

Name

Solaiappan Manimaran

Email

mani2012@bu.edu

Institutional Affiliation

Boston University

Campus

Medical Campus

School

School of Public Health

Department

Biostatistics

Position Held at Institution

Graduate

Poster Submissions

Poster Title

BatchQC: Interactive software framework for evaluating sample and batch effects in genomic data

Authors and their Affiliation

Solaiappan Manimaran(Boston University), Evan Johnson (Boston University)

Abstract Submission

- [2015-GSI Mani-Abstract.docx](#)

Would you like your abstract to be considered for an oral presentation (students and post docs only)?

Yes

BatchQC: Interactive software framework for evaluating sample and batch effects in genomic data

Authors: Solaiappan Manimaran, Evan Johnson

Sequencing and microarray samples often need to be collected or processed in multiple batches or at different times. This often leads to known and unknown technical biases that unless properly adjusted for, can lead to incorrect results in the downstream analysis. There are several batch adjustment tools that are currently available, but none of them can indicate a priori whether batch adjustment need to be conducted and how it should be applied before proceeding with the analysis. We present a new software pipeline called BatchQC that addresses these issues by visually and statistically evaluating whether batch effects exist and can automatically adjust for these effects using the best possible methods. We have used the BatchQC pipeline on both simulated data and real data samples to establish the efficiency of the software pipeline. Another issue with batch adjusted data is the presence of correlation of samples, which has to be properly accounted in all downstream analysis of expression data. Here, we have developed a function to automatically deduce the number of batches present in batch adjusted data based on the presence of correlation in batch adjusted data. It also constructs the batch design matrix of the samples based on the correlation matrix. Once the batch design is inferred, all downstream analysis of the expression data is accurate without any interference of the batch effects. To demonstrate this using an example, we have applied this technique on simulated and real data examples.

Keywords

- 1: BatchQC software
- 2: Batch Effects visualization
- 3: Next-gen Sequencing and microarray data analysis