Research

Genome-wide investigation of light and carbon signaling interactions in *Arabidopsis*

Karen E Thum*, Michael J Shin*, Peter M Palenchar*, Andrei Kouranov*† and Gloria M Coruzzi*

Addresses: *Department of Biology, New York University, New York, NY 10003, USA. †Current address: Center for Bioinformatics, University of Pennsylvania, 423 Guardian Drive, Philadelphia, PA 19104, USA.

Correspondence: Gloria M Coruzzi. E-mail: gloria.coruzzi@nyu.edu

Published: 27 January 2004

Genome ***Biology*** 2004, **5:**R10

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2004/5/2/R10

Received: 26 August 2003
Revised: 25 November 2003
Accepted: 12 December 2003

Abstract

Background: Light and carbon are two essential signals influencing plant growth and development. Little is known about how carbon and light signaling pathways intersect or influence one another to affect gene expression.

Results: Microarrays are used to investigate carbon and light signaling interactions at a genome-wide level in *Arabidopsis thaliana.* A classification system, 'InterAct Class', is used to classify genes on the basis of their expression profiles. InterAct classes and the genes within them are placed into theoretical models describing interactions between carbon and light signaling. Within InterAct classes there are genes regulated by carbon (201 genes), light (77 genes) or through carbon and light interactions (1,247 genes). We determined whether genes involved in specific biological processes are over-represented in the population of genes regulated by carbon and/or light signaling. Of 29 primary functional categories identified by the Munich Information Center for Protein Sequences, five show over-representation of genes regulated by carbon and/or light. Metabolism has the highest representation of genes regulated by carbon and light interactions and includes the secondary functional categories of carbon-containing-compound/carbohydrate metabolism, amino-acid metabolism, lipid metabolism, fatty-acid metabolism and isoprenoid metabolism. Genes that share a similar InterAct class expression profile and are involved in the same biological process are used to identify putative *cis* elements possibly involved in responses to both carbon and light signals.

Conclusions: The work presented here represents a method to organize and classify microarray datasets, enabling one to investigate signaling interactions and to identify putative *cis* elements *in silico* through the analysis of genes that share a similar expression profile and biological function.

Background

Carbon and light are two important signals regulating plant growth and development. They exert their effects primarily through their ability to affect the expression of a large number of genes. While much is known about the way in which plants respond to and transduce light signals [1-4], less is known about the perception and transduction of carbon signals [5-7]. More recently, it has been demonstrated that carbon and

light signaling pathways influence one another, suggesting crosstalk between these pathways exists [8-11]. In comparison to what is understood about individual signaling pathways of carbon and light, much less is known about the way in which carbon and light signaling pathways intersect or influence one another.

To investigate the interaction of light quality (red, far-red, blue and white light) and quantity (low versus high fluence) with carbon with respect to gene regulation, previous studies from our lab have focused on a 'systems' analysis of genes involved in amino-acid metabolism, known to be regulated by both carbon and light [11]. Other studies have focused on the isolation of mutants involved in carbon sensing and signaling [6] or have focused more specifically on the influence of carbon on phytochrome signal transduction pathways [8-10]. Other analyses have demonstrated that some genes encoding proteins involved in or relating to photosynthesis are strongly induced by light yet repressed by carbon (for example, chlorophyll *a*/*b* binding protein, plastocyanin, small subunit of Rubisco) [5]. Other genes are induced by carbon in dark-adapted plants [5,11-13], but not induced by carbon in light-treated plants (for example, glutamine synthetase 2, asparagine synthetase 2) [11]. More specific interactions between carbon and light have been observed by the ability of carbon to suppress a phytochrome-A-specific, far-red light-induced block of greening. Hence, carbon may antagonize or suppress the phytochrome A signaling pathway(s) in this case [8]. Thus, from these studies and the varying effects that light and carbon have on the expression of genes, it is clear that there are different modes of regulation by these two signals on specific subsets of genes and that the interactions between light and carbon are complex.

Whereas the studies outlined above have provided insight into the interactions between light and carbon signaling, they are limited in the fact that only a few genes have been analyzed at a time. Thus, the general picture of how interactions between light and carbon affect gene regulation on a genomic scale overall cannot be gleaned from these studies. Fortunately, with the use of cDNA microarrays, a semi-quantitative analysis of genome-wide patterns of gene expression can be carried out [14]. Microarray analysis has been used to investigate gene-expression patterns modulated by circadian rhythms [15,16], light [17-19], nutrients [20,21], as well as responses to drought and cold stress [22] in plants.

In this study, we sought to investigate the interactions between carbon and white-light signaling pathways by examining the expression profiles of genes responding to carbon and/or light at the genome-wide level using *Arabidopsis thaliana* gene chips from Affymetrix. Here, we identify and classify those genes responsive to carbon and/or light according to specific models presented herein. We also identify specific biological processes involving genes that are subject to a significant degree of regulation by carbon and/or light compared to the general population of genes on the chip. Other studies have investigated changes in gene expression using single comparisons, and therefore a single input, coupled with statistical analysis for global identification of putative *cis* elements in yeast [23-25]. We use a similar approach to identify *cis* elements involved in the regulation of two inputs (carbon and light). Our approach is more restrictive because genes that are both similarly regulated and functionally related are used to initiate this analysis. Here, our analysis led to the identification of putative *cis*-regulatory elements that are proposed to function downstream of carbon and light signaling pathways. Finally, the work in this study presents a method to organize and classify microarray datasets that are smaller than can typically be used for clustering analysis [26], and enables one to investigate biological interactions and to further identify putative *cis*-regulatory elements *in silico* through the analysis of genes that share both a similar expression profile and biological function.

Results

To investigate light and carbon signaling interactions at the genome-wide level, genes responding to either white light and/or sucrose were analyzed using microarrays. Affymetrix AG gene chips containing 8,000 unique genes from *A. thaliana*, approximately one-third of the entire genome, were analyzed after hybridization with cRNA from 14-day-old plants subjected to four different light and/or carbon treatment conditions as shown in Figure 1a (-C-L, -C+L, +C-L and +C+L) (see Materials and methods for experimental details). Comparative analyses of these RNA samples by microarrays were used to identify genes responding to light only or carbon only, and genes that are regulated by both light and carbon, as described below.

Models for predicting light and carbon signaling interactions

Light and carbon signaling may interact in the regulation of a hypothetical gene (for example, X, Y or Z) by one of three models shown in Figure 1b. In model 1, carbon and light are shown to act independently of one another to affect expression of gene X, in either a positive or negative manner in various combinations. In this example, both carbon and light are shown to act in a positive manner. However, for another gene, light may induce transcription, while carbon may independently repress (depress) transcription. Alternatively, for another gene, carbon and light may independently act to repress transcription. Model 2 predicts that carbon and light are two signals that are dependent, in that they are both required to affect expression of gene Y, where either signal alone has no effect on gene expression (Figure 1b). This could happen in a positive or negative manner. Model 3 predicts that carbon and light signals act both independently of and dependent on one another to influence the expression of gene Z. In the example shown in Figure 1b for model 3, gene Z is induced to a certain level in the presence of carbon alone and

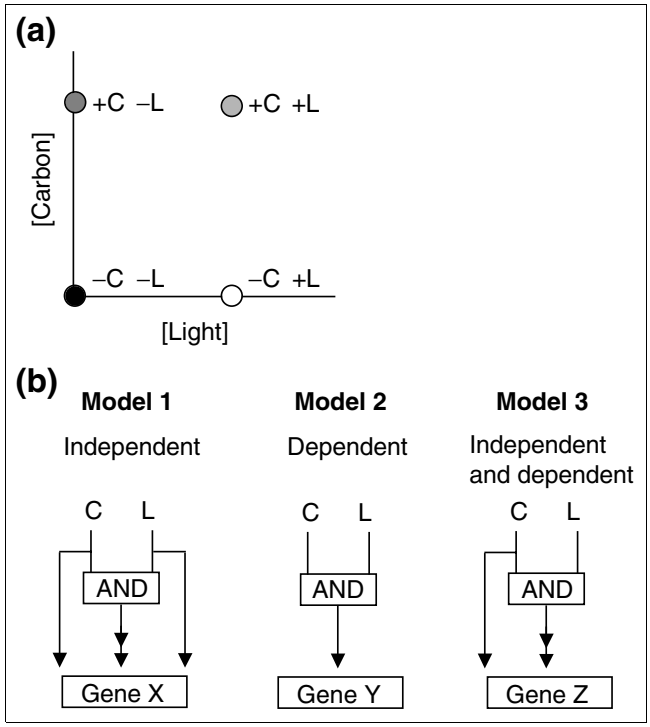

Figure 1
Light and carbon treatments and their effects on gene expression.
(a) Diagrammatic representation of various carbon and light treatments given to plants before microarray analyses. -C-L, no carbon and no light; -C+L, no carbon and 70 μE/m²/sec white light (8 h); +C-L, 1% (w/v) sucrose and no light; +C+L, 1% (w/v) sucrose and 70 μE/m²/sec white light (8 h). **(b)** Three models showing theoretical interactions between carbon (C) and light (L) that influence the expression of a hypothetical gene (X, Y, or Z). Model 1 shows carbon and light acting independently, model 2 shows the actions of carbon and light dependent on one another, and model 3 shows both dependent and independent interactions between carbon and light. The arrows designate any effect of carbon and/or light on the expression of gene X, Y, or Z, whether it is repressive or inductive.

is unresponsive to light alone (in the absence of carbon). However, in the presence of both carbon and light, the expression level of that particular gene exceeds that observed in the presence of carbon alone. Thus, the carbon and light regulation of gene Z is responsive to carbon alone, yet requires carbon for induction by light. This is but one of many examples of interactions between carbon and light in model 3. We validate the existence of all three models for carbon and light interactions through the analysis of genome-wide responses to carbon and/or light, as detailed below.

Using InterAct Class to classify genes regulated by carbon and/or light on the basis of their expression profiles

RNA samples obtained from plants subjected to four different light and/or carbon conditions (see Figure 1a and Materials and methods) were hybridized to Affymetrix AG gene chips. Gene-expression analysis was carried out using Affymetrix's Microarray Suite 5.0 software. Microarray data obtained from plants treated with no carbon and no light (-C-L) were used as the control background (Table 1, -C/-L) to which data from all other treatments (that is, +C/-L, +C/+L and -C/+L) were compared (see Table 1). Using this comparative approach, genes responding to carbon or light, as well as genes responding to both carbon and light, were identified at the genomic level as I, D or NC (I, induced; D, depressed; NC, no change). To analyze further the relative response of a gene under each of the three treatments, a classification system termed InterAct Class [27] was used. InterAct Class allowed us to classify genes on the basis of their relative regulation by carbon and/or light. The InterAct Class program takes Affymetrix difference calls of treatment to control (NC, I, D), and further classifies a gene on the basis of its relative expression in each of the three treatments: carbon alone; carbon and light combined; or light alone. For example, in Table 1, row 1, gene X is induced by carbon alone (fivefold), and by light alone (fivefold), and is induced 10-fold in the presence of both carbon and light. Gene X thus falls into InterAct class 121 (1 C/ 2 C+L/ 1 L; Table 1, row 1). In another example, shown in Table 1, row 2, gene Y shows no change in expression in the presence of carbon or light alone, yet is induced threefold in the presence of both carbon and light. As such, gene Y is placed in the 010 InterAct class (0 C/ 1 C+L/ 0 L). In another example, gene Z shows a similar level of induction by carbon alone, or light alone, or by carbon plus light (Table 1). Gene Z falls into InterAct class 111 (1 C/ 1 C+L/ 1 L).

For this study, InterAct Class ranked gene-expression values from -3 to 3, with -3 being the most depressed value and 3 being the highest expression value. Therefore, for each comparison any one gene was assigned one of seven possible values: -3, -2, -1, 0 (no change), 1, 2 or 3. Taking into consideration all possible combinations, genes were theoretically categorized into 7^3 or 343 permutations of gene regulation by carbon and/or light. However, the actual number of possible InterAct classes was lower, because InterAct Class ranks the change in expression levels of the conditions relative to one another for each gene. Thus, in order to have a value of 3, there must be a 2 in the InterAct class, and in order to have a value of 2, there must also be a 1 in the InterAct class. Using this criteria, 72 potential InterAct classes were generated rather than the theoretical 7^3 permutations. Microarray expression data was used to place *A. thaliana* genes into one of 72 potential InterAct classes. Thus, the *in vivo* significance of various models proposed for gene responses to light and carbon (Figure 1b) were validated by identifying genes whose expression profiles fit into one of these 72 InterAct classes.

Out of 72 potential InterAct classes, there were 43 classes whose existence was validated by the expression pattern of at least one gene. Of the 8,000 genes present on the partial *A. thaliana* genome array, 24.9% (1,997) unique genes were placed into InterAct classes on the basis of their expression profiles that met certain criteria. For a gene to be placed into

Table 1

Example of how genes are assigned an InterAct class

Row	Control	Treatments			InterAct class			L/C interactions	
		+C/-L	+C/+L	-C/+L	C	C+L	L	Theoretical	Example
1	-C/-L	Induced (5x)	Induced (10x)	Induced (5x)	1	2	1	Independent (model 1)	Inductive
2	-C/-L	No change (0x)	Induced (3x)	No change (0x)	0	1	0	Dependent (model 2)	Synergistic
3	-C/-L	Induced (5x)	Induced (5x)	Induced (5x)	1	1	1	Independent and dependent (model 3)	Equal effect

an InterAct class, it must have been called present on both gene chip replicates, and have a similar expression profile in both biological replicates under all analyses (see Materials and methods for more details). For instance, if a gene was induced to similar levels by carbon or light alone, and was induced even further by carbon and light together in one replicate, but a similar expression profile was not observed in the second replicate, this gene would be discarded and not placed into an InterAct class. Because partial *A. thaliana* genome chips were used, the number of genes that were placed in an InterAct class is an underestimate. The number of genes that validate specific InterAct classes is shown in Table 2. Of the 43 InterAct classes that contained genes, 39 are shown in Table 2. The four InterAct classes that contained only one or two genes are not shown in Table 2.

InterAct classes that were validated by chip data were further used to categorize genes into specific models of carbon and light regulation (Table 2). Table 2 shows the number of genes regulated by carbon alone or light alone, and the number of genes whose regulation is affected by the interactions between carbon and light. Those genes that were categorized as being subject to regulation by carbon and light were further placed under one of the three models predicting carbon and light interactions presented in Figure 2. Within these models, the response of genes to carbon and/or light can be described as inductive or repressive (model 1 or model 2), antagonistic (model 1), suppressed, enhanced, equally affected by either signal, or dominated by either light or carbon (model 3). These responses of genes to carbon or light and the interaction between carbon and light become evident upon closer examination of the numbers of genes validating a particular InterAct class shown in Table 2. For instance, InterAct class 121 (40 genes), fits within model 1, because genes within this class are likely to be induced independently by both carbon and light (Table 2). By contrast, InterAct class -1-2-1 (129 genes), is also under model 1, but the expression of genes within this class are likely to be repressed independently by both carbon and light (Table 2).

Two hundred and seventy-eight genes are classified as being responsive to carbon alone or to light alone (Table 2). In contrast, 1,247 genes comprise the InterAct classes that show regulation by carbon and light and fall under one of the three models presented in Figure 1b (Table 2). The responses to carbon and light and possible interactions between carbon and/or light within such InterAct classes are described in Table 2 (carbon and light-response column), according to the various general models proposed in Figure 1b.

Identifying primary functional categories of genes regulated by carbon and light

We next determined whether the signals - carbon and light - or interactions between carbon and light signaling regulate specific functional classes of genes involved in particular biological functions. For this analysis, the classification system of the Munich Information Center for Protein Sequences (MIPS) [28,29] was used in conjunction with InterAct Class to determine whether gene function correlates with specific patterns of gene regulation by carbon and/or light. The MIPS system sorts genes according to functional categories called 'funcats'. MIPS classifications are separated into 29 major (primary) funcats that are further subdivided into minor (secondary, tertiary, and so on) functional categories [28,29]. We tested whether the genes within a specific primary funcat were significantly ($p < 0.05$) over- or under-represented in a particular InterAct class, as compared to the general population of genes on the chip that were assigned both a funcat and an InterAct class. This was determined by calculating the percentage of genes expected to be in a InterAct class on the basis of the general population, versus the percentage of genes in a specific funcat that were grouped into a single InterAct class. Using all chip data, there were 1,898 Affymetrix IDs that fell into a specific funcat and that were also assigned to a specific InterAct class (Table 3). InterAct class 000 was used as an example for this analysis. The genes that fell into InterAct class 000 were not regulated by light and/or carbon. Therefore, if genes in a particular funcat were under-represented in the 000 class, this indicated that the particular funcat contained more genes that are regulated by carbon and/or light than expected from the general population of genes on the chip. In addition, a funcat that contains genes that were over-represented within the 000 class indicates that genes in that particular funcat are less regulated by carbon and/or light

Table 2

Number of genes validating InterAct classes

	Class (39 validated/72 theoretical)			Number of genes (1,997/8,000)	Model	C+L response
Row	C	C+L	L			
1	0	0	0	467	None	No effect
2	1	1	0	143	C-only	C-only
3	-1	-1	0	58	(201 genes)	
4	0	1	1	65	L-only	L-only
5	0	-1	-1	12	(77 genes)	
L- and C-independent: model 1						
6	-1	-2	-1	129	Repressive	
7	-2	-3	-1	70	(model 1)	
8	-1	-3	-2	55	(254 genes)	
9	1	2	1	40	Inductive	
10	2	3	1	43	(model 1)	
11	1	3	2	14	(97 genes)	
12	-1	0	1	27	Antagonistic	
13	-1	1	2	23	(model 1)	
14	-2	-1	1	6	(56 genes)	
L- and C-dependent: model 2						C + L responses (1,247 genes)
15	0	1	0	19	Dependent	
16	0	-1	0	13	(model 2) (32 genes)	
L- and C-dependent and independent: model 3						
17	2	2	1	154	C-dominates	
18	-2	-2	-1	53	(model 3)	
19	0	0	1	15	(235 genes)	
20	0	0	-1	13		

Table 2 *(Continued)*

Number of genes validating InterAct classes						
21	1	0	0	54		
22	1	2	2	45	L-dominates	
23	-1	0	0	36	(model 3)	
24	-1	-2	-2	27	(173 genes)	
25	-1	1	1	6		
26	2	1	1	5		
27	1	1	1	86	Equal effect	
28	-1	-1	-1	40	(model 3) (126 genes)	
29	2	1	0	92		
30	3	2	1	46		C + L
31	0	1	2	47	Suppressed	responses
32	-2	-1	0	20	(model 3)	(1,247 genes)
33	1	2	3	14	(229 genes)	
34	-3	-2	-1	6		
35	0	-1	-2	4		
36	-1	-2	0	29	Enhanced	
37	1	2	0	6	(model 3)	
38	0	-2	-1	6	(45 genes)	
39	0	2	1	4		

compared to the general population of genes on the chip. Thus, focusing on the 000 InterAct class and determining which funcats were under-represented directed us to the funcats that are significantly regulated by carbon and/or light.

For example, it was observed that in the general population, 440 out of 1,898 genes were placed into the InterAct class 000 (Table 3). Thus, 23.1% (440/1,898) of all the genes on the chip that were assigned both a funcat and an InterAct class are in class 000. Genes in nine of these funcats showed statistically significant under- or over-representation (-S or +S, respectively) in InterAct class 000 compared to the general population (Table 3). Among these was the primary funcat metabolism (45 genes) which was found to be statistically under-represented within the 000 InterAct class (45/274; 16.4%), compared to the general population (440/1,898; 23.2%). Additionally, a number of other primary funcats were also under-represented in InterAct class 000, indicating these particular funcats contain more genes that are regulated by carbon and/or light than expected in the general population of genes on the chip. These primary funcats which contain a significant number of genes regulated by carbon and/or light include protein fate, protein synthesis, energy, cell rescue, defense and virulence (see Table 3, -S). Conversely, other funcats were over-represented in the 000 InterAct class. This indicates that these processes are more immune to regulation by carbon and light compared to the general population of genes on the chip that are assigned both a funcat and an InterAct class. These funcats included unclassified proteins, transcription, cellular communication/signal transduction, cell cycle and DNA processing (Table 3). Other primary funcats that showed significance within any other InterAct class (besides 000) are shown in Table 3. The actual number of additional InterAct classes in which a particular funcat was under- or over-represented is shown in the last column of Table 3.

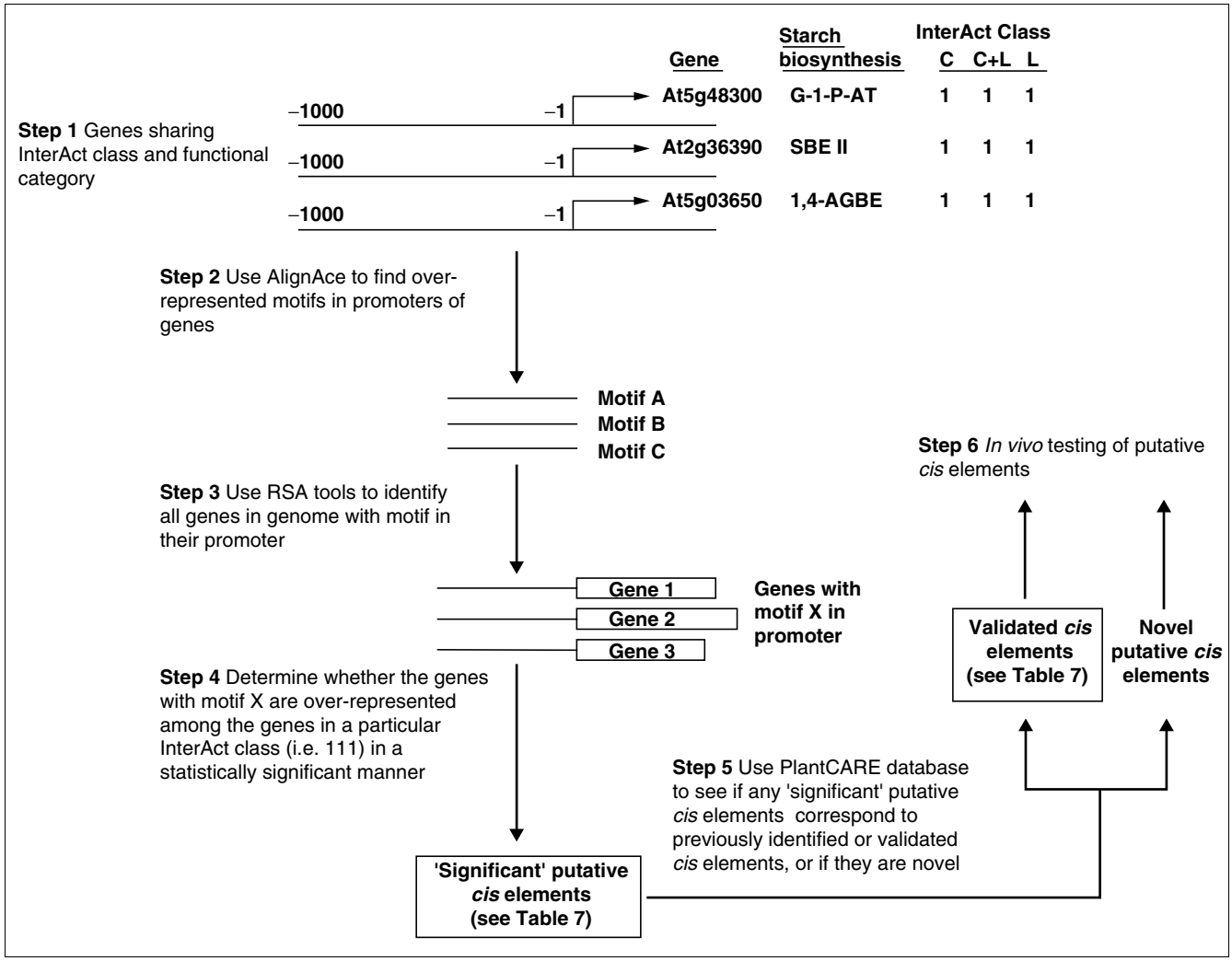

Figure 2
Flow chart of the procedure used to identify common potential light- or carbon-responsive (LCR) *cis* elements in genes that share both a similar expression profile and a functional category.

Further analysis of the regulation of genes in the primary funcat metabolism was carried out because metabolism was one of the primary funcats that had the largest number of genes that was assigned an InterAct class and because 45 genes in this funcat were significantly under-represented in the 000 InterAct class. This suggests that metabolism is a process highly regulated by light and/or carbon. Metabolism was used as a working example of the type of further analysis that could also be performed for any of the funcats listed in Table 3 that are also regulated by light and carbon.

Secondary funcats of metabolism and significance of all models of carbon and light interactions

To identify the relative dominance and interaction of carbon and light in regulating genes in the metabolism funcat, we first examined whether any secondary funcats of metabolism were under- or over-represented in a particular InterAct class. A breakdown of the secondary funcats of metabolism according to their distinct carbon and/or light regulation models (Figure 1b) is shown in Table 4. In each column the models noted above correspond to InterAct classes shown in Table 2. Each row shows the number of genes in a secondary funcat of metabolism and their significance (over- or under-representation, +S or -S, respectively) in distinct models of carbon and/or light regulation. Bold type in the table indicates the funcats with a statistically significant number of genes that fell into an InterAct class that corresponds to a particular mode of carbon and/or light regulation. Funcats in plain type are not statistically significant. Upon further dissection of the primary funcat metabolism into secondary funcats, it was observed that the secondary funcats C-compound (carbon-containing compound)/carbohydrate metabolism and amino-acid metabolism were each under-represented in 000 InterAct class (Table 4). This analysis indicates that

Table 3

Primary functional categories whose genes show over-representation or under-representation in InterAct class 000

Primary funcat	Total number of genes in a funcat assigned an InterAct class	Number of genes in InterAct class 000	Significance of class 000 in funcat	Number of other significant InterAct classes in funcat
Total	1,898	440	N/A	N/A
Metabolism*	274	**45**	**-S**	2
Protein fate (modification, folding, destination)	149	**22**	**-S**	2
Protein synthesis	106	**12**	**-S**	5
Energy	81	**3**	**-S**	1
Cell rescue, defense and virulence	76	**10**	**-S**	1
Unclassified proteins	632	**167**	**+S**	2
Transcription	135	**45**	**+S**	3
Cellular communication/signal transduction	130	**43**	**+S**	2
Cell cycle and DNA processing	49	**17**	**+S**	1
Cell fate	47	16	NS	2
Subcellular localization	387	76	NS	1
Transport facilitation	122	30	NS	1
Control of cellular organization	61	10	NS	2
Development (systemic)	51	17	NS	1
Systemic regulation of/interaction with environment	39	10	NS	2
Regulation of/interaction with cellular environment	25	4	NS	3
Protein activity regulation	14	4	NS	1
Organ differentiation	4	0	NS	1

The table indicates those funcats that are responding to light and/or carbon (for example, under-represented in InterAct class 000). -S or +S means that the genes in a particular functional category are significantly ($p < 0.05$) under-represented or over-represented in InterAct class 000, respectively. N/A, not applicable. Numbers in bold type are statistically significant; NS, non-significant. * The funcat that is the focus of this paper.

genes in these processes are significantly more regulated by carbon and/or light than the general population of genes on the chip that were assigned both a funcat and an InterAct class. Indeed, genes in these secondary funcats were over-represented in InterAct classes corresponding to model 3, carbon and light interactions (Table 4).

For further analysis, we focused on secondary funcats of metabolism whose genes were over-represented in the InterAct classes that suggest regulation by carbon and light (Table 4). In the general population, 386/1,898 genes (20.3%) validated the existence of model 1 (carbon and light independent) (Table 4) of which 91/1,898 genes (4.8%) fell under the 'inductive' classification of model 1 (Table 4). The primary funcat metabolism (274 genes) contained 21 genes whose expression patterns fit this model. Thus, the primary funcat metabolism was over-represented within model 1, 21/274 (7.6%) genes compared to the general population (Table 4). Out of those 21 metabolism genes, seven genes within the secondary funcat lipid, fatty-acid and isoprenoid metabolism were over-represented within the InterAct classes belonging to this model, and two genes within the secondary funcat nitrogen and sulfur metabolism also showed over-representation in InterAct classes belonging to model 1 (Table 4). This suggests that genes involved in lipid, fatty-acid, and isoprenoid metabolism are highly regulated by light and carbon independently. Whereas five genes in metabolism fell into model 2 (carbon and light are dependent), no specific funcat was under- or over-represented in this model, compared to the general population (Table 4).

Model 3 was validated by the general population of genes in which 776/1,898 (40.8%) fell into model 3 (Table 4). For the 'C-dominates' classification of model 3, 223 genes or 11.7% of the population (223/1,898) were found under this model (Table 4). Fifty-one genes in the metabolism primary funcat were over-represented in InterAct classes that belonged to 'C-dominates', model 3 (51/274; 18.6%). Out of those 51 genes, the funcats C-compound/carbohydrate metabolism (17 genes) and amino-acid metabolism (12 genes) were each

Table 4

Number of genes in the metabolism primary and secondary funcats in each InterAct class that are regulated by light only, carbon only and light and/or carbon

							L+C interactions									
	Total	No effect	C-only		L-only		Model 1-independent			Model 2-dependent		Model 3-dependent and independent				
			I	R	I	R	I	R	A	I	R	C-dominates	L-dominates	EE	S	E
Total	1,898	440	135	56	62	12	91	242	53	19	13	223	166	124	218	45
Metabolism*	274	**45-S**	20	3	13	0	**21+S**	30	5	4	1	**51+S**	27	24	24	7
C-compound/ carbohydrate metabolism	84	**12-S**	5	1	4	0	6	8	2	1	0	**17+S**	8	**10+S**	9	1
Amino-acid metabolism	44	**2-S**	4	1	4	0	3	7	1	0	0	**12+S**	4	2	3	1
Lipid, fatty-acid, isoprenoid, metabolism	40	9	2	0	0	0	**7+S**	5	0	2	0	5	5	3	2	0
Metabolism of vitamins, co-factors, prosthetic groups	20	8	1	1	**3+S**	0	1	0	0	0	0	2	0	2	2	0
Nitrogen, sulfur metabolism	13	2	2	0	**2+S**	0	**2+S**	0	0	0	0	3	0	1	0	**2+S**

-S and +S indicate statistical ($p < 0.05$) under-representation or over-representation of genes in a particular funcat for a mode of regulation by carbon and/or light. *Metabolism is a primary funcat. Numbers in bold type are statistically significant. I, inductive; R, repressive; A, antagonistic; EE, equal effect; S, suppressed; E, enhanced.

over-represented in the InterAct classes that correspond to 'C-dominates', model 3 (Table 4). There were 124 genes in the general population (124/1,898; 6.5%) that fell into another variation of model 3, where light and carbon have equal effects on gene expression, (Table 4). Genes in the primary funcat metabolism were not over-represented in the InterAct classes that belonged to this model. However, there were 10 genes in the secondary funcat metabolism C-compound/carbohydrate metabolism that were over-represented in the InterAct classes which belonged to 'equal effect', model 3 (Table 4). Genes in InterAct classes that belonged to 'equal effect', model 3 were equally affected by light, carbon and carbon plus light (for example, 111 or -1-1-1). Thus, genes that fell under these two InterAct classes may contain a single *cis* element that responds to either light or carbon. Alternatively, carbon regulation of these genes may be due to an indirect effect of light. This would occur through an increase in carbon skeletons generated through photosynthesis. Below we show how a *cis*-analysis of genes in the 111 InterAct class (model 3) was used to identify putative *cis* elements responsive to either carbon or light. This analysis is an example of the type of data mining that can be carried out for other funcats that show genes over-represented in an InterAct class, implicating various models for carbon and light interactions.

Breakdown of model 3 shows over-representation of secondary funcats of metabolism in InterAct class 111

Model 3 (equal effect of carbon and/or light) comprises the two InterAct classes, 111 (85 genes) and -1-1-1 (39 genes) (Table 5). Out of 274 genes that are categorized as being involved in metabolism and placed into any InterAct class, 22 fell into class 111. Thus, metabolism genes were over-represented within this InterAct class (22/274; 8%) compared to the general population (85/1,898; 4.5%) (Table 5). Of these 22 metabolism genes, nine in the secondary funcat, C-compound/carbohydrate metabolism also showed over-representation in InterAct class 111 (9/84; 10.7%) compared to the general population (85/1,898; 4.5%) (Table 5). This suggests that C-compound/carbohydrate metabolism has a significant number of genes that are equally regulated by light and/or carbon according to model 3. Although there were other genes from some of the metabolism secondary funcats that fell into InterAct classes 111 and -1-1-1, they were not statistically significant compared to the general population.

Genes within C-compound/carbohydrate metabolism and InterAct Class 111

The genes categorized under the secondary funcat C-compound/carbohydrate metabolism were further subdivided according to the specific processes in which they are involved and by InterAct class. One subset of these genes, belonging to C-compound/carbohydrate metabolism and 111 InterAct class, all encode proteins involved in starch biosynthesis and include glucose-1-phosphate adenylytransferase, starch branching enzyme II and 1,4-alpha-glucan branching enzyme protein isoform SBE2 (Table 6). Another subset of genes that belong to C-compound/carbohydrate metabolism and InterAct class 111 are involved in cell-wall metabolism/

Table 5

Significance of the number of genes in metabolism secondary funcats in InterAct classes 111 and -1-1-1 in model 3

		L+C interactions under model 3 (equal effect of carbon and light)					
		C	C+L	L	C	C+L	L
		1	1	1	-1	-1	-1
Funcat	Total number of genes	Number of genes			Number of genes		
Total	1,898	85			39		
Metabolism*	274	**22+S**			2		
C-compound/carbohydrate metabolism	84	**9+S**			1		
Amino-acid metabolism	44	2			0		
Lipid, fatty-acid and isoprenoid metabolism	40	3			0		
Nitrogen and sulfur metabolism	13	1			0		
Secondary metabolism	47	4			1		
Nucleotide metabolism	26	3			0		
Metabolism of vitamins, co-factors and prosthetic groups	20	2			0		

+S means the particular funcat is significantly ($p < 0.05$) over-represented in InterAct class 111 or -1-1-1. Significance is only determined for the number of genes in InterAct classes 111 and -1-1-1 and not for the total number of genes in each functional category. * Metabolism is a primary funcat. Numbers in bold type are statistically significant.

biosynthesis. These genes encode the cellulose synthetase catalytic subunit (AthA) and a putative pectate lyase A11 (Table 6). Finally, five genes with unconfirmed C-compound/carbohydrate metabolism functions that fell into 111 InterAct class were grouped as 'other' (Table 6). In addition to C-compound/carbohydrate metabolism, other secondary funcats were identified that were over-represented in the InterAct class 111 (see Additional data file 1). These genes corresponded to the secondary funcats for cell-wall biosynthesis genes (five genes), and genes involved in electron transport and membrane associated energy conservation (three genes) (Table 6). By using InterAct class and MIPS to group genes according to similar expression profiles and biological function, we were able to find over-represented promoter motifs which may identify previously unknown shared *cis* elements involved in gene regulation by both carbon and light, as described below.

Putative co-regulated genes were used to identify shared *cis*-regulatory motifs responding to light and carbon

The InterAct class 111 was selected for *cis*-analysis because this class of genes showed equal responses to carbon, light, and carbon plus light. This suggests there may be a single *cis* element that responds to either carbon or light (among other possibilities, see Discussion). Three secondary funcats showed over-representation of genes in the InterAct class 111. These corresponded to metabolism: C-compound/carbohydrate metabolism (10 genes); control of cellular organization: cell wall (five genes); and energy: electron transport (three genes) (Table 6).

Of the 10 genes in InterAct class 111 that are found in C-metabolism, three genes were initially selected for *cis*-analysis because they are involved in a specific biochemical process, starch metabolism (At2g36390, At5g03650, and At5g48300) (see Table 6 and Figure 2, step 1). These three starch-metabolism genes were used as the basis for the discovery of carbon- and light-responsive *cis* elements because they satisfied two criteria that suggested they may be 'co-regulated'. First, they are functionally related (metabolism: C-compound/carbohydrate metabolism: starch metabolism); and second, they have similar expression patterns under varying carbon and light treatments (InterAct class 111).

The program AlignAce [30,31] was used to find elements that were over-represented in the proximal 1,000 base pairs (bp) (-1 to -1,000) of the promoters of these three genes (Figure 2, step 2). Using this method, 59 degenerate motifs were found to be over-represented among the promoters of these InterAct class 111 genes encoding starch-metabolism enzymes. The promoters of all genes in the *A. thaliana* genome were retrieved using the program RSA Tools [32,33] (Figure 3, step 3). Finally, for each motif, we determined whether the genes containing any given motif were significantly over-represented ($p < 0.05$) in all of the genes of a particular InterAct class (that is, InterAct class 111) (Figure 3, step 4). As a negative control, these motifs were also analyzed to determine

Table 6

Genes over-represented in primary and secondary metabolism funcats in InterAct class 111, model 3 where equal effects of carbon and light are observed

Primary funcat	Secondary funcat	Genes in InterAct Class 111 where over-representation of metabolism primary and secondary funcat are observed		InterAct class		
				C	C+L	L
Metabolism	C-compound/ carbohydrate metabolism	**Starch biosynthesis**				
		Glucose-1-phosphate adenylyltransferase	At5g48300			
		*Starch branching enzyme II	At2g36390			
		1,4-alpha-glucan branching enzyme protein isoform SBE2	At5g03650			
		Cell wall				
		Cellulose synthetase catalytic subunit (AthA)	At4g39350	1	1	1
		Unknown protein (putative pectate lyase A11)	At1g04680			
		Other				
		Glucose phosphomutase	At1g23190			
		Ribose-5-phosphate isomerase	At2g01290			
		Enoyl (ACP) reductase [NAD]	At2g05990			
		Lactate dehydrogenase (LDH1)	At4g17260			
		Unknown protein	At1g04680			
Control of cellular organization	Cell wall	Spermine synthase and spermine	At1g23820			
		Expansin-like protein	At1g69530	1	1	1
		Cellulose synthase catalytic subunit (Ath A)	At4g39350			
		Putative cinnamoyl CoA reductase	At2g33590			
		Unknown protein	At1g04680			
Energy	Electron transport and membrane associated energy conservation	Nitrate reductase [NADH/NADPH]	At1g77760	1	1	1
		Unknown protein	At1g68790			
		Putative protein	At1g64900			

*This gene, At2g36390, had two probe sets on the chip that were both assigned an InterAct class and a funcat. One probe was classified in InterAct class 111 and the other in InterAct class 122.

whether they were also over-represented in all genes in InterAct class 000 (these are genes that are not regulated by carbon, light, or carbon plus light). Any motifs that were over-represented in genes in InterAct class 111 but not in genes in InterAct class 000 were identified as 'significant' motifs (Figure 3, step 4).

Thus, in this analysis the promoters of genes that shared both a similar expression profile and were involved in similar biochemical processes were used to identify putative shared *cis* elements. Using genome-scale data, those motifs that also shared the carbon and light regulation on a genome-wide level with similarly regulated genes (same InterAct class) were selected for further study. Using this analysis, eight putative *cis* elements (of the 59 motifs originally found to be over-represented in the promoters of the three starch-metabolism genes) were identified because they were statistically over-represented in all genes belonging to InterAct class 111 when the whole chip data were analyzed (Table 7).

This analysis was also carried out using the five genes (At1g04680, At1g23820, At1g69530, At2g33590, At4g39350) in the funcat 'Control of cellular organization: cell wall', that was also found to be over-represented in InterAct class 111 as compared to the general population (Table 6). Forty-five motifs were found to be over-

Table 7

LCR *cis*-motifs are proposed to be responsive to light or carbon

Basis for motif discovery		Motifs discovered				Over represented in InterAct						
InterAct	Funcat	Name	LCR	*p*-value*	Sequence†	111	110	011	121	000	Example of element	Validated/ proposed function (PlantCARE database)
111	**Metabolism**	Motif55§	N/A	0.0001	GGHNGGNDCR	**Y**	N	**Y**	N	N	GCCCG-motif	N/A
	C-metabolism										RE1	Repressing element
											Unnamed 1	TEF 1 factor binding site
											Unnamed 17	N/A
	Starch biosynthesis	**Motif45**	**LCR1**	0.0013	**SDHTHGTGDKTG**	**Y**	N	N	N	N	None	N/A
		Motif23	**LCR2**	0.0020	**YGGGTTC**	**Y**	N	N	N	N	GCCCG-motif	N/A
	(3 genes)	**Motif58**	**LCR3**	0.0020	**ANKNGRNNGSNHC**	**Y**	N	N	N	N	Chs-unit 1	Light response
		Motif5§	N/A	0.0066	GNKNCGGWG	**Y**	**Y**	N	N	N	PE3	Light response
											RE1	Light response
											RbcS-CMA7c	Light response
		Motif2	**LCR4**	0.0161	**CAAAAWT**	**Y**	N	N	N	N	DREP-module‡	Light response
		Motif10	**LCR5**	0.0281	**GNKNYGGWG**	**Y**	N	N	N	N	PE3	Light response
											RE1	Light response
											ACE‡	Light response
											RbcS-CMA7c	Light response
		Motif51	**LCR6**	0.0287	**ATTKTGVA**	**Y**	N	N	N	N	HSE	Heat stress response
	Cellular organization	**Motif12**	**LCR7**	0.0006	**ANNCCANANHHMNNA**	**Y**	N	N	N	N	H-box module‡	Light response
		Motif13	**LCR8**	0.0102	**GDGDTKGTG**	**Y**	N	N	N	N	GT-1 motif‡	Light response
	Cell wall										ACE‡	Light response
											RbcS-CMA7c	Light response
	(5 genes)	**Motif20**	**LCR9**	0.0135	**GGTCNBNGWG**	**Y**	N	N	N	N	G-box‡	Light response
		Motif37	**LCR10**	0.0179	**GNGRNNGNNNNWGNNY**	**Y**	N	N	N	N	Endosperm-box‡	Endosperm regulation
		Motif6	**LCR11**	0.0207	**GHTTNGAWTBDG**	**Y**	N	N	N	N	None	N/A
		Motif24	**LCR12**	0.0239	**GDGDTDGTG**	**Y**	N	N	N	N	Endosperm-box‡	Endosperm regulation
		Motif16	**LCR13**	0.0256	**GARANNNARA**	**Y**	N	N	N	N	CAG-motif	Light response
											Prolamin-box‡	Endosperm regulation
											MPE	N/A
		Motif11	**LCR14**	0.0427	**ANVNAANAAGHNG**	**Y**	N	N	N	N	MPE	N/A

Table 7 *(Continued)*

LCR *cis*-motifs are proposed to be responsive to light or carbon										
Motif21	**LCR15**	0.0427	**ARAGAGA**	**Y**	N	N	N	N	GAG-motif	Light response
									3-AF1 binding site‡	Light response
Motif36	**LCR16**	0.0455	**GGTYNNNGWG**	**Y**	N	N	N	N	O2-site‡	Zein regulation

*p-value designates the significance level of genes with a motif being over-represented among all InterAct 111 genes used in our analysis. For each motif, p-values were calculated for genes containing: any copy number, less than two copies, more than one copy, more than three copies, more than five copies, and more than eight copies. The p-value with the greatest significance is shown here. †Sequences used to query the PlantCARE database are underlined. ‡Elements that have known binding factors. §Motifs originally found that did not pass the further statistical tests that they were not overrepresented in additional InterAct classes (see text for more details).

represented in the promoters of these five cell-wall genes that showed similar expression profiles. Genes with 10 (out of 45) of these motifs were found to be statistically over-represented in all genes that belonged to InterAct class 111 (but not in genes that belonged to InterAct class 000) when whole chip data were analyzed (Table 7), which suggests they are putative *cis* elements.

The expression profiles of the eight genes (Table 6) that fell into InterAct class 111 and that were used for *cis*-analysis were checked using quantitative polymerase chain reaction (PCR) (Additional data file 2). The absolute expression levels were translated into fold changes and were classified according to the InterAct Class classification system. Using quantitative PCR data, two (At5g48300 and At2g36390) out of three genes involved in starch metabolism fell into InterAct class 111 (Additional data file 2). The third gene, At5g03650, did not fall into any InterAct class because a criterion necessary to place a gene in a class was not met, which in this case was overlapping standard deviations. Similarly, using quantitative PCR to measure gene regulation, three (At2g33590, At4g39350, At1g23820) out of five genes that fell into the cell-wall biosynthesis funcat were placed into InterAct class 111. The other two cell-wall biosynthesis genes, At1g69530 and At1g04680, fell into InterAct class 221, a class similar to InterAct class 111 (Additional data file 2). Overall, the expression profiles of five out of the eight genes used for *cis*-analysis were verified using quantitative PCR. Upon closer examination of the fold changes and standard deviations that were calculated using quantitative PCR versus microarray data, quantitative PCR was found to follow the same trends as the microarray data. Differences are likely to be due to the small standard deviations for the expression levels of these genes in -C+L as determined through quantitative PCR analysis (Additional data file 2).

Motifs identified in genes regulated by light or carbon correspond to known light-response elements

The eight putative *cis* elements identified from the starch-metabolism genes that both share a similar expression profile and are found to be statistically over-represented in all InterAct class 111 genes (Table 7) were compared with known *cis*-regulatory elements found in the PlantCARE database [34,35] (Figure 3, step 5). All but one of these putative *cis* elements that were identified in genes equally regulated by carbon and light were found to be similar to validated light-response *cis* elements in PlantCARE. The exception was motif45/LCR1, which did not match any known elements in the database (Table 7). One putative *cis* element, motif2/LCR4, was found to have an exact match to sequences in the DREP-module, an element shown to be involved in light responses as well as other processes [36-38].

Ten putative *cis* elements were identified from the cell-wall genes in InterAct class 111 that responded equally to light and carbon (Table 7). Three of these putative *cis* elements were found to be similar to validated light-response elements in PlantCARE (Table 7). It is significant that some elements in the PlantCARE database (ACE and RbcS-CMA7c), validated as involved in light responses [38,39], were found to match putative *cis* elements that we identified using the promoters of both starch metabolism and cell-wall genes in the 111 InterAct class. This indicates that the motifs we identified using the protocol outlined in Figure 3 may correspond to light- or carbon-responsive *cis* elements of a more global nature, which are not limited to the regulation of one specific process such as starch metabolism or cell-wall synthesis.

Candidate *cis* elements responsive to either light or carbon are over-represented in InterAct class 111 but not 110

We identified a total of 18 putative *cis* elements (eight identified from starch-metabolism genes and 10 from cell-wall genes) that were over-represented in genes in the InterAct class 111 (Table 7). Genes in InterAct class 111 are induced equally by carbon, light, and carbon plus light. However, what is perceived as a direct response to light could actually be an indirect response through carbon. That is, genes in InterAct class 111 could be induced either by light or carbon, or by carbon alone. We next sought to further classify the putative

cis elements we identified according to one of these two models, as either a carbon-only response element, or a carbon-or-light response element, that is, a single *cis* element responsive to either carbon or light.

To investigate this question, we reanalyzed the putative *cis* elements that were identified using the promoters of the starch-metabolism and cell-wall genes in the 111 InterAct class. We now asked whether any of these putative *cis* elements were also over-represented in InterAct class 110 (carbon-only, model 1), a class that responds to carbon and carbon-plus-light, but not to light alone. Genes in InterAct class 110 are proposed to be regulated by carbon only, and not by light. The rationale for this analysis was that putative *cis* elements over-represented in InterAct class 111, but not 110, are candidate light-or-carbon responsive (LCR) *cis* elements. Motifs over-represented in both InterAct classes 111 and 110 are candidates for putative *cis* elements that are responsive to carbon only.

Our analyses found that seven of the putative *cis* elements identified using the starch-metabolism genes were over-represented in InterAct class 111 genes, but were not over-represented in InterAct class 110 genes (Table 7). (One motif, motif5, was over-represented in genes in both InterAct class 111 and 110.) We further eliminated an additional motif (motif55) by testing for over-representation in genes in InterAct class 011 (Table 7; see also Discussion). The remaining six putative *cis* elements are candidate LCR *cis* elements that confer responsiveness to either light or carbon. When these six putative LCR *cis* elements were compared to known elements in the PlantCARE database, four were similar to elements that had already been identified as involved in light responsiveness. Motif2/LCR4, one of the LCR *cis* elements identified from the starch-metabolism genes, corresponded exactly to sequences in the DREP module, a known element in the PlantCARE database (Table 7) [34]. This *cis* element has been experimentally shown to be involved in light regulation and other processes [36-38]. None of the 10 putative *cis* elements that we identified using the cell wall genes was significantly over-represented in genes in either InterAct class 110 or 011, and all 10 motifs are classified as LCR elements (Table 7). We identified 16 LCR elements from the analysis of starch-metabolism and cell-wall genes in InterAct class 111 and most corresponded to known light-response elements including GT-1-binding sites, G-box, H-box and RE1 (Table 7).

Thus, candidates for LCR *cis* elements that are responsible for responses to either carbon or light are proposed by the use of a combination of InterAct class analysis, MIPS analysis and statistical methods as outlined in Figure 3. Many of these candidate LCR *cis* elements have already been shown to be involved in light regulation. Our analyses suggest that these LCR *cis* elements may also be able to confer carbon responsiveness. Experiments are under way in our lab to test these hypotheses.

Discussion

In this study we undertook a genome-wide investigation of the interactions between light and carbon signaling. Here we identify and classify not only those individual genes but also specific biological processes that are subject to a high degree of regulation by carbon and/or light. With this information, genes with similar expression profiles that also share a similar biological function are used for *cis* element discovery. Using this approach, putative *cis*-regulatory elements deemed to be potential downstream targets of carbon or light signaling are identified. This study represents the first genome-wide analysis of carbon-regulated genes, as well as the first genome-wide study of both carbon- and light-regulated genes in *A. thaliana*.

A comparative analysis of Affymetrix *A. thaliana* partial genome DNA chips hybridized with RNA from plants treated under four different conditions was carried out. To aid in the analysis of microarray expression data, a classification system, termed InterAct Class [27] was used to classify genes on the basis of their relative expression level in each treatment. Three categories were identified: one unaffected by carbon and/or light (InterAct class 000); a second primarily regulated by carbon only or light only; and a third regulated by both carbon and light. Genes in the last category are further grouped into one of three models distinguishing possible types of interactions between carbon and light. Thirty-nine validated InterAct classes, each containing at least three genes, are placed under one of these three models that hypothesize how carbon and/or light signaling may influence one another to ultimately affect gene expression. All of the Affymetrix IDs, their gene descriptions and the InterAct class in which they are classified can be found in Additional data file 3.

The InterAct Class classification system accurately predicts expression profiles

The expression of many genes has been described as being either carbon or light responsive. A subset of these genes are regulated by both carbon and light [5-7] where there are no obvious mechanisms of interaction between light and carbon signaling pathways controlling gene regulation. In this study, using the InterAct Class system to categorize gene-expression profiles, 201 genes (10%) are shown to be regulated by carbon only and 77 (3.8%) by light alone (Table 2). The percent gene regulation determined here by either of these signals may seem low based on what is known about light regulation of gene expression [1-4,17-19]. However, it is possible that these genes originally shown to be light regulated are in fact also affected by carbon. For example, a previous microarray analysis determined that approximately 25% of the genes in *A. thaliana* are regulated by light [18]. However, all analyses were carried out with plants given a constant carbon source (1% sucrose) [18], which could potentially complicate the analysis of light-regulated genes.

Overall, the InterAct classification system accurately predicts the regulation of carbon- and/or light-regulated genes as determined though the comparison of known expression profiles of a subset of genes with their InterAct classification in this study. For example, a number of genes involved in nitrogen assimilation, a process partly controlled by light and carbon, are also regulated by light and carbon. Glutamine synthetase, encoded by *GLN2*, and ferredoxin-glutamate synthase (Fd-GOGAT), encoded by *GLU1*, are two enzymes involved in the assimilation of ammonia into glutamine and glutamate whose genes are induced by light and carbon [13,39,40]. The gene *NR1*, encoding NADH/NADPH-dependent nitrate reductase, is also positively regulated by light and carbon [5,40]. In contrast, light and carbon repress the genes *ASN1* and *GDH1*, which encode asparagine synthetase and glutamate dehydrogenase, respectively [12,40,41]. In this study, the InterAct classification of *GLU1* (InterAct class 221, At5g04140), *NR1* (InterAct class 111, At1g77760) and *ASN1* (InterAct class -2-3-1, At3g47340) remains consistent with what is already known about the regulation of these genes (see Additional data file 3 for gene classifications). Other genes involved in nitrogen assimilation that are regulated by light and carbon, such as *GLN2* and *ASN2*, did not meet the criteria to be placed into an InterAct class or were not present on the AG array.

Some genes encoding proteins involved in or relating to photosynthesis are strongly induced by light yet repressed by sucrose [5-7]. In this study, we find that three photosynthetic structural genes - *PSBO* (33 kDa polypeptide of oxygen evolving complex, At5g66570), *OEC33* (putative protein 1 of photosystem II oxygen evolving complex, At3g50820) and *PSBY* (putative photosystem II core complex protein, At1g67740) - follow this mode of regulation by carbon and light and these genes fall into InterAct classes -101, -112 and -112, respectively. Similarly, the expression of a number of genes for chlorophyll *a*/*b* binding proteins remains consistent with their known regulation by carbon and light. These genes include *LHCA3.1* (PSI type III chlorophyll *a*/*b* binding protein, At1g61520), *CAB4* (light harvesting chlorophyll *a*/*b*, At3g47470), *CAB* (chlorophyll *a*/*b* binding protein, At3g54890), *LHCB1b2* (photosystem II type I chlorophyll *a*/*b* binding protein, At2g34420) and chlorophyll *a*/*b* binding protein CP29 (At5g01530). These genes fall into the InterAct classes -100, -101, -111, -111, -2-10, respectively (see Additional data file 3 for gene classifications).

However, as InterAct Class is a quantitative method used to classify genes on the basis of a semi-quantitative analysis method (microarrays) [14], a subset of genes may fall into more than one InterAct class or may be incorrectly classified. For example, although the gene-expression profiles of the majority of genes in InterAct class 111 used for *cis*-analysis were verified using quantitative PCR, there were two genes that fell into InterAct class 221, a class similar to yet distinct from 111. However, as quantitative PCR is a quantitative method to detect changes in mRNA levels and microarrays are semi-quantitative, differences in expression levels detected using the two methods are to be expected.

Models for mechanisms of light and carbon signaling

In our study we find 1,247 genes (62.4%) that fall into InterAct classes in which light and carbon both have a significant role in regulating gene expression. To understand the possible mechanisms by which light and carbon interact to regulate gene expression, we propose three models of this interaction. Model 1 describes the interaction between light and carbon as independent. The direction of regulation by carbon and light for model 1 can be further described as being repressive or inductive (see Table 2). In addition, light and carbon may act to regulate the expression of genes in an opposite manner (for example, light induces and carbon represses). In this study, we find that 407 genes (20%) fit this model of regulation where light and carbon serve as two independent signals (see Additional data file 4).

Model 2 predicts that both light and carbon are required for gene regulation, and either signal alone is ineffective. Our study finds 32 genes (1.6%) whose expression profiles fit this model (see Additional data file 4).

Model 3 predicts that the interaction between light and carbon contains both a dependent and independent component in regulating gene expression. The interactions between these two signals, or the effect of one of these signals on the other, that ultimately affect gene expression can be described in various ways: first, either carbon or light is dominant; second, either signal has an equal effect, indicating a certain threshold has been attained; third, carbon or light act to suppress or enhance the opposite signal (Table 2). The expression of 808 genes (40%) is influenced by interactions between light and carbon such as those described by model 3 (see Additional data file 4). As this study was carried out using the partial *A. thaliana* genome array (containing approximately 8,000 genes that represent 25% of the genome), it is likely that the number of genes in each InterAct class is an underestimate.

Correlating gene expression data with biological function

To identify specific biological processes that are affected by carbon and/or light signaling, genes that are both given a functional category classification by MIPS [28,29] and fall into an InterAct class are used to determine whether biological function can be correlated with specific patterns of gene regulation by carbon and light. Statistical analyses are used to determine whether genes sharing a biological function (same funcat) show significant over- or under-representation within an InterAct class.

In this study, we first focus on the funcats that are under-represented in InterAct class 000 because these particular funcats are over-represented in another InterAct class that responds to carbon and/or light regulation. Besides the

primary funcat metabolism, a number of other primary funcats such as protein fate, protein synthesis and cell rescue, defense and virulence are all under-represented in InterAct class 000 (Table 3). This indicates that these biological processes are also subject to more carbon and/or light regulation than expected. These findings remain consistent with what has been shown about the carbon regulation of genes involved in plant defense [5]. Processes such as protein fate and protein synthesis require energy, so it is to be expected that these biological processes are, in general, regulated by carbon and/or light. To our knowledge, however, this is the first study documenting that these biological processes are subject to carbon and/or light regulation.

It is interesting to note that the funcats transcription, cellular communication/signal transduction and cell cycle and DNA processing are all processes requiring a substantial amount of energy, yet they show over-representation within InterAct class 000. This suggests that, overall, these primary funcats or biological processes are less regulated by carbon and/or light than expected.

We focus our detailed analysis on the primary funcat metabolism because it is under-represented in InterAct class 000 (no regulation by carbon and/or light). This indicates that the metabolism funcat is over-represented in another InterAct class whose genes show regulation by carbon and/or light. Indeed, further analysis of the metabolism secondary funcats shows the secondary funcats C-compound/carbohydrate metabolism, cell wall, and electron transport and membrane-associated energy conservation are over-represented in InterAct class 111. Genes that fall into InterAct class 111 (model 3) are of particular interest because they show equal regulation by carbon or light or by both light and carbon. Thus, the genes in both InterAct class 111 and in the funcats C-compound/carbohydrate metabolism and cell wall are used as the basis for *cis* element discovery.

Cis element discovery in genes with similar expression profiles and biological function

We are interested in investigating not only regulation by carbon and light, but also the interactions between these two factors in regulation. Therefore, we focus on genes in InterAct class 111 (model 3, equal effect of carbon or light) for further *cis* element analysis. Although genes in InterAct class -1-1-1 are also in model 3 (equal effect), because only two genes share a similar expression profile and are within the same funcat (Table 5) they are not used for *cis* element analysis. Genes in InterAct class 111 are genes that are equally induced by carbon, light, and carbon plus light, and thus display a potentially interesting interaction of carbon and light. Here, either input alone is sufficient (but neither is absolutely necessary) to induce expression. However, having both inputs does not result in any greater level of induction, suggesting that these two factors are not acting in an independent or additive manner.

One model for the regulation of InterAct class 111 genes is that they are induced by carbon alone, and according to this model the perceived light induction is an indirect effect through carbon (photosynthate). Another model, however, is that these genes are induced by either carbon or light in such a manner that carbon and light signals are perceived independently and later converge at a downstream junction. In this model, carbon and light signals would ultimately converge upon shared transcription factors and *cis*-regulatory elements to mediate their similar regulation. These two models are distinguished by determining whether a motif that is over-represented among InterAct class 111 genes is also over-represented among InterAct class 110 genes (which are induced by carbon alone, or carbon-plus-light, but not by light alone). Motifs that are over-represented in InterAct class 111 genes, but not in class 110 genes, are candidates for LCR elements. An LCR element responsive to either light or carbon is not expected to be over-represented in InterAct class 110 genes, which are not induced by light alone, because genes with an LCR motif should also be responsive to light treatments. Conversely, an element responsive only to carbon may or may not be sensitive enough to respond to the endogenous levels of photosynthate produced in response to light exposure depending on additional factors (that is, depending on motif copy number in the promoter). Such an element thus might be expected to be over-represented both in InterAct class 111 (as a result of genes with a higher sensitivity to endogenous carbon, sufficient to respond to photosynthate) and in InterAct class 110 (as a result of genes with a lower sensitivity to endogenous carbon, insufficient to respond to photosynthate). A *cis* element responsive only to carbon might therefore be expected to be over-represented in both InterAct classes 111 and 110. In contrast to such putative carbon-only response elements, we are instead interested in motifs that are over-represented among InterAct class 111, but not InterAct class 110, genes. Motifs satisfying these criteria are characterized as candidate LCR motifs, putative *cis*-regulatory elements that are involved in a convergent light and carbon pathway. Such putative *cis*-regulatory elements would be multifunctional and are proposed to be responsive to either light or carbon. There is precedent for such multifunctional *cis*-regulatory elements. For example, sequences in the DREP and H-box modules are proposed to be involved in multiple regulatory pathways [34,42-47].

Another possible model is that genes in InterAct class 111 are regulated by distinct light and carbon *cis* elements (with their corresponding factors), but that factor binding to the *cis* elements is mutually exclusive and therefore competitive. According to this model, the factor for either the light element or the carbon element can bind at any given time, but both cannot bind simultaneously because of steric hindrance. Such genes might thus be responsive to light, carbon, or light plus carbon equally via two separate but competitive *cis*-regulatory elements (and their respective pathways). Such distinct light and carbon elements might be expected to be found

together and in close proximity, and therefore functionally dependent, in the promoters of genes in InterAct class 111. Conversely, they could be found independently, both physically and functionally, in the promoters of other genes that are regulated by either light or carbon alone (that is, genes in light only or carbon only categories), or by both factors in an independent, additive manner (that is genes corresponding to model 1). Consequently, such elements might be expected to be over-represented not only in InterAct class 111 genes, but also in either InterAct class 110 genes (if regulated by carbon only, as discussed above) or InterAct class 011 genes (if regulated by light only). Thus, we further tested the motifs that we identify in our analyses to determine if any motifs that are over-represented in InterAct class 111 genes are also over-represented in InterAct class 110 genes (as discussed above) or in InterAct class 011 genes (Table 7).

Of the eight motifs identified from the promoters of InterAct class 111 starch-metabolism genes that are over-represented in genes in InterAct class 111, one motif is found also to be over-represented in genes in InterAct class 110 (see Results) and a second motif is found also to be over-represented in genes in InterAct class 011 (Table 7). Therefore, six of the motifs identified from the promoters of starch-metabolism genes are proposed to be LCR *cis* elements (Table 7). When this analysis is repeated for the motifs identified from the promoters of cell-wall genes in InterAct class 111, none of the 10 motifs that are found to be over-represented in InterAct class 111 genes is found to be also over-represented in genes in either InterAct class 110 or 011; thus, all are proposed to be LCR *cis* elements (Table 7). We also test whether our identified motifs are over-represented in InterAct class 121 genes to further distinguish the LCR *cis* elements from monofunctional light elements or carbon elements which operate together in an independent, additive manner in the regulation of genes (that is, genes corresponding to model 1). None of the proposed LCR *cis* elements is found to be over-represented in InterAct class 121 genes (Table 7).

Many of the LCR elements we identify correspond to previously validated light-response elements (DREP, GT-1 motif, G-box, H-box, ACE, RE1, PE3, chs unit-1, RbcS-CMA7c) ([34] and references therein), some of whose binding factors have already been identified (see Table 7). The genes we use as the basis for motif identification are from InterAct class 111, which signifies that these genes are induced by light, carbon, or light plus carbon. Therefore, it is expected that any motifs we identify are involved in light or carbon induction. However, one of the motifs, motif10/LCR5, corresponds to the known element RE1, an element that has been shown to function in light repression [48,49]. However, there is precedent for *cis* elements that function in both activation and repression of gene expression. For example, the GT-1 element has been shown to be required for light induction in some contexts, and for light repression in others [50]. The putative LCR elements are found in genes placed in class 111, but are not involved in starch or cell-wall metabolism, suggesting that the analysis uncovered *cis* elements globally involved in light and carbon regulation despite being initiated with genes involved in particular processes.

Finally, the approach we present here is a general method for organizing and analyzing microarray data for datasets containing multiple input signals and for smaller datasets than are normally required for clustering analyses [23]. In addition, our work presents a way of investigating biological signaling interactions and identifying putative *cis*-regulatory elements *in silico*, through the analysis of genes that both share similar expression profiles and are involved in similar biological processes. To our knowledge, this is the first genome-wide analysis of carbon-regulated genes as well as the first study of both carbon- and light-regulated genes. This work should contribute to our ability to integrate the responses of genes to multiple input signals into models designed for testing.

Materials and methods

Plant growth and treatment for analysis

A. thaliana seeds of Columbia ecotype were surface-sterilized, plated on designated media and vernalized for 48 h at 8°C. Plants were grown semi-hydroponically under 16-h-light (70 μE/m²/sec) 8-h-dark cycles at a constant temperature of 23°C on basal Murashige and Skoog (MS) medium (Life Technologies, NY) supplemented with 2 mM KNO_3, 2 mM NH_4NO_3, and 30 mM sucrose [13]. Two-week-old seedlings were transferred to fresh MS media without nitrogen and carbon (sucrose) and dark-adapted for 48 h. To perform specific metabolic treatments, dark-adapted plants were transferred to fresh MS medium containing 0% or 1% (w/v) sucrose and either placed back into the dark or illuminated with white light for an additional 8 h (70 μE/m²/sec). Following light treatments, whole plants were harvested, immediately frozen in liquid nitrogen and stored at -80°C until further use.

RNA isolation and microarray analysis

RNA was isolated from whole plants using a phenol extraction protocol as previously described [51]. Double stranded cDNA was synthesized from 8 μg total RNA using a T7-Oligo(dT) promoter primer and reagents recommended by Affymetrix, (Santa Clara, CA). Biotin-labeled cRNA was synthesized using the Enzo BioArray HighYield RNA Transcript Labeling Kit (Enzo, NY). The concentration and quality of cRNA was estimated through an $A_{260/280}$ nm reading and running 1:40 of a sample on 1% (w/v) agarose gel. cRNA (15 μg) was used for hybridization (16 h at 42°C) to the partial *Arabidopsis* genome AG array (Affymetrix). Washing, staining, and scanning were performed as recommended by the Affymetrix instruction manual. Expression analysis was performed with the Affymetrix Microarray Suite software (version 5.0) set at default values with a target intensity set to 150. Two biological replicates for each treatment were carried out.

Assigning Affymetrix IDs to InterAct classes

Probes from the Affymetrix AG chips were assigned an InterAct class if they were stably expressed across all conditions. To be considered stably expressed, a probe must meet two criteria: first, the absolute call made by Microarray Suite 5.0 must be 'present' (P) for each replicate and each treatment, indicating that the observed expression values can be trusted; and second, in the comparison of the treatments to the background (-L-C), the biological replicates must give a similar change in expression. To determine if the change in expression was similar for biological replicates, the Affymetrix difference call was used. For this analysis, in order for a probe to be assigned an InterAct Class, the biological replicates had to result in the same difference call (calls of MI were counted as being I, induced, and MD was counted as being D, depressed). For example, a probe that was called present, P, in every replicate and condition and had the same difference call for the biological replicates for two of the conditions in the comparisons to the background (for example, +C+L was called I, induced, both times and -C+L was called NC, no change, both times), but was called I, induced, for one replicate when +C-L was compared to the background and NC in the other replicate would not have been assigned a class. Also, a probe that had the same difference call for each biological replicate for each treatment, but had an absolute call of absent (A) for one treatment, would not be assigned an InterAct class.

Verification of gene-expression profiles using quantitative PCR

RNA was isolated from whole plants using a phenol extraction protocol as previously described [48]. The RNA samples used for quantitative PCR are the same as those used for the microarray analysis in this study. Samples that were used for checking the expression profiles of genes At1g23820, At1g69530 and At1g04680 were precipitated with LiCl prior to cDNA synthesis. cDNA synthesis using 1.0 μg total RNA was carried out according to Invitrogen (catalog number 11146-024). Subsequent real-time quantitative PCR was carried out with a LightCycler (Roche Diagnostics, Mannheim, Germany). PCR amplification in a 20 μl reaction volume consisted of a master mixture containing DNA Taq polymerase, dNTP mixture and buffer (LightCycler DNA Master SYBR Green 1, Roche Diagnostics, catalog number 2158817), 4 mM $MgCl_2$, 0.9 μM of each primer and cDNA in a glass capillary tube. Primers spanned at least one intron for each gene analyzed and were designed using the LightCycler probe design software (Roche). The primers were synthesized at Invitrogen Life Technologies (Carlsbad, CA). The following primers were used for amplification: At5g48300 5' CATTGCTAGTATGGGT3' (forward primer), 5' CCTATGGGTACACTTC3' (reverse primer); At5g03650 5' ATAAATGCCGACGTAG 3' (forward primer), 5' GTAGGTAAACTCGGAC 3' (reverse primer); At2g36390 5' AGGAATAGCTTTGCAC 3' (forward primer), 5' GGTCGTTCGTCGTATA 3' (reverse primer); At4g39350 5' GTTCATATCCATAGCAGT 3' (forward primer), 5' CAAGCATTCCCTTGAG 3' (reverse primer); At1g04680 5' CATCTCTAACAACCACT 3' (forward primer), 5' GTAAGCACCGTTCTGA 3' (reverse primer); At1g69530 5' TACCCTTGGAGCAATG 3' (forward primer), 5' GTGTCCGTTTATCGTAA 3' (reverse primer); At2g33590 5' TCCCTAATCCTGAGGT 3' (forward primer), 5' CAGGCTACTAGCATTT 3' (reverse primer); At1g23820 5' TTCTATCCCTAACCCTAAG 3' (forward primer), 5' CCACTGGGGTATGTTG 3' (reverse primer) and At4g24550 (putative clathrin coat assembly protein), 5' AGTCTGTTCGTCTGGATAGC 3' (forward primer), 5' TCCAGCCTCTTCAATCAAGG 3' (reverse primer). Thermal cycling was performed as follows: initial denaturation at 95°C for 2 min, followed by 45 cycles of denaturation at 95°C for 0 sec, annealing at 55°C for 5 sec (*GLN2*) and extension at 72°C for 15 sec. Melting-curve analyses were carried out for all amplification reactions as follows: one cycle of an initial denaturation at 95°C for 0 sec, annealing at 50°C for 5 sec and another denaturation at 95°C for 0 sec with a slope of 0.1°C/sec. Standards were prepared with a 10-fold serial dilution (10^{-4} to 10 pg) of the PCR products and were run under the same PCR conditions used for the samples. The absolute amount of mRNA (in ng/ml) for At5g48300, At5g03650, At2g36390, At4g39350, At1g04680, At1g69530, At2g33590 and At1g23820 was corrected/normalized according to the amount of At4g24550, a putative clathrin coat assembly protein. Comparisons of treatments (+C-L, +C+L, -C+L) to their respective backgrounds (-C-L) were carried out. Fold changes in gene expression were subjected to InterAct class analysis and compared with those obtained using microarrays (Additional data file 2).

Determination of *P*-values

P-values were calculated as described in [52] and related pages, the only difference being that the exact binomial probability test was used when n was less than or equal to 170, not 179. For the funcat analysis, the number of genes assigned to the funcat being analyzed and any class was used as n; p was the number of genes assigned to the class or model being analyzed divided by the number of genes assigned a class and a funcat (1,898 genes); k was the number of genes in the funcat being analyzed assigned to the model or class being analyzed. The one tailed p-value was considered when the Poisson approximation of binomial probabilities was used. For the binomial-ratio and the exact binomial probability test, the p-value for k or more out of n was used when k was larger than the expected value ($n \times p$) and for k or less out of n when k was smaller than the expected value.

Analysis of *cis* elements

Genes used as the basis for *cis* element identification were selected based on InterAct class analysis and MIPS analysis. Genes that were in the same InterAct class (that is, InterAct class 111) and in the same secondary funcat (that is, C-metabolism or cell wall) were selected for analysis (minimum of

three genes). Promoter sequences for genes (-1 to -1,000 bp, unless there was overlap with an upstream open reading frame) were retrieved using the RSA Tools [32,33] 'Retrieve sequence' function. Motifs over-represented in the selected promoters were identified using AlignAce 3.0 [30,31] using the following parameters: number of columns to align = 7; number of sites to expect = 5, 6, or 10; fractional background GC content = 0.364. AlignAce 3.0 was run with three different values for 'number of sites to expect' - 5, 6 and 10 - and was run in triplicate for each value. Motifs identified in this way were collapsed into degenerate sequences. RSA Tools was used to retrieve the promoters of all genes in the genome [32,33]. We then determined whether genes containing a particular motif, in various copy numbers, were statistically over-represented in genes in a particular InterAct class, as compared to the general population. *P*-values were calculated as described above with *n* being the number of genes with a particular motif in its promoter (all genes or genes with a specified copy number); *p* being the number of genes in the InterAct class (that is, 111) divided by the total number of genes assigned a funcat and an InterAct class ('total' = 1,898 genes); and *k* being the number of genes whose promoters contain a particular motif (in a specified copy number) and that are in the particular InterAct class (that is, 111). Any significant *cis*-motifs identified by this process were compared to known validated elements in the PlantCARE database [34,35].

The microarray data for this article have been deposited in the ArrayExpress database, accession number E-MEXP-71.

Additional data files

The following additional data files are included with the online version of this article: a list of the number of genes and their *p*-value that are in each functional category that fall under a specific InterAct class (Additional data file 1); a table and graphs showing quantitative PCR verification of the expression profiles of the eight genes in InterAct class 111 that were used for *cis*-analysis (Additional data file 2); a list of Affymetrix IDs, their corresponding At gene number, gene description and their InterAct Class designation (Additional data file 3); and a list of the number of genes and their *p*-values in each functional category that fall under a particular model (shown in Table 2) (Additional data file 4).

Acknowledgements

This work was supported by the Department of Energy (grant number DEFG02-92-20071 to G.M.C.) and by the National Institutes of Health (grant number GM32877 to G.M.C.), National Research Service Award (grant number GM63350 to K.E.T. and GM65690 to P.M.P).

References

1. Fankhauser C, Chory J: **Light control of plant development.** *Annu Rev Cell Dev Biol* 1997, **13:**203-229.
2. Neff MM, Fankhauser C, Chory J: **Light: an indicator of time and place.** *Genes Dev* 2000, **14:**257-271.
3. Moller SG, Ingles PJ, Whitelam GC: **The cell biology of phytochrome signaling.** *New Physiol* 2002, **154:**553-590.
4. Nagy F, Schafer E: **Phytochromes control photomorphogenesis by differentially regulated, interacting signaling pathways in higher plants.** *Annu Rev Plant Biol* 2002, **53:**329-355.
5. Koch KE: **Carbohydrate-modulated gene expression in plants.** *Annu Rev Plant Physiol Plant Mol Biol* 1996, **47:**509-540.
6. Rolland F, Moore B, Sheen J: **Sugar sensing and signaling in plants.** *Plant Cell Supp* 2002, 14 Suppl:S185-S205.
7. Sheen J, Zhou L, Jang J-C: **Sugars as signaling molecules.** *Curr Opin Plant Biol* 1999, **2:**410-418.
8. Barnes SA, Nishizawa NK, Quaggio RB, Whitelam GC, Chua NH: **Far-red light blocks greening of *Arabidopsis* seedlings via a phytochrome A-mediated change in plastid development.** *Plant Cell* 1996, **8:**601-615.
9. Dijkwel PP, Huijser C, Weisbeek PJ, Chua NH, Smeekens S: **Sucrose control of phytochrome A signaling in *Arabidopsis*.** *Plant Cell* 1997, **9:**583-595.
10. Short T: **Over-expression of *Arabidopsis* phytochrome B inhibits phytochrome A function in the presence of sucrose.** *Plant Physiol* 1999, **119:**1497-1506.
11. Thum KE, Shasha DE, Lejay LV, Coruzzi GM: **Light and carbon signaling pathways. Modeling circuits of interactions.** *Plant Physiol* 2003, **132:**440-452.
12. Lam HM, Hsieh MH, Coruzzi GM: **Reciprocal regulation of distinct asparagine synthetase genes by light and metabolites in *Arabidopsis thaliana*.** *Plant J* 1998, **16:**345-353.
13. Oliveira IC, Coruzzi GM: **Carbon and amino acids reciprocally modulate the expression of glutamine synthetase in *Arabidopsis*.** *Plant Physiol* 1999, **121:**301-309.
14. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270:**467-470.
15. Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA: **Orchestrated transcription of a key pathway in *Arabidopsis* by the circadian clock.** *Science* 2000, **290:**2110-2113.
16. Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E: **Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*.** *Plant Cell* 2001, **13:**113-123.
17. Tepperman JM, Zhu T, Chang HS, Wang X, Quail PH: **Multiple transcription-factor genes are early targets of phytochrome A signaling.** *Proc Natl Acad Sci USA* 2001, **98:**9437-9442.
18. Ma L, Li J, Qu L, Hager J, Chen Z, Zhao H, Deng XW: **Light control of *Arabidopsis* development entails coordinated regulation of genome expression and cellular pathways.** *Plant Cell* 2001, **13:**2589-2607.
19. Ma L, Gao Y, Qu L, Chen Z, Li J, Zhao H, Deng XW: **Genomic evidence for COP1 as a repressor of light-regulated gene expression and development in *Arabidopsis*.** *Plant Cell* 2002, **14:**2383-2398.
20. Wang R, Guegler K, LaBrie ST, Crawford NM: **Genomic evidence of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate.** *Plant Cell* 2000, **12:**1491-1509.
21. Wang YH, Garvin DF, Kochian LV: **Rapid induction of regulatory and transporter genes in response to phosphorus, potassium, and iron deficiencies in tomato roots. Evidence for cross talk and root/rhizosphere-mediated signals.** *Plant Physiol* 2002, **130:**1361-1370.
22. Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Caminic P, Hayashizaki Y, Shinozaki K: **Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray.** *Plant Cell* 2001, **13:**61-72.
23. Bussemaker HJ, Li H, Siggia ED: **Regulatory element detection using correlation with expression.** *Nat Genet* 2001, **27:**167-171.
24. Keles S, Van Der Laan M, Eisen MB: **Identification of regulatory elements using a feature selection method.** *Bioinformatics* 2002, **18:**1167-1175.
25. Conlon EM, Liu XS, Lieb JD, Liu JS: **Integrating regulatory motif discovery and genome-wide expression analysis.** *Proc Natl Acad Sci USA* 2003, **100:**3339-3344.
26. Quackenbush J: **Computational analysis of microarray data.** *Nat Rev Genet* 2001, **2:**418-427.
27. **Details on microarray gene expression classification using InterAct Class** [http://www.cs.nyu.edu/~pathexp]

28. **MAtDB entry page** [http://mips.gsf.de/proj/thal/db/]
29. Schoof H, Zaccaria P, Gundlach H, Lemcke K, Rudd S, Kolesov G, Arnold R, Mewes HW, Mayer FX: **MIPS *Arabidopsis thaliana* database (MAtDB): an integrated biological knowledge resource based on the first complete plant genome.** *Nucleic Acids Res* 2002, **30:**91-93.
30. **AlignACE homepage** [http://atlas.med.harvard.edu]
31. Roth FP, Hughes JD, Estep PW, Church GA: **Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation.** *Nat Biotechnol* 1998, **16:**939-945.
32. **Regulatory sequence analysis tools** [http://rsat.ulb.ac.be/rsat/]
33. van Helden J: **Regulatory sequence analysis tools.** *Nucleic Acids Res* 2003, **31:**3593-3596.
34. **PlantCARE, a database of plant promoters and their *cis*-regulatory elements** [http://oberon.fvms.ugent.be:8080/PlantCARE/]
35. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S: **PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences.** *Nucleic Acids Res* 2002, **30:**325-327.
36. Bolle C, Kusnetsov VV, Herrmann RG, Oelmuller R: **The spinach AtpC and AtpD genes contain elements for light-regulated, plastid-dependent and organ-specific expression in the vicinity of the transcription start sites.** *Plant J* 1996, **9:**21-30.
37. Kusnetsov V, Bolle C, Lubberstedt T, Sopory S, Herrman RG, Oelmuller R: **Evidence that the plastid signal and light operate via the same *cis*-acting elements in the promoters of nuclear genes for plastid proteins.** *Mol Gen Genet* 1996, **252:**631-639.
38. Kusnetsov V, Landsgerger M, Meurer J, Oelmuller R: **The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids.** *J Biol Chem* 1999, **274:**36009-36014.
39. Coschigano KT, Melo-Oliveira R, Lim J, Coruzzi GM: ***Arabidopsis gls* mutants and distinct Fd-GOGAT genes: Implications for photorespiration and primary nitrogen assimilation.** *Plant Cell* 1998, **10:**741-752.
40. Lam HM, Coschigano K, Oliveira IC, Melo-Oliverira R, Coruzzi GM: **The molecular-genetics of nitrogen assimilation into amino acids in higher plants.** *Annu Rev Plant Physiol Plant Mol Biol* 1996, **47:**569-593.
41. Melo-Oliveira R, Oliveira IC, Coruzzi G: ***Arabidopsis* mutant analysis and gene regulation define a non-redundant role for glutamate dehydrogenase in nitrogen assimilation.** *Proc Natl Acad Sci USA* 1996, **93:**4718-4723.
42. Arguello-Astorga GR, Herrera-Estrella LR: **Ancestral multipartite units in light-responsive plant promoters have structural features correlating with specific phototransduction pathways.** *Plant Physiol* 1996, **112:**1151-1166.
43. Hartman U, Valentine WJ, Christie JM, Hays J, Jenkins GI, Weisshaar B: **Identification of UV/blue light response elements in the *Arabidopsis thaliana* chalcone synthase promoter using a homologous protoplast transient expression system.** *Plant Mol Biol* 1998, **36:**741-754.
44. Terzaghi WB, Cashmore AR: **Light-regulated transcription.** *Annu Rev Plant Physiol Plant Mol Biol* 1995, **46:**445-474.
45. Bezhani S, Sherameti I, Pfannschmidt T, Oelmuller R: **A repressor with similarities to prokaryotic and eukaryotic DNA helicases controls the assembly of the CAAT box binding complex at a photosynthesis gene promoter.** *J Biol Chem* 2001, **276:**23785-23789.
46. Yu LM, Lamb CJ, Dixon RA: **Purification and biochemical characterization of proteins which bind to the H-box *cis*-element implicated in transcriptional activation of plant defense genes.** *Plant J* 1993, **3:**805-816.
47. Loake GJ, Faktor O, Lamb CJ, Dixon RA: **Combination of H-box [CCTACC(N)7CT] and G-box (CACGTG) *cis* elements is necessary for feed-forward stimulation of a chalcone synthase promoter by the phenylpropanoid-pathway intermediate p-courmaric acid.** *Proc Natl Acad Sci USA* 1992, **89:**9230-9234.
48. Kuno N, Furuya M: **Phytochrome regulation of nuclear gene expression in plants.** *Semin Cell Dev Biol* 2000, **11:**485-493.
49. Bruce WB, Deng XW, Quail PH: **A negatively acting DNA sequence element mediates phytochrome-directed repression of phyA gene transcription.** *EMBO J* 1991, **10:**3015-3024.
50. Zhou DX: **Regulatory mechanism of plant gene transcription by GT-elements and GT-factors.** *Trends Plant Sci* 1999, **4:**210-214.
51. Jackson AO, Larkins BA: **Influence of ionic strength, pH and chelation of divalent metals on isolation of polyribosomes from tobacco leaves.** *Plant Physiol* 1976, **57:**5-10.
52. **Exact binomial probability calculator** [http://faculty.vassar.edu/lowry/binom_stats.html]