

DNA Microarrays in the Undergraduate Microbiology Lab: Experimentation and Handling Large Datasets in as Few as Six Weeks

David B. Kushner*

Department of Biology, Dickinson College, Carlisle, Pennsylvania 17013

DNA microarrays have significantly impacted the study of gene expression on a genome-wide level but also have forced a more global consideration of research questions. As such, it has become critical to introduce undergraduate students to genomics approaches to research. A challenge with performing a DNA microarray experiment in the teaching lab is determining the time required for the study and how to handle the voluminous data generated. At an unexpectedly low cost, a 6-week, project-based lab module has been developed that provides 3 weeks for wet lab (hands-on work with the DNA microarrays) and 3 weeks for dry lab (analyzing data, using databases to help with data analysis, and considering the meaning of data within the large dataset). Options exist for extending the number of weeks dedicated to the project, but 6 weeks is sufficient for providing an introduction to both experimental genomics and data analysis. Students indicate that being able to both perform array experiments and thoroughly analyze data enriches their understanding of genomics and the complexity of biological systems.

It has been just over a decade since the first report about a DNA microarray was published (14). While the first few years of this period featured technical challenges and high costs, improvements have allowed for greater ease in performing experiments using DNA microarrays. Widespread use of this technology, its applications in numerous research areas, and connections to genomics and bioinformatics predicate the need to introduce this concept into the undergraduate biology curriculum. Several recent publications have addressed the benefits of incorporating microarray work into the lab curriculum (1–4; 6). Notably, the Genome Consortium for Active Teaching (GCAT) provides resources and information for faculty who teach microarrays to undergraduate students (1, 2, 4). In addition, as per the recommendations of BIO2010 (13), a microarray study can easily be an aspect of a project-based laboratory.

Challenges with a DNA microarray study include determining an appropriate course in which to incorporate it, the timeframe required to complete the study, and what types of data analysis to perform following the microarray experimentation. Trying to complete all of this can be daunting not just for the students, but also the professor. Nevertheless, these challenges can all be met, based on the goals one has for students performing the exercise. I have incorporated a DNA microarray experiment into an upper-level course in microbiology. Prerequisites for this course include 1 year of Introductory Biology or equivalent AP credit. Of the 68 students enrolled in this course between 2003 and 2005, there were 18 sophomores, 19 juniors, and 31 seniors with majors as diverse as biology, biochemistry and molecular biology, chemistry, physics, environmental science, psychology, and

French. The 14-week course includes a weekly 3-hour lab meeting. Because other lab exercises (e.g. identification of unknowns using traditional staining and molecular methods, etc.) are important to include in a microbiology course, only about half of the semester is available for a DNA microarray experiment.

This study is based on three hypotheses and three goals. It was hypothesized that in a short time (half a semester) it is possible to perform both wet and dry aspects of a DNA microarray study, with student understanding of experimental design, execution of method, and data analysis involving the ability to consider large datasets. Second, it was hypothesized that students become invested in this extensive exercise based on its open-ended, project-based nature. Finally, it was hypothesized that DNA microarrays serve as an effective introduction to genomics and illustrate the complex biology resulting from a single change in a biological system. As for goals, the first was for students to perform a DNA microarray experiment for hands-on experience with array technology and to be introduced to genomic approaches to project-based research. Another goal was for students to analyze data from such an experiment to better understand how to work with a large amount of data. A final goal was to have students use databases so as to understand and appreciate their value as a scientific resource, and see how they can be used to help consider how parts of a large dataset may biologically connect—a foray into the realm of the complex biology of a system. In this report, I demonstrate that these hypotheses were validated and assess achievement of these goals.

MATERIALS

Saccharomyces cerevisiae DNA microarrays with DNA oligonucleotides (arrayed in duplicate), each representing a unique 70-nucleotide portion from each gene of the genome, were obtained from GCAT (www.bio.davidson.edu/projects/gcat/gcat.html). The 3DNA 900 kit (Genisphere, Hatfield, Pa.) was used for conversion of total yeast RNA to cDNA

*Corresponding author. Mailing address: Department of Biology, Dickinson College, P. O. Box 1773, Carlisle, Pennsylvania 17013. Phone: (717) 245-1328. Fax: (717) 245-1130. E-mail: kushnerd@dickinson.edu

and the sequential hybridizations of cDNA and fluorochromes (Cy3 and Cy5 dyes) to the DNA microarrays. Microarrays were scanned at Davidson College. Microarray data were analyzed using MAGIC Tool (7) and Microsoft Excel. Various databases linked from the *Saccharomyces* Genome Database (www.yeastgenome.org) were employed during data analysis of genes found to have strongly altered levels of transcription. Details on working with *S. cerevisiae* (including media preparation for, growth of, transformation of plasmids into, and use of heated acidic phenol to extract RNA from) and on performing the microarray experiment can be obtained from the Yeast Resources portion of the Protocols section of the GCAT webpage (download document by Kushner and Tiede, www.bio.davidson.edu/projects/gcat/GCATprotocols.html#phenol).

METHODS

A significant challenge with setting up a DNA microarray experiment in the teaching lab is determining the amount of time that can be afforded for such work (without taking time away from other relevant, and important, lab exercises within a course), and therefore what elements of the study can be performed given time constraints. Work with arrays can be divided into four components: (i) experimental design, (ii) wet labwork (which in this study involved manipulation and growth of *S. cerevisiae*, RNA extraction, cDNA synthesis, and then sequential hybridizations of cDNA and fluorochromes to the microarray), (iii) determination of up- and down-regulated genes (which in this study primarily involved use of MAGIC Tool (7)), and (iv) working with large data sets (which in this study involved using information from databases to learn about the genes, analyzing that information, and proposing biological connections of gene products from within the dataset). One of the purposes of this report is to present an overview about the options for performing the aforementioned four components to microarray experimentation and to note the time required for each step. In my course, students performed components (ii) through (iv) with the bulk of the work taking 6 weeks (Table 1). In the lab, six groups of three to four students each used one microarray; for the second and third years I taught this lab, three of the six groups performed one of two different experiments, therefore each of the two experiments was completed in triplicate (an ideal setup). As noted in the Materials section, a detailed protocol for using *S. cerevisiae* and performing arrays can be found at the GCAT webpage; these approaches are applicable for model organisms other than *S. cerevisiae*.

(i) Experimental design—multiple options for consideration. Experimental design offers numerous options for the teacher and the members of the class. The key to designing a DNA microarray experiment is to ensure that only one variable will exist between the two conditions to be tested. In three separate offerings of my microbiology course (2003, 2004, and 2005), I designed the experiments to be performed, since I used these opportunities to ask questions that might be helpful for my research program. Although this removed the option of the students designing their experiment, the project

was open-ended as the result was unknown at the onset of the study. As such, the students were intrigued to perform original research, there was immediate “ownership” of the experiment, and there was a time benefit in that no lab time was required for experimental design. Alternatively, several weeks prior to the experiment, but after an introduction to the theory underlying uses of microarrays, lab groups could have been given homework directing them to design a microarray experiment. Proposed experiments could have been discussed during a class meeting with no loss of lab time. Doing this several weeks in advance would allow the instructor time to obtain necessary materials and to help students modify their proposals to ensure they would be testable (include only one variable). Another option would be for students to mimic previously performed DNA microarray experiments, allowing students to compare their results to published results; an example of this is comparison of gene expression in steady-state log-phase yeast grown in the presence of glucose versus galactose (11).

TABLE 1. Microarray timeline

When conducted	Activity
2 weeks in advance ^a	Transform yeast with plasmids (30' 2x during nonarray lab)
1 week in advance ^a	Grow and harvest yeast (few days)
Week 1	Extract RNA from yeast (3 hr)
Week 2	cDNA synthesis (3 hr)
Week 3	cDNA and Cy dye hybridizations ^b
Week 4	MAGIC Tool; gridding (3 hr)
Week 5	Identify induced and repressed genes (3 hr)
Week 6	Data analysis and database exploration ^c

^aOptional for students.

^bSequential; requires meeting over 2 days (see text).

^cAdditional nonlab time is required for lab report preparation.

(ii) Wet labwork—3 weeks of lab. In my microbiology course, the main aspects of this portion of the experiment were RNA extraction from yeast, cDNA synthesis, and sequential hybridizations and washes of cDNA and Cy dyes to the DNA microarray. This portion of the experiment required 3 weeks of lab time. However, in order for students to start at the point of RNA extraction, yeast cells were previously

grown and harvested. I preferred to grow and harvest the yeast, since I had time to carefully monitor the growth of the yeast so that all cultures of the cells were simultaneously harvested while at the same part of their logarithmic growth phase, reducing variability between samples (the theory underlying growth and division of unicellular organisms was discussed in lecture).

It should be noted that one way to create differences between populations of yeast cells is to transform yeast with a plasmid(s) that confers foreign gene expression (with parallel transformation of an “empty” plasmid(s) into the control yeast). This was done, as part of my labs, 2 weeks prior to the RNA extraction lab (Table 1). This provided time to grow and select yeast transformants on solid media, grow the yeast in liquid media, and harvest the cells prior to the lab where RNA was extracted. Benefits for the students included having time for me to introduce the long-term microarray project and learning how to transform competent cells. Of course, experiments could be designed not requiring transformations (e.g., reference 11).

Students in my course spent one 3-hour lab period extracting RNA from yeast using heated acidic phenol. Students learned principles of how phenol is used to separate protein and nucleic acid and, by default, how to carefully work with hazardous material. In cases where working with phenol is not ideal, the MasterPure RNA purification kit (Epicentre Biotechnologies, Madison, Wis.) can be used (C. J. Alvarez, personal communication; www.bio.davidson.edu/projects/GCAT/protocols/RNA/RNA_methods.html).

The second of the three wet lab weeks involved using the Genisphere 3DNA 900 kit to reverse transcribe the isolated RNA into cDNA. Each RNA was subjected to reverse transcription with either a primer for Cy3 or Cy5 to enable dye reversal (see next paragraph). The 2-hour incubation allowed for further explanation of the background and rationale behind the experiment and details of microarray technology.

The third week of wet lab work involved using the 3DNA 900 kit for sequential hybridization (described below) of the cDNA and Cy dyes to the DNA microarray. As each yeast DNA microarray obtained from GCAT has a top portion and a duplicate bottom portion, a dye reversal was performed on the single array (Fig. 1; www.bio.davidson.edu/projects/gcat/GCATprotocols.html#phenol) so that, in total, the control and the experimental samples were each ultimately labeled with either Cy3 or Cy5. This accounted for variability in strength of signal and is important since Cy5 signal is easily degraded in the presence of ozone (5).

In general, there are two approaches for hybridizing cDNA to the DNA microarrays. One option is to incorporate the fluorochromes into the cDNA as it is being synthesized, with subsequent hybridization of the fluorochrome-labeled cDNAs to the array followed by a series of washes. Use of the 3DNA 900 kit, with sequential hybridizations of cDNA and the Cy dyes, required some creativity in scheduling lab time. Students spent about 2 hours the evening before the scheduled lab period pretreating the DNA microarray, preparing the cDNA hybridization mixes, and adding the cDNA to

the microarray for an overnight incubation. The next day, the students spent about 90 minutes washing the array, preparing the Cy dye hybridization mixes, and adding the Cy dyes to the washed array for a 4- to 5-hour incubation. Later that day, the students returned for an hour to wash and dry the slides. Since the nature of the work did not fit neatly into the scheduled afternoon lab time, in order to ensure each student gained hands-on experience with method, each student was asked to be present for at least one hybridization solution prep, hybridization of cDNA or Cy dyes to the array, and one series of washing. Since I am aware of these scheduling challenges as the semester starts, the class and I selected meeting times so that the greatest number of students were present during all of the steps. The students were willing to meet during the extra time since they had already spent a lot

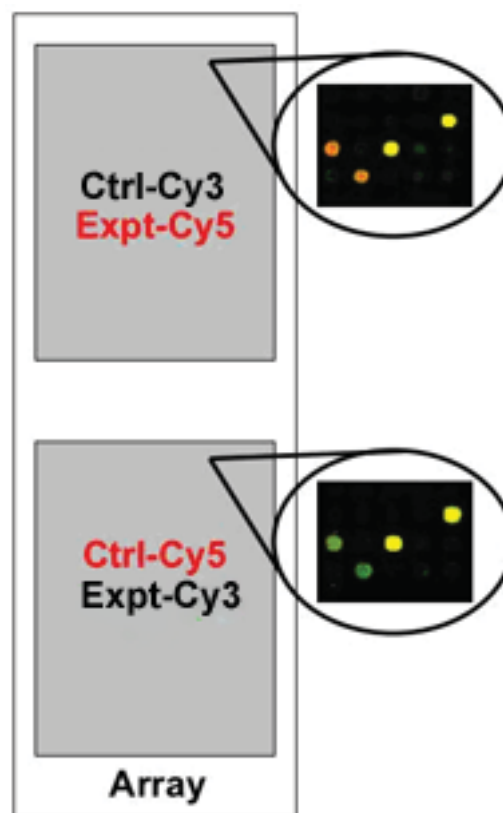


FIG. 1. DNA microarray experimental setup. Control (ctrl) and experimental (expt) RNA were each reverse transcribed with Cy3 or Cy5 tagged primers (four total reactions) in week 2 of the wet lab work. Distinct pairs of cDNAs were hybridized to distinct portions of the microarray (as indicated in the figure). A sequential incubation with Cy dyes generated signal. The results of the dye reversal can be seen in the sample array image (right). The two yellow spots in the center and upper right of both halves of the sample array indicate similar expression of the gene in both conditions; no signal indicates no expression of the gene in either condition; expression of the two genes in the experimental condition is seen using both dyes via the dye reversal.

of time and energy on the project and were willing to put in the extra time with the aim of obtaining good array data.

(iii) Dry labwork—3 weeks of lab. The first week of data analysis employed MAGIC Tool (7) to determine the relative gene expression of each yeast gene under each of the two tested conditions (control versus experimental). First, students were introduced to MAGIC Tool by uploading sample red and green .tif image files which feature four grids from an array and then “gridding” the individual data spots (www.bio.davidson.edu/projects/MAGIC/MAGIC.html). Students then uploaded their personal red and green .tif images into a new MAGIC Tool file. The yeast microarrays from GCAT (currently) feature 16 grids on each of two identical halves of the array; therefore 32 grids, encompassing 13,544 data spots, were generated. Working in groups of three or four, the students took turns gridding to complete this task in about 75 to 90 minutes. “Segmentation” of the gridded data was performed, generating red-green signal ratios for each spot on the array. Powerful computers with a lot of memory should be used; for example, Macintosh G5 computers segmented the data in 1 to 2 minutes, whereas older G4 machines required about 20 minutes. According to the MAGIC Tool program installation guide (www.bio.davidson.edu/projects/magic/magic.html#down), MAGIC Tool runs on any operating system with Sun Java; 512MB of RAM is minimal and 1 to 2GB of RAM is optimal. Following segmentation, red-green ratios were saved for the second week of dry lab work.

The second week of dry lab entailed identifying genes with either enhanced or repressed expression on both halves of the array. This allowed one to account for different raw red-green ratios that were generated on each half of the array. Red-green ratios were \log_2 transformed; then using Microsoft Excel spreadsheets, the \log_2 transformed red-green ratios from each part of the array were normalized using the “standardize” function, which generates a normalized value based on a distribution characterized by the mean and standard deviation of the dataset. Therefore, for each half of the array, the genes in the experimental condition can be ordered from “most enhanced expression” to “most repressed expression.” Since the goal of the array is to identify the genes with the most enhanced or repressed expression in the experimental condition, 6.25% of the most enhanced genes from both halves of the array (about 422 genes from 6,552 nonmitochondrial genes (some genes are duplicated on each half of the array)) were compared. Genes on both lists were considered to be enhanced and were retained for further study (the same held true for repressed genes). In general, this resulted in groups needing to further analyze about 80 total enhanced and repressed genes.

The third week of dry lab featured two components. The first was to introduce students to some of the databases they could use to help with analysis of their genes (see part iv, below). The second was a short exercise (Table 2) to verify

that the students understand how enhanced and repressed gene expression relates to the original experimental design.

TABLE 2. Exercise to help verify understanding of experimental setup and meaning of results

Question	Answer
Originally, which samples were labeled green and red, and which green-red combinations were hybridized to each respective half of the array?	To verify that if the experimental cDNA was tagged with Cy5 (red) and added to array half “A,” that Cy3 (green)-tagged experimental cDNA was added to array half “B” (and the converse for the control sample).
For array half “B,” we did a dye reversal, but then “flipped” the red and green files when loading them into MAGIC Tool. What does this effectively do to Grid B in terms of what red and green represent?	MAGIC Tool only computes red-green ratios. Since a dye reversal effectively is a “flip” of the samples, MAGIC Tool needs to be instructed about this so that for array half “B” the raw red-green ratio is properly computed.
From the log transformed data, positive and negative numbers were obtained; ultimately, how were positive and negative numbers derived from red-green ratios?	Positive red-green ratios indicate that the signal from Cy5 (red) was greater than the signal from Cy3 (green); negative ratios indicate greater signal from Cy3 (green).
An array signal for a gene features strong red signals in array half “A” and strong green signals in array half “B” (and, similarly, weak green in “A” and weak red in “B”). In which yeast cell population (control or experimental) was the gene highly expressed?	Assuming that cDNA from the experimental cells was labeled with Cy5 in array half “A” (and Cy3 in “B”), then the gene expression was greatly enhanced in the experimental cells (also, the gene expression was repressed in the control cells).

(iv) Analysis of large data sets—incorporation into student lab report preparation. As an introduction to engaging students to think about the cell as a dynamic entity, that of a “system,” students used the *Saccharomyces* genome database (SGD; www.yeastgenome.org) and databases linked from SGD to obtain information about their genes and gene products. For example, gene name, open reading frame designation, and gene ontology information are provided on the SGD web page specific to each yeast gene. As for databases linked from SGD web pages that are specific to each yeast gene, the yeast GFP-fusion localization database (GFP DB

at UCSF; yeastgfp.ucsf.edu; available via the “Localization Resources” link) was accessed for protein localization information; the General Repository for Interaction Datasets (BioGRID; www.thebiogrid.org; available via the “Interactions” link) lists protein-protein interactions, etc. This information was used to help compose a table for a lab report listing each gene identified during the second week of the dry lab (Fig. 3). The table included gene name, ORF designation, gene ontology annotation, essential or nonessential gene, protein localization, and known protein-protein interactions. Students were encouraged to work on this as a team, since an enormous amount of information was generated. Information in the table was then used by the students as a vehicle to help them look for potential connections between genes with up-regulated expression and also between genes with down-regulated expression. Such connections were reported in the discussion section of lab reports.

RESULTS

Over 3 years of microbiology class, 18 groups of three or four students each utilized one array per group. Sixteen of the 18 groups (89%) obtained data that could be gridded and analyzed. Since array projects were generally performed in triplicate (as described in Methods), student groups who did not perform the array successfully used data from another group for data analysis, so that they could participate in that element of the laboratory exercise.

Assessment of student understanding was performed for various aspects of the lab exercise. The three main areas

assessed were (i) understanding of experimental design, (ii) understanding trends in gene expression, and (iii) ability to propose connections between genes with either up- or down-regulated expression, as an initial exercise in considering biology of an entire system.

(i) Understanding of experimental design and general microarray use. Student understanding of general principles in design of microarray experiments was assessed through GCAT-based assessment tools and/or questions during an in-class exam. In 2003, the students ($n = 22$) were asked three questions (Table 3) at the start of the semester and again as part of an exam administered late in the term (students knew that any material from lecture or lab could be part of the exam). In response to how one could measure a genome-wide response to a virus, 20 students (91%), following their experience in lab, knew to employ a DNA microarray (the other two students suggested looking at immune response, which while a clue to virus infection, does not involve measuring genome-wide response) compared to only three students (14%) knowing this at the start of the term. When asked about controls to use, 17 students (77%) replied that, at the least, cells lacking virus should be used (two other students (9%) proposed other kinds of controls, but uninfected cells are best since this control directly tests the variable of the virus being present); only nine students (41%) thought of this before experimentation. In terms of technical limitations for microarray use, 19 students (86%) were able to list at least two potential issues for use of microarrays (common answers included needing to know the genome sequence of the organ-

TABLE 3. Assessment questions

Question Number	Open-ended questions (2003)	Online multiple-choice questions ^a (2004 and 2005)
1	Design an experiment to measure the genome-wide response when a cell becomes infected with a virus.	You are interested in a bacterium that grows in high temperature conditions. One interesting observation is that it is green when grown at the normal high temperature, but colorless when grown at a temperature that is 20 degrees below the optimal growth temperature. Which of the following would be the best description of the way you could investigate this difference using microarrays? (Isolate total RNA from the organism grown at the two temperatures, reverse transcribe it with dyes to cDNA and incubate with the microarray)
2	What controls would you need for this experiment?	A smoker and a coal miner both got lung cancer. The cancers are histologically identical, but there might be functionally important differences in gene expression in the two cancers which would allow personalized treatment. Which of the experiments is the best design to determine appropriate treatment for the two lung cancers? (Compare the gene expression profiles of the two lung cancers to lung samples without cancer)
3	What are the technical limitations to the method(s) you chose?	You are investigating the changes in gene expression in cancer cells compared to normal cells by labeling of RNA samples. When you look at your entire microarray after performing the scan in both dye channels, you observe spots with several shades of green, but no red spots. Which of the following best explains the data? (Poor labeling of one of the samples)

^a Correct multiple-choice answer provided after the question, in parentheses.

ism so the array can be generated, challenges with accurate quantitation of gene expression, and not knowing if changes in protein levels correlated with mRNA levels), when, at the start of the term, only two students (9%) described one technical limitation.

In 2004 and 2005, student understanding of microarrays was determined via a new online assessment tool designed by members of GCAT. Two multiple-choice questions on this assessment addressed application of knowledge in terms of experimental design (Table 3). In 2004 ($n = 24$ students; 22 students took both the pre- and posttests) for question 1, two of 22 students (9%) answered correctly on the pretest; for the posttest that improved to 16 students (73%). For question 2, four of 22 students (18%) answered correctly on the pretest; 10 students (45%) answered correctly on the posttest. In 2005 ($n = 22$ students; 14 students took both the pre- and posttests) for question 1, three of 14 students (21%) answered correctly on the pretest; 11 students (79%) did so on the posttest. For question 2, five of 14 students (36%) answered the pretest question correctly; 11 students (79%) did so on the posttest. Finally, a third question, asked each year, addressed what could have happened during the microarray experimental procedure to result in no red signal on the microarrays. Two of 22 students (9%; 2004) and two of 14 students (14%; 2005) knew the answer to this question prior to microarray work; after array work, 13 (59%;

2004) and 6 (43%; 2005) students answered this question correctly.

Because the assessment used to ascertain student understanding of experimental design changed after 2003, a two-way analysis of variance with Student-Newman-Keuls comparisons test (factors: pre- versus posttest and course year) was performed to determine (i) if, each year, there was a statistically significant increase in student understanding, and (ii) if this was independent of the year the students took the course (and which assessment was performed). The results are summarized in Table 4. The P value of <0.001 indicates that the improvement in understanding was significant. In addition, there was no effect due to the year the students took the course (and which assessment was performed) based on the P value of 0.128.

(ii) Understanding trends in gene expression data. In 2003, as described above, 22 students completed both the pre- and posttests. Here, a fictitious diagram (Fig. 2a) showing gene expression changes (via fold induction and fold repression) for each of six different genes over a 75-minute time course (time 0 and every 15 minutes thereafter) was provided, and students were asked to comment on any aspect of gene expression they chose. On the pretest, six of 22 students (27%) were able to apply the concept of the “map” of gene expression changes to accurately comment on gene expression changes of some or all of the six genes shown. After doing the experiment

TABLE 4. Statistical^a analysis of percentage of students correctly answering assessment questions

Year	Question(s)	Pretest (%)	Posttest (%)	P value ^b
	Involved in understanding of microarray experimental design			<0.001
2003	Table 3 question 1	14	91	
	Table 3 question 2	41	77	
	Table 3 question 3	9	86	
2004	Table 3 question 1	9	73	
	Table 3 question 2	18	45	
	Table 3 question 3	9	59	
2005	Table 3 question 1	21	79	
	Table 3 question 2	36	79	
	Table 3 question 3	14	43	
	Involved in understanding trends in gene expression data			0.103
2003	Figure 2a question	27	100	
2004	Figure 2b question	18	45	
2005	Figure 2b question	7	36	

^a Two-way analysis of variance with test interval and year as Student-Newman-Keuls pairwise comparisons; analysis performed using Sigmasat 2.0.

^b Pre- and posttest information for each set of questions was analyzed as a group via two-way analysis of variance; the P values shown are for the interval of learning. No significant year-to-year effect was observed (see text in Results).

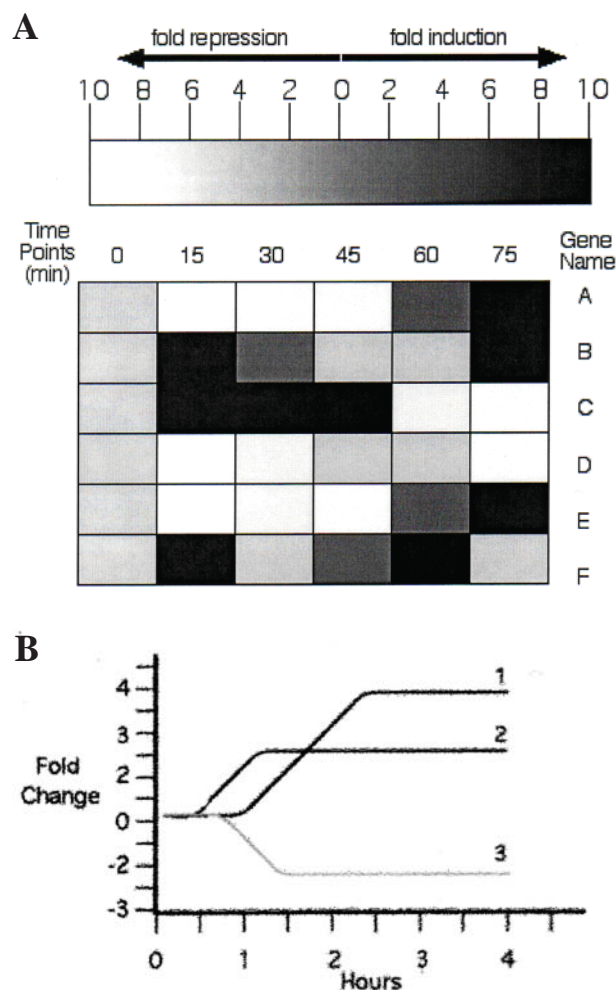


FIG. 2. Figures provided to students to quiz them on understanding trends in gene expression data. (A) This figure was used in 2003 when asking students to consider trends in gene expression over time (reprinted from *Cell Biology Education* (3) with permission of the publisher. (B) This figure was used in 2004 and 2005 when asking students to directly consider relationships between genes and their levels of expression over time.

and discussing graphical representation of array data, all 22 students (100%) accurately commented on several aspects of gene expression.

In 2004 and 2005, a different diagram was shown (Fig. 2B) and students were asked to consider potential relationships between the three genes shown. In 2004, at the start of the term, four of 22 students (18%) were able to correctly note that gene 2 likely induces expression of gene 1 (a specific answer from a panel of multiple-choice options); whereas 10 of 22 students (45%) did so at the end of the term. In 2005, one of the 14 students (7%) who took both the pre- and post-tests knew the answer at the start of the term, and five of 14 students (36%) did so at the end of the term. Of note, this style of data representation was not discussed at any time in the course.

As noted above, because the assessment design changed after 2003, a two-way analysis of variance was performed to determine (i) if there was a statistically significant increase in student understanding in trends in gene expression, and (ii) if this was independent of the year the students took the course (and which assessment was performed). The results are summarized in Table 4. There was a trend towards improvement in understanding ($P = 0.103$), yet there was no effect due to the year the students took the course (and which assessment was performed) based on the P value of 0.26.

(iii) Working with large datasets: proposing biological connections between gene products. One of the greatest challenges for students performing microarrays is the sheer volume of data that is generated and knowing what to do with it. The discussion section of the lab report offered an opportunity for students to consider the entirety of the data and select some elements of it to discuss. Guidelines for the lab report included prompts such as “Discuss any genes that are functionally related...focus on gene ontology, localization, and protein-interaction information.” The goal was for each student to peruse the information in the table that their group generated (Fig. 3) and identify sensible biological connections between subsets of genes in the microarray analysis, as an initial foray into considering the complex nature of the cell. The ability of each student to do this was assessed upon reading the lab report. Fifty-eight of the 68 students (85%) in the course from 2003 to 2005 described several potential connections between subsets of genes that had been found to be enhanced or repressed in their expression levels on the microarray. Six students (9%) only noted one possible connection; two students (3%) listed groups of genes but did not explain the connections between them; two students (3%) simply did not attempt this part of the discussion section of the lab report.

DISCUSSION

Two factors that concern faculty in regard to introducing a DNA microarray experiment into a lab portion of their course are time required and cost of materials. I have described a protocol where wet and dry lab exercises require 6 weeks of lab (7 weeks if the option of transforming yeast with plasmids is performed; however, since it does not take much time this can be combined with another lab earlier in the term or, alternatively, could simply be performed by the lab instructor and students given cell pellets harvested from cultures of transformed yeast cells). Some students have noted that they would like more time to perform the microarray experiment and more time to spend with the data; while this would be ideal, the results from the assessment and, importantly, the lab report discussions suggest that an initial understanding of the main aspects of project-based genomics research and how microarrays can be used in that regard is obtained. At Dickinson College, since another faculty member in our department teaches a course on Genomics, Proteomics, and Bioinformatics, I can suggest that students enroll in this course for additional instruction in this and related areas. Admittedly, for some students, more time discussing arrays,

ORF	Standard gene name	Feature type	Gene description	Molecular function	Biological process	Cellular compartment	Viability of a deletion strain	Protein localization	Protein-protein interactions
YKR034W	DAL80	Verified	Negative regulator of genes in multiple nitrogen degradation pathways; expression is regulated by nitrogen levels and by Gln3p; member of the GATA-binding family, forms homodimers and heterodimers with Deh1p	Transcription factor activity	Regulation of nitrogen utilization; transcription	Nucleus	Viable	N/A	YJL110C, YER047C, YKR034W, YLR376C, YDR520C, YNL021W, YHR060W
YKR073C	N/A	Dubious	Unlikely to encode a protein	Unknown	Unknown	Unknown	Viable	N/A	YER081W
YKR105C	N/A	Uncharacterized ORF	Hypothetical protein	Unknown	Unknown	Unknown	Viable	N/A	YDR155C, YJR091C
YLR184W	N/A	Dubious	Hypothetical protein	Unknown	Unknown	Unknown	Viable	N/A	N/A
YYL226W	BUR2	Verified	Cyclin for the Sgv1p (Bur1p) protein kinase; Sgv1p and Bur2p comprise a CDK-cyclin complex involved in transcriptional regulation through its phosphorylation of the carboxy-terminal domain of the largest subunit of RNA polymerase II	Cyclin-dependent protein kinase regulator activity	Mitotic sister chromatid segregation; transcription	Nucleus	Viable	N/A	YPR161C, YBR135W, YMR125W, YJL158C
YLR236C	N/A	Dubious	Hypothetical protein	Unknown	Unknown	Unknown	Viable	N/A	YLR438W, YJR091C
YML009W-B	N/A	Dubious	Hypothetical protein	Unknown	Unknown	Unknown	Viable	N/A	N/A
YML080W	DUS1	Verified	Dihydrouridine synthase, member of a widespread family of conserved proteins including Smm1p, Dus3p, and Dus4p; modifies pre-tRNA (Phe) at U17	tRNA dihydrouridine synthase activity	tRNA modification	Nucleus	Viable	N/A	N/A
YML095C-A	N/A	Dubious	Hypothetical protein	Unknown	Unknown	Unknown	Viable	N/A	YOR025W, YNL298W, YJL030W, YCL029C, YDR332W, YER016W, YFR162C, YBR109C, YJL020C, YNL250W,
YML095C	RAD10	Verified	Single-stranded DNA endonuclease	Single-stranded DNA specific endodeoxyribonuclease activity (TAS)	Double-strand break repair via single-strand annealing, removal of nonhomologous ends (TAS)	Nucleotide-excision repair factor 1 complex (TAS)	Viable	N/A	YPL262W, YJL138C, YEL003W, YGR063C, YER043C, YDL185W, YAL027W, YCR086W, YJL026W, YGL105W, YHR008C, YDR300C, YDR155C, YOR260W, YPL022W
YML123C	PHO84	Verified	High-affinity inorganic phosphate (Pi) transporter and low-affinity manganese transporter	Inorganic phosphate transporter activity (IDA, IMP, ISS), Manganese Ion Transporter activity (IMP)	Manganese ion transporter activity(IMP), phosphate transport (IDA, IMP, ISS)	Integral to plasma membrane (IDA, ISS)	Viable	N/A	YER125W, YDR388W, YDR523C, YLR238W, YNL061W, YNL116W, YJR091C, YDR200C, YGR040W, YOR181W

FIG. 3. Portion of a student's table (from fall 2005) containing information obtained from online databases. The molecular function, biological process, and cellular compartment categories relate to gene ontology. Protein localization information came from the yeast GFP-fusion localization database, and protein-protein interactions information came via BioGRID. As seen in the gene ontology columns, traceable author statement (TAS), inferred from direct assay (IDA), inferred from sequence or structural similarity (ISS), and inferred from mutant phenotype (IMP) are evidence codes for how the annotation was generated.

how the data are generated (mathematical analysis thereof), and more ways the data can be analyzed would be beneficial and might enable improved scoring on the multiple-choice assessment tools, which stress application of knowledge. However, considering that this experiment is performed in a relatively short time, the hypothesis—that the compressed timeframe still allows students to clearly understand how to use microarrays and why one would do so—is supported.

The experiment is (potentially surprisingly) cost-effective. Since the dry lab aspects of the course are cost free (MAGIC

Tool is shareware available via GCAT) and if reagents for general work with microbes are already available (in my case, solid and liquid media for growth of yeast, chemicals for transformation of plasmids into yeast, etc.), then the one-time-per-course costs come from the DNA microarrays (via GCAT, six yeast arrays cost \$150) and the 3DNA kit (to generate the cDNA, hybridize the cDNA to the array, and add the Cy dyes to the bound cDNA, I use two kits, which cost about \$700 with shipping). Therefore, a 6-week lab for six groups of three to four students only costs about \$850.

Notably, the lab exercise described herein is not restricted to working with *S. cerevisiae*; parallel approaches can be taken with other species. Currently, GCAT can provide instructors of undergraduate students with DNA microarrays from 10 species, including *Escherichia coli*, human, mouse, and *Arabidopsis*. For data analysis, instead of the *Saccharomyces* genome database, the *E. coli* Genome Project (www.genome.wisc.edu/sequencing/k12.htm) can be accessed by students using *E. coli* arrays in the lab (e.g., to obtain gene annotation).

For the students, a challenge of the DNA microarray experiment is that it occurs over a long period of time. Therefore, being able to recollect details of the early parts of the project can be difficult. I stress to the students that it is important for them to keep good notes so that they are able to recall all elements of their work on the project. Also, in my Microbiology course the DNA microarray study is the second of two multiweek experiments, so a subset of the students may be helped by having experience in a lab project prior to the DNA microarray work.

Another challenge for the students is that, as designed in my course, the project-based lab lacks a “correct” answer—I like the open-ended nature of this work since it is analogous to real research experiences; others have also commented on the value of project-based labs (13). My observations of students becoming deeply engaged in the DNA microarray work and subsequent data analysis supports the hypothesis that, even with the half-semester timeframe for the study, this is a successful vehicle for students to experience an open-ended, project-based lab.

Another important concept for the students to appreciate is that generation of DNA microarray data is really only a starting point in what would be a more extensive study; for example, the microarray should only be considered to be a qualitative-semiquantitative assessment of changes in transcription and resulting levels of steady-state RNA. For quantitation of gene expression, genes from microarray studies would need to be analyzed by quantitative, real-time PCR. Following that, biological significance should be investigated to understand why changes in gene expression occurred. Furthermore, while I do not stress statistical analysis of the data in my course, other faculty may find this valuable; indeed, there are many dry lab exercises that could be performed following generation of the microarray data based on the goals that one has for his/her students.

The assessment performed makes it clear that doing an experiment using a DNA microarray helps the students understand how to perform an experiment on a genome-wide level. In addition, during informal conversations following completion of the microarray exercise, students often remark that while it is difficult to ponder the enormous dataset generated, looking for connections between genes in the dataset really helps them realize the complexity within the cell and helps them move away from consideration of genes and gene products functioning in isolation. This validates the hypothesis concerning the use of microarrays as a valid tool for engaging students in considerations of biological com-

plexity. In some ways, this may be one of the most important results of the array work, since the genomic era forces us to think about biology in a more complex manner. Furthermore, several students gain interest in the computer science side of the project, from the power of databases to the realization that computer programs are invaluable in analysis of large data sets. Such realization spurs some students to appreciate that in order to help comprehend biological systems, approaches in systems biology, an emerging area of biology that relies on math and computer science (8–10; 12), will need to be pursued. This is a wonderful illustration of the need for collaborative, interdisciplinary science as we continue with investigation of complex biological questions. An experience working with DNA microarrays and microarray data therefore serves as an exceptionally effective introduction to this arena.

ACKNOWLEDGMENTS

This work was partially presented as a poster at the 2005 American Society for Microbiology Conference for Undergraduate Educators in Atlanta, Georgia (Kushner D. B. and B. J. Tiede, Microarray analysis in the microbiology laboratory: completing both experimentation and data analysis in as little as six weeks). I thank the GCAT community, especially A. Malcolm Campbell, Laurie Heyer, Todd Eckdahl, Laura Hoopes, Mary Lee Ledbetter, Anne Rosenwald, and Consuelo Alvarez for providing information on microarrays, MAGIC Tool, and protocols to help me design my laboratory exercise. Ben Tiede (Dickinson College class of 2005) helped refine the data analysis process. Tom Arnold and Rachel Sanders assisted with statistical analysis of assessment data. I thank Amy Cheng Vollmer for critical review of the manuscript. Special thanks go to the 68 Dickinson College students who took my microbiology course between 2003 and 2005 and performed the microarray studies described herein.

REFERENCES

1. **Brewster, J. L., K. B. Beason, T. T. Eckdahl, and I. M. Evans.** 2004. The microarray revolution. *Biochem. Mol. Biol. Educ.* **32**:217–227.
2. **Campbell, A. M.** 2002. Meeting report: genomics in the undergraduate curriculum—rocket science or basic science? *Cell Biol. Educ.* **1**:70–72.
3. **Campbell, A. M.** 2003. Public access for teaching genomics, proteomics, and bioinformatics. *Cell Biol. Educ.* **2**:98–111.
4. **Campbell, A. M., T. T. Eckdahl, E. Fowlks, L. J. Heyer, L. L. M. Hoopes, M. L. Ledbetter, and A. G. Rosenwald.** 2006. Genome Consortium for Active Teaching (GCAT). *Science* **311**:1103–1104.
5. **Fare, T. L., E. M. Coffey, H. Dai, Y. D. He, D. A. Kessler, K. A. Kilian, J. E. Koch, E. LeProust, M. J. Marton, M. R. Meyer, R. B. Stoughton, G. Y. Tokiwa, and Y. Wang.** 2003. Effects of atmospheric ozone on microarray data quality. *Anal. Chem.* **75**:4672–4675.
6. **Hancock, D., L. L. Nguyen, G. S. Denyer, and J. M. Johnston.** 2006. Microarrays for undergraduate classes. *Biochem. Mol. Biol. Educ.* **34**:432–437.

7. **Heyer, L. J., D. Z. Moskowitz, J. A. Abele, P. Karnik, D. Choi, A. M. Campbell, E. E. Oldham, and B. K. Akin.** 2005. MAGIC Tool: integrated microarray data analysis. *Bioinformatics* **21**:2114–2115.
8. **Ideker, T.** 2004. Systems biology 101—what you need to know. *Nature Biotechnol.* **22**:473–475.
9. **Kirschner, M. W.** 2005. The meaning of systems biology. *Cell* **121**:503–504.
10. **Kling, J.** 2006. Working the systems. *Science* **311**:1305–1306.
11. **Lashkari, D. A., J. L. DeRisi, J. H. McCusker, A. F. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, and R. W. Davis.** 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. USA* **94**:13057–13062.
12. **Lok, C.** 2005. Thinking outside the cell. *Nature* **438**:706–707.
13. **National Research Council, Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century.** 2003. *BIO2010: transforming undergraduate education for future research biologists.* National Academy Press, Washington, D.C.
14. **Schena, M., D. Shalon, R. W. Davis, and P. O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**:467–470.

BUILD YOUR FUTURE!

The ASM Kadner Institute

July 21 - 25, 2007

University of Colorado - Boulder

The ASM Kadner Institute is one week of intensive, hands-on training in:

- Grantsmanship
- Scientific presentations
- Teaching
- Career opportunities
- Scientific ethics

A premiere institute for graduate students and postdoctoral scientists!

Since 2000, more than 150 graduate students and postdoctoral scientists have participated; over 70 percent have moved onto permanent employment or entered postdoctoral positions.

Space is limited to 36 participants. Applications are due June 15, 2007.

www.ASMGAP.org



AMERICAN
SOCIETY FOR
MICROBIOLOGY

GET A HEAD START ON YOUR FUTURE!

Opportunities for Students and Postdoctoral Scientists

- ASM Annual Biomedical Research Conference for Minority Students (ABRCMS)
- ASM Kadner Institute
- ASM Fellowship Programs
 - ASM/CCID Postdoctoral Research Fellowship
 - ASM Robert D. Watkins Graduate Fellowship
 - Microbiology Undergraduate Research Fellowship (MURF) Program
 - Undergraduate Research Fellowship (URF) Program

Let ASM help prepare you for your future in microbiology. Find out which program is best for you.

www.asm.org/education

fellowships-careerinformation@asmusa.org



AMERICAN
SOCIETY FOR
MICROBIOLOGY