Project Title:

Microarray Data Analysis, an Interdisciplinary Course

Applicant(s):

<table>
<thead>
<tr>
<th>Name</th>
<th>Telephone</th>
<th>Dept.</th>
<th>College</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yolande Tra</td>
<td>5-4497</td>
<td>School of Mathematical Sciences</td>
<td>Science</td>
<td></td>
</tr>
<tr>
<td>Irene Evans</td>
<td>5-2491</td>
<td>School of Life Sciences</td>
<td>Science</td>
<td></td>
</tr>
</tbody>
</table>
Microarray Data Analysis, an Interdisciplinary Course

1. **Summary and new course description**

The field of biological sciences has been transformed in recent years by the development of microarray technology enabling the analysis of genome-wide patterns of gene expression. Enormous amounts of raw information on DNA sequences are available and the information is increasing exponentially. The need for computational approaches to obtain important biological information from these raw data is critical. The most widely used technique for the analysis of gene expression on a global scale is DNA microarray analysis. An increasing number of laboratories have established it as a tool for monitoring tens of thousands of genes in parallel to help diagnose diseases or find drugs to cure them.

- There is a growing need for expertise in this area. There is also a need to enhance undergraduate education and training at the intersection of the biological and mathematical sciences so as to better prepare undergraduate biology or mathematics students to pursue graduate study and careers in fields that integrate the mathematical and biological sciences.
- This proposal seeks support to design a model course where students learn to conduct computational analysis using the most up-to-date methods. This course should provide students exposure to contemporary mathematics and biology, addressed with modern research tools and methods.
- The course that is envisaged is a research course. Students will work on a research problem that has not yet been solved. After generating the data, the students will use various computational methods to analyze the data. These computational methods are drawn from the latest published papers. No textbook is required. The course will give the students a “hands-on” introduction to this emerging field and challenge them to critically think about the results they obtain when they use different methods of analysis. The analysis of microarray data involves a wide variety of disciplines including biology, bioinformatics, biochemistry, computer sciences, mathematics and statistics. This interdisciplinary feature makes the proposed course unique.

2. **Targeted learners**

The targeted audiences are third and fourth year students from the following disciplines: Biology, Bioinformatics, Biotechnology, Biochemistry, Biomedical Sciences, Mathematics and Statistics. It will be an elective course (Credit 4) for the students in the listed majors, during the pilot program. Any students in related disciplines interested in the subject can take the course as well. The goal is to draw all undergraduate students from different disciplines to be trained in the analysis of the data from the microarray technology and to be knowledgeable about high dimensional biology data analysis. The course will help to minimize a major bottleneck in the utilization of these data. It is aimed at familiarizing the participants with the use of advanced tools and methodologies. A background in microarray data
analysis is tremendously useful for a job search or graduate study endeavor. It makes a difference in a resume and will get the students challenging and well-paying jobs.

3. **Anticipated impact on teaching and/or learning**

Teaching this course poses many challenges. First, many students have a fear of statistics and would rather be out in the biology lab doing experiments to obtain data from a microarray rather than sitting in class learning about estimating the intensity of the signal from the scanned images and analyzing complex genetic information. We intend to address this challenge by spending some time in classes discussing the fields of statistics and biology and asking the students how they feel about studying these fields. We will record the students’ responses and revisit them as the course progresses to see if there are changes in attitudes. We will also highly recommend that students taking the course will have taken at least Data Analysis I. Second, microarray data analysis is an emerging field; new methods to handle such massive data are not taught in traditional classes at an undergraduate level. Even when such courses are taught, they seldom allow the students to have a genuine research experience in which the students generate the data, analyze the data using computational methods, evaluate the results and design new experiments.

The objective is to create a special environment where after lecture on some selected topics, students get hands on practice in a computer-equipped laboratory and use analysis-related software to extract and explore the data. To pass the course, each student is assigned to a group to work on a small project based on a provided dataset given the first week of class. Each group will apply the methods presented in the course, present their findings to the class, and critically assess the results obtained. The class will consist of modules. In each module, students will learn a new approach using a simple dataset with the methodology demonstrated by the instructor. After practicing on the simple dataset, the students will apply the approach to a research dataset generated at RIT. Also, in each module, students will read recently published papers and present these papers using a journal club approach in which a different student or group of students leads the discussion of the paper.

4. **Impact on student success/retention**

The lab part requires

- Interactive activity through the use of computer analysis will demonstrate the concepts and techniques being discussed in lectures. Hands-on exercises ease understanding of statistical concepts usually considered difficult by students and make the ideas more memorable. They also help to retain the student’s interest and develop computational skills.

- Student’s active involvement: learning by doing is achieved. This should aid student retention. The data focused approach is a successful technique in teaching statistical concepts.
We expect
• To have students being able to define the biological problem, understand how the
data are generated, and then see how the computational analysis helps to analyze
and hopefully solve the problem.
• Students with limited mathematical background and who are uncomfortable
dealing with image analysis and statistics will learn the materials by example and
less use of mathematical or statistical notation
• Students trained as a new generation of scientists who can bridge the gap between
the biological data and the computational methods needed to maximize the
information extracted from the data. This will enhance the learning experience
and deepen the understanding of each step of the analysis.
• Students getting experience with data analysis and proficiency in statistical
software (MAGIC, Scanalyze, R and Bioconductor packages) to conduct the
analysis. With the use of software, each student will master the technique of
locating the gene within each pixel and going through each step of statistical
analysis.
• Student from each discipline listed above should benefit from this course. For
example, after completion of the course
✓ A student majoring in biology gains an understanding of the structure of the data
and the statistical procedures used to remove the inherent noise.
✓ A student majoring in statistics learns new methods to describe observations made
on a biological phenomenon and to make inferences on the biological significance
of the results.

5. Measurement and Dissemination

The impact will be measured by evaluations done after each module. Students and
instructors will evaluate themselves and each other for how well they performed using a
checklist made up by the instructors. After doing the evaluation, each person will suggest
several changes they can make to improve their performance and their group
effectiveness. More specifically, the evaluation will survey the understanding of the
material, the effectiveness of the teaching environment and the techniques taught.

Students will be also evaluated based on written and oral presentation of a group class
project. A written final report to PLIG committee is also considered. Other usual ways to
disseminate the results include: publication in peer-reviewed education journal,
presentation at an undergraduate education theme conference, a web-based site for the
model course for public view

6. Rationale

A. Interdisciplinary, relevance to mathematical biology

Faculties from the School of Mathematical Sciences and Biological Sciences are
currently developing a proposal to NSF to support an interdisciplinary program that trains
students in mathematical (or quantitative) biology. The proposed course provides a model
for such an interdisciplinary program. It is broken into modules related to biology, bioinformatics and statistics. Dr. Irene Evans in the Biological Sciences Department will teach modules covering the biology materials. Dr. Tra will cover modules related to Bioinformatics and Statistics.

B. Faculty credentials

Dr. Tra is an Assistant Professor of Statistics in the School of Mathematical Sciences. Her research interests are statistics and data analysis for medicine and genetics. Currently Dr. Tra teaches and maintains a website of Data Analysis II (1016-320) http://people.rit.edu/~smam320 with structure similar to the proposed course. The latter class, taught as a pilot course two years ago, is now successfully implemented and offered as a 6 credit hour lab based course.

Dr. Irene Evans is a Professor of Biological Sciences and her research area is using microarray technology to analyze gene expression patterns in different strains of yeast containing prion proteins. Dr. Tra and Dr. Evans began a collaboration in which Dr. Tra was analyzing microarray data from Dr. Evans research group. They jointly mentored research scholars and students from Biology, Bioinformatics and Biotechnology in 2006-7. These students went on to graduate schools, bringing to their graduate programs the benefit of the microarray research experience from RIT. One of the students, John Brothers, an Outstanding Undergraduate Scholar while at RIT, has chosen as his Ph.D. thesis topic to analyze microarray gene expression data gathered from smokers with and without lung cancer.

7. Innovation

This new approach of tying together several disciplines in the college of science is unique and innovative. It is relevant since it provides students exposure to research projects at the intersection of contemporary mathematics and biology. Students will address these projects with modern research tools and methods. Thus, this course provides an innovative approach of teaching material using the available emergent technology. Using available academic free software, this course intends to teach image analysis techniques, data preprocessing, feature selection and high level information analysis.

8. Timeline

| Summer 0704 | 1. Develop the course: create modules, web-based site, lab assignment, search for available public data, coordinating the coherence of biology, bioinformatics and statistics contents
| Winter 0802 | 2. Two students perform microarray experiment, collect data for class analysis
| | 3. Train the two students for being a TA for the course
| Spring 0803 | 1. Statistical Analysis of the assessment
| | 2. Present Results to an undergraduate education theme conference