# Accepted Manuscript

Killing Two Birds with One Stone: Model Plant Systems as a Tool to Teach the Fundamental Concepts of Gene Expression While Analyzing Biological Data

Irina Makarevitch, Betsy Martinez-Vaz

Please cite this article as: Irina Makarevitch, Betsy Martinez-Vaz, Killing Two Birds with One Stone: Model Plant Systems as a Tool to Teach the Fundamental Concepts of Gene Expression While Analyzing Biological Data, *BBA - Gene Regulatory Mechanisms* (2016), doi: 10.1016/j.bbagrm.2016.04.012

**Killing Two Birds with One Stone: Model Plant Systems as a Tool to Teach the Fundamental Concepts of Gene Expression While Analyzing Biological Data**

Irina Makarevitch[1*] and Betsy Martinez-Vaz[1]

[1]Department of Biology, Hamline University, Saint Paul, MN 55104

[*]corresponding author (e-mail: imakarevitch01@hamline.edu)

*Highlights*

• Incorporating gene expression concepts in undergraduate classroom allows for engaging students in authentic research

• Plants are an effective and convenient model to engage students in learning about gene expression

• Integrating bioinformatics approaches, data mining and quantitative data analysis with wet lab modules in plant models lead to successful learning of gene expression concepts

• Various resources for teaching gene expression concepts using plant systems are developed and could be utilized by the undergraduate instructors

*Keywords*

Regulation of gene expression, undergraduate classroom, student research experiences

*Abstract*

Plants are ideal systems to teach core biology concepts due to their unique physiological and developmental features. Advances in DNA sequencing technology and genomics have allowed scientists to generate genome sequences and transcriptomics data for numerous model plant species. This information is publicly available and presents a valuable tool to introduce undergraduate students to the fundamental concepts of gene expression in the context of modern quantitative biology and bioinformatics. Modern biology classrooms must provide authentic research experiences to allow developing core competencies such as scientific inquiry, critical interpretation of experimental results, and quantitative analyses of large dataset using computational approaches. Recent educational research has shown that undergraduate students struggle when connecting gene expression concepts to classic genetics, phenotypic analyses, and overall flow of biological information in living organisms, suggesting that novel approaches are necessary to enhance learning of gene expression and regulation. This review describes different strategies and resources available to instructors willing to incorporate authentic research experiences, genomic tools, and bioinformatics analyses when teaching transcriptional regulation and gene expression in undergraduate courses. A variety of laboratory exercises and pedagogy materials developed to teach gene expression using plants are discussed.

*1 Introduction*

The National Science Foundation (NSF) report, *Vision and Change: A Call to Action*, included flow, exchange, and storage of biological information in the list of the five core concepts required for biological literacy [1]. Gene transcription and its regulation in response to various internal and external stimuli are recognized as one of the key principles of the "information flow" concept [2,3] and questions aimed at testing students' understanding of gene expression are consistently included in biology assessment instruments [4-6]. Despite the widespread emphasis on teaching gene expression and regulation, students of different levels continue to struggle when presented with questions on gene expression. Twenty to forty percent of the students, in some cases even over 50% of the students, failed to answer these questions correctly when assessed using detailed open-ended and interview-based evaluations [7] or in multiple choice question format [4,5]. Many students struggle to connect classical genetics and outcomes of the crosses with the framework of gene transcription and regulation and frequently fail to see how misregulation of gene expression could result in phenotypic outcomes. The assessment of student learning of the gene expression concepts underscores the need to develop novel nontraditional approaches to teaching and learning the fundamental aspects of gene regulation. Approaches highlighting the connections between classical and molecular genetics and emphasizing regulation of gene expression in response to environment or during development are ideal to enhance students' learning of genetic switches and transcriptional regulation.

One of such approaches involves engaging students in research experiences focusing on investigating various aspects of gene expression (Figure 1). Several laboratory modules and tutorials integrating research experiences have shown their effectiveness in teaching the concepts

of gene expression and regulation, while simultaneously increasing students' skills in core competencies such as data interpretation, quantitative skills, oral and written communication, experimental design, interdisciplinary nature of modern biology, and student experience with large datasets [8-12]. In efforts to improve science education, instructors frequently turn to course-based inquiry and research experiences in the undergraduate classrooms [1, 13,14]. Benefits from early and consistently engaging undergraduates in research were demonstrated by many studies [15-17]. Students participating in research through research internships or course-based undergraduate research experiences (CUREs) report cognitive gains such as the development of knowledge and skills [18]; affective gains, such as satisfaction with their research experience [19]; and psychosocial gains, such as feeling like a scientist and finding research exciting [17, 19, 20]. CUREs involve students in addressing research problems or questions in the context of a class (Figure 1; reviewed in [15]), are available to students early in their undergraduate careers, and were shown to influence student's decisions in choosing a science-related major or considering applying to graduate programs [17, 21, 22]. In addition, CUREs aimed at non-science majoring students and freshmen have the potential to involve students who might not otherwise have access to science research opportunities [22, 23]. Several national projects implementing course-based research experiences in the undergraduate curriculum reported promising results in engaging students with high-impact learning experiences in biology research [16, 24-27].

The recent prevalence of big data in various fields of science, including biology, demands integration of data analysis and quantitative reasoning into the skill set required of science graduates [1, 13, 28-30]. "Data scientists" have been described as holding the "sexiest job of the

21st century" [31].  Unfortunately, many biology graduates lack sufficient skills in data analysis

and visualization, mathematical and quantitative reasoning, and cross-disciplinary approaches

required to solve biological problems [30, 32].  Biology students lack mathematical and

computational skills necessary for data analysis and perceive mathematics as irrelevant to their

field [33].  As plant researchers develop guidelines for complex and integrated statistical analysis

of large data sets [34], developing opportunities to directly involve undergraduate students in

quantitative data analysis, especially in analysis of large data sets, is recognized as an important

and pertinent problem by the community pursuing development of curricular materials that

infuse computational and mathematical training into biology courses [30, 35-38].

In this review we will discuss resources available to instructors who would like to engage their

students in analysis of gene expression data while learning the concepts of gene expression and

regulation, with an emphasis on plant systems (Figure 2; Table 1).  We will also review

pedagogical materials developed to incorporate these resources into the undergraduate

curriculum in the light of providing students with inquiry and research-based experiences and

emphasizing quantitative reasoning, data analysis, and communication skills.

*2 Green Classroom: Teaching Biology Using Model Plants Organisms*

Plants are frequently neglected in teaching biological sciences [39].  Many students prefer

studying animals [40], instructors and textbooks frequently use animal examples demonstrating

important biological concepts, plant-focused graduate programs have harder time recruiting

strong students.  The proposed causes for this observed lack of interest for plants termed "plant

blindness" [41] range from the convenience and ease of manipulation of microbial and animal

models and zoocentric life science curriculum [42] to the fundamental differences in how the human visual system processes plants [43, 44]. Nevertheless, plants offer a variety of developmental and physiological traits that make them unique when compared to their bacterial and animal counterparts. The unique nature of plants, opportunities they provide for teaching biology, and plant biology and plant behavior resources available to educators were recently featured in CBE-Life Science Education [45, 46]. Several projects implementing plant-centered curriculum infused with research opportunities and engaging learning activities reported a change in student perception of plants, highlighting the need for development of novel approaches to integrate plant sciences into undergraduate curriculum [47 - 49]. Many model plant species could be easily and inexpensively grown, manipulated and stored (as seeds) making them ideally suited as experimental systems for large-scale undergraduate teaching laboratories [46]. Other benefits of using plant systems when teaching undergraduate laboratory courses include very limited space required for propagation of many plants including *Arabidopsis thaliana,* a rapid life cycle from seed to seed for many plant models under defined growth conditions, highly predictable patterns of growth and differentiation, accessible systems for genetic transformations with "reporter" genes [50, 51], and the possibility to avoid ethical and safety constraints [46]. Plant systems have been used successfully in the classroom to illustrate a variety of biological concepts in various disciplines and to engage students in research opportunities [50, 52 - 54].

*2.1 In- Silico Gene Expression:  Databases to Learn About Transcriptional Regulation and Promoter Structure*

Transcriptional regulation and promoter structure are two of the core concepts covered when teaching the fundamental principles of gene expression (Figure 2). In order to understand that gene expression changes in response to environmental and physiological cues, students must be familiar with promoter structure, transcriptional units, and the role of transcription factors in the regulation of gene expression. While many students easily memorize the definition and function of a promoter, they often struggle when linking the role of these genetic elements to gene regulation and cellular phenotypes [55]. Moreover, students are aware that gene expression changes in response to environmental conditions but often forget that transcription factors interact with cellular metabolites and DNA sequences to mediate gene regulation in response to specific physiological and environmental stimuli [56].

The availability of genome sequences of model organisms such as *Arabidopsis thaliana, Medicago truncatula,* and *Escherichia coli* have led to the development of databases and tools to visualize promoters, learn about transcription factors, and identify motifs associated with the binding of these proteins to DNA (Table 1). The Arabidopsis Information Resource (*TAIR)* contains a series of plant promoter and regulatory elements resources that can be incorporated into undergraduate laboratory exercises to help students learn about transcription factors, small RNAs and their role in gene expression. For example, TAIR's AthaMap creates a map of potential transcription factors binding sites and small RNA targets in the genome of *Arabidopsis thaliana.* This tool also allows users to search a genome region of interest for gene targets of approximately 115 different transcription factors that have been characterized in this model plant organism [57]. Resources such as the "Plant Transcription Factor Database (PlnTFDB)" provide complete sets of transcription factors (TFs) and other transcriptional regulators (TRs) in plant

species whose genomes have been completely sequenced and annotated (Table 1; [58]). This tool also links different species to each other by using orthologous genes thereby facilitating sequence comparisons and the study of evolutionary relationships.

Many promoter databases and prediction tools have been developed and successfully used for plant genome research [59 - 62]. The PlantProm database contains an annotated, non-redundant collection of proximal promoter sequences for RNA polymerase II with experimentally determined transcription start sites [61]. Identification of functional promoter elements in plant genome sequences can be accomplished with Regsite DB. This tool allows recognition of plant regulatory motifs in a given sequence (Nsite-PL) as well as promoters conserved in several sequences (NsiteM-PL) [62].

The information and analysis tools contained in transcription factor and promoter databases offer an easy and low cost alternative to incorporate bioinformatics in the study of gene expression and regulation in undergraduate courses. Currently, most laboratory exercises designed to combine these concepts are based on model bacterial systems [8, 55, 56]. For example, Martínez-Vaz developed a series of inquiry based laboratory modules to study gene regulation and promoter structure in *Escherichia coli* using the publicly available databases, Regulon DB and Ecocyc site [55, 63, 64]. As part of these exercises, students examined the structure of promoters associated with genes involved in amino acid metabolism and identified the binding sites of the transcription factors known to regulate these genes. The data gathered was used to make predictions and formulate hypotheses about the environmental and physiological stimuli that triggered the activity and regulation of a given gene. The students concluded their projects

by designing and carrying out assays to test their hypotheses and investigate gene expression

under several experimental conditions [56]. Another set of laboratory activities that effectively

incorporated authentic undergraduate research on promoter structure and gene expression were

developed by Campbell and Eckdahl [8, 55]. During this multi-week project, the participants

searched the primary literature to choose and clone short promoters and investigate their activity

using fluorescent reporters. These exercises utilized the pClone Red and pClone Green vectors to

facilitate cloning of bacterial promoters and functional activity assays using fluorescent proteins

such as GFP (green fluorescent protein) and RFP (red fluorescent protein). Following control

experiments, the students created promoter mutants and formulated hypotheses about their effect

on cellular phenotypes and promoter activity [8, 55].  Both laboratory modules showed

substantial gains in students' learning of the core concepts of promoter structure and gene

regulation.  The activities presented in these inquiry based laboratory exercises can be easily

adapted to plant systems by using the resources and information provided by PlnTFDB, TAIR,

PlantProm, Regsite and Athamap amongst others (Table 1).


*2.2 Lighting Up the Room:  Measuring and Quantifying Gene Expression With Reporter*

*Genes*

Reporter constructs are one of the most convenient and powerful tools employed to detect and

measure gene expression. These genes encode proteins products that can be easily visualized and

quantified [65].  Traditionally, reporter genes such a β-galactosidase (β-gal) and β-glucuronidase

(GUS) have been used to introduce biology students to the fundamental concepts of gene

expression and regulation.  While microbial systems continue to be one of the main choices

when teaching about gene expression in the undergraduate classroom, several educators have

developed exercises that utilize model plant organisms to learn about transcriptional regulation. Gelvin and Karcher pioneered the use of reporter genes and transgenic plants to teach about gene expression in response to environmental signals in *Arabidopsis thaliana* [50, 51]. In these series of laboratory exercises, the authors utilized transgenic plants containing several promoters involved in plant development and stress responses fused to the gene encoding the enzyme glucuronidase (GUS). Using this approach, students worked with transgenic plants containing *gusA* fused to the promoter of GH3, a soybean gene that encodes a small auxin up RNA (SAUR) [50]. This reporter construct provided an ideal way to show students how gene expression changes during plant development in response to auxins. Another set of experiments in these series consisted of investigating plant responses to stress using a *cor15a-gusA* reporter. The *cor15a* gene encodes a polypeptide that confers tolerance to freezing and other stresses in *Arabidopsis thaliana.* The availability of *cor15a-gusA* transcriptional reporters allowed students to formulate hypotheses and design experiments to observe and quantify changes in *cor15a* expression in plant tissues upon exposure to various conditions such as temperature, drought and high salt [51]. Currently, GUS reporter constructs continue to be a popular choice to study and quantify transcriptional regulation in plant systems. Bargmann and collaborators recently described a laboratory exercise based on the use of GUS reporter fusions to study the regulation of genes involved in lateral root formation in *Arabidopsis thaliana* [66]. As part of this module, students designed experiments to follow the expression of LBD (lateral organ boundary domain containing protein) genes in response to auxin at different stages of plant development. The authors reported learning gains in students' ability to interpret current scientific knowledge of the factors that control lateral root initiation in plants as well as in determining the effect of auxin concentrations on gene expression by using reporter gene assays [66]. Since GUS reporters have

been successfully incorporated in the undergraduate classroom, the Arabidopsis Biological

Resource Center offers a variety of educational kits for laboratory exercises on gene expression

using these constructs (http://abrcoutreach.osu.edu/educational-kits; Table 1). These kits allow

the study of promoter activity and differential gene expression in plant tissues during growth and

development.

When studying plants, the use of fluorescent proteins in laboratory modules designed to measure

gene expression is less common compared to bacterial and animal systems. Several factors limit

the use of fluorescent reporters in plants. For example, some fluorescent proteins are not suitable

due to weak fluorescence signals and high levels background arising from leaf and stem tissues.

Moreover, access to fluorometers, fluorescence microscopes and digital equipment to quantify

fluorescence images might be limited at some educational institutions. Baker addressed these

limitations by developing an epifluorescent attachment to improve imaging and quantification of

the red-shifted green fluorescent protein in Arabidopsis tissues [67]. This technology offers an

alternative for low cost imaging and quantification of fluorescence with potential applications for

the study of transcriptional regulation in plant tissues using GFP and similar proteins. Despite

the potential limitations of using fluorescent reporters to study gene expression in plants, the

Arabidopsis Biological Resource Center have several educational kits suitable for investigating

gene regulation in various seed lines and phyto tissues using the green fluorescent protein (Table

1).

Experiments involving the use of reporter genes were one of the first activities that incorporated

quantitative analysis in undergraduate biology courses. When measuring gene expression using

β-galactosidase (β-gal) and β-glucuronidase (GUS) reporters, students learn how to calculate

Miller units and select the correct controls to normalize their experimental data. These

calculations often involve replicas and are suitable to introduce basic statistics such as average,

standard deviations and significance analyses using t-tests. Regarding fluorescent reporters, gene

expression measurements involving these constructs often involve correction for background

fluorescence and normalization of signals to account for differences in growth rate and cell

counts [68]. These calculations are ideal tools to incorporate quantitative analysis and selection

of proper experimental controls in undergraduate laboratory exercises. This approach showed

significant students' learning gains in a module designed to study transcriptional regulation in

bacteria using GFP-promoter fusions [56]. The study demonstrated that students were able to

select appropriate controls and effectively calculate differences in gene expression after

background subtraction and normalization of fluorescence data. The results of this activity

suggest that similar exercises can be developed to use fluorescent- protein gene reporter in

undergraduate courses to investigate transcriptional regulation in plants.


### 2.3 An Old Tale with a New Twist: RT-qPCR as a Tool to Teach Gene Expression and Regulation

The polymerase chain reaction (PCR) is one of the fundamental techniques of modern molecular

biology.  Students are introduced to PCR and its applications early during high school and

undergraduate biology courses.  Investigative laboratory exercises involving PCR to detect

genes, mutations and to carry out genotyping experiments are commonly used to teach basic

genetics and molecular biology [69, 70].  The development of real time PCR instruments has

provided a tool to quantify allelic DNA sequences and to measure the levels of specific mRNA

transcripts under different cellular conditions.  When combined with reverse transcriptase, quantitative real time PCR (RT-qPCR) is one of the most powerful techniques to monitor and quantify gene expression in living organisms.  Modern biology is increasingly focused on genome analyses and gene expression profiling.  Therefore, it is imperative that undergraduate students are exposed to the theory and applications of real-time PCR as part of their coursework.

Despite the scientific relevance and widespread use of real-time PCR in clinical, academic and industrial settings, efforts to develop teaching activities involving this technique have been limited by the cost of reagents and accessibility to instrumentation [71, 72].  Real time PCR machines are mostly available in research laboratories at large universities and many primarily undergraduate colleges do not have access to these instruments.  While the cost of real time PCR reagents is affordable to many researchers, purchasing real time PCR supplies for an undergraduate class often means spending more than half or the entire course budget in one experiment.  Due to these limitations, published reports on teaching activities and laboratory exercises involving real time PCR are scarce.  A recent literature search performed using the keywords "real-time PCR", "undergraduate" and "gene expression" returned only a handful of studies describing the implementation of this approach to teaching transcriptional regulation concepts in the undergraduate courses [73-79].  Hancock and collaborators recognized the need of incorporating real time-qPCR when teaching transcriptional regulation in undergraduate biology courses [76].  As part of a series of laboratory exercises, students measured changes in the expression of three genes (β-globin, amino levulinate synthase, and carbonic anhydrase-1) during the development of murine erythroleukemia cells.  These activities effectively introduced undergraduates to the fundamentals of the qPCR technique, as well as data analysis and

interpretation through $\triangle\triangle$Ct method [76]. Following Hancock's work, several educators developed similar activities to study transcriptional regulation in different tissue culture models and viral systems using RT-qPCR [73 – 77, 79]. For example, a human osteoclast differentiation model was to introduce molecular biology and biotechnology students to the fundamental concepts of transcriptional regulation using RT-qPCR, bioinformatics analyses, and RNAi [78]. The project consisted of analyzing expression-profiling data to identify genes that changed significantly during osteoclast differentiation. These analyses were followed by gene knockouts using RNAi and RT-qPCR to confirm the levels of mRNA in mutant and wild type cell lines. Surveys administered after the completion of these laboratory activities showed positive feedback of the project and a high level of satisfaction amongst the participants [78]. More recently, Hardagon designed and implemented a multi-week laboratory project to study the levels of expression an immunosuppressive cytokine (*Tgfb1*) in two murine carcinoma cell lines with different degrees of tumorigenesis [73]. The students participating in this activity performed traditional and quantitative RT-PCR combined with flow cytometry and in situ hybridization to investigate the level of expression of the *Tgfb1* gene in several cancer cell lines. The assessment results obtained during and after the completion of the project revealed that the laboratory exercises enhanced students' knowledge of the basic concepts of transcriptional regulation, data analysis skills and methods to investigate gene expression [73].

Currently, only one educational activity to investigate transcriptional regulation in plants using RT-qPCR has been developed and published [77]. This exercise, designed by Eickelberg, investigated the regulation of plant genes in response to environmental stimuli [77]. The students examined the effects of environmental conditions on the expression of the *FLOWERING LOCUS*

*C* gene, a transcriptional repressor that controls floral transitions in *Arabidopsis thaliana.* After completing this multi-week project, students were able to demonstrate their understanding of RT-qPCR by creating a laboratory report containing: (1) tables with Ct-values and fold changes in gene expression, (2) melting curves and (3) representative amplification plots. Pre and post laboratory exercises supported students' understanding of the theory and applications of RT-qPCR to study gene expression in plants [77]. This laboratory activity provides a base for the development of open-ended laboratory exercises in which students can design their experiments to study transcriptional regulation in model plant species.

The availability of resources to teach biology using model plant species will facilitate the development of additional laboratory exercises involving RT-qPCR to investigate transcriptional regulation. The Arabidopsis Biological Resource Center has several educational kits that can be adapted for the study of gene expression by RT-qPCR. For example, the "Expression Analysis of Light-Regulated Genes in the Det1-1 Mutant Kit" is an educational kit created to study plant gene expression in response to light. Educators with accessibility to a RT-qPCR instrument can easily adapt this module to introduce students to gene expression analysis by q RT-PCR. Similarly, the following kits: "Same Genes, Different Fates," "Who Turn the Lights Off?" and "GFP Plant Anatomy Set" were constructed to study differential gene expression in *Arabidopsis thaliana* using reporter constructs and mutant cell lines. The resources contained within these kits provide ideas and starting materials that undergraduate educators can modify to implement a plant - focused RT-qPCR unit in introductory and upper level biology courses.

In summary, although implementing real-time PCR as a standard teaching tool might still be a challenge at some institutions, it is worth noting that biotechnology companies are developing affordable real time PCR products for educational purposes. Hopefully, the development of more commercial kits and educational resources will allow RT-qPCR technology to be used in every undergraduate classroom. Employing real time PCR as a teaching tool will enable life science educators to introduce and further enhance key concepts in transcriptional regulation as well as engage students in research.

*2.4 Big Data in Biology: Integrating RNA-Seq Analysis into Undergraduate Classroom*

Easy access to genome information resulted not only in a quantitative, but also in a qualitative shift in biological research, fundamentally changing approaches not only to research in, but also to teaching of biology [80, 81]. A decade ago, most hypotheses about molecular genetic variation were limited by the cost and difficulty of obtaining large amounts of sequence data. Now, next generation sequencing (NGS) presents a virtually data unlimited platform in which hypotheses are often derived from the sequence data itself [82]. Students entering the world of biological graduate studies more and more are expected to be familiar with next generation sequencing and approaches to NGS data analysis. Therefore, integrating analysis of newly generated and publicly available next generation sequencing data as research experiences for undergraduate students provides great opportunities for training future scientists.

RNA-Seq, a technique used to quantify the amount of RNA transcribed from each gene of the organism via massively parallel next generation sequencing technology, has become a method of choice for researchers investigating gene expression changes in a variety of biological systems,

including plants (for example, [83-86]).  The availability of nearly unlimited RNA-Seq data

through the Short Read Archive (SRA) of NCBI [87] and access to powerful bioinformatics

analyses from shared servers offer students the opportunity to develop into scientists while

enrolled in undergraduate biology courses [81].  Despite these opportunities, integration of RNA-

Seq analysis into the undergraduate curriculum is complicated by the steep learning curve the

educators themselves have to encounter.  In addition, the RNA-Seq analysis relies on running

several software packages on large files, complicating logistics of giving students access to the

files and analysis tools and requiring extensive support provided by the instructor to students

along the way.


Several resources have been developed to ease in the way into RNA-Seq data analysis for the

students (Table 1).  Although some of the materials developed for integrating RNA-Seq analysis

into the classroom are not plant-specific, they could be used for the analysis of student-derived or

publicly available data by students interested in a variety of biological questions in different

model systems, including plants.  The Genome Consortium for Active Teaching using Next-

Generation Sequencing (GCAT-SEEK) as well as the DNA Learning Center offer workshops

that aim at helping educators gain necessary experience in the RNA-Seq analysis and develop

educational tools for the students ([88, 89], http://www.rnaseqforthenextgeneration.org/).  The

GCAT-SEEK network of faculty continues to design educational resources, including tutorials

and sample data sets for use in undergraduate classrooms [11, 89].  The Faculty Group led by the

DNA Learning Center (http://www.rnaseqforthenextgeneration.org/) develops a series of

curricular materials aimed at different audiences and utilizing the Green Line of the DNA

Subway as their platform (http://dnasubway.iplantcollaborative.org/).  Standalone series of

instructional materials intended to integrate RNA-Seq into the classroom have been implemented and assessed [10].  All of these approaches use various software platforms to streamline the data analysis and navigate students through the software packages employed in the analysis.

### 2.4.1 "Point and Click": Green Line of the DNA Subway

Green Line of the DNA subway built on the platform of the iPLANT collaborative [90] utilizes a standard research-grade TUXEDO protocol for RNA-Seq data analysis [91] implemented in a series of steps - from sequence quality control to identification of differentially expressed genes and building gene expression plots.  The software packages are run "under closed doors" using mostly default parameters with a few modifications available to users.  Data storage is integrated with iPlant accounts, so the users are able to seamlessly store the data, assess the quality of the data, and run the data analysis from the original RNA-Seq sequence read files to a list of differentially expressed genes.  Although, the users are limited to the list of pre-loaded genomes, over 30 genomes from a variety of organisms, including such plant species as maize, *Brassica rapa*, *Arabidopsis thaliana*, *Brachypodium*, soybean, barley, rice, potato, and others are currently available.  When steps of the analysis are completed, the users have access to the reports produced at each step, including the limited number of graphs, data quality measures, and the list of differentially expressed genes (http://dnasubway.iplantcollaborative.org/).

Various approaches to teaching RNA-Seq gene expression using Green Line of the DNA Subway were implemented in the undergraduate classrooms by the members of the workshops lead by the DNA Learning Center.  Although the detailed analysis of student learning assessment as the result of these implementations has not been published yet, several common themes

emerge from conference presentations, workshop discussions, and materials presented at the DNA Learning Center website. Compared to running the complete analysis on the Discovery Environment iPlant platform or using command line, the use of Green Line limits the amount of confusion and frustration from the students and drastically decreases the likelihood of the failure. Students with limited programming and command line skills are able to integrate several analysis packages and concentrate on the results and follow up analysis and conclusions. Using Green Line as a platform for RNA-Seq analysis in the undergraduate classroom provides students with the opportunities to conduct independent or guided research projects, analyze data in the form of graphs, ask and answer scientific questions, and presents their results to peers.

Instructors who would like to utilize the Green Line should consider the following limitations. File sharing / acquisition should be set up by instructors through the instructor's iPlant account or through student-generated DNA Subway user accounts for public data. Since the analysis is run on a supercomputer at the Texas Advanced Computing Center, the users are not able to control the sequencing of the tasks and the time it takes to analyze the data, so careful planning of the activities is required to fit the activities into time-limited teaching lab settings. As with any "black box" approaches, employing Green Line "shields" students from the details of algorithms used in RNA-Seq data analysis. Some of the graphs and outputs produced by the Green Line could be difficult for students to interpret independently. Thus, supplemental pedagogical approaches are required if deeper understanding of the RNA-Seq analysis is among the learning outcomes of the teaching activities. The faculty group led by the DNA Learning Center continues to develop Green Line and RNA-Seq curricular materials for classroom use, including educational resources on RNA-Seq analysis and data analysis protocols. However,

such resources are still limited and development of effective teaching approaches for RNA-Seq remains an important hurdle to integration of RNA-Seq analysis into undergraduate curriculum.

**2.4.2 Learning Command Line: GCAT-SEEK Tutorial for RNA-Seq Analysis**

GCAT-SEEK faculty group recently developed a tutorial that can be used to introduce students and faculty to the concepts and skills required to begin RNA-Seq bioinformatic analyses [11]. This tutorial uses a command line interface and relies heavily on an R package (rnaseqWrapper) developed specifically to ease novel computer users into bioinformatics analysis. The package provides simple functions to encompass many of the complicated steps of RNA-seq analysis, from accessing the software on the UNIX server to user-friendly functions for visualization and output of data analysis [11]. The tutorial starts with explaining how to set up a computer and access the UNIX server, load the data, install required packages, and run jobs on Unix servers using scripts, and continues with detailed background information and a tutorial on how to run the subsequent analysis steps. The authors of the tutorial implemented it as a course-based research experience project in an experimental undergraduate course in bioinformatics at a small liberal arts school with students who took at least two prior biology courses [11]. In this setting, most of the students successfully completed the analysis and increased their skills in data analysis and communication, as assessed by a variety of approaches, despite initial frustration and push back from students due to the intimidation of transitioning from graphical user interfaces to command line. Since this tutorial was designed to serve as a technical supplement for a course rather than as a syllabus for a standalone course or a course module, when used in a classroom, it needs to be supplemented with in-class lectures and discussions to match student needs depending on student background, level, and interests. Although the tutorial is loaded

with the specific sample data set, it could be implemented for the analysis of any system that the instructor or students choose provided that they have access to the corresponding genome files. In addition, the use of this tutorial requires the instructor to set up the system for the students to have access to a UNIX server with the required software and data for analysis and to provide consistent and detailed support for the students at various stages of the project.

**2.4.3 Looking for Middle Ground: Laboratory Modules Integrating RNA-Seq Analysis**

In an effort to develop curricular materials allowing students to analyze genome scale gene expression changes in response to environmental conditions, a series of authentic research laboratory exercises incorporating a large data RNA-Seq analysis into an introductory undergraduate classroom was recently implemented [10]. This laboratory series is focused on analyzing gene expression changes in response to abiotic stress in maize seedlings and highlights the use of plant model systems in investigating consequences of climate change. During the course of four to six weeks, students are engaged in learning about the concepts of gene expression changes in response to environmental conditions and in research analysis of an RNA-Seq data set utilizing a variety of pedagogical approaches, including worksheets, models, guided inquiry, and programming tutorials, as well as various analysis platforms, including the Green Line [88], an R-based package DE-Seq [92], and R-based data visualization functions. The approach chosen by Makarevitch et al. [10] allowed students to focus on asking experimental questions, selecting appropriate data, running the analysis, and visualizing the data and avoid time-consuming steps required to prepare RNA-Seq raw data for differential gene analysis.

A connection to climate change and abiotic stress served as a great way to introduce plant genetics to the medically-oriented students and show them the relevance of plant genetics research, adding to the "instructional toolbox" of plant science educators.  Analysis of student comments in the online course evaluations suggested that students were excited to participate in the real research project using a plant model system, highlighting the need for a careful choice of the RNA-seq data set [10].  As the costs for library construction and sequencing, the most expensive steps of generating RNA-Seq data, continue to decrease, the possibility of running RNA-Seq experiments designed and conducted by students in undergraduate biology courses will be within the reach for many institutions.

**3 Conclusion**

Integrating investigations of gene expression into undergraduate curriculum brings multiple benefits to students, including but not limited to understanding one of the core biological concepts, engaging in authentic research, as well as improving skills in science communication. The quantitative nature of gene expression naturally leads to developing students' skills in quantitative reasoning, data visualization and analysis, and basic programming.  The approaches developed to integrate gene expression analysis into undergraduate classroom range from wet lab laboratories to implementing "research-grade" big data analysis and could be adapted to fit a variety of courses based on the biological systems of interest and level of the student preparation. Using plants as model systems for investigating gene expression provides ample opportunities to engage students with interesting biological questions: from climate change and environmental stress to fundamental questions of plant development and the future of elite crops. Undergraduate students often believe that biological research is limited to understanding human

diseases.  They are frequently not aware of other areas of biological research.  Engaging in interesting and exciting research projects as a part of their courses may broaden their understanding of biology as a field of study and prepare them for careers involving non-medically related biological research.

**Acknowledgements**

**Figure Legends**

Figure 1.  Components of research-based laboratory modules

Figure 2.  Considerations when incorporating gene expression in undergraduate curriculum

**Tables**

Table 1. Resources developed to engage undergraduate students in the research of gene expression

**References**

1. American Association for Advancement of Science (2011) Vision and Change: A Call to Action, Washington, DC
2. Brownell SE, Freeman S, Wenderoth MP, Crowe AJ (2014) BioCore Guide: A Tool for Interpreting the Core Concepts of Vision and Change for Biology Majors. CBE Life Sci Educ 13: 200–211

3. Tansey JT, Baird T, Cox MM, Fox KM, Knight J, Sears D, Bell E (2013) Foundational concepts and underlying theories for majors in "biochemistry and molecular biology." Biochem Mol Biol Educ 41:289–296

4. Couch BA, Wood WB, Knight JK (2015) The molecular biology capstone assessment: a concept assessment for upper-division molecular biology students. CBE Life Sci Educ 14:ar10

5. Shi J, Wood WB, Martin JM, Guild NA, Vicens Q, Knight JK (2010) A diagnostic assessment for introductory molecular and cell biology. CBE Life Sci Educ 9:453-461

6. Smith MK, Knight JK (2012) Using the Genetics Concept Assessment to Document Persistent Conceptual Difficulties in Undergraduate Genetics Courses. Genetics 191:21-32

7. Wright LK, Fisk JN, Newman DL (2014) DNA → RNA: What Do Students Think the Arrow Means? CBE Life Sci Educ. 13: 338–348

8. Campbell AM, Eckdahl T, Cronk B, Andresen C, Frederick P, Huckuntod S, Shinneman C, Wacker A, Yuan J (2014) pClone: Synthetic Biology Tool Makes Promoter Research Accessible to Beginning Biology Students. CBE Life Sci Educ 13:285–296

9. Kuldell NH (2006) How Golden is Silence? Teaching Undergraduates the Power and Limits of RNA Interference. CBE Life Sci Educ 5:247–254

10. Makarevitch I, Frechette C, Wiatros N (2015) Authentic Research Experience and "Big Data" Analysis in the Classroom: Maize Response to Abiotic Stress. CBE Life Sci Educ. 14:ar27

11. Peterson MP, Malloy JT, Buonaccorsi VP, Marden JH (2015) Teaching RNAseq at Undergraduate Institutions: A tutorial and R package from the Genome Consortium for Active Teaching. CourceSource.

12. Tra YV, Evans IM (2010) Enhancing interdisciplinary mathematics and biology education: a microarray data analysis course bridging these disciplines. CBE Life Sci Educ 9:217–226

13. National Research Council (2003) BIO2010: Transforming undergraduate education for future research biologists. National Academies Press (Washington, DC USA)

14. President's Council of Advisors on Science and Technology (2012) Report to the President: Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics

15. Corwin LA, Graham MJ, Dolan EL (2015) Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. CBE Life Sci Educ 14:es1

16. Laursen S, Hunter AB, Seymour E, Thiry H, Melton G (2010) Undergraduate Research in the Sciences: Engaging Students in Real Science, San Francisco: Jossey-Bass.

17. Lopatto D, Alvarez C, Barnard D, Chandrasekaran C, Chung HM, et al. (2008) Undergraduate research. Genomics Education Partnership. Science 322: 684-685

18. Kardash CM (2000) Evaluation of an undergraduate research experience: perceptions of undergraduate interns and their faculty mentors. J Educ Psychol 92:191-201

19. Thiry H, Laursen SL (2011) The role of student-advisor interactions in apprenticing undergraduate researchers into a scientific community of practice. J Sci Educ Technol 20:771-784

20. Adedokun OA, Bessenbacher AB, Parker LC, Kirkham LL, Burgess WD (2013) Research skills and STEM undergraduate research students' aspirations for research careers: mediating effects of research self-efficacy. J Res Sci Teach 50:940-951

21. Dolan EL, Lally DJ, Brooks E, Tax FE (2008) PREPping students for authentic science. Sci Teach 75:38-43

22. Wei CA, Woodin T (2011) Undergraduate research experiences in biology: alternatives to the apprenticeship model. CBE Life Sci Educ 10:123-131

23. Bangera G, Brownell SE (2014) Course-based undergraduate research experiences can make scientific research more inclusive. CBE Life Sci Educ 13:602-606

24. Campbell AM, Ledbetter MLS, Hoopes LLM, Eckdahl TT, Heyer LJ, Rosenwald A, Fowlks E, Tonidandel S, Bucholtz B, Gottfried G (2007) Genome Consortium for Active Teaching: Meeting the Goals of BIO2010. CBE Life Sci Educ 6:109–118

25. Ditty JL, Kvaal CA, Goodner B, Freyermuth SK, Bailey C, Britton RA, Gordon SG, Heinhorst S, Reed K, Xu Z, Sanders-Lorenz ER, Axen S, Kim E, Johns M, Scott K, Kerfeld CA (2010) Incorporating genomics and bioinformatics across the life sciences curriculum. PLoS Biol 8:e1000448

26. Hanauer DI, Jacobs-Sera D, Pedulla ML, Cresawn SG, Hendrix RW, Hatfull GF (2006) Inquiry learning. Teaching scientific inquiry. Science 314:1880-1881

27. Shaffer CD, Alvarez C, Bailey C, Barnard D, Bhalla S, et al. (2010). The genomics education partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. CBE Life Sci Educ 9, 55-69

28. Bialek W and Botstein D (2004) Introductory science and mathematics education for 21st century biologists. Science 303: 788-790

29. Feser J, Vasaly H, Herrera J (2013) On the edge of mathematics and biology integration: improving quantitative skills in undergraduate biology education. CBE Life Sci Educ 12: 124-128

30. AAMC/HHMI Report (2009) Scientific Foundations for Future Physicians

31. Davenport TH and Patil DJ (2012) Data scientist: the sexiest job of the 21st century. Harvard Business Review *October*.

32. Magana AJ, Taleyarkhan M, Alvarado DR, Kane M, Springer J, Clase K (2014) A survey of scholarly literature describing the field of bioinformatics education and bioinformatics educational research. CBE Life Sci Educ 13:607-623

33. Zan R, Brown L, Evans J, Hannula MS (2006) Affect in mathematics education: an introduction. Educ Stud Math 63, 113–121

34. Brady SM, Burow M, Busch W, Carlborg Ö, Denby KJ, Glazebrook J, Hamilton ES, Harmer SL, Haswell ES, Maloof JN, Springer NM, Kliebenstein DJ (2015) Reassess the t Test: Interact with All Your Data via ANOVA. Plant Cell 27:2088-2094

35. Caudill L, Hill A, Hoke K, Lipan O (2010) Impact of interdisciplinary undergraduate research in mathematics and biology on the development of a new course integrating five STEM disciplines. CBE Life Sci Educ 9:212-216

36. Hester S, Buxner S, Elfring L, Nagy L (2014). Integrating quantitative thinking into an introductory biology course improves students' mathematical reasoning in biological contexts. CBE Life Sci Educ 13:54-64

37. Milton JG, Radunskaya AE, Lee AH, de Pillis LG, Bartlett DF (2010) Team research at the biology-mathematics interface: project management perspectives. CBE Life Sci Educ 9:316-322.

38. Sorgo A (2010). Connecting biology and mathematics: first prepare the teachers. CBE Life Sci Educ 9:196-200

39. Uno GE (2009) Botanical literacy: What and how should students learn about plants? American Journal of Botany 96:1753-1759

40. Marbach-Ad G (2004) Expectations and difficulties of first year college students in biology. Journal of College Science Teaching, 33, 18-23

41. Wandersee J, Schussler E (1999) Preventing plant blindness. The American Biology Teacher 61:82–86

42. Schussler EE, Link-Pérez MA, Weber KM, Dollo VH (2010) Exploring plant and animal content in elementary science textbooks. J Biol Educ 44:123–128

43. Schussler EE, Olzak L (2008) It's not easy being green: student recall of plant and animal images. J Biol Educ. 42:112–119

44. Balas B, Momsen JL (2014) Attention "blinks" differently for plants and animals. CBE Life Sci Educ. 13:437-443

45. Liu DWC (2014) Plant Behavior. CBE Life Sci Educ 13:363-368

46. Ebert-May D and Holt E (2014) Seeing the Forest and the Trees: Research on Plant Science Teaching and Learning. CBE Life Sci Educ. 13: 361–362

47. Ward JR, Clarke HD, Horton JL (2014) Effects of a research-infused botanical curriculum on undergraduates' content knowledge, STEM competencies, and attitudes toward plant sciences. CBE Life Sci Educ. 13:387-396

48. Kirchoff BK, Delaney PF, Horton M, Dellinger-Johnston R (2014) Optimizing Learning of Scientific Category Knowledge in the Classroom: The Case of Plant Identification. CBE Life Sci Educ. 13:425 – 436

49. Batzli JM, Smith AR, Williams PH, McGee SA, Dósa K, Pfammatter J (2014) Beyond Punnett squares: Student word association and explanations of phenotypic variation through an integrative quantitative genetics unit investigating anthocyanin inheritance and expression in *Brassica rapa* Fast plants. CBE Life Sci Educ. 13:410-424

50. Gelvin SB, Karcher SJ (1996) Reporter genes and transgenic plants to study response to environmental signals. pp:71-84, in *Tested studies for laboratory teaching*, Volume 17

51. Karcher SJ (2002) Presented at the Conference of the Association for Biology Laboratory Education (ABLE), Toronto, CA.

52. Chrispeels HE, Klosterman ML, Martin JB, Lundy SR, Watkins JM, Gibson CL, Muday JK (2014) Undergraduates Achieve Learning Gains in Plant Genetics through Peer Teaching of Secondary Students. CBE Life Sci Educ 13:641–652

53. Hacisalihoglu G, Hilgert U, Nash EB, Micklos DA (2008) An Innovative Plant Genomics and Gene Annotation Program for High School, Community College, and University Faculty. CBE Life Sci Educ 7:310 –316

54. Wendell DL, Pickard D (2007) Teaching human genetics with mustard: rapid cycling *Brassica rapa* (fast plants type) as a model for human genetics in the classroom laboratory. CBE Life Sci Educ 6:179–185

55. Eckdahl TT, Campbell AM (2015) Using synthetic biology and pClone Red for authentic research on promoter function: genetics (analyzing mutant promoters). CourseSource

56. Martinez-Vaz BM, Makarevitch I, Stensland S (2010) Studying gene expression: database searches and promoter fusions to investigate transcriptional regulation in bacteria. J Microbiol Biol Educ 11:42-49

57. Bulow L, Brill Y, Hehl R (2010) AthaMap-assisted transcription factor target gene identification in *Arabidopsis thaliana*. Database (Oxford) 2010:baq034

58. Perez-Rodriguez P, Riano-Pachon DM, Correa LG, Rensing SA, Kersten B, Mueller-Roeber B (2010) PlnTFDB: updated content and new features of the plant transcription factor database.  Nucleic Acids Res 38:D822-827

59. Chow C, Zheng HQ, Wu NY, Chien C, Huang H, Lee T, Chiang-Hsieh Y, Hou P, Yang T, Chang W (2016) PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucleic Acids Res 44: D1154– D1160

60. Hieno A, Nazhin HA, Hyakumachi M, Sakurai T, Tokizawa M, Koyama H, Sato N, Nishiyama T, Hasebe M, Zimmer AD, Lang D, Reski R, Rensing SA, Obokata J, Yamamoto YY (2014) Nucleic Acid Research 42:D1188-1192

61. Shahmuradov IA, Gammerman AJ, Hancock JM, Bramley PM, Solovyev VV (2003) PlantProm: a database of plant promoter sequences. Nucleic Acids Res 31:114-117

62. Shahmuradov IA, Solovyev VV (2015) Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements. Bioinformatics 31:3544-3545

63. Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muniz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP, Karp PD (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acid Res 39:D583-D590

64. Salgado H, Gama-Castro S, Martínez-Antonio A, Díaz-Peredo E, Sánchez-Solano F, Peralta-Gil M, Garcia-Alonso D, Jiménez-Jacinto V, Santos-Zavaleta A, Bonavides-Martínez C, Collado-Vides J (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. Nucleic Acids Res 32:D303-306

65. Sambrook J, Russel DW (2001) Molecular Cloning: A Laboratory Manual. 3$^{rd}$ edition, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York

66. Bargmann BOR, Birnbaum KD, Brenner ED (2014) An undergraduate study of two transcription factors that promote lateral root formation. Biochem Mol Biol Educ 42:237–245

67. Baker SS, Vidican CB, Cameron DS, Greib HG, Jarocki CC, Setaputri AW, Spicuzza CH, Burr AA, Waqas MA, Tolbert DA (2012) An epifluorescent attachment improves whole-plant digital photography of *Arabidopsis thaliana* expressing red-shifted green fluorescent protein. AoB Plants. pls003.

68. Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U (2004). Just-in-time transcription program in metabolic pathways. Nat. Genet. 36:486–491

69. Imperial S, Boronat A (2005) Determination of the Rh factor: A practical illustrating the use of the polymerase chain reaction. Biochem Mol Biol Educ 33:50-53

70. Martin P, Muruke M, Hosea K, Kivaisi A, Zerwas N, Bauerle C (2004) A rapid PCR-RFLP method for monitoring genetic variation among commercial mushroom species. Biochem Mol Biol Educ 32:390-394

71. Elkins KM, Kadunc RE (2012) An undergraduate laboratory experiment for upper-level forensic science, biochemistry, or molecular biology courses: human DNA amplification using STR single locus primers by Real-Time PCR with SYBR green detection. J. Chem. Educ. 89:784–790

72. Sulzinski MA, Wasilewski MA, Farrell JC, Glick DL (2009) Undergraduate virology exercises demonstrate conventional and real-time PCR using commercially available HIV primers and noninfectious target. Biochem Mol Biol Educ. 37: 232-235

73. Hargadon KM (2016) A model system for the study of gene expression in the undergraduate laboratory. Biochem Mol Biol Educ. (ahead of print)

74. McCauslin CS, Gunn KE, Pirone D, Staiger J (2015) qPCR for second year undergraduates: A short, structured inquiry to illustrate differential gene expression. Biochem Mol Biol Educ. 43:273–282

75. Weinlander KM, Hall DJ, De Stasio EA (2010) RFLP analysis and allelic discrimination with real-time PCR using the human lactase persistence trait: A pair of molecular genetic investigations. Biochem Mol Biol Educ 38:167-171

76. Hancock D, Funnell A, Jack B, Johnston J (2010) Introducing undergraduate students to real-time PCR. Biochem Mol Biol Educ 38:309-318

77. Eickelberg GJ, Fisher AJ (2013) Environmental regulation of plant gene expression: an RT-qPCR laboratory project for an upper-level undergraduate biochemistry or molecular biology course. Biochem Mol Biol Educ 41:325-333

78. Birnbaum MJ, Picco J, Clements M, Witwicka H, Yang M, Hoey MT, Odgren PR (2009) Using osteoclast differentiation as a model for gene discovery in an undergraduate cell biology laboratory. Biochem Mol Biol Educ. 38:385-392

79. Weinlander KM, Hall DJ (2010). Designing laboratory exercises for the undergraduate molecular biology/biochemistry student: techniques and ethical implications involved in personalized medicine. Biochem Mol Biol Educ, 38:180–187.

80. Boyle, MD (2010) "Shovel-ready" Sequences as a Stimulus for the Next Generation of Life Scientists. Journal Microbiology and Biology Education 11:38.

81. Micklos D, Lauter S, Nisselle A (2011) Essays on science and society. Lessons from a science education portal. Science 334:1657-1658.

82. Kahvejian A, Quackenbush J, Thompson JF (2008) What would you do if you could sequence everything? Nat Biotechnol 26:1125-1133

83. Alvarez M, Schrey AW, Richards CL (2015) Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Mol Ecol 24:710-725

84. DeBiasse MB, Kelly MW (2016) Plastic and Evolved Responses to Global Change: What Can We Learn from Comparative Transcriptomics? J Hered 107:71-81

85. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM (2015) Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress. PLoS Genet 11(1): e1004915

86. Weber AP (2015) Discovering New Biology through Sequencing of RNA. Plant Physiol 169:1524-1531

87. Kodama Y, Shumway M, Leinonen R; International Nucleotide Sequence Database Collaboration (2012) The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res 40:D54-56

88. Buonaccorsi VP, Boyle MD, Grove D, Praul C, Sakk E, Stuart A, Tobin T, Hosler J, Carney SL, Endgle MJ, Overton BE, Newman JD, Pizzorno M, Powell JR, Trun N (2011) GCAT-Sequence: genome consortium for active teaching of undergraduates through increased faculty access to next-generation sequencing data. CBE Life Sci Educ 10:342-345

89. Buonaccorsi VP, Peterson M, Lamendella G, Newman J, Trun N, Tobin T, Aguilar A, Hunt A, Praul C, grove D, Roney J, Roberts W (2014) Vision and change through the genome consortium for active teaching using next-generation sequencing (GCAT-SEEK). CBE Life Sci Educ 13:1-2

90. Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, et al. (2011) The iPlant Collaborative: Cyberinfrastructure for Plant Biology.  Front Plant Sci 2:34

91. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105-1111

92. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc 8:1765-1786

Table 1. Resources developed to engage undergraduate students in the research of gene expression

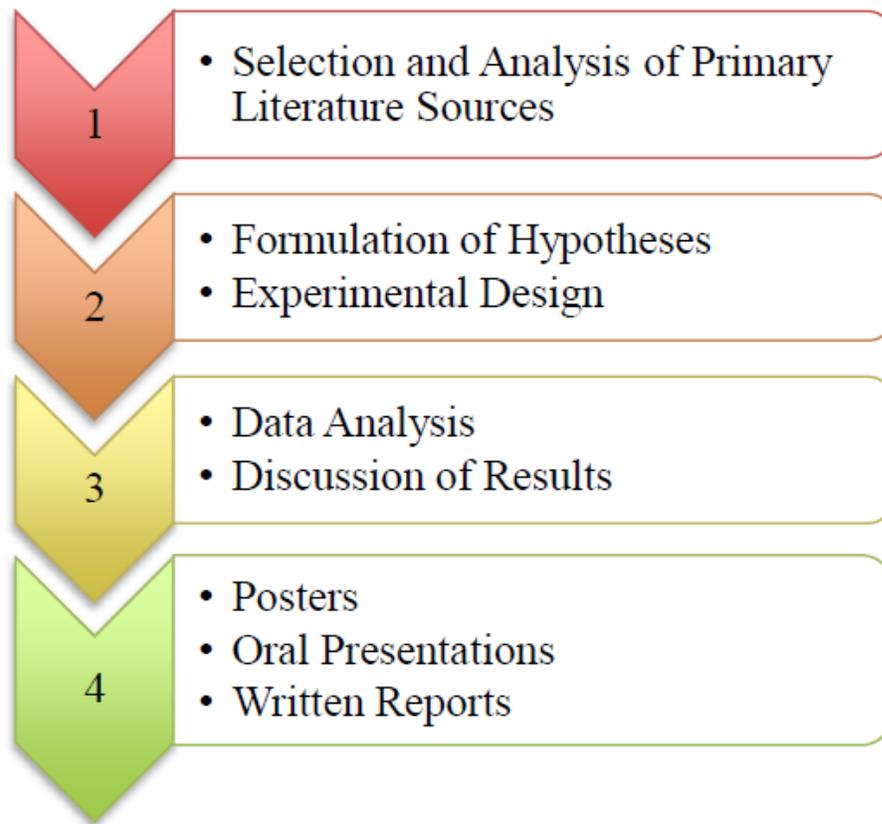| Research approaches | Resources | Teaching materials | Skills |
|---|---|---|---|
| RT-PCR / qRT-PCR | Kits:<br>- Arabidopsis Biological Resource Center<br>-Biorad<br>-Thermo | Detailed manuals and teaching materials | Wet lab skills, data analysis, experimental design |
| Promoter / Transcription Factor Analysis | Databases:<br>Plant Transcription Factor Database; AthMap, PlantProm, Regsite  (Nsite-PL & Nsitem-PL) | Manuals | Data mining, bioinformatics |
| | Kits:<br>-pClone (Carolina) | Detailed manual and teaching materials | Wet lab skills, data analysis, experimental design |
| RNA Sequencing | DNA Subway / Green Line | Manual and some teaching materials, sample data | Quantitative data analysis, bioinformatics |
| | UNIX server command line analysis | GCAT RNA-Seq tutorial, sample data | Programming in R, command line interface, quantitative data analysis |
| | Lab module on abiotic stress in maize | Lab manual, worksheets, tutorials, assessment tools, sample data | data visualization, programming in R, experimental design, quantitative data analysis |
| Reporter genes | Kits<br>-Arabidopsis Resource Center | Seed lines, transgenic plants, laboratory manuals | GUS and GFP gene reporters, data analysis |

**Figure 1**

**Figure 2**