jmp®
Statistical Discovery.™ From SAS.

Susan Singer, PhD

Now they see

At Carleton College, data visualization aids 'huge paradigm shift' to genome-scale studies

"We're excited," says Susan Singer, PhD, the Laurence McKinley Gould Professor of Natural Sciences at Carleton College, a private liberal arts college in Northfield, MN. It's that feeling teachers at all levels get when students make a major breakthrough, and Singer has just experienced it with an undergraduate genetics class, into which she recently introduced JMP Genomics statistical software from SAS.
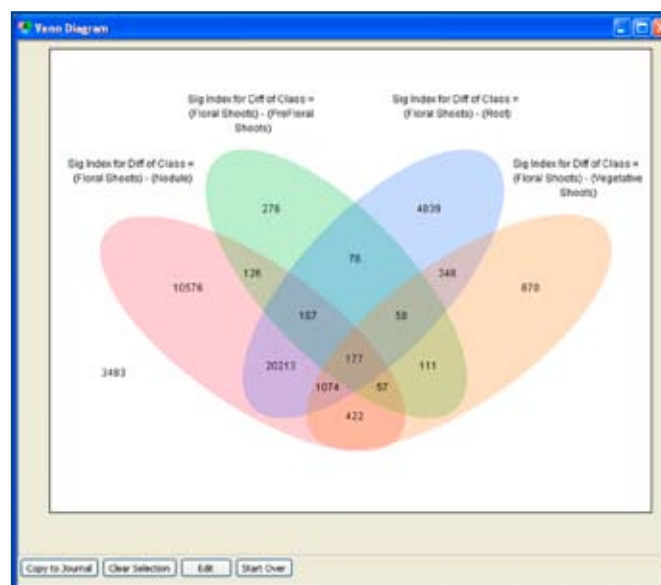
What a world of difference it's made.

"The two biggest challenges I've found in teaching genomics have been getting students to truly think and ask questions at the level of the whole transcriptome and to effectively use data visualizations," Singer says. "JMP Genomics solved both those challenges. The nature of the questions students are now asking are substantially different than in the past and they're already thinking much more visually."

Singer and several of her colleagues – Jeff Doyle (Cornell University), Greg May (National Center for Genome Resources) and Steven Cannon (Iowa State University) – are leading a Big Science at Small Schools Collaboration, a project funded by the National Science Foundation. They are studying the genomics of *Chamaecrista*

Comparisons between transcripts from floral shoots and other tissues identify clusters of unique and shared sets of genes. Comparative visualizations of large data sets encourage students to ask questions at a genomics scale, expanding beyond their prior experiences with a single gene focus.

fasciculata, a native prairie plant with potential for use in biofuels. The goal, Singer has written, "is to engage students in authentic genomic research, in a formal teaching setting, where the questions they ask are both original and of value to the research community."

"I think we're part of a wave of educators who are trying to bring genomics into the undergraduate classroom in a way that prepares students to go on and participate in the new biology," she says, "and that's really exciting."

The collaborative approach

The Big Science project has both an educational and a scientific component (*Plant Physiology*, 2009, 151:1041). The impetus of the educational component is to help students begin to study genetics data at the scale of entire genomes, which Singer calls a "huge paradigm shift." Undergraduates, she says, are more accustomed to thinking on a gene-by-gene basis.
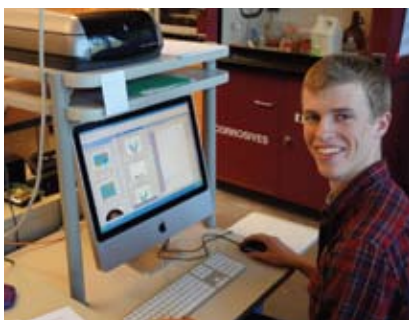
Kai Knutson took genetics at Carleton in 2009 and then helped develop the JMP Genomics instructional materials for the 2010 class.

"This project has developed as a way to help students think critically and take advantage of the scale of data that's now available," she says, "and to look at data in an authentic situation."

The objective is to work with a number of students in a collaborative environment, as opposed to supervising independent research. So Singer and her colleagues at Carleton's Science Education Resource Center designed a term-long laboratory experience, whereby students are introduced to plants from a number of different ecotypes and then delve into the data to learn more about them – in this case, looking at how the plants flower at different times and why.

The experience provides a means for students to choose strategies for asking and addressing interesting questions using a range of genomics tools. Students read selected literature and keep journals throughout the term, and a Web-based interface helps them frame the choices they can make in mapping their research.

"The first year we did this, the students did pretty well," Singer says. "But it was really very hard for them to visualize the data or know where to start, and a lot of students just defaulted to hunting for their favorite gene or trying to imagine what an interesting gene would be.

"But this term, we introduced them to JMP Genomics, and it's just been incredibly helpful in allowing them to visualize the data."

Students quickly learned how to navigate the software.

"I'd never used JMP Genomics before I was assigned to create a lab module incorporating it for the students in Professor Singer's genetics course," says Kai Knutson, a student who worked very closely with Singer throughout the term. "I found the user interface very intuitive and within weeks was able to design a set of instructions to guide the students through a full expression analysis of a transcriptome data set generated by Dr. Singer's lab group.

"Visualizing data in the way that JMP Genomics allows deepened my own understanding of how to interpret expression at a genomic level and relationships between various genes. The ease of creating such complex models as volcano plots and using them to determine good candidate genes for further study is remarkable."

Singer concurs that the most beneficial tool in JMP Genomics has been the volcano plots.

"They can compare two different stages of development and they actually can see and understand which data points are statistically significant and different from each other." Students then can click on individual points in the plots and get transcript IDs.

"Once they find an area that's interesting, they can collect a number of transcripts. From there, some of them took candidate gene approaches; others

were interested in looking for families of genes that were differentially expressed.

"JMP allowed them to begin to frame questions in a much more meaningful way than they were able to do when just handed a spreadsheet that went on for tens of thousands of lines of raw data."

Assessing success

Singer and her colleagues take a mixed-methods approach in assessing students' progress, monitoring their online journals and giving them assignments in which they have to show how they're thinking through the issues. At the end of the term, they give group PowerPoint presentations, get feedback and then have another

week to revise their work and turn in a journal-style paper.

"The quality of those papers is incredibly good," Singer says.

"At the beginning of the term, I ask them to draw a model of how flowering is regulated in these different ecotypes of *Chamaecrista* that flower at different times and may potentially go extinct because of global change. And at the end of the term, I have them turn in another model. When they start out, the majority say, 'What do you mean by a model? I don't know how to make a picture of this. Can't I just write it down in words?'
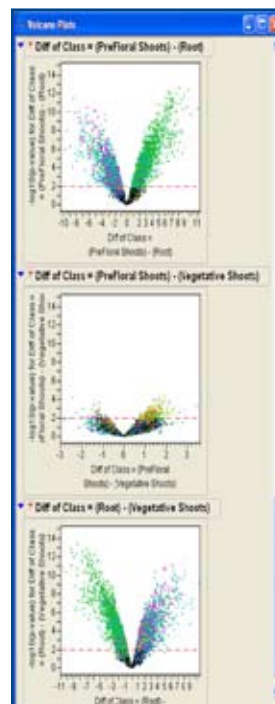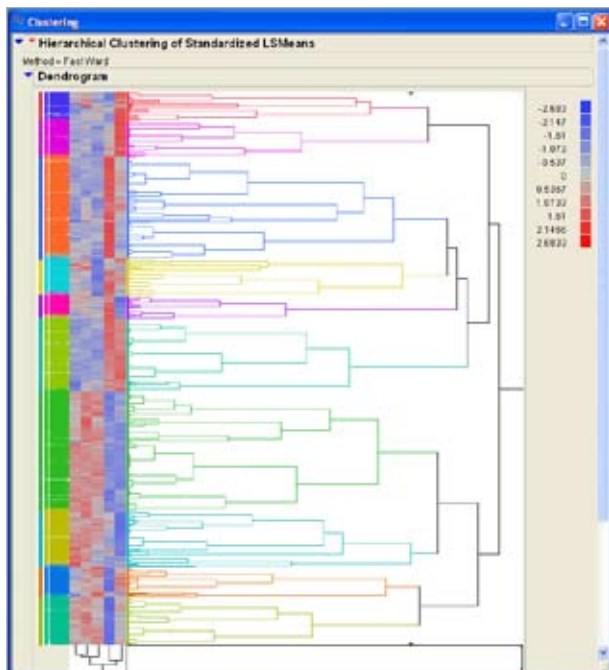
"Then at the end of the term, the models are really quite sophisticated;

they're insightful. They're really good, and they're much better this year than the first – mainly, I think, because JMP shows them ways to begin to visualize data, and it emphasizes that visualization is really a helpful tool with large data sets."

Far left: Heat map and cluster analysis of thousands of transcripts simultaneously shows relationships among transcripts and tissue types of *Chamaecrista fasciculata* at different developmental stages.

Left: Volcano plots allow students to select a few genes of interest for further bioinformatic and wetlab analysis.

Singer says the students are able to synthesize information much more effectively than they were before they began using JMP Genomics. "It was fascinating to me how compelling the volcano plots were."

True partners

Meanwhile, the students are engaged as true partners in real science.

"A neat hook with this work is the research that's come out showing that if you make synfuel from mixed-prairie plants it's much more energy-efficient than, say, corn ethanol, and you do less harm to the environment," Singer says. "So understanding more about this plant has implications for biofuels." The students' research may provide insight on whether delaying flowering increases the biomass of the plant.

In accordance with the stipulations of this NSF-funded project, Singer is collecting extensive assessments on the students' progress. "In addition, I have observers who have been coming to observe classes for the past two years. This year, they've found the change in student higher-order thinking, hypothesis generation and strategy choices to be substantially different from the pre-JMP days.

"I'm thrilled."

Carleton genetics students (from left) Anna Newman, Margaret Taylor and Kelly Mayo look at research results with senior research scientist Sonja Maki.

From left, Madelyn Lenhard, an undergraduate research student in the Singer lab at Carleton, examines *Chamaecrista fasciculata* plants with genetics students May Dixon, Anna Newman and Margaret Taylor.