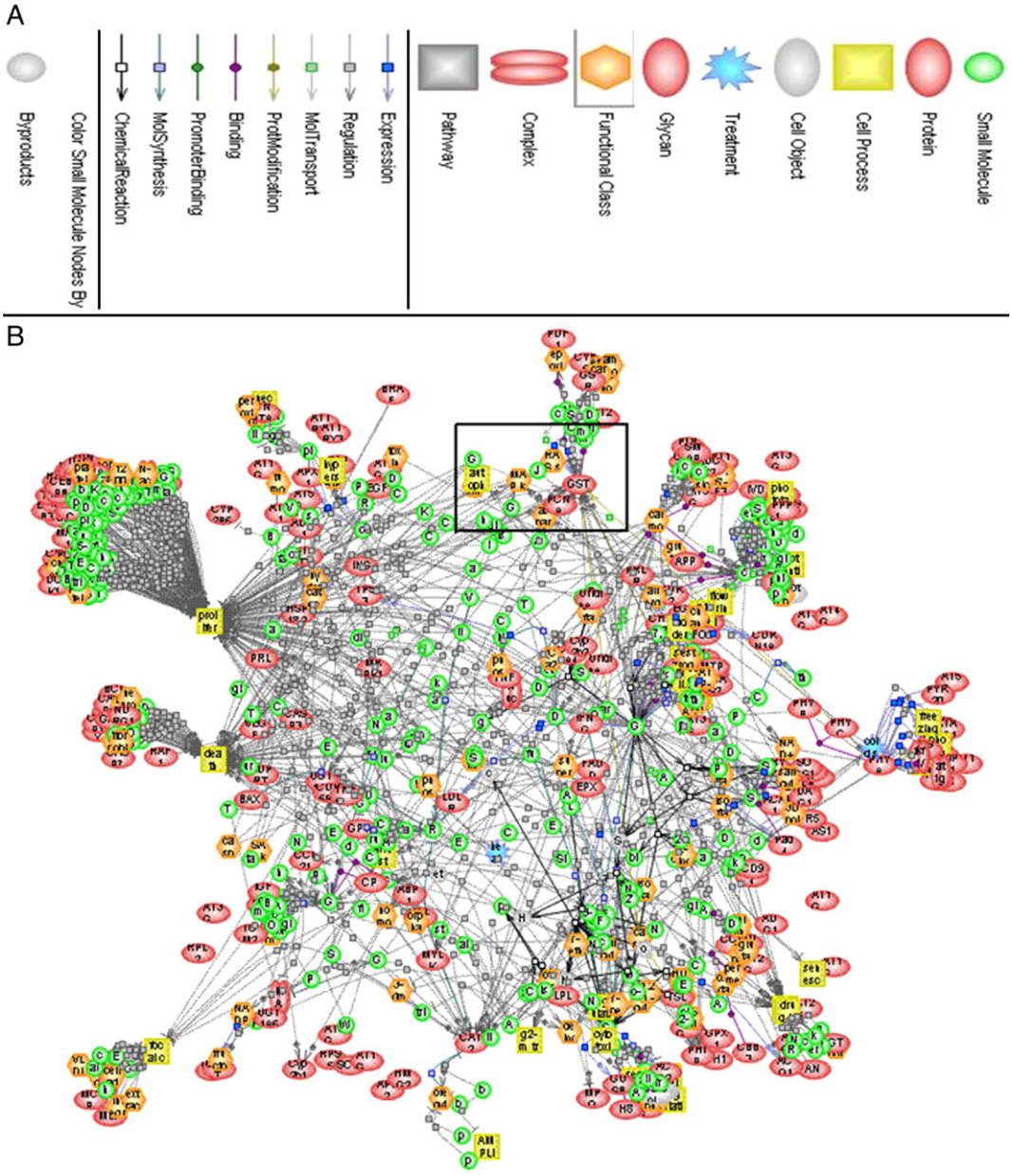

Fig. 1. The putative molecular interaction regulation network during somatic embryogenesis. A model organism dataset (*Arabidopsis* database) was used to transfer information on genes between genomes based on orthologic relationships. (A) Visual representation styles for nodes and controls. (B) The putative network of various cellular pathways that may be involved, directly or indirectly, in regulating somatic embryogenesis, containing the entire list of nodes and control types.

these genes affect specific cell types and fates during SE initiation and cellular differentiation.

Nodes of damage, including ROS, cytotoxicity/toxin, and stress (cold shock, hypersensitive response), found in our draft network further confirmed the hypothesis that SE is an adaptation process of in vitro cultured plant cells [10]. Cell death and cell proliferation are considered in general to be mutually exclusive processes: in somatic cells, when cell death occurs cell proliferation is arrested, and the live cell's fate is linked with the capacity for sustained division both in vitro and in planta. During SE, however, the relationship between cell death and cell proliferation is much more complex. Cell cycle progression during proliferation and programmed cell death are related but divergent processes in animal cells as well, indicating similar mechanisms in the related but divergent processes of plant cell division and differentiation [11]. Moreover, the regulatory mechanisms of developmental genes in plant SE appear to resemble those in *Drosophila* [3,4,12]. Although additional evidence is required, it appears that several key regulators of cell cycle/differentiation during cell death and cell proliferation play key roles in SE through coordinated interactions with hormonal and other developmental signaling pathways (for example, calcium, MAP kinase, calmodulin, transcription factors, et al., shown in Supplementary Materials 10, 12, and 1), cell–cell cross talk, and environmental.

A complex molecular system predicted by SE association network

Expression profiling data can be combined with an interactome to generate a network of functional relationships for early SE. The draft network suggests that the molecular machines acting in early SE are highly interconnected and are likely to operate together through regulatory molecules that coordinate their activities. The integrated network is a potential reservoir for hundreds of testable predictions about cellular processes in the early somatic embryo.

Fig. 1 and Supplementary Material 1 show the genes involved in SE, which are part of a complex molecular system that controls transcription/posttranscription, signal transduction, exocytosis (extracellular proteins/transporters), sugar metabolism, glycolysis, protein synthesis and storage, heat-shock proteins (HSPs), defense, intracellular traffic, programmed cell death/senescence, cell division, proteolysis, photorespiration, secondary metabolism, autophosphorylation, ROS/detoxification, motility, cell wall synthesis, and phototaxis.

Candidate SE-related genes discovered in SE association networks and validation of these molecular pathways/proteins in plants

We tested predictions from this integrated network by extensively checking both gene expression data and published literature.

Extracellular proteins/transporters

Secretion of proteins into the growth medium of suspension cultures has been reported in several plant species. Extracellular proteins play a significant role in the development of somatic embryos, although their exact roles are unclear. These proteins and changes in their expression patterns are associated with induction and initiation of SE [13,14]. One of the extracellular proteins that promote somatic embryo development in embryogenic cultures has been identified as a glycosylated acidic endochitinase [15]. Likewise, an endochitinase from sugar beet stimulated early development of somatic embryos in *Picea abies* [16]. In embryogenic cultures of *Daucus carota*, an endochitinase gene was related to a cell population that plays a nursing role in SE [17]. Most of the extracellular proteins involved in SE are associated with early stages of embryo development and also play a significant role in providing the nursing conditions for somatic embryo induction and cell wall degradation.

Liposaccharides are a class of signaling molecules that promote division of plant cells. Several studies have revealed the involvement of liposaccharides in the regulation of somatic embryo development: they were found to stimulate *D. carota* somatic embryos to proceed to the late globular stage [18], and they promoted the development of larger proembryogenic masses from small cell aggregates in *P. abies* [16,19]. Lipooligosaccharides can substitute for chitinases in their effect on early somatic embryo development in both *D. carota* and *P. abies* embryogenic systems [16,18], which indicates convergence of the activated signaling pathways. Recently, the application of a new generation of growth regulators, such as jasmonates and polyamines, has proven to be useful for initiating SE in many plant species [20–22].

Endogenous hormone and hormone-responsive genes

Hormones are the most likely candidates as regulators of developmental switches, and it has been proposed that hormones play a central role in mediating the signal transduction cascade leading to the reprogramming of gene expression. The requirement for a wide variety of hormones for SE is largely determined by the developmental stage of the explant tissue. For example, SE is induced by auxin and cytokinin in early stages such as proembryogenic masses [3,11], by ethylene in preglobular embryos [4], by GA in globular embryos [9], and by ABA in mature embryos [3,11]; in addition, brassinolide is necessary for acquisition of embryogenic competence and embryogenic tissue initiation [23,24]. Thus, a dynamic balance in the interactions among all hormones (auxin, ethylene, brassinolide, gibberellins, cytochalasin, and ABA) that regulate (positive, negative, or unknown) the distinct processes of cell death and cell proliferation suggests they are pivotal in switching cell fate during the developmental plasticity of SE through coordinated interactions with many developmental signaling pathways (for example, calcium, MAP kinase, calmodulin, transcription factors, et al.). For detailed effects and interactions, see Supplementary Materials 10, 12, and 1.

HSPs are expressed during somatic embryo development in response to hormones such as 2,4-D [25,26]. Auxin belongs to the plant glutathione *S*-transferase (GST) family, and a close relationship exists between somatic embryo induction and auxin-induced gene expression [26,27]. The HSP genes are expressed differentially in somatic embryos; for example, heat-shock treatment can arrest the growth of globular embryos but not somatic embryos at other developmental stages [28–30].

Housekeeping of the cell

Genes associated with important cellular activities and housekeeping genes play significant roles at various stages of embryo differentiation. During embryogenesis, an increase in the expression of actin and tubulin genes results in enhanced cell wall and membrane formation [31]. The ability to control cell expansion and proliferation is associated with polysaccharides of the cell wall and corresponding hydrolytic enzymes [18,32,33]. Similarly, histone-coding genes show enhanced expression during embryogenesis [34]. Translation of a gene encoding elongation factor-1a is regulated in the actively dividing cells of somatic embryos [35]. Balestrazzi et al. [36,37] reported increased topI gene expression during cellular proliferative activities such as mitotic divisions during SE. Glutamine synthetase was verified to play a key role in mediating SE [38,39], highlighting the involvement of a common regulatory system for nitrogen metabolism in embryogenesis.

Signal transduction pathways

Environmental stimuli or secondary messengers such as calcium may trigger signal transduction cascades, as seen in

various kinases, such as kinesin-like calmodulin-binding protein, which plays multiple roles in cell division and cell growth in flowering plants [40]. These protein kinases often undergo autophosphorylation during their activation, suggesting they have a significant role in the signal transduction pathway during SE [41–43]. For example, the rice SE receptor kinase 1 (SERK1) gene positively regulated SE of cultured cells [44]. Calcium was essential for morphogenesis of undifferentiated cells into somatic embryos [45,46]. Calmodulin (CaM) is a prominent protein involved in the mediation of calcium signaling in plants, and CaM levels increase upon induction of somatic embryos [45,46].

Maturation and protein storage

Differential expression of lectins during various stages of somatic embryo development highlights their importance in SE in alfalfa, as well as that of the globulin-1 gene in regenerable *Zea mays* callus [47]. Lectins are likely critical for alfalfa embryo development, and they may also be involved in growth regulation during embryogenic pattern formation [48]. Germins are developmentally regulated proteins, first discovered in germinating wheat, that are resistant to denaturation and proteases. Germins' oxidase activity confers them a role in plant development and defense responses, and they have been proposed to play a significant role during somatic and zygotic embryogenesis [49–51]. Identification of the process of germination in our putative network and the involvement of germins in SE suggest that germins are overlapping molecular products during both SE and germination.

Programmed cell death/senescence

Two waves of programmed cell death occur during the formation and development of somatic embryos in the gymnosperm of Norway spruce [52]. Coincidental with somatic embryo formation, the primary structure degenerated and redifferentiated through massive programmed cell death [52,53]. These cell suicide events ensure normal progression of SE, that is, transition from proembryogenic masses to somatic embryos and correct embryo pattern formation.

Damage, stress, and developmental program switch

While emphasizing the endogenous signal and epigenetic reprogramming during the developmental program switch, environmental factors and cell–cell cross talk [54] are also fundamental. Several researchers have proposed that in vitro conditions such as damage and stress play a central role in mediating the signal transduction cascade leading to the reprogramming of gene expression during SE [10,55–57]. Lower levels of stress enhance metabolism and induce adaptation mechanisms [58], including the reprogramming of gene expression as well as physiological and metabolic changes in the cells.

Defense and antioxidation

Mono-oxygenase [59], SERK, and lipid-transfer proteins may serve as markers to distinguish embryogenic cells, enabling an early diagnosis of embryogenic potential. Catalase activity is related to H_2O_2 concentration and was associated with embryogenic potential of the calli during SE [60]. Genes expressed during antioxidation are key factors in SE and control the regeneration ability in rice [61].

Proteolysis and protein modification

Redifferentiation processes during SE are involved in the general reprogramming of gene expression (chromatin remodeling, transcription machinery), and they require complex changes in the protein pattern and/or proteolysis [62,63]. Protein-folding is controlled by specific proteins exhibiting chaperone activities during SE, including HSPs [26,64].

Cell division and expansion

Two mechanisms appear to be important for in vitro formation of embryogenic cells: asymmetric cell division and control of cell elongation [32,65]. Plant cell enlargement is typically the result of the organized secretion of Golgi vesicles, the organized deposition of cell wall materials, and the modification of the existing cell wall [66]. Regulators of the cell cycle are key factors in the transition from somatic to embryogenic cell. Cell cycle genes also play a key role in SE [67,68]. During SE, active cell wall synthesis in embryogenic tissues is always accompanied by increased activity of the genes involved in cell cycle regulation. Dudits et al. [64] reported the expression of cell-cycle-related *cdk* and cyclin genes resulting in the formation of somatic embryos.

Obviously, the overall reprogramming of gene expression has to be governed by regulator genes, including those encoding transcription factors. Our SE association network revealed candidate SE-related transcription factors (for detailed examples, see Supplementary Material 1 or 10). There is evidence that most of the transcription factors noted above are involved in developmental regulation, but there are few reports of their involvement in SE.

Discussion

Somatic embryogenesis is the developmental reprogramming of somatic cells toward the embryogenesis pathway, which is a notable illustration of cell totipotency. SE is a unique developmental pathway and has been viewed as a potential model system for the study of the basic mechanisms of development reprogramming among the higher eukaryotic organisms; nevertheless it attracts wide interest in understanding the totipotency of cells. Despite very extensive tissue culture research, we are far away from understanding the key molecular events leading to SE. It is a very difficult task to extract the experimental findings and provide a comprehensive view.

To identify plant SE-related genes efficiently and study the molecular and cellular bases of this important reprogramming process, it is important to understand the mechanisms involving multiple cellular pathways and the relationships among molecular events. In our previous investigation, transcriptome profiling during the whole somatic embryogenesis process in cotton was comprehensively analyzed. A broad repertoire of

SE genes was identified, which is an important resource for understanding the genetic interactions underlying the SE regulatory network. Ideally, such knowledge should be gained through the construction of a logical molecular interaction network representing an adequate number of regulation control and molecular markers specifying distinct developmental profiles. The molecular interaction network, which shows the whole metabolome and interactome during SE, will provide the basis for further analyses of specification, induction, and patterning of the cell totipotency during SE. We convincingly show that important features of this association network are in good accordance with the published data (for detailed information, see Results). The result shows that our approach is feasible and that the predictions are successful, although an accurate estimate of the SE association network requires the accumulation of more published documents. The integrated network is a potential reservoir for hundreds of testable predictions about cellular processes in the early SE. The approach presented here is scalable and can be extended to include additional data types. In particular, with more information from studies and pathway databases becoming available, including those in databases and documents, we expect this effective system approach to promote the further development of a biology association network and be adapted to various targeted gene networks in the future. Namely, this general strategy is applicable to other biological processes and many other organisms.

Several important features in our approach are worth mentioning. Because of space limitations, the integrated draft network in Fig. 1 appears very dense and complicated. In fact, the network provides a clear picture of the features of the biological association network. For example, Supplementary Material 2 illustrates a legible part of the complete network. The legible and magnified illustration of the whole draft somatic embryogenesis association network is shown in Supplementary Material 8. The object nodes can also be distributed in the pathway diagram on the images of the cell organelles according to the cell localization parameter (data not shown). By constructing this draft network and incorporating it into our analytical workflow, we have converted a vast array of information into biological knowledge and thereby accelerated the development and validation of SE and cell totipotency research. Compiling a network can help us to devise a hypothesis to explain the molecular mechanisms of their regulation. By using bioinformatics tools, which combine experimental data, the results of published studies, and information from public databases, we can gain a broad-scale view of how genes, proteins, and small molecules interact to mediate cellular processes, thereby improving the process of converting information to biological knowledge and accelerating research.

In the evaluation of discovered genes in a complex molecular system and validation of molecular pathways/protein states for SE in plants based on published documents and microarray experiments, we faced several problems. First, very few microarray experiments provided the full dataset. Second, some studies in the literature were custom-made for special purposes, sometimes making the evaluation nonrepresentative. For these reasons, it is impossible to make any accurate estimate about the SE association network, except that the spectrum of such genes is broad.

The findings of this study raise the question of whether somatic cells acquire the embryonic ability and complex phenotype at the whole-cell level or via individual components. The complexity and coordination within this system suggest that SE may represent a concerted transformation of the whole cell involving multiple cellular pathways, rather than a piecemeal transformation of proembryogenic masses. More likely, somatic cells acquire the embryogenic phenotype by turning on the entire molecular system. This idea is supported by evidence that SE in plant cells resembles animal embryogenesis in terms of signaling and behavior in vivo [4,12] and by findings that mitogen-activated protein kinase phosphorylation cascades may link oxidative stress responses to auxin signaling and cell cycle regulation [69]. The histone H1 phosphorylating activity of cyclin-related kinase was dependent on posttranslational modifications that required the presence of cytokinin in embryogenic alfalfa protoplasts [70]. GSTs can have diverse roles, including detoxifying xenobiotics, targeting their substrates for transmembrane transport, and protecting against oxidative stress; in addition, some plant GSTs can bind, and probably carry and store, natural auxin [27]. All this evidence suggests the system is active when embryonic ability is induced but is turned off in nonembryonic somatic cells. Due to the probable single-cell origin of the somatic embryo, SE is well suited to this kind of analysis, in which filtered expression profiling data are functionally analyzed in the context of known pathways, and this method of analysis is especially effective when it is applied to homogeneous cell populations.

The molecular pathways related to extracellular proteins, endogenous hormones and hormone-responsive genes, cell housekeeping, signal transduction pathways, protein maturation and storage, programmed cell death/senescence, HSPs, stress and the developmental program switch, defense, antioxidation, proteolysis and protein modification, cell division and expansion, and transcription/posttranscription have been validated by independent laboratories (see Results). However, it is unclear if these models, which were based on somatic cells cultured in vitro, are accurate for in vivo embryogenesis in plants.

During the past few years, a considerable number of studies have investigated SE in different culture systems, particularly carrot, chicory, cotton, alfalfa, and conifer tissue cultures. Despite this accumulation of experimental data, due to tedious and expensive experiments, a limited number of plant genes are functionally well characterized. It is still unclear how many and which genes are highly specific SE association genes. We are still far from understanding the key events underlying the transition of differentiated somatic cells to the totipotent and embryogenic cell state. We demonstrate that the SE association network constructed by combining our experimental data with high-throughput pathway datasets provides important clues to SE and allows prediction of the partial genes involved in SE. However, only a subclass of the SE association network is covered by our method. The limitation of this approach is that the network construction is dependent on the development and

update of the pathway database. With more information from SE studies and pathway databases becoming available, including those in databases and documents, we expect this effective and accurate system approach to promote the further development of the SE association network and be widely applied to various targeted gene networks in the future.

The broad repertoire of genes and complex expression patterns in SE suggests that multiple cellular pathways are controlled by a concerted gene regulatory network. SE is an ideal model system for investigating developmental flexibility and stem cells. This report presents a comprehensive analysis of genes involved in plant SE association networks. The integrated network is a potential reservoir for hundreds of testable predictions about cellular processes in the early somatic embryo. Combined with the information generated through various genome projects, it will provide a systematic basis and an effective approach for understanding the precise mechanisms of SE in higher plants and, ultimately, the detailed steps by which these genes direct the specific processes involved in SE. The present work is a successful attempt to describe a putative interaction regulatory network that can underlie the SE. It ultimately will provide deeper insight into understanding the enigmatic reprogramming of cells in higher plants. This work could provide a useful test for modeling of a plant systems network and may have merit as a study presenting an advanced technology application in biology due to its biological and economical importance.

Materials and methods

Isolation of specific transcriptionally regulated SE-related genes

By incorporating the multiform Excelsior differential screening procedure and macroarray into the classical and highly powerful SSH protocol, numerous genes preferentially expressed during SE were successfully identified in our series of real experiments [4]. After subsequent confirmation by reverse Northern blot analysis, the cDNA library was composed of a broad array of SE genes, which is an important resource for understanding the genetic interactions underlying SE signaling and regulation. In addition, we used cDNA arrays to investigate the expression patterns of these genes in differentiating gradient culture, ranging from proembryogenic masses to somatic embryos at every stage. The SE cDNAs in the isolated clones were confirmed to be expressed in specific manners; all transcriptionally regulated SE-related genes were described by Zeng et al. [4]. Specific transcriptionally regulated ESTs associated with cotton SE are shown in Supplementary Material 3. Representative differentially expressed genes associated with cotton SE are presented in Supplementary Material 4.

SE association network pathway analysis and SE-related nodes prediction

Before giving details of this method, we define several important terms.

Controls (or functional links, relationships, and links) represent how proteins regulate or control one another on the graph, including events of functional regulation, chemical reactions, and protein–protein interactions. *Controls* are identified by mechanism and effect type. *Control* properties store information about tissues, cell types, experimental conditions, and other biologically meaningful details. Controls are subdivided into two types: direct controls are physical interactions between objects and indirect controls represent mainly logical relationships between entries when exact details of the regulation mechanism are unknown (e.g., A inhibits B). These include: (1) *mol synthesis*, describing events of cellular synthesis or degradation; (2) *expression*, describing various gene expression events, with the mechanism being transcriptional, posttranscriptional, or expression (unspecified); (3) binding, describing direct protein interactions; (4) *prot modification*, describing protein modification events, with mechanisms including about 40 different types of protein modifications (e.g., phosphorylation, geranylation); (5) *mol transport*, describing events of protein or cell object transport in cells, such as secretion and mitochondrial transport, with mechanisms being import, export, or unknown; and (6) *cell object control*, describing various events occurring with cellular objects in cells (e.g., actin polymerization, chromosome assembly), with mechanisms being assembly, disassembly, biogenesis, or movement.

Nodes represent proteins, small molecules, and other cell objects. *Node properties* store information such as functional class, localization, and chemical structure. Property values, such as node type, can be represented by color or shape coding of nodes in the visual model. The following biological object types (nodes) are supported in this draft: proteins, protein complexes, small molecules, enzymes, protein functional classes, cellular components, cellular objects, treatments, glycan, processes, complexes, functional classes, and pathways. Visual representation styles for nodes and controls are displayed in Fig. 1A.

SE association networks and pathway analysis during SE—examining, for instance, how genes, proteins, and small molecules interact to mediate cellular processes—were built using the PathwayAssist tool following the detailed manual and analysis guide (which is presented in the supplementary material or can be downloaded from http://www.ariadnegenomics.com/). It was described in brief by Nikitin et al. [71], and abbreviated tool instructions are presented in Supplementary Material 9. Using the analysis guide, readers will be able to look up and find descriptions of specific features and functions easily.

The SE association networks are built based on a wave-propagation algorithm developed for navigation through complex networks, which is used in computerized microchip design, for example. The algorithm begins by building waves (neighborhoods) from all input nodes. The waves represent the nodes neighboring the initial center node. The wave propagation continues until it intersects a wave from a different center.

To start building the putative molecular interaction pathway, we imported a list of gene names or IDs (Supplementary Materials 5 and 6) for putative orthologs of specific transcriptionally regulated SE-related genes discovered in our lab, which can be searched on the Web site http://www.arabidopsis.org/ and were identified according to the description by Lehner and Fraser [72]. We used model organism datasets (*Arabidopsis* database) to transfer information on the orthologic relationships of these genes. To create a pathway we used the Expand Pathway option of the Build Pathway tool. The Expand Pathway option finds all nodes in the database (ResNet *Arabidopsis* 2.0, 09/04/2005) linked directly to the nodes selected for expansion. That is, the Expand Pathway option (also known as Find All Nodes Connected to Selected Nodes, Maximum Number of Steps is 2) allows us to find neighbors of the selected objects in the database.

This draft network resolves multiple interaction relationships in a single view, bringing signal transduction, biochemical interactions, transcriptional regulation, and metabolic diagrams together, allowing for translation of a vast array of information into biological knowledge. We combined the Build Pathway tool with *Arabidopsis* gene model annotation programs and then coupled this with PubMed literature searches and microarray dataset searches for validation.

Evaluation of candidate SE-related nodes

We searched several published microarray datasets for candidate genes involved in early SE, and we used literature searches to find out if a candidate gene was related directly to SE. Candidate genes were subjected to a double evaluation process. First, to find microarray expression data evidence, we investigated all major microarray gene expression and proteome analysis datasets [9,63,73–77] to evaluate our prediction of a gene being related to SE. We considered a gene as related to SE if it showed a greater than twofold (either up or down) differential change in its expression level during SE in at least one experiment. The cutoff value of greater than twofold is considered in many microarray studies as sufficient to indicate reaction of the gene [78]. Second, published research on candidate genes was searched for in the PubMed database for all nonredundant candidate genes.

Acknowledgments

This work was supported by a grant from the National Basic Research Program of China (2004CB117301).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.07.007.

References

[1] F.C. Steward, M.O. Mapes, J. Smith, Growth and organized development of cultured cells. I. Growth and division of freely suspended cells, Am. J. Bot. 45 (1958) 693–703.
[2] J.L. Zimmerman, Somatic embryogenesis: a model for early development in higher plants, Plant Cell 5 (1993) 1411–1423.
[3] A. Chugh, P. Khurana, Gene expression during somatic embryogenesis: recent advances, Curr. Sci. 83 (2002) 715–730.
[4] F. Zeng, X. Zhang, L. Zhu, L. Tu, X. Guo, Y. Nie, Isolation and characterization of genes associated to cotton somatic embryogenesis by suppression subtractive hybridization and macroarray, Plant Mol. Biol. 60 (2006) 167–183.
[5] T.A. Wilkins, K. Rajasekaran, D.M. Anderson, Cotton biotechnology, Crit. Rev. Plant Sci. 19 (2000) 511–550.
[6] T.A. Wilkins, R. Mishra, N.L. Trolinder, Agrobacterium-mediated transformation and regeneration of cotton, J. Food Environ. Agric. 2 (2004) 179–187.
[7] R. Mishra, H.Y. Wang, N.R. Yadav, T.A. Wilkins, Development of a highly regenerable elite acala cotton (*Gossypium hirsutum* cv. Maxxa): a step toward genotype-independent regeneration, Plant Cell, Tissue Organ Cult. 73 (2003) 21–35.
[8] A.B. Arpat, M. Waugh, J.P. Sullivan, M. Gonzales, D. Frisch, D. Main, T. Wood, A. Leslie, R.A. Wing, T.A. Wilkins, Functional genomics of cell elongation in developing cotton fibers, Plant Mol. Biol. 54 (2004) 911–929.
[9] F. Thibaud-Nissen, R.T. Shealy, A. Khanna, L.O. Vodkin, Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean, Plant Physiol. 132 (2003) 118–136.
[10] D. Dudits, J. Györgyey, L. Bögre, L. Bakó, Molecular biology of somatic embryogenesis, in: T.A. Thorpe (Ed.), *In Vitro* Embryogenesis in Plants, Kluwer Academic, Dordrecht, 1995, pp. 267–308.
[11] A. Fehér, T.P. Pasternak, D. Dudits, Transition of somatic plant cells to an embryogenic state, Plant Cell, Tissue Organ Cult. 74 (2003) 201–228.
[12] T. Kiyosue, H. Shiota, K. Higashi, H. Kamada, K. Shinozaki, A chromobox gene from carrot (Daucus carota L.): its cDNA structure and expression during somatic and zygotic embryogenesis, Biochim. Biophys. Acta 1398 (1998) 42–46.
[13] J.L. Hilbert, T. Dubois, J. Vasseur, Detection of embryogenesis-related proteins during somatic embryo formation in Cichorium, Plant Physiol. Biochem. 30 (1992) 733–741.
[14] C. Boyer, J.L. Hilbert, J. Vasseur, Embryogenesis-related protein synthesis and accumulation during early acquisition of somatic embryogenesis competence in *Cichorium*, Plant Sci. 93 (1993) 41–53.
[15] A.J. de Jong, J. Cordewener, F. Lo Shiavo, M. Terzi, J. Vandekerckhove, A. van Kammen, S.C. de Vries, A *Daucus carota* somatic embryo mutant is rescued by chitinase, Plant Cell 4 (1992) 425–433.
[16] U. Egertsdotter, S. von Arnold, Development of somatic embryos in Norway spruce, J. Exp. Bot. 49 (1998) 155–162.
[17] A.J. Van Hengel, F. Guzzo, A. Van Kammen, S.C. De Vries, Expression pattern of the carrot EP3 endochitinase genes in suspension cultures and in developing seeds, Plant Physiol. 117 (1998) 43–53.
[18] A.J. de Jong, R. Heidstra, H.P. Spaink, M.V. Hartog, E.A. Meijer, T. Hendriks, F. Lo Shiavo, M. Terzi, T. Bisselin, A. van Kammen, S.C. de Vries, *Rhizobium* lipo-oligosaccharides rescue a *Daucus carota* somatic embryo variant, Plant Cell 5 (1993) 615–620.
[19] J.V. Dyachok, A.E. Tobin, N.P.J. Price, S. von Arnold, Rhizobial Nod factors stimulate somatic embryo development in *Picea abies*, Plant Cell Rep. 19 (2000) 290–297.
[20] J. Singh Yadav, M. Venkat Rajam, Temporal regulation of somatic embryogenesis by adjusting cellular polyamine content in eggplant, Plant Physiol. 116 (1998) 617–625.
[21] D. Bertoldi, A. Tassoni, L. Martinelli, N. Bagni, Polyamines and somatic embryogenesis in two Vitis vinifera cultivars, Physiol. Plant. 120 (2004) 657–666.
[22] I. Ruduś, E. Kępczyńska, J. Kępczyński, Comparative efficacy of abscisic acid and methyl jasmonate for indirect somatic embryogenesis in *Medicago sativa* L, Plant Growth Regul. 48 (2006) 1–11.
[23] G.S. Pullman, Y. Zhang, B.H. Phan, Brassinolide improves embryogenic tissue initiation in conifers and rice, Plant Cell Rep. 22 (2003) 96–104.
[24] R. Karlova, S. Boeren, E. Russinova, J. Aker, J. Vervoort, S. de Vries, The Arabidopsis somatic embryogenesis receptor-like kinase 1 protein complex includes brassinosteroid-insensitive 1, Plant Cell 18 (2006) 626–638.
[25] M.A. Coca, C. Almoguera, J. Jordano, Expression of sunflower low-molecular-weight heat-shock proteins during embryogenesis and persistence after germination: localization and possible functional implications, Plant Mol. Biol. 25 (1994) 479–492.
[26] E. Kitamiya, S. Suzuki, T. Sano, T. Nagata, Isolation of two genes that were induced upon the initiation of somatic embryogenesis on carrot hypocotyls by high concentrations of 2,4-D, Plant Cell Rep. 19 (2000) 551–557.
[27] K.A. Marrs, The functions and regulation of glutathione *S*-transferases in plants, Annu. Rev. Plant Physiol. Plant Mol. Biol. 47 (1996) 127–158.
[28] J. Pitto, F. Lo Schiavo, G. Guiliano, M. Terzi, Analysis of the heat-shock protein pattern during somatic embryogenesis of carrot, Plant Mol. Biol. 2 (1983) 231–237.
[29] J. Kanabus, C.S. Pikaard, J.H. Cherry, Heat-shock proteins in tobacco cell suspension during growth cycle, Plant Physiol. 75 (1984) 639–644.
[30] J.L. Zimmerman, N. Apuya, K. Darwish, C. O'Caroll, Novel regulation of heat shock genes during carrot somatic embryo development, Plant Cell 1 (1989) 1137–1146.
[31] V. Raghavan, Molecular Embryology of Flowering Plants, Cambridge Univ. Press, New York, 1997.
[32] A.M.C. Emons, Somatic embryogenesis: cell biological aspects, Acta Bot. Neerl. 43 (1994) 1–14.
[33] S.C. Fry, Polysaccharide-modifying enzymes in the plant-cell wall, Annu. Rev. Plant Physiol. Mol. Biol. 46 (1995) 497–520.
[34] T. Kapros, L. Bögre, K. Nemeth, L. Bakó, J. Gyorgyey, S.C. Wu, D. Dudits, Differential expression of histone H3 variants during cell cycle and somatic embryogenesis in alfalfa, Plant Physiol. 98 (1992) 621–625.
[35] R. Kawahara, S. Sunabori, H. Fukuda, A. Komamine, A gene expressed preferentially in the globular stage of somatic embryogenesis encodes elongation-factor 1 alpha in carrot, Eur. J. Biochem. 209 (1992) 157–162.
[36] A. Balestrazzi, I. Toscano, G. Bernacchia, M. Luo, S. Otte, D. Carbonera, Cloning of a cDNA encoding DNA topoisomerase I in *Daucus carota*, and expression analysis in relation to proliferation, Gene 183 (1996) 183–190.
[37] A. Balestrazzi, G. Bernacchia, L. Pitto, G. Luccarni, D. Carbonera, Spatial expression of DNA topoisomerase I genes during cell proliferation in *Daucus carota*, Eur. J. Histochem. 45 (2001) 31–38.
[38] K. Higashi, H. Shiota, H. Kamada, Patterns of expression of the genes for glutamine synthase isoforms during somatic and zygotic embryogenesis in carrot, Plant Cell Physiol. 39 (1998) 418–424.
[39] R. Galland, B. Randoux, J. Vasseur, J.L. Hilbert, A glutathione-S-transferase cDNA identified by mRNA differential display is upregulated during somatic embryogenesis in Cichorium, Biochim. Biophys. Acta 1522 (2001) 212–216.
[40] M.L. Preuss, D.P. Delmer, B. Liu, The cotton kinesin-like calmodulin-binding protein associates with cortical microtubules in cotton fibers, Plant Physiol. 132 (2003) 154–160.
[41] E. Lindzen, J.H. Choi, A carrot cDNA encoding an atypical protein kinase homologous to plant calcium-dependent protein kinases, Plant Mol. Biol. 28 (1995) 785–797.

[42] R.W. Giroux, K.P. Pauls, Characterization of somatic embryogenesis-related cDNAs from alfalfa (Medicago sativa), Plant Mol. Biol. 33 (1997) 393–404.
[43] V.S. Anil, A.C. Harmon, K.S. Rao, Spatio-temporal accumulation and activity of calcium-dependent protein kinases during embryogenesis, seed development, and germination in sandalwood, Plant Physiol. 122 (2000) 1035–1043.
[44] H. Hu, L. Xiong, Y. Yang, Rice SERK1 gene positively regulates somatic embryogenesis of cultured cell and host defense response against fungal infection, Planta 222 (2005) 107–117.
[45] M.A.K. Jansen, H. Booij, J.H.N. Schel, S.C. de Vries, Calcium increases the yield of somatic embryos in carrot embryogenic suspension cultures, Plant Cell Rep. 9 (1990) 221–223.
[46] P.J. Overvoorde, H.D. Grimes, The role of calcium and calmodulin in carrot somatic embryogenesis, Plant Cell Physiol. 35 (1994) 135–144.
[47] D.R. Duncan, A.L. Kriz, R. Paiva, J.M. Widholm, Globulin-1 gene expression in regenerable Zea mays (maize) callus, Plant Cell Rep. 21 (2003) 684–689.
[48] L.M. Brill, C.J. Evans, A.M. Hirsch, Expression of MsLEC1- and MsLEC2-antisense genes in alfalfa plant lines causes severe developmental and reproductive abnormalities, Plant J. 25 (2001) 453–461.
[49] J.M. Domon, B. Dumas, E. Laine, Y. Meyer, A. David, H. David, Three glycosylated polypeptides secreted by several embryogenic cell cultures of pine show highly specific serological affinity to antibodies directed against the wheat germin apoprotein monomer, Plant Physiol. 108 (1995) 141–148.
[50] G. Neutelings, J.M. Domon, N. Membre, F. Bernier, Y. Meyer, A. David, H. David, Characterization of a germin-like protein gene expressed in somatic and zygotic embryos of pine (*Pinus caribaea* Morelet), Plant Mol. Biol. 38 (1998) 1179–1190.
[51] D. Patnaik, P. Khurana, Germins and germin like proteins: an overview, Indian J. Exp. Biol. 39 (2001) 191–200.
[52] L. Filonova, P. Bozhkov, V. Brukhin, G. Daniel, B. Zhivotovsky, S. von Arnold, Two waves of programmed cell death occur during formation and development of somatic embryos in the gymnosperm, Norway spruce, J. Cell. Sci. 113 (2000) 4399–4411.
[53] A.P. Smertenko, P.V. Bozhkov, L.H. Filonova, S. von Arnold, P.J. Hussey, Re-organisation of the cytoskeleton during developmental programmed cell death in Picea abies embryos, Plant J. 33 (2003) 813–824.
[54] K. Takahata, M. Takeuchi, M. Fujita, J. Azuma, H. Kamada, F. Sato, Isolation of putative glycol-protein gene from early somatic embryo of carrot and its possible involvement in somatic embryo development, Plant Cell Physiol. 45 (2004) 1658–1668.
[55] A. Fehér, T. Pasternak, P. Miskolczi, F. Ayaydin, D. Dudits, Induction of the embryogenic pathway in somatic plant cells, Acta Horticult. 560 (2001) 293–298.
[56] A. Fehér, T. Pasternak, K. Ötvös, P. Miskolczi, D. Dudits, Induction of embryogenic competence in somatic plant cells: a review, Biologia 57 (2002) 5–12.
[57] T.P. Pasternak, E. Prinsen, F. Ayaydin, P. Miskolczi, G. Potters, H. Asard, H.A. Van Onckelen, D. Dudits, A. Fehér, The role of auxin, pH, and stress in the activation of embryogenic cell division in leaf protoplast-derived cells of alfalfa, Plant Physiol. 129 (2002) 1807–1819.
[58] H.K. Lichtenthaler, The stress concept in plants: an introduction, Ann. N. Y. Acad. Sci. 851 (1998) 87–198.
[59] E. Charbit, T. Legavre, L. Lardet, E. Bourgeois, N. Ferriere, M.P. Carron, Identification of differentially expressed cDNA sequences and histological characteristics of *Hevea brasiliensis* calli in relation to their embryogenic and regenerative capacities, Plant Cell Rep. 22 (2004) 539–548.
[60] M. Libik, R. Konieczny, B. Pater, I. Slesak, Z. Miszalski, Differences in the activities of some antioxidant enzymes and in H2O2 content during rhizogenesis and somatic embryogenesis in callus cultures of the ice plant, Plant Cell Rep. 23 (2005) 834–841.
[61] A. Nishimura, M. Ashikari, S. Lin, T. Takashi, E.R. Angeles, T. Yamamoto, M. Matsuoka, Isolation of a rice regeneration quantitative trait loci gene and its application to transformation systems, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 11940–11944.
[62] E. Gianazza, P. De Ponti, A. Scienza, P. Villa, L. Martinelli, Monitoring by two-dimensional electrophoresis somatic embryogenesis in leaf and petiole explants from *Vitis*, Electrophoresis 13 (1992) 203–209.
[63] A. Sallandrouze, M. Faurobert, M.E. Maataoui, H. Espagnac, Two-dimensional electrophoretic analysis of proteins associated with somatic embryogenesis development in *Cupressus sempervirens* L, Electrophoresis 20 (1999) 1109–1119.
[64] D. Dudits, L. Bögre, J. Györgyey, Molecular and cellular approaches to the analysis of plant embryo development from somatic cells *in vitro*, J. Cell. Sci. 99 (1991) 475–484.
[65] A.J. de Jong, E.D.L. Schmidt, S.C. de Vries, Early events in higher plant embryogenesis, Plant Mol. Biol. 22 (1993) 367–377.
[66] L. Lu, Y.R. Lee, R. Pan, J.N. Maloof, B. Liu, An internal motor kinesin is associated with the Golgi apparatus and plays a role in trichome morphogenesis in Arabidopsis, Mol. Biol. Cell 16 (2005) 811–823.
[67] S. Hata, H. Kouchi, I. Suzuka, T. Ishii, Isolation and characterization of cDNA clones for plant cyclins, EMBO J. 10 (1991) 2681–2688.
[68] H. Hirt, A. Páy, J. Györgyey, L. Bakó, K. Németh, L. Bögre, R.J. Schweyen, E. Heberle-Bors, D. Dudits, Complementation of a yeast cell cycle mutant by an alfalfa cDNA encoding a protein kinase homologous to $p34^{cdc2}$, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 1636–1640.
[69] H. Hirt, Connecting oxidative stress, auxin, and cell cycle regulation through a plant mitogen-activated protein kinase pathway, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 2405–2407.
[70] T. Pasternak, P. Miskolczi, F. Ayaydin, T. Mészáros, D. Dudits, A. Fehér, Exogenous auxin and cytokinin dependent activation of CDKs and cell division in leaf protoplast-derived cells of alfalfa, Plant Growth Regul. 32 (2000) 129–141.
[71] A. Nikitin, S. Egorov, N. Daraselia, I. Mazo, Pathway Studio—the analysis and navigation of molecular networks, Bioinformatics 19 (2003) 2155–2157.
[72] B. Lehner, A.G. Fraser, A first-draft human protein-interaction map, Genome Biol. 5 (2004) R63.
[73] L. van Zyl, P.V. Bozhkov, D.H. Clapham, R.R. Sederoff, S. von Arnold, Up, down and up again is a signature global gene expression pattern at the beginning of gymnosperm embryogenesis, Gene Expr. Patterns 3 (2003) 83–91.
[74] N. Imin, F. de Jong, U. Mathesius, G. van Noorden, N.A. Saeed, X.D. Wang, R.J. Rose, B.G. Rolfe, Proteome reference maps of *Medicago truncatula* embryogenic cell cultures generated from single protoplasts, Proteomics 4 (2004) 1883–1896.
[75] N. Imin, M. Nizamidin, D. Daniher, K.E. Nolan, R.J. Rose, B.G. Rolfe, Proteomic analysis of somatic embryogenesis in *Medicago truncatula*: explant cultures grown under 6-benzylaminopurine and 1-naphthaleneacetic acid treatments, Plant Physiol. 137 (2005) 1250–1260.
[76] D. Lippert, J. Zhuang, S. Ralph, D.E. Ellis, M. Gilbert, R. Olafson, K. Ritland, B. Ellis, C.J. Douglas, J. Bohlmann, Proteome analysis of early somatic embryogenesis in *Picea glauca*, Proteomics 5 (2005) 461–473.
[77] T. Winkelmann, D. Heintz, A. Van Dorsselaer, M. Serek, H.P. Braun, Proteomic analyses of somatic and zygotic embryos of *Cyclamen persicum* Mill. reveal new insights into seed and germination physiology, Planta 224 (3) (2006) 508–519.
[78] G. Zhu, L. Reynolds, T. Crnogorac-Jurcevic, C.E. Gillett, E.A. Dublin, J.F. Marshall, D. Barnes, C. D'Arrigo, P.O. Van Trappen, N.R. Lemoine, I.R. Hart, Combination of microdissection and microarray analysis to identify gene expression changes between differentially located tumour, Oncogene 22 (2003) 3742–3748.