Volume 5, Issue 4, 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Gene Expression Profiling of Transcription Factors of Arabidopsis Thaliana using Microarray Data Analysis

**Nivedita*, Pramod Kumar Yadav, Budhayash Gautam**
Jacob School of Biotechnology and Bioengineering, Department of Computational Biology and Bioinformatics,
Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad, Uttar Pradesh, India

***Abstract - Transcription factors (TFs) regulate the expression of genes at the transcriptional level. Modification of TF activity dynamically alters the transcriptomes, which leads to metabolic and phenotypic changes. Thus, functional analysis of TFs using 'omics-based' methodologies is one of the most important areas of the post-genome era. We have studied the gene expression profiles of transcription factors of Arabidopsis thaliana at the two specific flower maturation stages (stage 12 and stage 13). The stage 12 is the oldest closed bud and stage 13 is the youngest open flowers, which were considered to study the role of ARF and MYB transcription factors. This study is based on the microarray data analysis and functional analysis for the characterization of gene expressions in flowers before and after they open, and to determine major biological processes of ARF6 ARF8 and MYB21 MYB24 transcription factors. Our data analyses showed, total 27199 genes of transcription factors of Arabidopsis thaliana. Using the criteria of fold change (FC ≥ 2.0 and FC ≤ -2.0), we identified 515 and 203 upregulated and 493 and 302 downregulated genes of ARF transcription factor of stage 12 and 13 respectively. Likewise for MYB transcription factor, we identified 50 and 253 upregulated and 177 and 325 downregulated genes for stage 12 and 13, rest of many genes shows none significant expression of these two specific flower maturation stages of Arabidopsis thaliana. There were only 206 and 65 upregulated genes of ARF and MYB and 334 and 156 downregulated genes of ARF and MYB that were overlapped among both stages (12 and 13) respectively. We studied, Arabidopsis TFs and introduce strategies for the functional analysis of plant TFs. These strategies can be assigned to three categories: bioinformatics analysis; data analysis; expression analysis and functional analysis. The data was analyzed using dChip (tool).***

***Key words - bioinformatics, microarray, Arabidopsis thaliana, transcription factors, gene expression profiling.***

## I. INTRDUCTION

*Arabidopsis thaliana* is a small flowering plant that is widely used as a model organism in plant biology. Arabidopsis is a member of the mustard (*Brassicaceae*) family, which includes cultivated species such as cabbage and radish. Although not of major agronomic significance, Arabidopsis offers important advantages for basic research in genetics and molecular biology. The small size of its genome, and the fact that it is diploid, makes *Arabidopsis thaliana* useful for genetic mapping and sequencing — with about 157 mega base pairs and five chromosomes, *Arabidopsis* has one of the smallest genomes among plants. It is marked that transcriptional regulation plays a pivotal role in the control of gene expression in plants. Thorough studies of plant mutants have revealed that informative phenotypes are often caused by mutations in genes for transcription factors (TFs), and a number of TFs have been identified that act as key regulators of various plant functions. TFs, which regulate the first step of gene expression, are usually defined as proteins containing a DNA-binding domain (DBD) that recognize a specific DNA sequence. In addition, proteins without a DBD, which interact with a DNA-binding protein to form a transcriptional complex, are often categorized as TFs **[1].** In 2000, the entire genome sequence of *Arabidopsis thaliana* was determined and the genome was predicted to contain 25,498 protein-coding genes (Arabidopsis Genome Initiative 2000). Based on sequence conservation with known DBDs, **[2]** reported that around 1,500 of these genes encode TFs, and more recent analyses have recognized > 2,000 TF genes in the Arabidopsis genome **[3], [4], [5], [6].** .*Arabidopsis* has been extensively studied as a model for flower development and maturation. Flower maturation consists of several events that contribute to reproductive success as flowers open, including petal expansion, stamen filament elongation, pollen release, nectary maturation, stigma growth, and gynoecium maturation to support pollen tube growth. The Arabidopsis transcription factors ARF6 (Auxin Response Factor 6) and ARF8 regulate all of these processes, in part by activating jasmonate biosynthesis. Jasmonates in turn activate genes encoding the transcription factors MYB21 and MYB24, which mediate a subset of the processes controlled by ARF6 and ARF8. Perfect flowers have both male organs that produce and release pollen and female organs that make and harbor seeds. Flowers also often attract pollinators using visual or chemical signals. So that male, female, and pollinator attraction functions occur at the right time, flower organs must grow and mature in a coordinated fashion **[7].** In the model self-pollinating plant Arabidopsis, a transcriptional network regulates genes that ensure coordinated growth of different flower organs, as well as pollen release and gynoecium (female) competence to support pollination. This network also regulates nectary development and production of volatile chemicals that may attract or repel insects **[8].** The most commonly used technology to profile the expression of thousands of transcripts simultaneously is microarrays **[9], [10]** Microarrays are one of the latest breakthroughs in experimental molecular biology that allow monitoring the

 

expression levels of tens of thousands of genes simultaneously. Arrays have been applied to studies in gene expression, genome mapping, SNP discrimination, transcription factor activity, toxicity, pathogen identification and many other applications **[11].**Due to recent advances in microarray technology, it is now feasible to obtain gene expression profiles of samples. Basically, it is used to characterize complex biological circumstances and diseases. Transcriptional profiling is a tool that provides unique data about disease mechanisms, regulatory pathways, and gene function. This technology not only allows comparison of gene profiles in normal and pathological tissues or cells, but also helps us establish interrelationships among genes **[12].** Hence, using these techniques, the aim of this study was to characterize the gene expression profiles in flowers before and after they open, and to determine major biological processes of arf6 arf8 and myb21 myb24 transcription factors, by microarray data analysis.

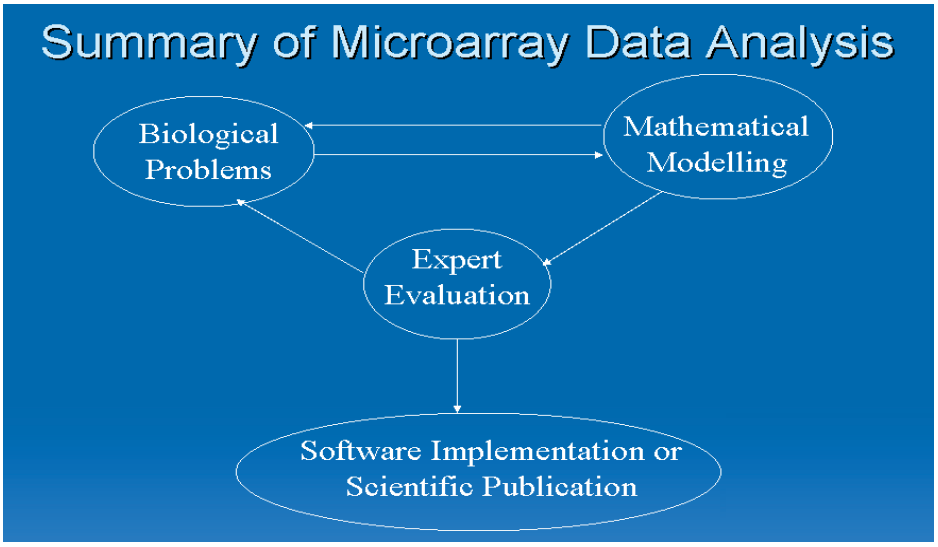


Fig 1. Summary diagram of microarray data analysis.

## II. MATERIAL AND METHODS

**Data Collection-** The dataset, consisting of the microarray samples of *Arabidopsis thaliana* at two flower maturation stages. The two specific flower maturation stages are stage 12 and stage 13. The stage 12 is the oldest closed bud and stage 13 is the youngest open flowers. A Profile consists of the expression measurements for an individual gene across all samples in a DataSet. Profiles can be searched using the GEO Profiles interface. Used GEO DataSet for accession no is GSE32193. As in Table 1, there were three biological replicates at each of two developmental stages, stage 12 (oldest closed buds) and stage 13 (youngest open flowers), for three genotypes (Wild type, arf6-2 arf8-3, and myb21-5 myb24-5), completing total 18 samples.

Table I List of datasets used

| | | | |
|---|---|---|---|
| GSM797716 | Wild type stage 12 rep A | GSM797725 | Wild type stage 13 rep A |
| GSM797717 | Wild type stage 12 rep B | GSM797726 | Wild type stage 13 rep B |
| GSM797718 | Wild type stage 12 rep C | GSM797727 | Wild type stage 13 rep C |
| | | | |
| GSM797719 | arf6 arf8 stage 12 rep A | GSM797728 | arf6 arf8 stage 13 rep A |
| GSM797720 | arf6 arf8 stage 12 rep B | GSM797729 | arf6 arf8 stage 13 rep B |
| GSM797721 | arf6 arf8 stage 12 rep C | GSM797730 | arf6 arf8 stage 13 rep C |
| | | | |
| GSM797722 | myb21 myb24 stage 12 rep A | GSM797731 | myb21 myb24 stage 13 rep A |
| GSM797723 | myb21 myb24 stage 12 rep B | GSM797732 | myb21 myb24 stage 13 rep B |
| GSM797724 | myb21 myb24 stage 12 rep C | GSM797733 | myb21 myb24 stage 13 rep C |

**Data analysis:** For analysis we used dchip (version 2005) microarray analyzer tool, which is freely available on web. dChip *(http://www.hsph.harvard.edu/cli/complab/dchip/)* is one of the most commonly used tool for normalizing methods for estimating expression signals. It is free for academic use, though it is not open source. dChip used model-based estimate for gene expression indeces (MBEI), which was proposed by Li and Wong (2001*)*. dChip require CDF files for *Arabidopsis thaliana*. And to correctly annotate the probe sets, it's necessary to download Affymetrix annotation files from *www.affymatrix.com* which is an American company that manufactures DNA microarray chip. To use dChip, it is needed to provide Affymetrix array data files in CEL format. Now cell files of all the 18 samples were loaded to dChip

 

microarray data analyzer and samples were normalized by RMA (Robust Microarray Analysis) method to remove errors and redundancies for make data accurate and correct. As in following plots, blue dots represent the experimental probes and is red dots are the selected probes in invariant set. In this part of the plot shows the deviation of blue and red dots which indicates the need of normalization.

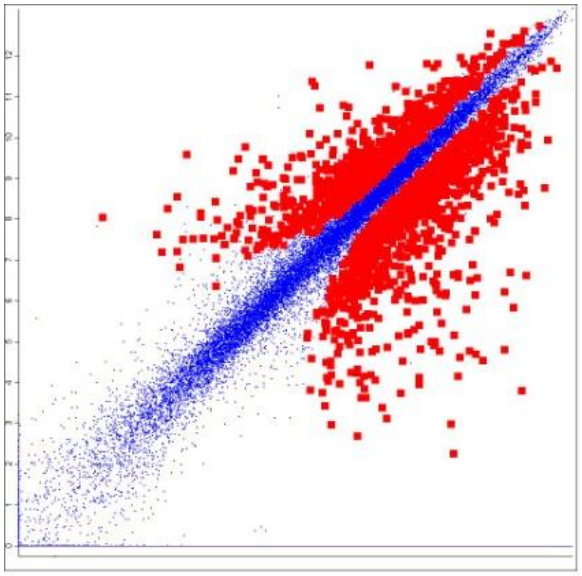

Fig 2. Scatter plot between control and experiment (ARF) of stage 12

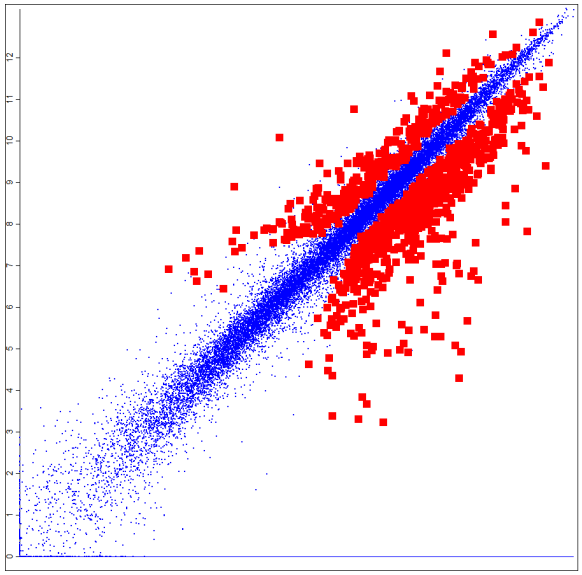

Fig 4. Scatter plot between control and experiment (MYB) of stage 12

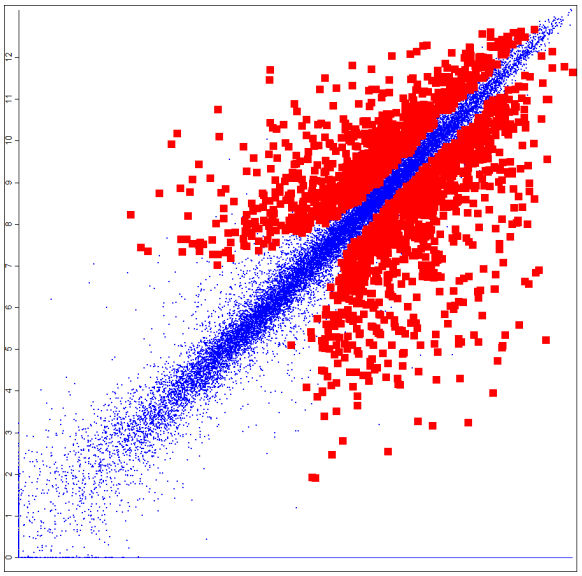

Fig 3. Scatter plot between control and experiment (ARF) of stage 13

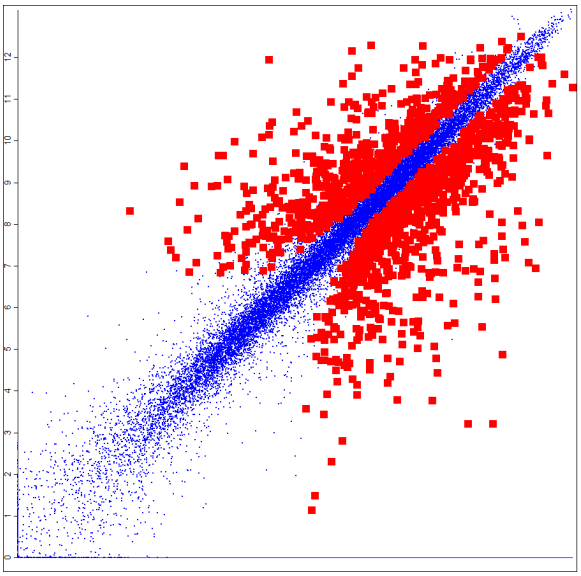

Fig 5. Scatter plot between control and experiment (MYB) of stage 13

After normalization of 18 samples were compared as – Arf regulatory factors were compared with control factor of stage12, Myb regulatory factors were compared with control factor of stage 12, Arf regulatory factors were compared with control factor of stage13 and Myb regulatory factors were compared with control factor of stage 13. These 4 Comparison results were further categorized into up regulated value (FC ≥2.0) performs more expressions, down regulated value (FC ≤ -2.0) performs low expressions and non significant values (FC < 2.0 and FC > -2.0) Selections of probe sets according to fold change are like this-

- Up regulated values of Arf regulatory factor and control factor of stage12
- Down regulated values of Arf regulatory factor and control factor of stage12
- Up regulated values of Myb regulatory factor and control factor of stage12
- Down regulated values of Myb regulatory factor and control factor of stage12
- Up regulated values of Arf regulatory factor and control factor of stage13
- Down regulated values of Arf regulatory factor and control factor of stage13
- Up regulated values of Myb regulatory factor and control factor with of stage13
- Down regulated values of Myb regulatory factor and control factor of stage13

 

Data analysis, with above mentioned combinations were used for future analysis. Venn diagram of up and down regulated probe sets was made from Venny tool *(www.bioinfogp.cnb.csic.es/tools/venny*), which shows all possible combinations between mentioned data sets.

**Functional analysis:** The functional analysis of TFs using bioinformatics techniques has become an important and effective strategy. For functional analysis of above datasets, PLEXdb (plant expression database) *www.plexdb.org* and TAIR10 (The Arabidopsis Information) *www.arabidopsis.org*, databases were used. The locus ids of all probsets were retrieved from PLEXdb which were used to retrieve GO (gene ontology) annotation from TAIR 10 database. Each list of differentially expressed genes was analyzed in the background of gene ontology (GO) to identify groups of genes with similar functions or processes.

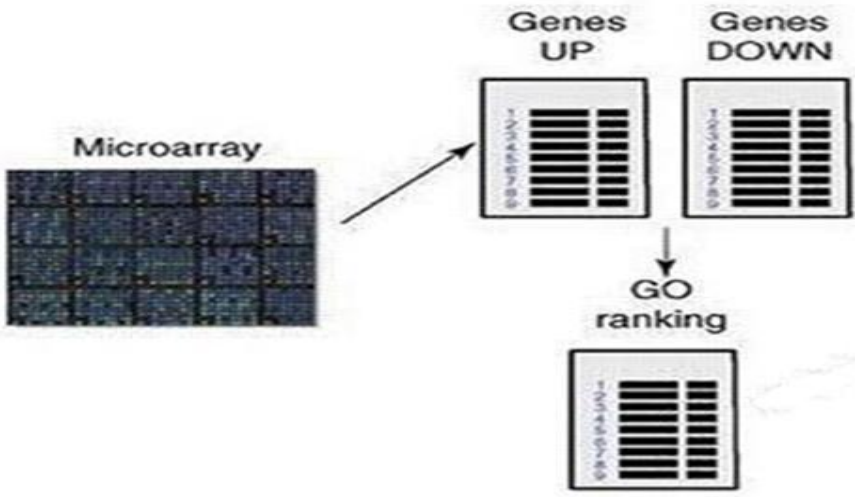


Fig 6. It Represents microarray to biological interpretation of results. Genes up Regulated, Genes down Regulated, and GO ranking symbolize lists, which are essentially one- dimensional annotation of the data, though GO ranking already contains groups of genes on the basis of their P-value significance. The arrows indicate that there are many ways to develop the more complex structures from the initial data.

## III. RESULTS AND DISCUSSION

**Differential Gene Expressions of transcription factors at two specific flower maturation stages of *Arabidopsis thaliana:***

Microarray data sets are commonly very large, and analytical precision is influenced by a number of variables. So it is extremely useful to reduce the dataset to those genes that are best distinguished between the two cases or classes (e.g. normal vs. diseased). Such analyses produce a list of genes whose expression is considered to change and known as differentially expressed genes. Identification of differential gene expression is the first task of an in depth microarray analysis. Analysis focused on identifying differential gene expression among transcription factors at two specific flower maturation stages of *Arabidopsis thaliana.* In this prediction, the two stages are stage 12, which is an oldest closed bud and stage 13 is the youngest open flower, both two stages contains ARF and MYB transcription factors. After normalization of all dataset we have found that how many genes are over expressed, under expressed and non significant genes, on the basis of fold change, in all combinations of regulatory factors at stage 12 and stage 13. Table 2 describes the total number of up regulated genes, down regulated genes and non significant genes of transcription factors at two specific flower maturation stages of *Arabidopsis thaliana.*

Table 2 Up regulated; down regulate and non significant genes in all four combinations of regulatory factors of stage 12 and 13

| **Control - arf of stage 12** | **No of genes** |
|---|---|
| up regulated genes | 780 |
| non significant genes | 21074 |
| down regulated genes | 892 |
| **control - myb of stage 12** | |
| up regulated genes | 119 |
| non significant gene | 500 |
| down regulated gene | 344 |
| **control - arf of stage 13** | |
| up regulated genes | 425 |
| non significant gene | 900 |

 

| | |
|---|---|
| down regulated gene | 654 |
| **control - myb of stage13** | |
| up regulated genes | 330 |
| non significant gene | 686 |
| down regulated gene | 495 |

**Venn diagram-** Venn diagram or set diagram of up and down regulated probe sets was made from Venny tool. With the help of Venn diagram we have found all possible relations between eight combinations of regulatory factors of stage 12 and 13, which contains over expressed genes and under expressed genes. Fig 7 describes as colored circles. The blue circle represents myb and arf regulatory factors containing over and under expressed genes of stage 12, and yellow circle represents myb and arf regulatory factors containing over and under expressed genes of stage 13. Each separate type of regulatory factor of different stage can be imagined as a point somewhere in the diagram, area where the blue and yellow circles overlap. That area contains all common genes present in both stage (stage 12 and 13). Table 3, represents the total no of up regulated and down regulated genes present in stage 12 stage 13 and common genes present in both stages of all combinations of regulatory factors. With this table, we can say that there are 1292 genes, which regulate the Arabidopsis flower maturation cycle highly and 1787 genes down regulated.

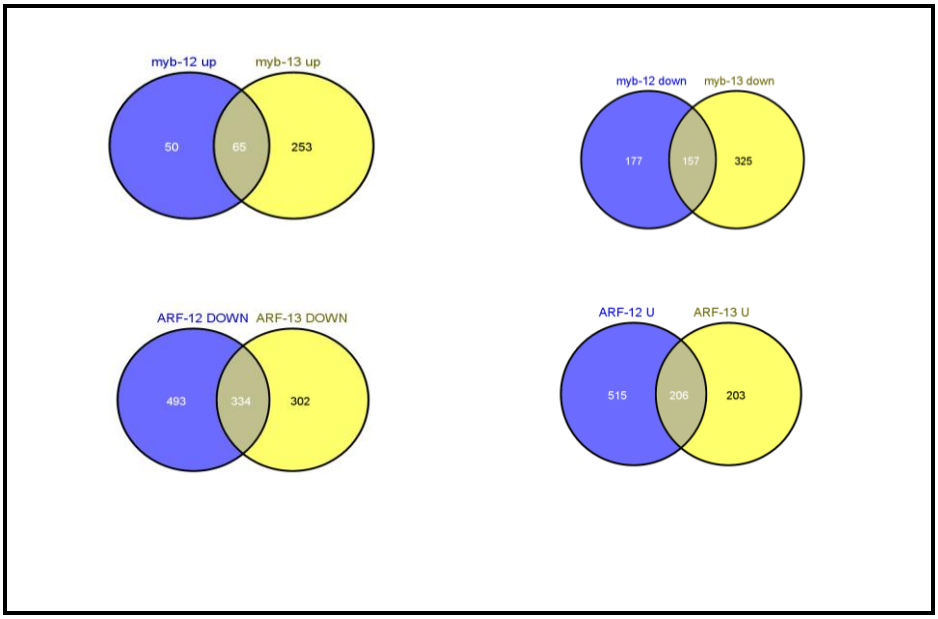


Fig 7. Venn diagram analysis of differentially expressed gene sets of comparison ARF and MYB transcription factors of both stages

Table 3Total number of genes present in all stages

| **Up regulated genes** | **No of genes present** | **Down regulated genes** | **No of genes present** |
|---|---|---|---|
| control-arf (only stage 12) | 515 | control-arf (only stage 12) | 493 |
| control-arf (common to stage 12 and 13) | 206 | control-arf (common to stage 12 and 13) | 334 |
| control-arf (only stage 13) | 203 | control-arf (only stage 13) | 302 |
| control-myb (only stage 12) | 50 | control-myb (only stage 12) | 177 |
| control-myb (common to stage 12 and 13) | 65 | control-myb (common to stage 12 and 13) | 156 |
| control-myb (only stage 13 ) | 253 | control-myb (only stage 13) | 325 |
| Total | 1292 | Total | 1787 |

Up-regulation is a process that occurs within a biological system triggered by a signal (originating internal or external to the system), which results in increased expression of one or more genes and as a result the protein(s) encoded by those genes. On the converse, down-regulation is a process resulting in decreased gene and corresponding protein expression. By above table, there were 1021 and 1297 total genes which are highly and under expressed in both two stages of transcription factors of *A. thaliana*.

 

**Functional classification of the differentially expressed gene of two specific transcriptomes of *Arabidopsis thaliana* at two stages:**

Very often, the gene expression data are most richly understood and most valuable when related to other types of information at the protein, DNA or functional level. In the next step, to further classify the genes into various biological categories, plant expression database and The Arabidopsis Information Resource databases were used. Each list of differentially expressed genes was analyzed in the background of gene ontology to identify gene expression data for various major biological processes and to assign up and down regulated genes to biological process categories in a systematic manner.This analysis will allow us to obtain an overview on the biological functions of the differentially expressed genes identified in four comparisons for up regulated genes of ARF and MYB at stage 12 and 13 likewise down regulated genes of ARF and MYB at stage 12 and 13. As fig. 8 and 10 describes distribution of all up regulated genes of ARF and MYB transcription factors at stage 12 and 13 into major biological processes and fig. 9 and 11 describes distribution of down regulated genes of ARF and MYB transcription at stage 12 and 13 into major biological processes. As shown in given figures, there are total four comparisons. Up and downregulated gene sets were grouped into the 19 most abundant categories. From following graphs, up and down regulated genes of ARF transcription factor of *Arabidopsis thaliana* for stage 12 and 13 were taken. There were total 515 and 203 up regulated genes and 993 and 301 down regulated genes for stage 12 and 13 identified respectively. ARF encodes a member of the auxin response factor family and mediates auxin response via expression of auxin regulated genes. Acts redundantly with ARF8 to regulate stamen elongation and flower maturation. From figure, high percentages of genes from the biological functions are response to stress, response to abiotic and biotic stimulus and transport and also genes are in high percentage in stage 13 which is a youngest open flower as compare to stage 12 which is a closed bud, this explains flower development is occurring fine with these biological processes.

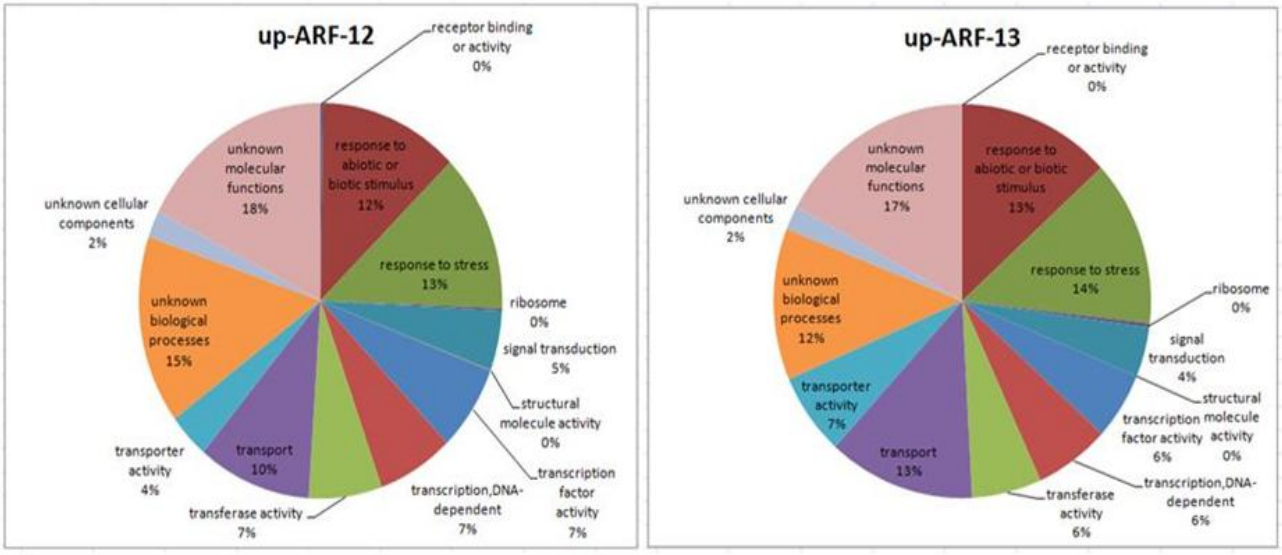


Fig 8: distribution of up regulated genes of ARF transcription factor of stage 12 and 13, into major biological processes. GO was used to classify the genes into functional category. Gene numbers are displayed next to the term.

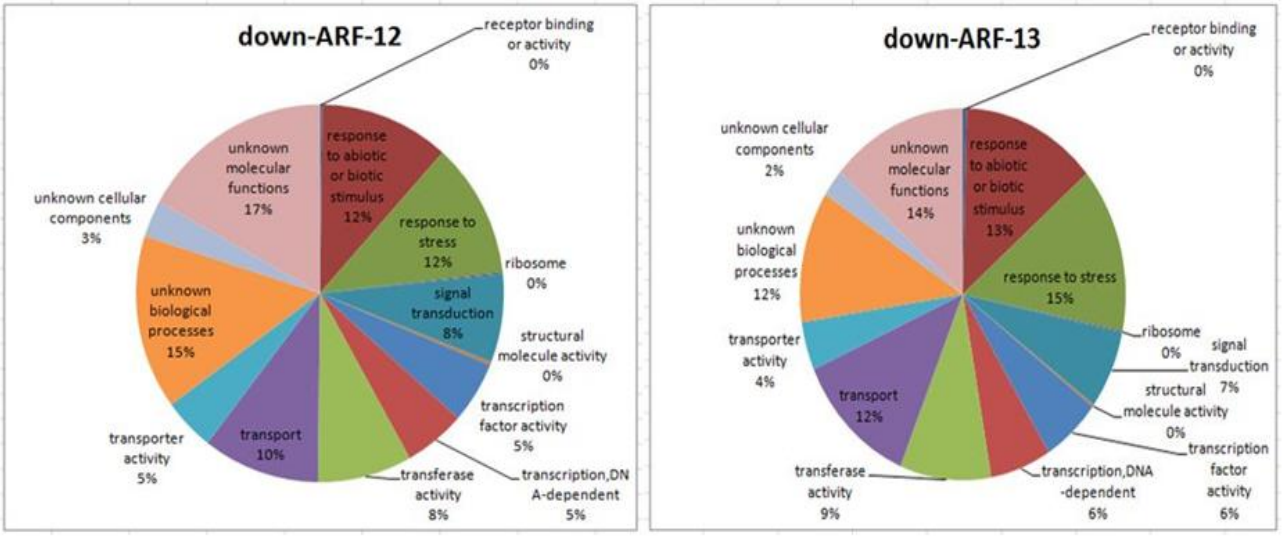


Fig 9: distribution of down regulated genes of ARF transcription factor of stage 12 and 13, into major biological processes**.**

 

MYB proteins are key factors in regulatory networks controlling development, metabolism and responses to biotic and abiotic stresses. From following graphs 10 and 11, up and down regulated genes of MYB transcription factor of *Arabidopsis thaliana* for stage 12 and 13 were taken. There were total 50 and 253 up regulated genes and 177 and 325 down regulated genes for stage 12 and 13 identified respectively. Here also, high percentage of gene were identified from the biological functions are response to stress, response to abiotic and biotic stimulus, transferase activity and transport.

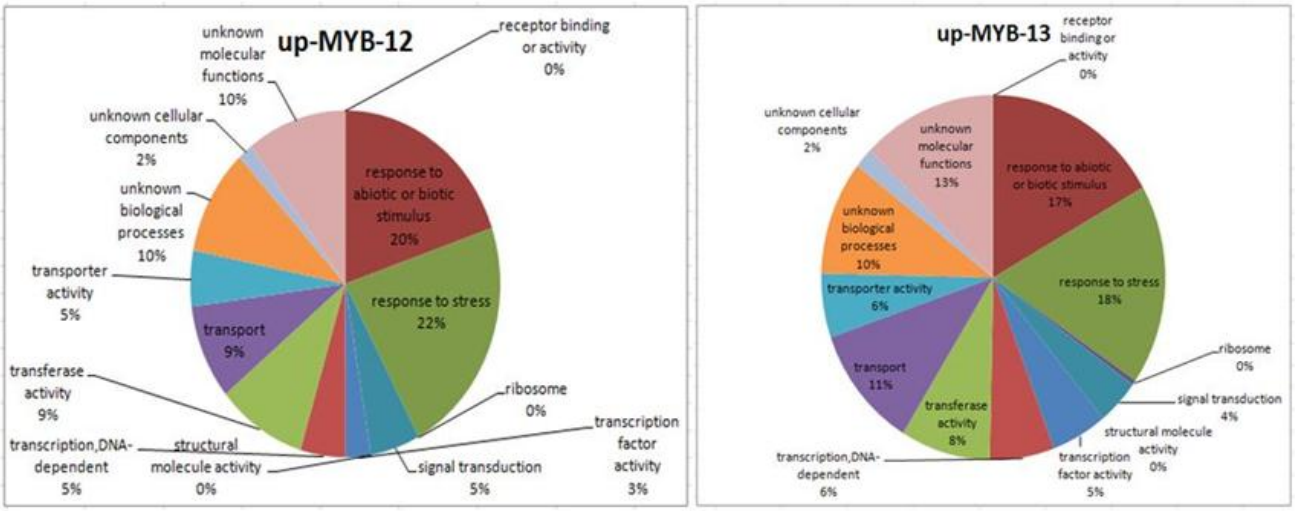


Fig 10: distribution of up regulated genes of MYB transcription factor of stage 12 and 13, into major biological processes.

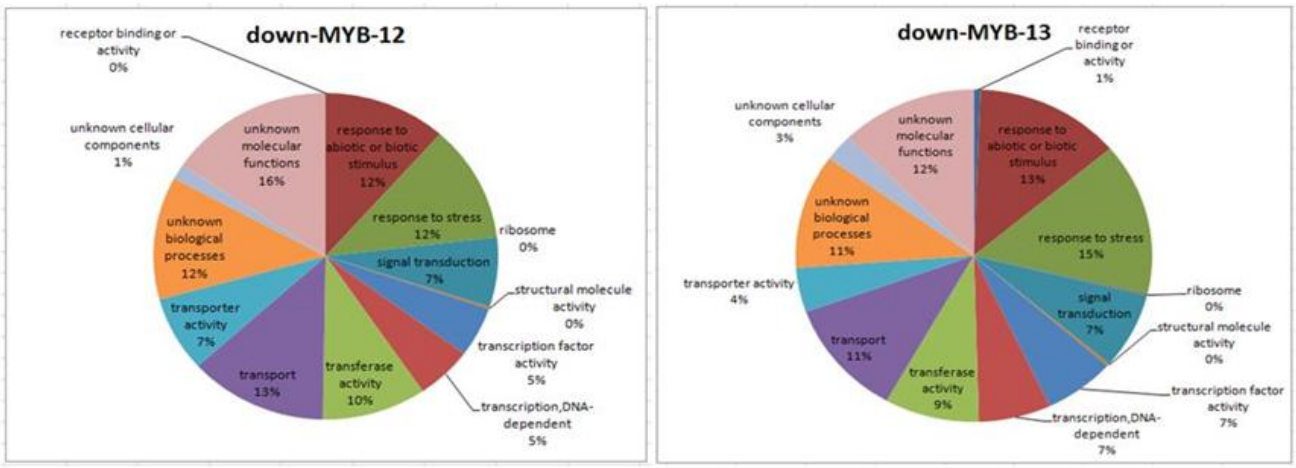


Fig 11: distribution of down regulated genes of MYB transcription factor of stage 12 and 13, into major biological processes.

Hence, from above method we have been successful to indentify major biological processes of transcription factors of *Arabidopsis thaliana.* Table 4, describes top five up and down regulated genes with their probe set, baseline mean, experimental mean, fold change, locus identifier and annotations, of ARF and MYB transcription factor for both stages. Table 4 describes the top five up – down regulated genes of ARF and MYB transcription factor of *A. thaliana*, from both stages, with related information like probe set id, baseline mean, experiment mean, fold change, locus identifier and its annotations.

Table 4 Top five genes

| **C-ARF stage 12 Down regulated** | | | | | |
|---|---|---|---|---|---|
| **probe set** | **baseline mean** | **experiment mean** | **fold change** | **Locus Identifier** | **Annotation** |
| 257220_at | 3974.79 | 10.32 | -385.07 | AT3G27810 | ATMYB21 (MYB DOMAIN PROTEIN 21); DNA binding / transcription factor |
| 249349_at | 1267.61 | 4.24 | -299.29 | AT5G40350 | MYB24 (myb domain protein 24); DNA binding / transcription factor |
| 258237_at | 1173.44 | 8.91 | -131.65 | AT3G278 | ATMYB21 (MYB DOMAIN |

 

| | | | | | |
|---|---|---|---|---|---|
| | | | | 10 | PROTEIN 21); DNA binding / transcription factor |
| 264430_at | 2895.67 | 25.86 | -111.99 | AT1G61680 | terpene synthase/cyclase family protein |
| 256636_at | 2101.09 | 23.97 | -87.65 | AT3G12000 | S-locus related protein SLR1, putative (S1) |
| **C-ARF stage 12 up regulated** | | | | | |
| **probe set** | **baseline mean** | **experiment mean** | **fold change** | **Locus Identifier** | **Annotation** |
| 245622_at | 4.1 | 265.57 | 64.78 | AT4G14080 | MEE48 (maternal effect embryo arrest 48); hydrolase, hydrolyzing O-glycosyl compounds |
| 262083_at | 4.56 | 273.08 | 59.84 | AT1G56100 | pectinesterase inhibitor domain-containing protein |
| 265342_at | 1.25 | 73.93 | 59.2 | AT2G18300 | basic helix-loop-helix (bHLH) family protein |
| 251772_at | 1.41 | 40.85 | 28.95 | AT3G55920 | peptidyl-prolyl cis-trans isomerase, putative / cyclophilin, putative / rotamase, putative |
| 247462_at | 18.99 | 527.02 | 27.76 | AT5G62080 | protease inhibitor/seed storage/lipid transfer protein (LTP) family protein |
| **C-MYB stage 12 Down regulated** | | | | | |
| **probe set** | **baseline mean** | **experiment mean** | **fold change** | **Locus Identifier** | **Annotation** |
| 253240_at | 1150.67 | 19.44 | -59.18 | AT4G34510 | KCS2 (3-ketoacyl-CoA synthase 2); acyltransferase |
| 251716_at | 460.5 | 9.55 | -48.22 | AT3G55870 | anthranilate synthase, alpha subunit, putative |
| 256833_at | 1230.73 | 28.01 | -43.93 | AT3G22910 | calcium-transporting ATPase, plasma membrane-type, putative / Ca(2+)-ATPase, putative (ACA13) |
| 264263_at | 304.82 | 9.5 | -32.08 | AT1G09155 | ATPP2-B15 (Phloem protein 2-B15); carbohydrate binding |
| 256521_at | 1325.55 | 41.39 | -32.03 | AT1G66120 | acyl-activating enzyme 11 (AAE11) |
| **C-MYB stage 12 up regulated** | | | | | |
| **probe set** | **baseline mean** | **experiment mean** | **fold change** | **Locus Identifier** | **Annotation** |
| 252063_at | 94.78 | 1004.07 | 10.59 | AT3G51590 | LTP12 (LIPID TRANSFER PROTEIN 12); lipid binding |
| 259719_at | 10.53 | 110.87 | 10.53 | AT1G61070 | LCR66/PDF2.4 (Low-molecular-weight cysteine-rich 66) |
| 260001_at | 37 | 372.14 | 10.06 | AT1G67990 | caffeoyl-CoA 3-O-methyltransferase, putative |
| 253772_at | 15.7 | 147.79 | 9.41 | AT4G28395 | ATA7 (Arabidopsis thaliana anther 7) |
| 267440_at | 17.34 | 132.66 | 7.65 | AT2G19070 | transferase family protein |
| **C-ARF stage 13 Down regulated** | | | | | |
| **probe set** | **baselin** | **exper** | **fold** | **Locus** | **Annotation** |

 

| | e mean | iment mean | change | Identifier | |
|---|---|---|---|---|---|
| 251988_at | 2621.95 | 13.09 | -200.3 | AT3G53300 | CYP71B31 (cytochrome P450, family 71, subfamily B, polypeptide 31); oxygen binding |
| 251716_at | 1615.61 | 9.28 | -174.19 | AT3G55870 | anthranilate synthase, alpha subunit, putative |
| 264263_at | 943.95 | 7.93 | -119.11 | AT1G09155 | ATPP2-B15 (Phloem protein 2-B15); carbohydrate binding |
| 265221_s_at | 519.21 | 5.38 | -96.44 | AT2G02000;AT2G02010 | [AT2G02000, GAD3 (GLUTAMATE DECARBOXYLASE 3); calmodulin binding];[AT2G02010, GAD4 (GLUTAMATE DECARBOXYLASE 4); calmodulin binding] |
| 256521_at | 2882.46 | 32.44 | -88.86 | AT1G66120 | acyl-activating enzyme 11 (AAE11) |
| **C-ARF stage 13 up regulated** | | | | | |
| **probe set** | **baseline mean** | **experiment mean** | **fold change** | **Locus Identifier** | **Annotation** |
| 261943_at | 13.93 | 823.02 | 59.09 | AT1G80660 | AHA9 (Arabidopsis H(+)-ATPase 9); hydrogen-exporting ATPase, phosphorylative mechanism |
| 261532_at | 13.94 | 712.22 | 51.08 | AT1G71680 | amino acid transmembrane transporter |
| 265280_at | 27.29 | 1196.8 | 43.85 | AT2G28355 | LCR5 (Low-molecular-weight cysteine-rich 5) |
| 265511_at | 59.33 | 2326.79 | 39.22 | AT2G05540 | glycine-rich protein |
| 264016_at | 18.22 | 623.66 | 34.24 | AT2G21220 | auxin-responsive protein, putative |
| **C-MYB stage 13 Down regulated** | | | | | |
| **probe set** | **baseline mean** | **experiment mean** | **fold change** | **Locus Identifier** | **Annotation** |
| 251716_at | 1615.61 | 8.02 | -201.47 | AT3G55870 | anthranilate synthase, alpha subunit, putative |
| 251988_at | 2621.95 | 13.42 | -195.37 | AT3G53300 | CYP71B31 (cytochrome P450, family 71, subfamily B, polypeptide 31); oxygen binding |
| 258290_at | 132.17 | 1.67 | -79.3 | AT3G23460 | cyclopropane fatty acid synthase-related |
| 256521_at | 2882.46 | 37.14 | -77.61 | AT1G66120 | acyl-activating enzyme 11 (AAE11) |
| 264263_at | 943.95 | 12.99 | -72.69 | AT1G09155 | ATPP2-B15 (Phloem protein 2-B15); carbohydrate binding |
| **C-MYB stage 13 up regulated** | | | | | |
| **probe set** | **baseline mean** | **experiment mean** | **fold change** | **Locus Identifier** | **Annotation** |
| 259286_at | 74.1 | 2717.91 | 36.68 | AT3G11480 | BSMT1; S-adenosylmethionine-dependent methyltransferase |
| 252611_at | 15.51 | 484.0 | 31.21 | AT3G451 | LAS1 (Lanosterol synthase 1); |

 

| | | | | | |
|---|---|---|---|---|---|
| | | 9 | | 30 | lanosterol synthase |
| 265511_at | 59.33 | 1692.79 | 28.53 | AT2G05540 | glycine-rich protein |
| 256021_at | 27.39 | 735.61 | 26.86 | AT1G58270 | ZW9 |
| 261532_at | 13.94 | 351.89 | 25.24 | AT1G71680 | amino acid transmembrane transporter |

Overall, this study provides a basic foundation for further analyses of functions analysis of particular genes or pathways analysis for mentioned transcription factors. So that we can expose many more information regarding transcription factors of *A. thaliana*.

## IV. CONCLUDING REMARKS

The advent of technologies for expression profiling of multiple genes has launched a new era of biological research. Microarrays are better suited to analysis of many genes (tens of thousands) in fewer biological specimens. Given these characteristics microarrays have more often been applied in the discovery phase of biological research with the aim of identifying the most informative genes, functions etc. In this present study, we have obtainable the methods for analyzing microarray data with the help of microarray data analyzer (dChip) and different databases. Using microarrays is a powerful technique to monitor the expression of thousands of genes, and a key technique for biologists attempting to unravel the regulation mechanisms of genes in a system. We have studied the gene expression profiles of transcription factors of Arabidopsis thaliana at the two specific flower maturation stages (stage 12 and stage 13). The stage 12 is the oldest closed bud and stage 13 is the youngest open flowers, which were considered to study the role of ARF and MYB transcription factors. This study is based on the microarray data analysis and functional analysis for the characterization of gene expressions in flowers before and after they open, and to determine major biological processes of ARF6 ARF8 and MYB21 MYB24 transcription factors. Our data analyses showed, total 27199 genes of transcription factors of Arabidopsis thaliana. Using the criteria of fold change ($FC \geq 2.0$ and $FC \leq -2.0$), we identified 515 and 203 upregulated and 493 and 302 downregulated genes of ARF transcription factor of stage 12 and 13 respectively. Likewise for MYB transcription factor, we identified 50 and 253 upregulated and 177 and 325 downregulated genes for stage 12 and 13, rest of many genes shows none significant expression of these two specific flower maturation stages of Arabidopsis thaliana. There were only 206 and 65 upregulated genes of ARF and MYB and 334 and 156 downregulated genes of ARF and MYB that were overlapped among both stages (12 and 13) respectively. Along with we have also found major biological processes like response to abiotic and biotic stimulus, transport, transferase activity etc, of the differentially expressed genes identified in four comparisons of up and down regulated genes of ARF and MYB at for both stage 12 and 13.

Many have said that this century is the century for bioinformatics. No matter what this means to each of us, we can clearly see that many biological data have been generated and many biological facts are known, yet general principles are still lacking. As a bioinformatician with an interest in biotechnology and computer science, I feel being blessed because I can now taste the great biological fruits (i.e., analyzing their data) without having to sweat to "grow" them by myself! I feel that our field is also blessed by the high throughput biological data generation-bioinformatics has never been in so much demand from biologists. But it is also a challenge to all bioinformaticians. Indeed, if we bioinformatician do not proactively participate in the biotechnology revolution, other scientists (e.g., computer scientists, biostatistician) will learn and apply it whether we approve it or not. We clearly have an advantage, for now, and we still can control our own fate if we try.

## ACKNOWLEDGEMENT

Author is very grateful to Jacob School of Biotechnology and Bioengineering, Department of Computational Biology and Bioinformatics, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Deemed University, Allahabad and University Grant Commission Delhi, Government of India.

## REFERENCES

[1] Hall , D.A. , Zhu , H. , Zhu , X. , Royce , T. , Gerstein , M. and Snyder , M. (2004), Regulation of gene expression by a metabolic enzyme. Science, 306:482 – 484.

[2] Riechmann , J.L. , Krizek , B.A. and Meyerowitz , E.M. ( 1996 ), Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Proc. Natl Acad. Sci. USA, 93:4793 – 4798.

[3] Deyholos , M.K. and Sieburth , L.E. (2000), Separable whorl-specific expression and negative regulation by enhancer elements within the AGAMOUS second intron, Plant Cell, 12: 1799 – 1810.

[4] Guo , A. , He K. , Liu , D. , Bai , S. , Gu X. , Wei , L. , et al . (2005), DATF: a database of *Arabidopsis* transcription factors. *Bioinformatics* 21: 2568 – 256 9.

[5] Iida , K. , Seki , M. , Sakurai , T. , Satou , M. , Akiyama , K. , Toyoda , T. , et al . (2005), RARTF: database and tools for complete sets of *Arabidopsis* transcription factors, DNA Res. 12: 247 – 256.

[6] Riano-Pachon , D.M., Ruzicic, S. , Dreyer , I. and Mueller-Roeber , B. (2007), PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics, 8: 42.

 

[7] Smyth DR, Bowman JL, Meyerowitz EM., (1990), Early flower development in Arabidopsis. Plant Cell; 1:37 52.

[8] Paul H. Reeves, Christine M. Ellis, Sara E. Ploense, Miin-Feng Wu,Vandana Yadav, Dorothea Tholl, Aurore Chételat, Ina Haupt, Brian J. Kennerley, Charles Hodgens, Edward E. Farmer, Punita Nagpal, and Jason W. Reed, (2012), PLoS Genetics; 8(2): e1002506.

[9] Schena M, Shalon D, Davis RW, Brown PO. (1995), Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science New York, 20; 270(5235):467-70.

[10] Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. (1996) , Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnology, 14(13):1675-80.

[11] S. Selvaraj, J. Natarajan, (2011) Microarray Data Analysis and Mining Tools, Bioinformation , 6(3): 95-99.

[12] Musa H. Asyali, Dilek Colak, Omer Demirkaya and Mehmet S. Inan, (2006) Gene Expression Profile Classification: A Review, Current Bioinformatics, 1, 55-73.

[13] Shena M, Shalon D, Davis RW, Brown PO, (1995), Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science; 270: 467-70.

[14] Ideker,T., Thorsson, V., Siegel, A.F., Hood, L.E., (2000), Testing for differentially-expressed genes by maximum likelihood analysis of microarray data, Journal of Computational Biology, (7)805-817.

[15] Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression, Science 270, 484-487.

[16] Nagpal P, Ellis CM, Weber H, Ploense SE, Barkawi LS, et al. (2005), Auxin Response Factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. Development, 132:4107–4118.

[17] Wu M-F, Tian Q, Reed JW. (2006), *Arabidopsis microRNA167* controls patterns of *ARF6* and *ARF8* expression and regulates both female and male reproduction. Development; 133:4211–421.

[18] Narlikar L, Hartemink AJ. 2006, Sequence features of DNA binding sites reveal structural class of associated transcription factor. Bioinformatics, 22:157–63.

[19] Alvord, WG, Roayaei JA, Quinones OA, Schneider KT. 2007, A microarray analysis for differential gene expression in the soybean genome using Bioconductor and R. Brief Bioinform.;8:415–31.

[20] Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun S, Kulp, D, Siani-Rose MA. 2003, NetAffx: Affymetrix probesets and annotations. Nucleic Acids Res.; 31:82–6.

[21] Tan AC, Naiman DQ, Xu L, Winslow RL, 2005. Geman D. Simple decision rules for classifying human cancers from gene expression profiles, Bioinformatics. 21:3896–904.

[22] Eisen MB, Brown PO., 1999, DNA arrays for analysis of gene expression, Methods Enzymology, 303:179–205.

[23] Bolstad BCF, Brettschneider J. 2005m Quality assessment of Affymetrix GeneChip Data. New York: Springer.

[24] Liu XS. 2007, Getting started in tiling microarray analysis, PLoS Computational Biology; 3:1842–4.

[25] Parkinson , H. , Kapushesky , M. , Kolesnikov , N. , Rustici , G. , Shojatalab , M. , Abeygunawardena , N. , et al . (2009) ArrayExpress update — from an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Res. 37: D868 – D872.

[26] Shippy , R. , Fulmer-Smentek , S. , Jensen , R.V. , Jones , W.D. , Wolber , P.K. , Johnson , C.D. , et al . (2006) Using RNA sample titrations to assess microarray platform performance and normalization techniques. Nat. Biotechnology. 2: 1123–1131.

[27] Szemenyei, H., Hannon, M. and Long, J.A. (2008) TOPLESS mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. Science 319: 1384 – 1386.

[28] Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W., (2004), GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox, Plant Physiology, 136: 2621 – 26 32.

[29] Iida , K. , Seki , M. , Sakurai , T. , Satou , M. , Akiyama , K. , Toyoda , T. , et al . (2005) RARTF: database and tools for complete sets of *Arabidopsis* transcription factors. DNA Res. 12: 247 – 256.

[30] Ashburner, M. , Ball , C.A. , Blake , J.A. , Botstein , D. , Butler , H. , Cherry , J.M. , et al . (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.* 25: 25 – 29.

 