Independent Studies in Computational Biology

Randy Von Smith, Deborah McGann, Keith Shockley, Shirngwern Tsaih, James Nadeau, Gary Churchill

The Jackson Laboratory , Bar Harbor, Maine 04609 USA

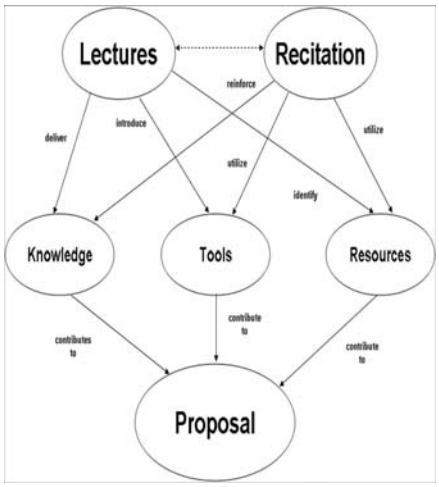
Abstract

The Center for Genome Dynamics created a course tailored to a group of Maine high school students that was most ready and willing to accept the challenge of learning about cross-disciplinary science. We took advantage of the Jackson Laboratory's strong and evolving relationship with Maine's elite public science and math boarding school: the Maine School of Science and mathematics (MSSM: www.mssm.org). The Jackson Laboratory has conducted a research oriented summer student program for over 77 years to encourage high school and college students to conduct independent, hypothesis-driven projects as a member of a sponsoring scientist's laboratory team. We adapted this structure for use throughout the regular academic year with video conferencing and web based tools so that our research oriented class could be taught remotely.

Our students were prepared for conducting team oriented independent research by a lecture series that covered the core processes of:

- **Introductory Statistics**
- **The R computer language**
- **Asking meaningful Scientific Questions**
- **Reading the scientific literature**
- **Writing a NIH grant proposal**
- **Genetics of the laboratory mouse**
- **Quantitative Trait Loci**
- **Microarray analysis**
- **Cluster analysis**

Concept Map for ISCB course



Methods

The students in the first year course attended the Maine School of Science and Mathematics (MSSM) in Limestone, Maine, a public residential magnet school chartered and funded by the Maine State Legislature and located 250 miles north of the Jackson Laboratory. Classes were conducted 4 days a week for a total of 6 hours of instruction. The staff at the Center for Genome Dynamics used video conferencing and web based e-meeting tools to conduct classes remotely. The Fall classes called for 2 hours a week of lectures and the Spring classes had 1 hour of lecture and 1 hour of recitation time. In addition, the students visited the Jackson Laboratory and our staff visited the MSSM campus.

The main philosophy behind the instruction was to teach students to ask meaningful and answerable questions. The lecture material provided a foundation to build on and specific techniques were introduced in a "just in time" format to meet the research goals. During the Spring term each class started with a repetition of the Big Question:

"Do different mouse inbred strains respond differently to high fat diet?"

Students were guided through the steps of writing and refining a research proposal. Two of the students worked on R based graphics function to display microarray data and another worked on the data analysis.

Table 1: Fall Lectures

The fall lecture introduced the fundamental techniques and concepts in computational biology that the students would be using in their research. Instruction in reading and writing scientific literature and in writing a grant proposal were provided. The first semester goal was to draft a research proposal.

	Fall Lectures
1	Introduction to Genetics Research
2	Introduction to Statistics with R (part 1)
3	What is a Gene Chip?.
4	Introduction to Statistics with R (part 2) Descriptive statistics
5	Introduction to Statistics with R (part 3) Correlation and Regression
6	R: How to use scripts (ANOVA, t-test, regression, Box plots)
7	"GO" Analysis
8	"GO" with sample data and scripts
9	How to Write a Grant Application
10	"GO" Analysis in R, the Gene List
11	How to write a research proposal, Specific Aims
12	Introduction to the Novartis 12 strain data set
13	Introduction to Graphical Modeling
14	Open discussion day
15	Recitation on using R scripts
16	Expression Quantitative Trait Loci
17	Graphical Modeling
18	Expression QTL
19	Experimental Strategies in Mouse Genetics
20	Structural Modeling Approach to Analysis of Complex Traits

Table 2: One day on site visits

Students visited the Jackson laboratory in the Fall and Spring semesters.

During the fall visit the students attended one day of lectures and then participated in a day at the short course on Complex Trait Analysis.

During the spring visit the students attended lectures and toured the Jackson Laboratory. The afternoon was a round table discussion with the students to review their research results and suggest new avenues of inquiry.

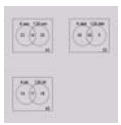
on site 1	Introduction to Genetic Research
	DNA Microarray Data Analysis
	A Brief Tour of R/qtl in Complex Trait Analysis
	Introductory Statistics using R - Linear Regression
	Classical and Modern Genetics
	The Origins of Experimental Strategies in Mouse Genetics
on site 2	Overview of Cholesterol Transport
	JGR and iPlots: interactive graphic in R
	GO terms and K-means clustering
	How to write a scientific paper
	Deliverables and Ideas: the students present their results followed by open discussion time

Table 3 : Spring Lectures

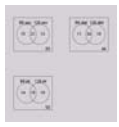
The spring lectures reviewed the research aims and concentrated on what the students needed in order to complete their research goals. Each week, time was set aside for open discussions and student interaction.

	Spring Lectures
1	Introduction to emeeting software
2	Graphical methods for displaying Microarray Data
3	Using loops in R: F test and Two-way Interaction Plots
4	The 99 gene list: Venn diagrams, clusters, histograms, and data integrity.
5	Cluster Analysis in R, Distance Metrics, Bi-clustering
6	K-means clustering
7	Software Life Cycle and documentation
8	Phenotypes for the Novartis data set
9	Research Recitation and Discussion, student results and discussion
10	Research Recitation and Discussion , student results and discussion
11	Research Recitation and Discussion , student results and discussion
12	Research Recitation and Discussion , student results and discussion

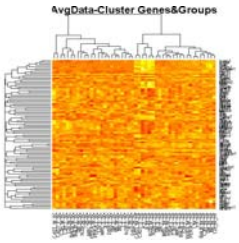
Figure 4: Sample graphs from student reports



Venn diagrams showing the number of genes with significant p-values for sex, diet, and interaction F-tests in both A/J and 129, in A/J only, in 129 only, and in neither of the two strains.

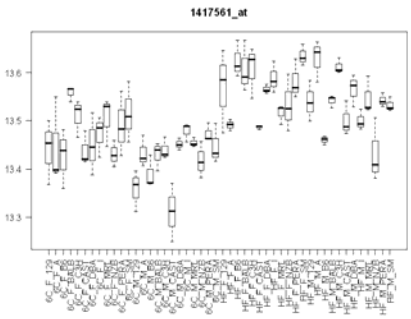


Venn diagrams showing the number of genes with significant p-values for sex, diet, and interaction F-tests in both B6 and 129, in B6 only, in 129 only, and in neither of the two strains.



Heat map showing the correlation in expression of 99 genes involved in Reverse Cholesterol Transport across 48 conditions of strain, sex, and diet.

Figure 5: Example output of student generated Box plot code




```
library(maanova);
raw.data <- read.madata("Affy061406rma.dat",
designfile="Novartis_12strains_Paigen_design_file_
corrected_030807.dat", cloneid=1, pmt=2, spot=F);
data <- createData(raw.data, n.rep=1, log.tran=F);
data$colmeans <- rep(0, length(data$data[1,]));
probe <- "1417561_at";
sort.array <- array(c("Diet", "Sex", "Strain"));
mabox(data, probe, sort.array);
```

This graph displays the expression values of apoc1, a gene that regulates the lipid metabolic process in humans and mice. The experimental groups (shown along the x-axis) are averages of three biological replicates. The experiment consisted of twelve unique strains, two sexes, and two different diets. The x-axis is arranged first by diet, then by sex, then by strain. At a glance, we can see that the apoc1 gene is expressed more in mice on the high-fat diet.

Conclusions

Two of our students were seniors this year and graduated on May 26th.

George Cooper (on left) plans to continue his studies at the University of Maine in the Fall.

David Witmer (on the right) will be a summer intern at the Jackson Laboratory this year. He plans to continue his studies at MIT in the Fall.