

A cloud computing system to quickly implement new microarray data normalization methods

Dajie Luo^{1*}, Prithish Banerjee^{1*}, James Harner¹ and Dongquan Chen²

Dept. of Statistics and Medicine²

¹West Virginia University, Morgantown, USA

²Univ. of Alabama at Birmingham, Birmingham, USA

The data normalization is crucial to the analysis of microarray related datasets. It takes time and efforts for newly-developed algorithms to be built into a traditional commercial software application or locally developed software tools with limited functionalities. While most of the new algorithms emerge in the form of sharable R packages in Bioconductor, e.g., it is difficult for biologists who may not possess programming skills to use them as soon as they are available. Currently, we rely on statistician and experienced programmer to develop code to access those R packages that are distributed and utilized locally. We need a quick deployment procedure to implement newly-developed normalization methods as they appear.

The newly emerged cloud computing concept in IT world direct us a new way to provide an easily accessible service to the biologists without requiring them any knowledge about R and programming skills to implement them. Based on our earlier Java-based software tool JavaStat, we developed such an internet-based application prototype that can be used to upload data, run normalization, run analysis, and download results. More importantly, R packages, e.g., for newly-developed normalization method, GC-RMA for Exon arrays, e.g., can be easily incorporated into the system without much imputes from a biologist or a programmer. The data are stored in the cloud and the R code will run on server. The similar system will speed up the accessibility of new algorithm including new normalization methods in R.

Chen@mail.dopm.uab.edu