

Ordinal Outcomes

Archer group develops methods for predicting diagnoses

Science can be expressed in a myriad of different ways, but in the end, it all comes down to data, and methods are needed to appropriately interpret this data. Kellie Archer, Ph.D., professor and chair of the biostatistics division in the College of Public Health at The Ohio State University, leads a research group that develops statistical methods and computational algorithms for analyzing genomic data in oncology where oftentimes categorizing stages of the disease is relevant to tumor characteristics.

“In pathology, most of the diagnoses are recorded on an ordinal scale, and there’s lots of different ordinal outcomes that people are familiar with, like stage of cancer,” Archer said. “There’s not a numerical relationship between pathological stages, but there’s an order to them.”

While many methods have been developed for continuous outcomes, such as quantitative traits like systolic blood pressure or level of prostate specific antigen (PSA), and for two-class outcomes (diseased or not diseased), there aren’t as many for ordinal outcomes. Archer is trying to take some of the guesswork out of predicting these types of ordinal data where samples are hard to come by and there are a high number of variables involved.

“A lot of methods have been developed for continuous outcomes... but there’s not as many for ordinal outcomes,” Archer said. “One of the problems with a lot of statistical methods is that they were developed in situations where you have a large sample size relative to the number of variables that you’re using to predict the outcome. Whereas in high-dimensional gene expression or other genomic data sets, you have the reverse problem.”

Archer’s group began with frequentist-based approaches to their statistical methods, which are useful in that they can identify variables important for modeling the ordinal outcome,

but they do not provide p-values, standard error estimates or confidence intervals. Archer is now working on a Bayesian approach which incorporates probability distributions for the model parameters. This will provide more useful information for each predictor variable so that researchers will have better information about how important a variable is in the model.

The researchers are running their own software on the Owens Cluster at the Ohio Supercomputer Center to perform their statistical methods and process huge data sets. They are also using simulation studies to evaluate and continuously improve their methods.

“We design a lot of simulation studies where we can vary different scenarios: we vary the number of samples, we vary the number of variables, we vary the strength of the relationship between the variables and the outcome,” Archer said. “Then we run our methods using a huge number of simulation studies to try to evaluate their performance under a wide variety of settings. And it would take forever if we just used our office computer to do that.” ●

PROJECT LEAD // Kellie Archer, Ph.D., The Ohio State University
 RESEARCH TITLE // Informatic methods for predicting an ordinal response
 FUNDING SOURCE // National Institutes of Health
 WEBSITE // cph.osu.edu/people/karcher

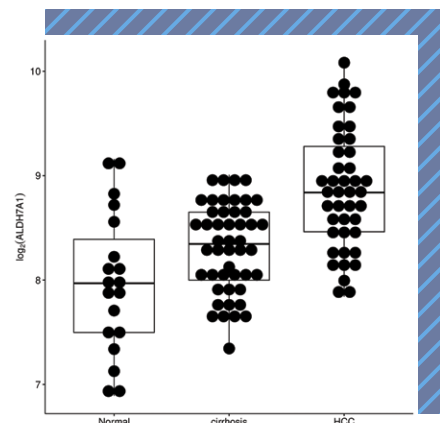


FIGURE // In examining genes involved in the progression from normal liver to cirrhosis to hepatocellular carcinoma (HCC), Kellie Archer’s group identified several whose expression is either increasingly or decreasingly dysregulated. This figure represents one such gene, ALDH7A1.