High throughput generation of promoter reporter (GFP) transgenic lines of low expressing genes in Arabidopsis and analysis of their expression patterns

Yong-Li Xiao, Julia C Redman, Erin L Monaghan, Jun Zhuang, Beverly A Underwood, William A Moskal, Wei Wang, Hank C Wu, Christopher D Town*

Abstract

Background: Although the complete genome sequence and annotation of Arabidopsis were released at the end of year 2000, it is still a great challenge to understand the function of each gene in the Arabidopsis genome. One way to understand the function of genes on a genome-wide scale is expression profiling by microarrays. However, the expression level of many genes in Arabidopsis genome cannot be detected by microarray experiments. In addition, there are many more novel genes that have been discovered by experiments or predicted by new gene prediction programs. Another way to understand the function of individual genes is to investigate their in vivo expression patterns by reporter constructs in transgenic plants which can provide basic information on the patterns of gene expression.

Results: A high throughput pipeline was developed to generate promoter-reporter (GFP) transgenic lines for Arabidopsis genes expressed at very low levels and to examine their expression patterns in vivo. The promoter region from a total of 627 non- or low-expressed genes in Arabidopsis based on Arabidopsis annotation release 5 were amplified and cloned into a Gateway vector. A total of 353 promoter-reporter (GFP) constructs were successfully transferred into Agrobacterium (GV3101) by triparental mating and subsequently used for Arabidopsis transformation. Kanamycin-resistant transgenic lines were obtained from 266 constructs and among them positive GFP expression was detected from 150 constructs. Of these 150 constructs, multiple transgenic lines exhibiting consistent expression patterns were obtained for 112 constructs. A total 81 different regions of expression were discovered during our screening of positive transgenic plants and assigned Plant Ontology (PO) codes.

Conclusions: Many of the genes tested for which expression data were lacking previously are indeed expressed in Arabidopsis during the developmental stages screened. More importantly, our study provides plant researchers with another resource of gene expression information in Arabidopsis. The results of this study are captured in a MySQL database and can be searched at http://www.jcvi.org/arabidopsis/qpcr/index.shtml. Transgenic seeds and constructs are also available for the research community.

Background

At the end of 2000, the first plant genome project, sequencing of the whole genome of Arabidopsis thaliana, was completed by a multinational collaborative effort [1]. Subsequently, the plant scientific community set the goal of understanding the function of each gene in the Arabidopsis genome, which is encapsulated in the National Science Foundation (NSF) 2010 program, and this challenge has been taken up by many Arabidopsis researchers [2]. Identification of each gene in the fully sequenced Arabidopsis genome and uncovering their function will provide crucial information for biologists to understand plant physiology, genetics, development and evolution. One way to understand the function of genes on a genome-wide scale is expression profiling by...
microarrays using formats that include complementary DNA [3], oligonucleotides [4,5] or amplicons [6]. Recently, whole genome tiling arrays have been developed and used to interrogate the gene structure and expression status of the entire Arabidopsis genome [7-9]. In addition, microarrays have been used to analyze genome features such as chromatin structure [10,11]; sites of DNA modifications [12,13]; and DNA-protein binding sites [14,15]. So far, the expression data from ATH1 chips covers a wide range of experimental treatments and conditions in the public domain http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl; http://arabidopsis.org/info/expression/ATGenExpress.jsp that collectively make a great contribution to understanding gene function in Arabidopsis. Another way to understand the function of individual genes is to investigate their in vivo expression patterns spatially, temporally, or conditionally by reporter construct methods that include 1) random integration of promoter-less reporter constructs into the genome and monitoring of their expression to identify genes, promoters, and enhancers [16]; 2) transformation of promoter-reporter fusions from a particular gene back into the organism and observing the reporter expression patterns [17-22].

Many studies on gene expression in plants have shown that the majority of elements necessary for and important for the regulation of expression lie immediately upstream of the transcriptional initiation site, usually within the first kilobase (17-21). Although there are reports showing that the gene regulatory signals may also be located in other regions (e.g. within introns or the 3′ UTR) [23-25], these are in a minority. Thus the traditional view that the majority of a plant promoter’s activity as being immediately upstream of the transcriptional initiation site is generally well supported and there is considerable evidence that the in vivo expression of reporter genes driven by such regions (1-2 kb upstream) does reflect the expression pattern of the native gene. It can thus provide preliminary but critical information on endogenous gene expression patterns and evidence of their biological or developmental functions. A transgenic system such as this can also reveal cell-type-specific patterns of gene expression (e.g. in trichomes, hydathodes or stomata) without a priori knowledge which would escape detection by microarray approaches.

In our current study, we focused on two groups of genes in Arabidopsis, 1) unannotated genes identified by our previous experiments [26] that were predicted by EuGene [27] and Twinscan [28], programs that incorporate comparative genomics information. 2) Annotated Arabidopsis genes whose expression was detected in less than 5% of ~1,400 ATH1 Affymetrix GeneChip experiments downloaded from The Arabidopsis Information Resource (TAIR) website at the start of this work. For both groups of genes, our knowledge of their function is extremely limited. In our previous study that focused on hypothetical gene structure in Arabidopsis, localized expression patterns of five exemplar genes were detected by cloning the promoter regions into green fluorescent protein (GFP) reporter constructs and transforming them into Arabidopsis [29]. For the present study, we developed a high throughput pipeline to create promoter-reporter (GFP) fusions of both gene groups, transformed them into Arabidopsis, and screened the expression of the reporter genes in positive transgenic plants at various developmental stages. Over 600 Arabidopsis genes were selected to make the promoter-reporter constructs and transgenic lines were obtained for just under half of these. The expression of promoter-reporter (GFP) constructs in transgenic plants was examined at 4 developmental stages: on the selection plate around 10 days after germination, at the rosette stage in soil, just before flowering, and at the flowering stage. GFP expression has been observed in constructs from 150 different promoters and is typically localized to a few tissues or cell types (e.g. hydathode, pedicel, socket cell, guard cell), consistent with the absence from or low abundance of transcripts from these genes in EST libraries. For the remaining target genes, no visible expression was detected, although PCR confirmed the presence of the transgene in all cases tested. All the expression patterns have been annotated according to plant ontology codes http://www.plantontology.org/ and stored in a MySQL database. To our knowledge, this is the first set of large-scale promoter-reporter expression data in Arabidopsis focusing on novel genes and genes with limited expression data, and thus should be a valuable resource for the plant research community. All our data are publicly available through the project website http://www.jcvi.org/arabidopsis/qpcr/index.shtml; transgenic seeds, and constructs are also available for research community through the Arabidopsis Biological Resource Center (ABRC).

Results
Selection of target genes for promoter-reporter analysis
The selection of target genes was based on the Arabidopsis annotation release 5 that was available when the project started and includes 1) previously identified unannotated genes; 2) genes represented on the Affymetrix ATH1 GeneChip showing no expression; 3) genes represented on the Affymetrix ATH1 GeneChip showing very limited expression. Un-annotated genes are genes in the intergenic regions whose expression was detected by RACE in our previous experiments and those predicted in intergenic regions by EuGene [27] and/or Twinscan [28]. Genes on the Affymetrix ATH1 GeneChip showing
no or limited expression were identified from the Agilent Metrix MAS calls of 1381 ATH1 arrays downloaded from TAIR, combined with massively parallel signature sequencing (MPSS) data available at that time [30] (for details see MATERIAL AND METHODS). The resulting set of genes (identifiers and targeted promoter sequences) is presented in additional file 1: Table S1. Many of these were not annotated at the time of this work. However, all except 27 are now present in the TAIR9 annotation (June 2009) (additional file 2: Table S2). Five of these 27 still unannotated genes showed GFP expression in transgenic plants.

Development of a high throughput pipeline
To handle the relatively large number of genes to be studied, we modified our original protocols [29,31] to produce a robust high throughput pipeline that included batch primer design, high throughput cloning and transformation, and a project-tracking LIMS implemented in a MySQL database (Figure 1). In this study, the Gateway™ Gene Cloning strategy [32] was used to make promoter-reporter constructs. PCR amplification of promoters and Gateway cloning were performed in 96-well plate format. Two colonies from each Gateway BP cloning reaction (from each promoter) were picked and sequenced to confirm target identity. DNA isolated robotically from the sequence-confirmed clones was used in Gateway LR reactions. The triparental mating method [33] was used to transfer promoter-reporter constructs from E. coli to Agrobacterium (GV3101). Bypassing the step of destination clone DNA isolation from E. coli made this step economical and large scale. In addition, we cultured only 50 ml of Agrobacterium for plant transformation and did the floral dip in 50 ml Falcon tubes. As shown in Figure 2, a total of 627 candidate genes were put through the construction and transformation pipeline. Gateway entry clones were obtained from 469 genes and sequence confirmed. Of these, 442 were successfully transferred into the destination vector, pYXT2 containing the GFP reporter gene [34]. Subsequently, Agrobacterium clones from 353 genes were obtained by tri-parental mating, and then transformed into Arabidopsis by the floral dip method [35]. Three independent transformations (floral dips with separate plants) were performed for each construct and 3 seedlings from each kanamycin selection plate (whenever possible) were transferred into soil for maximum possible set of 9 transformed plants per construct. Positive transgenic plants from 266 constructs representing promoters from 266 genes were obtained and GFP expression patterns have been observed from 150 constructs in at least one of the 4 developmental stages examined. This newly developed cloning and transformation pipeline greatly improved the throughput and lowered the effort compared with our previous work.

Additionally, we have developed a project specific MySQL database to record and track the large amount of data generated by this project, including target genes, promoter selection, primer design, different steps of clone tracking, transgenic plant tracking, and GFP image capture. All the GFP images were annotated according to Plant Ontology (PO) codes http://www.plantontology.org/index.html[36]. This MySQL database supports our publicly accessible website http://www.jcvi.org/arabidopsis/qpcr/index.shtml which users can search for GFP images by tissue type, locus name, or PO code.

Expression patterns of promoter-reporter (GFP) constructs
In total, 1,885 positive transgenic lines from 266 promoter-reporter constructs were obtained. However, only 1,457 transgenic plants from 239 promoters showed GFP expression. Of these 150 promoters, consistent GFP expression patterns were observed in two or more transgenic plants for 112 promoter constructs. For 19 promoter constructs, only one transgenic line was recovered and showed GFP expression. For 14 promoter constructs, more than one transgenic line was obtained, but only one line expressed GFP. In addition, plants from 5 promoters had inconsistent GFP expression patterns. These results are summarized in Figure 2. Among the 112 promoter constructs that displayed the same GFP expression pattern from more than one transgenic plant, promoters from 79 genes showed the same GFP expression patterns in transgenic plants obtained from different floral dip events, providing high confidence of the location of expression of the transgene. GFP expressing lines for 27 different constructs were randomly chosen for promoter re-amplification using flanking primers and the PCR products were sequenced. All were confirmed to be from the intended promoters as shown in additional file 2: Table S2. In addition, leaf PCR [37] using GFP primers was performed on 423 transgenic lines, from 83 promoters, that did not show GFP expression. All had positive amplification indicating the presence of the GFP construct in these lines. In total, we recorded 2,287 GFP expression images and described their expression patterns with Plant Ontology (PO) codes. A total of 3,371 PO codes were assigned, as more than one PO code referencing different parts of the plant could be assigned to one image, and collectively these represented eighty-one different expression patterns (Table 1).
chr1.16.7 promoter construct shows expression in vascular tissue from leaf (PO:0000036, Figure 3A), petal (PO: 0000054) and sepal (PO:0004723, Figure 3B). Examples of multiple PO codes/expression patterns from one construct include the novel_chr1_5915494 (a novel gene discovered in our previous project) which is expressed in root (PO:0009005, Figure 3C) and seed (PO:0009010, Figure 3D). Some expression patterns which we detected are very specific. The promoter-reporter construct from AT1G64820 is expressed throughout the root but excluding the root tip (Figure 3E), while the construct from AT4G13985 was expressed only in the root tip (Figure 3F). Many instances of GFP expression were detected in floral organs, such as petal expression from promoter-report construct of AT2G17845 (Figure 3G) and carpel...
expression from that of AT2G40250 (Figure 3H). A listing of all genes assigned to each PO code is provided in additional file 3: Table S3

Discussion
The aim of our study was to examine the in vivo expression of many Arabidopsis genes of unknown function that show little or no expression based on EST, MPSS, or microarray data using reporter (GFP) genes driven by their native promoter and thus learn something about their potential function. In this study, we selected 627 Arabidopsis genes with unknown function, based on TIGR5 genome annotation, including unannotated genes located in the intergenic regions revealed in our previous RACE experiments [26], intergenic genes predicted by EuGene [27] and/or Twinscan

Figure 2 Outcome of generation of promoter reporter (GFP) transgenic lines. The figure shows the numbers of genes for which primers were designed, the success rate of high throughput cloning, transformation and the outcome of transgenic lines.
Table 1 Number of expression patterns obtained categorized by PO codes

| PO_code   | PO_code_name                        | Number of expressed promoters |
|-----------|-------------------------------------|------------------------------|
| PO:0000036| Leaf vascular system                | 38                           |
| PO:00009005| root                                 | 37                           |
| PO:0005660| hydathode                            | 35                           |
| PO:0000013| cauline leaf                         | 32                           |
| PO:0003011| root vascular system                 | 32                           |
| PO:0000056| floral bud                           | 26                           |
| PO:0006504| leaf trichome                        | 25                           |
| PO:0009052| pedicel                              | 24                           |
| PO:0009032| petal                                | 22                           |
| PO:0006502| flower abscission zone               | 21                           |
| PO:0006325| inflorescence node                   | 20                           |
| PO:0000282| trichome                             | 20                           |
| PO:0009031| sepal                                | 19                           |
| PO:0000054| petal vascular system                | 17                           |
| PO:0009001| fruit                                | 14                           |
| PO:002139| midvein                              | 14                           |
| PO:0000293| guard cell                           | 12                           |
| PO:0009049| inflorescence                        | 12                           |
| PO:0009047| stem                                 | 12                           |
| PO:0006056| cotyledon epidermis                  | 11                           |
| PO:0006016| leaf epidermis                       | 11                           |
| PO:0000112| stem epidermis                       | 11                           |
| PO:0000115| socket cell                          | 10                           |
| PO:0000146| abscission zone                      | 9                            |
| PO:0004707| fruit dehiscence zone                | 9                            |
| PO:0004810| seed                                 | 9                            |
| PO:0004723| sepal vascular system                | 9                            |
| PO:00009067| filament                             | 8                            |
| PO:0002127| primary root                         | 8                            |
| PO:0006040| sepal epidermis                      | 8                            |
| PO:0000039| stem vascular system                 | 8                            |
| PO:0000034| vascular system                      | 8                            |
| PO:0006501| leaf abscission zone                 | 7                            |
| PO:0002128| leaf margin                          | 7                            |
| PO:0000332| pavement cell                        | 7                            |
| PO:0005003| stem trichome                        | 7                            |
| PO:0000035| cotyledon vascular system            | 6                            |
| PO:00005679| epidermis                            | 6                            |
| PO:0009025| leaf                                 | 6                            |
| PO:0006036| root epidermis                       | 6                            |
| PO:0009030| carpel                               | 5                            |
| PO:00000225| root tip                             | 5                            |
| PO:0000915| vascular tissue                      | 5                            |
| PO:0020030| cotyledon                            | 4                            |

Note that a single line may be characterized by more than one PO code.

Table 1 Number of expression patterns obtained categorized by PO codes (Continued)

| PO_code   | PO_code_name                        | Number of expressed promoters |
|-----------|-------------------------------------|------------------------------|
| PO:0004536| fruit pedicel                        | 4                            |
| PO:0009072| ovary                               | 4                            |
| PO:0000052| petiole vascular system             | 4                            |
| PO:0002566| root hair                           | 4                            |
| PO:0009046| flower                              | 3                            |
| PO:0004724| hypocotyl-root junction             | 3                            |
| PO:0009081| inflorescence branch                | 3                            |
| PO:0005028| inflorescence vascular system       | 3                            |
| PO:0005645| leaf mesophyll                      | 3                            |
| PO:0009053| peduncle                            | 3                            |
| PO:0005021| sepal margin                        | 3                            |
| PO:0000033| valve                               | 3                            |
| PO:0005011| anther dehiscence zone              | 2                            |
| PO:0005008| fruit septum                        | 2                            |
| PO:0008003| fruit vascular system               | 2                            |
| PO:0021000| hypocotyl                           | 2                            |
| PO:0006019| leaf abaxial epidermis              | 2                            |
| PO:0004006| mesophyll cell                      | 2                            |
| PO:0000051| petiole epidermis                   | 2                            |
| PO:0002031| radicle                             | 2                            |
| PO:00020141| stem node                           | 2                            |
| PO:0004711| axillary inflorescence bud          | 1                            |
| PO:0005019| carpel vascular system              | 1                            |
| PO:0006338| embryonic leaf                      | 1                            |
| PO:0008015| hypocotyl vascular system           | 1                            |
| PO:0006339| juvenile leaf                       | 1                            |
| PO:00020121| lateral root                        | 1                            |
| PO:0000017| leaf primordium                     | 1                            |
| PO:0006034| leaflet margin                      | 1                            |
| PO:00020091| male gametophyte                    | 1                            |
| PO:000205012| pedicel vascular system             | 1                            |
| PO:0006041| petal epidermis                     | 1                            |
| PO:0006081| primary root apical meristem        | 1                            |
| PO:0006085| root meristem                       | 1                            |
| PO:0000014| rosette leaf                        | 1                            |
| PO:0009073| stigma                              | 1                            |
| PO:00020041| stipule                             | 1                            |

[28], and annotated genes with zero or less than 5% “present” gene calls from the 1,381 ATH1 arrays available at the start of these experiments. We developed a high throughput pipeline to generate promoter-reporter (GFP) constructs and transform them into Arabidopsis.
Figure 3 Examples of the expression patterns from different promoter-reporter constructs. A. leaf vascular expression from promoter-report construct of gene AT.CHR1.167; B. sepal and petal expression from promoter-report construct of gene AT.CHR1.167; C. root expression from promoter-report construct of gene NOVEL.CHRI_5915494; D. seed expression from promoter-report construct of gene NOVEL.CHRI_5915494; E. root but not root tip expression from promoter-report construct of gene AT1G64820; F. root tip expression from promoter-report construct of gene AT4G13985; G. petal expression from promoter-report construct of gene AT2G17845; H. carpel expression from promoter-report construct of gene AT2G40250; I. pollen expression from promoter-report construct of gene AT2G24370; J. expression from promoter-report construct of gene AT4G18395 in the region of the flower abscission zone; K. hydathode expression from promoter-report construct of gene AT02EUG13430; L. pollen expression from promoter-report construct of gene AT2G24370 (lower silique) comparing to wild-type pollen (upper silique); some un-florescent pollens on the low silique are the wild-type pollens spreading from the upper one. M. wild-type leaf without GFP expression; only chlorophyll autofluorescence is visible. N. wild-type flower without GFP expression; chlorophyll autofluorescence is visible. O. wild-type seeds without GFP expression; chlorophyll autofluorescence is visible.
Selection of expression vector
At the outset, three Gateway compatible binary vectors were compared for the expression level of their reporter genes and their convenience for this project: 1) pBGWFS7 http://gateway.psb.ugent.be/search/index/transcriptional_reporters/any which contains dual reporter genes (GFP and GUS) with BASTA selection, 2) pYXT1 which contains a GUS reporter gene with kanamycin selection, and 3) pYXT2 which contains the eGFP reporter gene [38] with kanamycin selection. We used both pYXT1 and pYXT2 vectors successfully in our previous study [29]. For comparison, we cloned 5 promoters into each of these vectors and found that, although the dual reporter vector, pBGWFS7 that contains the BASTA selection marker is simplest for selection of the transgenic lines, only GUS expression and no GFP expression was detected. For pYXT1 and pYXT2 vectors, both GUS and GFP expression were detected equally well. However, we chose to use the pYXT2 vector because it eliminates the need for GUS staining that would increase the amount of effort required in these experiments.

High throughput procedure
A high throughput method was developed to facilitate the ease of cloning and transformation over traditional gene-by-gene methods such as that used previously [29]. It includes a Perl script to design Gateway-compatible primers for cloning the promoter regions of candidate genes, PCR amplification of promoters of candidate genes, Gateway BP cloning, and Gateway LR cloning all in 96-well format, Agrobacterium transformation by triparental mating and floral dipping in 50 ml Falcon tubes (see MATERIALS AND METHODS). The 5' ("left") primer is located at least 2000 bp upstream of the start position of the coding sequence of the candidate genes since many studies have shown that region 1-2 kb upstream of the translation start site of genes determines the specificity of gene expression [39-44]. The 3' ("right") primers for the promoters were all located between 50 bp and 150 bp downstream of the translation start site of the target genes allowing an in-frame fusion with the GAL4-VP16 component of the reporter construct that contained at least 16 amino acids of the studied gene’s coding sequence as well as the 8 amino acid linker from Gateway system. This script made the primer design process facile, reliable and consistent.

Positive BP clones were picked and sequenced to confirm the cloned promoter sequences. Residual BP DNA from the sequence confirmation step was used for the Gateway LR cloning reaction in 96-well plate format. The triparental mating method was chosen to transfer the Gateway expression clones from E. coli to Agrobacterium [33,45], eliminating the need for DNA isolation from the selected clones as well as the subsequent high throughput electroporation. For the plant transformation, we cultured only 50 ml Agrobacterium, which allowed all steps (precipitation and resuspension of Agrobacterium cells and floral dipping of Arabidopsis plants) to be carried out in single 50 ml Falcon tubes. Recently, Davis et al [46] successfully transformed Arabidopsis by dipping plants directly into Agrobacterium cultures supplemented with surfactant, eliminating the need for media exchange to a buffered solution and further simplifying the transformation process.

A very important feature of this pipeline was the creation of a project specific laboratory information management system (LIMS) in a MySQL relational database to track all stages of the pipeline: candidate genes, primers, different stages of clone construction, transgenic lines, GFP checking at different stages, PCR and sequencing results and annotation of all the GFP images.

Therefore, this high throughput cloning and data tracking pipeline made our project management efficient and robust. However, out of 627 targeted promoters, only 266 constructs were ultimately transferred into Arabidopsis plants. This overall success rate is due to a degree of failure at each experimental step. For example, the successful rate for obtaining entry clones by BP cloning is 74.8%, for expression clones from entry clones by LR cloning 94.4%, for Agrobacterium clones by triparental mating 77.7%, and for kanamycin-resistant transgenic plants by floral dip 75.4%. Because of the nature of this high throughput project, we have not yet repeated any experimental step. Certainly the overall successful rate will increase if the unsuccessful clones at each step are reprocessed.

Expression Pattern Analysis
In order to confirm that the expression patterns are from the intended cloned promoters, vector-based primers flanking the cloning site were used to amplify the cloned promoters from transgenic plants showing GFP expression and the PCR products were sequenced for confirmation. Of 27 different constructs showing GFP expression, all the promoters were verified as correct. However, since not all lines were tested, researchers may wish to perform their own confirmation before using our lines. There are a total of 112 promoter-reporter constructs that show the same expression in more than one plant. For 79 promoter-reporter constructs, the same GFP expression patterns were observed from transgenic lines derived from independent floral dips and the other 33 constructs produced the same GFP expression patterns from separate seed borne by a single dipped plant are thus the most reliable data set. It has been shown that female reproductive tissues are the
primary target of Agrobacterium-mediated transforma-
tion and that the transformants derived from the same
seed pod contain independent T-DNA integration
events [47].

The validity of the specific patterns of GFP expression
from a representative set of promoters was confirmed
by quantitative real-time PCR (qRT-PCR) on RNA sam-
ple from multiple tissues (additional file 4: Table S4).
In every case, the tissue showing the highest expression
(lowest Ct) by qRT-PCR was the one from which GFP
expression was observed, and in almost every case this
expression value was many times higher than any of
the other tissues examined.

In this study, we checked GFP expression at 4 differ-
ent stages: on the selection plate around 10 days after
germination, at the rosette stage in soil, just before flow-
ering, and at the flowering stage. These stages were
chosen both to cover several developmental stages and also
for the convenience of the large amount of GFP screen-
ing and to and minimize the stress for the T0 transgenic
plants (e.g. checking root GFP on the selection plates
and during transplanting to soil). If the kanamycin-
selected transgenic plants did not show any GFP expres-
sion at any of the stages examined, PCR with GFP spe-
ific primers was used to confirm the presence of the
reporter transgene. Out of 256 plants representing 89
promoter reporter constructs without GFP expression
that we tested by leaf PCR [37], all were positive with
GFP primers. There are several possible reasons for the
lack of detectable GFP expression in these lines. The
promoter might be active only under conditions or at
specific developmental stages not examined in this
study. Alternatively, in contrast to the localized expres-
sion seen with many of the promoters, those without
visible GFP expression may in fact be expressed in the
plant but at levels too low to be detected by this
method. It is also possible that some of the promoter-
reporter constructs were truncated or rearranged during
T-DNA integration [48], or that gene silencing occurred
[49]. In addition, the inconsistent GFP expression pat-
terns that were detected from different transgenic lines
of 5 promoter constructs may be due to position effects
or to truncation or re-arrangement of the constructs
during transformation as well as to human error.

The goal of this project was to use the expression of
promoter-reporter constructs in transgenic plants to
infer the function of these no/low expression genes. Pro-
moters from 35 genes tested had GFP expression in
hydathodes, a secretory structure on leaf margins. An
example of hydathode expression from promoter-report
construct of gene AT02EUG13430 is shown in Figure
3K. Studies have shown that some genes expressed in
hydathodes are related to plant tolerance to toxicity. For
example, the Bot1 gene in barley is responsible for
boron-toxicity tolerance [50], the MTP11 gene in Arabid-
opsis is associated with plant tolerance to manganese
[21], and AtHMA3, a P1B-ATPase protein plays a role in
the detoxification of heavy metals [51]. AtCML9, a cal-
modulin-like protein from Arabidopsis thaliana, can
alter plant responses to abiotic stress and abscisic acid
and the expression of its promoter-reporter construct
also included hydathodes [43]. In addition, hydathodes
are one of the expression locations of a promoter-reporter
construct from ECA3, a Golgi-localized P2A-type
ATPase that plays a crucial role in manganese nutrition
in Arabidopsis [44]. Thus, it is possible that some the
genes of unknown function analyzed in this study that
show hydathodes expression are also involved in toler-
ance or detoxification pathways, suggesting a direction
for further study. Motif search by Multiple Em for Motif
Elicitation (MEME) [52] for all promoter sequences with
hydathode expression found a motif of CTTAAGA (P =
8.67e-09). However, its function and specificity will
require experimental verification.

Twenty six promoter-reporter constructs expressed
GFP in the abscission zones of siliques, flowers and
leaves including expression around the flower abscission
zone from construct of gene AT4G18395 (Figure 3J).
Abscission is a physiological process that involves the
programmed separation of entire organs, such as leaves,
petals, flowers, and fruit, allows plants to discard non-
functional or infected organs, and promotes dispersal of
progeny [53]. Promoter-reporter constructs from a num-
ero of confirmed abscission related genes including
BOP1 [54], BFN1 [55,56], HAE, HSL2, MKK4,5 [53],
AtZFP2 [57] show similar expression at abscission
zones. Using MEME [52], the sequence TAACCCTCA
was the most significant motif found in the promoters
analyzed in this study.

Thirty-six promoter-reporter constructs are expressed
in trichomes or the socket cells that surround a tri-
chome and provide support, suggesting their possible
function in trichome development, expansion and
branching. Many promoter constructs in our study were
expressed in specific floral organs, including sepal, petal,
filament, anther, carpel, and pollen. For example, pollen
specific GFP expression was detected from the construct
of gene AT2G24370 (Figure 3I, L). In addition to pro-
viding the clues to their function, they may also provide
novel promoters for plant genetic engineering. For
example, it has been shown that completely sterile Ara-
bidopsis plants can be generated by engineering carpel
and stamen-specific expressed genes [58]. Use of the
Ory s1 promoter (pollen-specific promoter) with anti-
sense Lol p5A cDNA led to the production of hypoallergen-
cic rye grass (Lolium perenne) [59].

Overall, in our study, positive transgenic plants were
obtained from 266 promoter constructs derived from
our intergenic and non- or low-expressing genes of unknown function. Among them, about 56% of constructs showed GFP expression in Arabidopsis. Thus the in vivo expression data from promoter-reporter constructs generated in this study has provided insights into possible functions of many genes previously lacking both expression data and functional annotation as well as another great gene expression resource for the research community.

Methods

Selection of the genes on the Affymetrix ATH1 GeneChip showing no or very limited expression

ATH1 no- or low-expression genes were identified as follows. The results of 1381 ATH1 arrays were downloaded from TAIR and the Affymetrix MAS calls used to classify them as “expressed” (present call) or “non-expressed” (marginal or absent call) in each experiment. “ATH1 no expression genes” and “ATH1 low expression genes” are those showing expression in either none or in less than 5% of the experiments respectively. Candidate genes from the microarray analysis were excluded if they were shown by massively parallel signature sequencing (MPSS) [30] to be expressed.

Primer design

The Perl script for primer design for promoter cloning has following features 1) align the CDS of the candidate genes to the Arabidopsis genomic sequence to locate the ATG start codon using BLAT; 2) extract 3000 bp upstream of the ATG and 150 bp downstream of the ATG as the genomic target region; 3) invoke Primer3 http://primer3.sourceforge.net/ to design primers using ATG as the genomic target region; 4) check the specificity of potential primer pairs against the Arabidopsis genome using BLAST [60] using 80% identity and allowing a maximum of 3 hits in the genome per pair of primers (it will hit itself in the genome). The script iterates through primer pair design until these criteria are fulfilled. In addition, the user can directly input Gateway sequences at the ends of upstream and downstream primers (aaaagcagctcgtg is added to the 5’ end of upstream primers and agaagctgtcgtg to the 5’ end of downstream of primers). Thus, the output of this script will be the primers with gene specific sequences at 3’ ends and Gateway cloning sequences at 5’ ends in the format for plate ordering according to the primer manufacturers’ requirements. This primer design script is available request to the authors.

Promoter-reporter (GFP) construct production

The Gateway cloning strategy was used to make promoter-reporter constructs largely according to the protocols in the Gateway Cloning Technology booklet (Invitrogen, Carlsbad, CA). Promoter amplifications were performed in 96 well plate format and in 2 PCR stages in order to add Gateway cloning sequences at the ends. The first PCR reaction contains 13.4 μL H2O, 4.0 μL 5x HF buffer, 0.40 μL 10 mM dNTPs, 2.0 μL gDNA (10 ng/μL) and 0.2 μL Phusion/iProof enzyme (New England Bio Lab, Ipswich, MA). The PCR conditions are: 1×: 98°C 30 sec, 1×: 98°C 10 sec, 63°C 30 sec, 72°C 2 min, 1×: 98°C 10 sec, 62°C 30 sec, 72°C 2 min, 1×: 98°C 10 sec, 61°C 30 sec, 72°C 2 min, 1×: 98°C 10 sec, 60°C 30 sec, 72°C 2 min, 1×: 98°C 10 sec, 59°C 30 sec, 72°C 2 min, 1×: 98°C 10 sec, 58°C 30 sec, 72°C 2 min, 1×: 72°C 10 min, 1×: 4°C forever. The primers used in the 1st PCR reaction were from the primer design script and contain gene-specific sequence as well as the attB site sequence for BP cloning. The second PCR reaction contains: 18.8 μL H2O, 6 μL 5x HF buffer, 0.8 μL 10 mM dNTPs, 4.0 μL AttB Primer mix10 uM, 0.4 μL Phusion/iProof (New England Bio Lab, Ipswich, MA), and 10 μl from the PCR1 reaction was used as template. The second PCR reaction contains: 1×: 98°C 30 sec, 19×: 98°C 10 sec; 56°C 30 sec; 72°C 2 min, 1×: 72°C 10 min, 1×: 4°C forever. The universal Gateway adaptor primers used in the second PCR reactions are following: attB1adaptor_prime: 5′-GGGGCAAGTTGTACAAAAAAGGAGCTC-3′, attB2adaptor_prime: 5′-GGGGACACCTTTTGTACAAAAAAGGCTGTT-3′. Five μl from the 2nd PCR reaction is diluted using H2O into 25 μl and 1 μl of diluted PCR product is used in the Gateway BP reaction. Gateway BP reactions were done in 96 well plates. Each reaction contains: 1 μl of diluted PCR product, 0.75 μl TE buffer (pH 8.0), 0.25 μl pDONR207 (150 ng/μl), 0.5 μl BP Clonase II mix. The reaction mixture was incubated at 25°C for from 1 hour to overnight depending on the experimental schedule. After incubation, 0.3 μl proteinase K was added to each reaction and incubated at 37°C for 10 min. Transformation was done in 96 well plates as follows: TOP10 competent E. coli cells (Invitrogen, Carlsbad, CA) were thawed on ice and 24 μL of cells was added into each BP reaction and incubated on ice for 20 min. Cells were heat shocked at 42°C for 30 sec and put on ice for 2 min, then 175 μL room temp SOC (Invitrogen, Carlsbad, CA) was added to each sample that was then incubated at 37°C, 225 rpm for 1 hour. Fifty μL of each transformation culture was plated on LB plates containing gentamycin (7 μg/mL) using glass beads in 4 Well Rectangular Mutidish w/Lid (Thermo Fisher Scientific, Rochester, NY). For each reaction, two positive colonies were picked and sequenced to confirm the cloned promoter. One μl of DNA from each sequence-confirmed BP clone (robotically isolated in the sequencing
center) was used in a Gateway LR reaction using pYXT2 as destination vector in 96 well plates. Each reaction contained: 1 μl of entry clone DNA, 0.75 μl TE buffer (pH 8.0), 0.25 μl pYXT2 (150 ng/μl), 0.5 μl LR Clonase II mix. Incubation time at 25°C was 1 hour to overnight depending on the experimental schedule. LR clone selection was the same as BP clone selection except for the substitution of kanamycin (50 μg/mL) for gentamycin in the LB plates.

**Tri-parental mating with Agrobacterium**

For triparental mating, a 50 mL culture of Agrobacterium GV3101 in LB with rifampicin (50 μg/mL) and gentamicin (50 μg/mL) was grown overnight at 28°C, 250 rpm and a 50 mL culture of pRK2013 E. coli helper strain in LB with Kan (50 μg/mL) was grown overnight at 37°C, 250 rpm. At the same time, 250 μl cultures of LR clones were incubated overnight in LB medium with kanamycin (50 μg/mL) at 37°C in deep well blocks. The following day, 50 μL Agrobacterium culture, 50 μL E. coli helper strain containing pRK2013 plasmid, and 50 μL E. coli LR clone culture were plated together on LB agar in each compartment of 4 Well Rectangular MutiDish w/Lid (Thermo Fisher Scientific, Rochester, NY) for each LR clone. The following day, from the bacterial lawn, a small loopful of bacteria from each well was streaked onto LB plates containing rifampicin (50 μg/mL), gentamicin (50 μg/mL) and kanamycin (50 μg/mL) to select for Agrobacterium with the pYXT2 construct containing the target promoters. Two positive colonies from each reaction were inoculated into corresponding positions of two 96 well blocks with 250 μl. LB media containing rifampicin (50 μg/mL), gentamicin (50 μg/mL), kanamycin (50 μg/mL) grew at 28°C, 225 rpm for 2 days. PCR was then used to confirm the positively selected Agrobacterium colonies as follows: Twenty μl of 20 mM NaOH solution were added into a new PCR plate, then 3 μl Agrobacterium culture were added into each well. The PCR plate was sealed and incubated at 37°C for 5 minutes. Then 2 μl of NaOH treated Agrobacterium culture were transferred into new PCR plate and a PCR reaction was set up by adding 4.0 μl H2O, 10 μl 2× PCR mix (New England Bio Lab, Ipswich, MA), 4.0 μl promoter-specific primer pair (2.5 μM). The PCR conditions are: 1x: 95°C 2 min, 30×: 94°C 30 sec, 55°C 30 sec, 72°C 3 min, 1x: 72°C 5 min, 1x: 4°C forever. The positive colonies were used in plant transformation.

**Transformation and Plant Growth**

*Arabidopsis thaliana* ecotype Columbia-0 was used in all our experiments. Plants were grown on Redmix at 25°C and 24 hours light (cool white fluorescent; ~150 microeinsteins). Plant transformation and seed selection was done according to standard methods [35] except that only 50 ml of the Agrobacterium was cultured and used in floral dipping so that all subsequent manipulations up to and including floral dipping were done in 50 ml Falcon tubes (Becton Dickson Labware, Franklin Lakes, NJ). Three independent transformations were done for each construct (i.e. three plants were dipped separately in the same Agrobacterium tube) and up to 3 plants from each kanamycin (50 μg/mL) selection plate were transferred into soil.

**Microscopy**

Plants were observed using an Olympus SZX12 stereo-microscope equipped with a 100 W mercury lamp for epifluorescence and a parfocal 1.6× objective. Images were recorded with an Olympus DP71 digital camera.

**Quantitative real-time PCR**

Leaf RNA was isolated from Plant basal rosette leaves from plants about 21 days after germination. Flower RNA was isolated opened flowers and unopened flower buds. Young silique RNA was isolated from immature siliques; root RNA isolation has been described previously [61]. RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) as described by the manufacturer and then filtered using RNaseasy columns (Qiagen, Valencia, CA, USA). First-strand cDNA was synthesized by priming with oligo-dT using SuperScript III reverse transcriptase (Invitrogen Carlsbad, CA, USA) following the instructions of the provider. PCR reactions were carried out in an ABI PRISM® 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA). SYBR® Green was used to quantify dsDNA synthesis. Reactions (10 μl total volume) were amplified using the following standard PCR protocol: 50°C for 2 min; 95°C for 10 min; 40 cycles of 95°C for 15 sec and 60°C for 1 min, and SYBR® Green fluorescence was measured continuously. Three biological and 2 technical replicates were used for analysis. PCR efficiency was estimated using LinReg software with data obtained from the exponential phase of each individual amplification plot. Cycle time (Ct) values were taken at a threshold value of 0.2.

**Additional material**

| Additional file 1: Table S1 | Identifiers and promoter sequences of genes used in this study. |
|-----------------------------|---------------------------------------------------------------|
| Additional file 2: Table S2 | Listing of all genes used in this study with their TAIR identifiers (where available) and the success status at each step in the transgenic plant production pipeline. |
| Additional file 3: Table S3 | Listing of all genes assigned to each Plant Ontology (PO) code. |
| Additional file 4: Table S4 | Validation of GFP expression patterns by quantitative real time RT-PCR. |
Acknowledgements
We thank all members of the plant group at TIGR/JCVI for their help and Dr. Harry J Klee for providing us the E. coli helper strain for triparental matings. This work was supported by the National Science Foundation (grant no. OS20022).

Authors’ contributions
YX participated in the experimental work and wrote the manuscript. JCR, EJM, BAU, WAM, and WW performed the experiments. JZ and HCW carried out bioinformatic and database-related tasks. CDT conceived and coordinated the project and contributed to the manuscript. All authors have read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 4 December 2009 Accepted: 6 August 2010

Published: 6 August 2010

References
1. Initiative AG: Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 2000, 408:761-865.
2. Chory J, Ecker JR, Briggs S, Capeco M, Coruzzi GM, Cook D, Dangl J, Grant S, Guenniot ML, Henikoff S, et al: National Science Foundation-Sponsored Workshop Report: “The 2010 Project” functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. Plant Physiol 2000, 123:423-426.
3. Wisman E, Ohnoguro J: Arabidopsis microarray service facilities. Plant Physiol 2000, 124:1468-1471.
4. Zhu T, Wang X: Large-scale profiling of the Arabidopsis transcriptome. Plant Physiol 2000, 124:1472-1476.
5. Redman JC, Haas BJ, Tanimoto G, Town CD: Development and evaluation of an Arabidopsis whole genome Affymetrix probe array. J Plant 2004, 38:545-561.
6. Kim H, Snesrud EC, Haas B, Cheung F, Town CD, Quackenbush J: Gene expression analyses of Arabidopsis chromosome 2 using a genomic DNA amplification microarray. Genome Res 2003, 13:327-340.
7. Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, et al: Empirical analysis of transcriptional activity in the Arabidopsis genome. Science 2003, 302:842-846.
8. Stok V, Samanta MP, Tongprasit W, Sethi H, Liang S, Nelson DC, Hegeman A, Nelson C, Rancour D, Bednarek S, et al: Identification of transcriptional changes in Arabidopsis thaliana by using high-resolution genome tiling arrays. Proc Natl Acad Sci USA 2005, 102:4453-4458.
9. Labubinger S, Zeller G, Henn SR, Sachsenberg T, Widmer CK, Naouar N, Vuyveste M, Scholkopf B, Ransch G, Weigel D: AT-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis thaliana. Genome Biol 2008, 9:R112.
10. Schubeler D, MacAlpine DM, Scalzo D, Witelski C, Koopberg C, van Leeuwen F, Gottschling DE, O’Neill LP, Turner BM, Delrow J, et al: The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev 2004, 18:1263-1271.
11. Gendrel AV, Lippman Z, Martienssen R, Colot V: Profiling histone modification patterns in plants using genomic tiling microarrays. Nat Methods 2005, 2:213-218.
12. Kurdistan SK, Tavazoie S, Grunstein M: Mapping global histone acetylation. Cell 2004, 117:721-733.
13. Chang S, Pikaard CS: Transcript profiling in Arabidopsis reveals complex responses to global inhibition of DNA methylation and histone deacetylation. J Biol Chem 2005, 280:796-804.
14. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zelterling J, Schreiber J, Hennett N, Kanin E, et al: Genome-wide location and function of DNA binding proteins. Science 2000, 290:2306-2309.
15. Thibaud-Nissen F, Wu H, Richardson T, Redman JC, Johnson C, Green R, Anis J, Town CD: Development of Arabidopsis whole-genome microarrays and their application to the discovery of binding sites for the TGA2 transcription factor in salicylic acid-treated plants. Plant J 2006, 47:152-162.
16. Springer PS: Gene traps: tools for plant development and genomics. Plant Cell 2000, 12:1007-1020.
17. Li HM, Autschlindl M, Chory J: Arabidopsis mutants define downstream branches in the phototransduction pathway. Genes Dev 1994, 8:339-349.
18. Mannings JM, Penninckx IA, Vermaere K, Kazan K, Brown RL, Morgan A, Maclean DJ, Curtis MD, Cammue BP, Broekgaart WF: The promoter of the plant defensin gene PDF1.2 from Arabidopsis is systemically activated by fungal pathogens and responds to methyl jasmonate but not to salicylic acid. Plant Mol Biol 1998, 38:1071-1080.
19. Santamaria M, Thomson CJ, Read ND, Loake G: The promoter of a basic PR1-like gene, AtPR1, from Arabidopsis establishes an organ-specific expression pattern and responsiveness to ethylene and methyl jasmonate. Plant Mol Biol 2001, 47:641-652.
20. Vicente-Aguillo F, Riga S, Desbosses G, Dolan L, Hatzopoulos P, Grabov A: Potassium carrier TRH1 is required for auxin transport in Arabidopsis roots. Plant J 2004, 40:523-535.
21. Peiter E, Montanian B, Cogert A, Pedas P, Husted S, Maathuis FJ, Blaudez D, Chalot M, Sanders D: A secretory pathway-localized cation diffusion facilitator confers plant manganese tolerance. Proc Natl Acad Sci USA 2007, 104:8532-8537.
22. Bernal AJ, Yoo CM, Mutwil M, Jensen JK, Hou G, Blaukopf C, Sorensen L, Blancafor EB, Scheller HV, Williams WG: Functional analysis of the cellulose synthase-like genes CSLD1, CSLD2, and CSLD4 in tip-growing Arabidopsis cells. Plant Physiol 2008, 148:1238-1253.
23. Hong RP, Hanjaguchlu L, Busch MA, Weigel D: Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. Plant Cell 2003, 15:1296-1309.
24. Larkin JC, Oppenheimer DG, Pollock S, Marks MD: Arabidopsis GLABROUS1 Gene Requires Downstream Sequences for Function. Plant Cell 1993, 5:1739-1748.
25. Stamm M, Belote C, Dorweiler JE, Chandler VL: Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev 2002, 16:1906-1918.
26. Moskal WA Jr, Wu HC, Underwood BA, Wang W, Town CD, Xiao Y: Experimental validation of novel genes predicted in the un-annotated regions of the Arabidopsis genome. BMC Genomics 2007, 8:18.
27. Jethów TM, Rousted P: ElGene: an eukaryotic gene finder that combines several sources of evidence. Lect Notes in Comput Sci 2006, 11:125.
28. Korf I, Flicek P, Duan D, Brent MR: Integrating genomic homology into gene structure prediction. Bioinformatics 2001, 17(Suppl 1):S140-148.
29. Xiao YL, Smith SR, Ishmael N, Redman JC, Kumar N, Monaghan EL, Ayelle M, Haas BJ, Wu HC, Town CD: Analysis of the cDNAs of hypothetical genes on Arabidopsis chromosome 2 reveals numerous transcript variants. Plant Physiol 2005, 139:1323-1337.
30. Nakano M, Nobuta K, Veneramu K, Tej SS, Skogem JW, Meyers BC: Plant MPSS databases: signature-based transcriptional resources for analyses of miRNA and small RNA. Nucleic Acids Res 2006, 34:D731-735.
31. Underwood BA, Vanderhaegen R, Whitford R, Town CD, Hilson P: Simultaneous high-throughput recombinational cloning of open reading frames in closed and open configurations. Plant Biotechnol J 2006, 4:317-324.
32. Hartley JL, Temple GF, Brash MA: DNA cloning using in vitro site-specific recombination. Genes Res 2000, 10:1788-1795.
33. Van Haute E, Eoos H, Maes M, Warren G, Van Montagu M, Schell J: Intergeneretic transfer and exchange recombination of restriction fragments cloned in pBR322: a novel strategy for the reversed genetics of the Ti plasmids of Agrobacterium tumefaciens. Embo J 1983, 2:411-417.
34. Xiao YL, Malik M, Whitelaw CA, Town CD: Cloning and sequencing of cDNAs for hypothetical genes from chromosome 2 of Arabidopsis. Plant Physiol 2002, 130:2118-2123.
35. Clough SJ, Bent AF: Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 1998, 16:735-743.
36. Jaiswal P, Avraham S, Ilke K, Kellogg EA, McCouch S, Pujar A, Reiser L, Rixe SY, Sachs MM, Schaeffer M, et al: Plant Ontology (PO): A Controlled Vocabulary of Plant Structures and Growth Stages. Comp Func Genomes 2005, 6:388-397.
and perturbed cellular phenotypes in mature Arabidopsis embryos. *Plant J* 2000, 24:543-550.

39. de Boer GJ, Testenink C, Piela G, Nijkamp HJ, Stutte AR: Sequences surrounding the transcription initiation site of the Arabidopsis enoyl-acyl carrier protein reductase gene control seed expression in transgenic tobacco. *Plant Mol Biol* 1999, 39:1197-1207.

40. McKinney EC, Kandasamy MK, Meagher RB: Small changes in the regulation of one Arabidopsis profilin isoform, PRF1, alter seeding development. *Plant Cell* 2001, 13:1179-1191.

41. Stone JM, Liang X, Niel D, Stein JJ: Arabidopsis AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. *Plant J* 2005, 41:744-754.

42. Welchen E, Gonzalez DH: Differential expression of the Arabidopsis cytochrome c genes Cytc-1 and Cytc-2. Evidence for the involvement of TCP-domain protein-binding elements in anther- and meristem-specific expression of the Cytc-1 gene. *Plant Physiol* 2005, 139:88-100.

43. Magnan P, Ranty B, Charpentier M, Sotta R, Galaud JP, Aldon D: Mutations in AtCML9, a calmodulin-like protein from Arabidopsis thaliana, alter plant responses to abiotic stress and ascorbic acid. *Plant J* 2008, 56:575-589.

44. Mills RF, Doherty ML, Lopez-Marques RL, Weimar T, Dupree P, Palmgren MG, Pittman JK, Williams LE: ECAS, a Golgi-localized P2A-type ATPase, plays a crucial role in manganese nutrition in Arabidopsis. *Plant Physiol* 2008, 146:116-128.

45. Wise AA, Liu Z, Binns AN: Three methods for the introduction of foreign DNA into Agrobacterium. *Methods Mol Biol* 2006, 343:43-53.

46. Davis AM, Hall A, Millar AJ, Darrah C, Davis SJ: Protocol: Streamlined sub-protocols for floral-dip transformation and selection of transformants in Arabidopsis thaliana. *Plant Methods* 2009, 5:3.

47. Desfeux C, Clough SJ, Bent AF: Female reproductive tissues are the primary target of Agrobacterium-mediated transformation by the Arabidopsis floral-dip method. *Plant Physiol* 2000, 123:895-904.

48. Chilton MD, Que Q: Targeted integration of T-DNA into the tobacco genome at double-stranded breaks: new insights on the mechanism of T-DNA integration. *Plant Physiol* 2003, 133:956-965.

49. Francis KE, Spiker S: Identification of Arabidopsis thaliana transformants without selection reveals a high occurrence of silenced T-DNA integrations. *Plant J* 2005, 41:464-477.

50. Sutton T, Baumann U, Hayes J, Collins NC, Shi BJ, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M, Langridge P: ATPase, plays a crucial role in manganese nutrition in Arabidopsis. *AtHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in barley arising from efflux transporter amplification. *Plant Physiol* 2008, 146:116-128.

51. Wise AA, Liu Z, Binns AN: Three methods for the introduction of foreign DNA into Agrobacterium. *Methods Mol Biol* 2006, 343:43-53.

52. Davis AM, Hall A, Millar AJ, Darrah C, Davis SJ: Protocol: Streamlined sub-protocols for floral-dip transformation and selection of transformants in Arabidopsis thaliana. *Plant Methods* 2009, 5:3.

53. Desfeux C, Clough SJ, Bent AF: Female reproductive tissues are the primary target of Agrobacterium-mediated transformation by the Arabidopsis floral-dip method. *Plant Physiol* 2000, 123:895-904.

54. Bailey TL, Elkan C: Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994, 2:28-36.

55. Cho SK, Larue CT, Chevalier D, Wang H, Jinn TL, Zhang S, Walker JC: Identification of Arabidopsis thaliana transformants without selection reveals a high occurrence of silenced T-DNA integrations. *Plant J* 2005, 41:464-477.

56. Sutton T, Baumann U, Hayes J, Collins NC, Shi BJ, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M, Langridge P: AtHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in barley arising from efflux transporter amplification. *Science* 1994, 264:123-128.

57. Moor R, Crouzet J, Gravel A, Auroy P, Leonhardt N, Vavasseur A, Richard P: ATHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in Arabidopsis. *Plant Physiol* 2009, 149:894-904.

58. Bailey TL, Elkan C: Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994, 2:28-36.

59. Cho SK, Larue CT, Chevalier D, Wang H, Jinn TL, Zhang S, Walker JC: Identification of Arabidopsis thaliana transformants without selection reveals a high occurrence of silenced T-DNA integrations. *Plant J* 2005, 41:464-477.

60. Bhalla PL, Svoboda I, Singh MB: Reduction in allergenicity of grass pollen by genetic engineering. *Int Arch Allergy Immunol* 2001, 124:51-54.

61. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990, 215:403-410.

Submit your next manuscript to BioMed Central
and take full advantage of:

• Convenient online submission
• Thorough peer review
• No space constraints or color figure charges
• Immediate publication on acceptance
• Inclusion in PubMed, CAS, Scopus and Google Scholar
• Research which is freely available for redistribution

Cite this article as: Xiao et al.: High throughput generation of promoter reporter (GFP) transgenic lines of low expressing genes in Arabidopsis and analysis of their expression patterns. *Plant Methods* 2010 6:18.