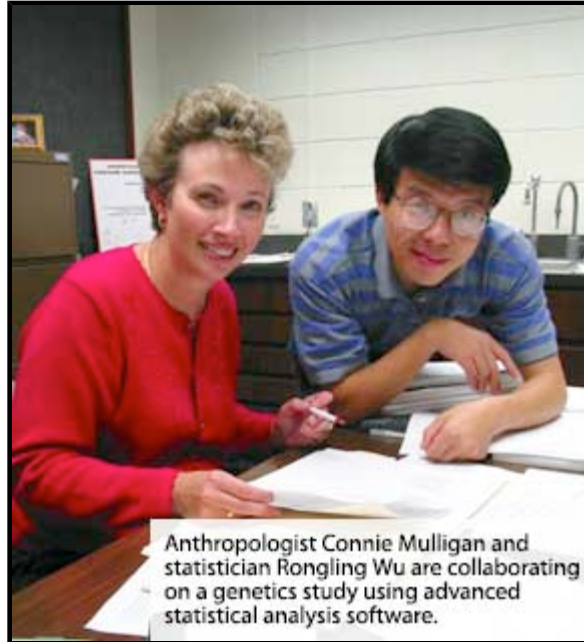


Summing Up the Genome

Statistical Genetics Collaboration Examines Wealth of New Data

"If you do the experiment right the first time, you don't need to use statistics" is an old adage among scientists that might make statisticians cringe. But while some scientists still choose to analyze their own data, many have realized they need a more sophisticated statistical approach to obtain better results. "Researchers might be looking to associate a trait, such as height, weight or growth, with a certain gene, but many geneticists cannot get by anymore by doing simple statistical t-tests," says George Casella, chair of UF's statistics department. "Now, we're dealing with much more complicated data sets, so a more complex analysis must be done, and this is where statistical genetics plays a role."



At UF, a group of more than 40 faculty members and students from across campus who work as geneticists and statisticians have formed the Statistical Genetics Group. "A few years ago, we started having a weekly seminar series where we would come together and talk about what we do and how we could assist each other," Casella says. "We've brought together folks from CLAS, IFAS and medicine, and sometimes we would have a scientist give a basic lecture about DNA and RNA, and then we would review some basic statistical analysis techniques. Now, some of us have started collaborating on various projects, and we're working with researchers at other universities in the US and internationally."

For the non-geneticist or non-statistician, a brief history lesson might help explain why this field has become even more important during the last 20 years. Genetics has its origins with Gregor Mendel (1822-1864) who derived basic laws of heredity such as: hereditary factors do not combine but are passed intact; each member of the parental generation transmits only half of its hereditary factors to each offspring (with certain factors "dominant" over others); and different offspring of the same parents receive different sets of hereditary factors. Mendel's work became the foundation for modern genetics.

Statistical genetics has its origins in the work of R.A. Fisher, S. Wright, and J.B.S. Haldane in the 1920s and 1930s. They realized that observable genetic variation

could be interpreted using probabilistic modeling, rigorous statistical analysis and well-founded scientific inference.

Statistical genetics has even more relevance today, since the Human Genome Project was completed in 2003. The project, which began in 1990, is a scientific effort to map and sequence the three billion chemical pairs that make up human DNA and identify the roughly 100,000 genes that comprise a person's genetic code. The challenge currently facing scientists is finding a way to organize and catalog this vast amount of information into a usable form. They are also trying to understand the genetic variation within and among individuals, populations and species. Both of these goals are intrinsically statistical and fall within the realm of statistical genetics.

"The completion of the Human Genome Project has resulted in a wealth of new data that must be carefully analyzed in order to reap the promised benefits of the project," says Connie Mulligan, an assistant professor of anthropology and associate director of UF's Genetics Institute. "It's complicated, but it's the next logical step if we're going to start determining relationships between certain genes and certain diseases." Mulligan, who worked at the National Institutes of Health (NIH) before coming to UF in 1999, has worked on several studies to determine which genes possibly increase or decrease the risk of alcoholism.

"When I was at NIH, we looked for genetic variants that increase or decrease the risk of developing alcoholism," she says. "Two variants, ADH1B and ALDH2, had been identified that appear to protect against alcoholism. These gene products have altered kinetic activity that results in the accumulation of acetaldehyde, which produces facial flushing, an accelerated heart rate and nausea, known as the 'flushing response.' These variants are present at high numbers in Asian populations, and the flushing response makes drinking unpleasant, so people don't drink, and there is a lower risk of alcoholism."

Now, Mulligan is looking at additional variants in the same two genes in a different population, American Indians, to determine if there are other variants that could lead to alcoholism. She is using a new statistical software package developed by UF statisticians to analyze the pile of clinical data. "This new program incorporates epistatic effects. Usually, we assume that each gene acts independently, when in fact that is probably not the case. Epistasis is when two genes interact, so their net effect is more or less than the total effect would be if you just added those two effects independently." Mulligan says a good example of this type of effect is evident in the recent research findings related to hormone replacement therapy, where estrogen in humans seems to have the opposite effect of estrogen in rats, in terms of heart disease and cancer. "In this case, it may be because in the clinical studies, an extra hormone was added for humans that may have interacted with the estrogen and modified its effects," explains Mulligan.

Associate Professor of Statistics Rongling Wu collaborated with Chang-Xing Ma, from the College of Medicine, and Casella to develop the model. Wu says the software took about six months to develop. "It was designed for high-resolution mapping of complex traits and can help geneticists precisely identify the location of genes (for diseases, plant size or milk yield) on the genome. This model is one of the most advanced in the genetics literature." Wu says traditional models for mapping complex traits are a combination of genetics and statistics, whereas this new model represents integration among genetics, statistics and general biology.

Mulligan says without the software, she would have stopped her study. "We published one paper this summer, but I thought I was finished with that data set," she says. "Now it's worth pursuing because there is a new way to analyze the data and possibly obtain more meaningful results. It would have been almost impossible to analyze these data further without a more sophisticated technique."

Another faculty member who has utilized the Statistical Genetic Group's consulting service is Assistant Professor of Zoology Marta Wayne. Wayne, who specializes in evolutionary genetics, is no stranger to collaborating with statisticians. Since she was a postdoctoral research fellow at North Carolina State University, Wayne has collaborated with a statistical geneticist on various projects. "I collaborate with Lauren McIntyre at Purdue University, and it's a longstanding collaboration, nearly 10 years. Even though I teach at NC State's Summer Institute in Statistical Genetics, I am an empirical geneticist, not a statistician. My specialty is applying the theory to the data, but to do the hard stuff, I need my collaborators!"

Wayne has brainstormed with Casella on a study she would like to pursue involving *Drosophila melanogaster*, or fruit flies. "There is an overall pattern we see in fruit flies of laying eggs over the course of their lifetime. The majority of female fruit flies have their peak of laying eggs earlier in life, but sometimes the flies lay eggs constantly, and sometimes it's reversed with the most eggs produced later in life. These exceptions appear to be genetic, but we need to develop a way to statistically evaluate this pattern and the variances within it." Since fruit flies are a model organism, Wayne's research on timing of reproduction could have implications for other organisms, including people.

In addition to collaborating with researchers from UF and other universities, Casella says another main goal is to establish a PhD program in statistical genetics at UF. "The new UF Genetics Institute will help us bring in faculty in this area, and we're already teaching some statistical genetics courses. A strong PhD program would put UF on the map as a place of research and teaching in this growing field." Casella and Wu also are writing a textbook, *Statistical Genetics of Complex Traits*, which will be published in 2004.

Within the next decade, Casella says he expects the field to advance even more. "We're starting to understand more and more about the genetic profile of humans and how this relates to health and disease. For example, one day, we'll be able to take a drop of blood from someone which contains their DNA and tell that person what medication would work best based on their genetic make-up. It's an important direction for scientists and statisticians to be moving in since the demand for this type of research will only increase, and much of it can only be accomplished using the expertise of each other."

--Allyson A. Beutke

Photo:
Jane Dominguez

[Return to Index](#)

<http://clasnews.clas.ufl.edu/news/clasnotes/0312-0401/genome.shtml>
Tuesday, 23-Dec-2003 14:33:48 EST