Statistical assessment of depth normalization methods for microRNA sequencing

Jian Zou

University of Pittsburgh
E-mail: jian.zou@pitt.edu

Abstract: Quality data is the foundational cornerstone for reliable scientific findings in evidence-based medical research. It is widely accepted that a crucial step to derive high-quality genomics data is to identify data artifacts caused by systematic differences in the processing of specimens and to remove these artifacts by data normalization. One major and unique aspect of RNA sequencing data normalization is the 'depth of coverage'. Statistical methods for depth normalization have been recently developed, including both simple rescaling-based methods and regression-based methods. Many of these normalization methods rely on the presupposition that variations in the assumed scaling factor or in the projection of the assumed regression function are solely due to data artifacts and should be removed. MicroRNAs are a unique class of small RNAs regulating gene expression and closely linked to carcinogenesis. They are low-complexity molecules (that is, a small number of molecules expressed dominantly) that tend to be expressed in a tissue-specific manner, especially in heterogeneous samples such as tumors. As a result, the assumption of depth normalization methods may not hold for microRNA sequencing. We performed a study to assess the performance of existing depth normalization methods on identifying disease-relevant microRNAs using both a pair of datasets on the same set of tumor samples and data simulated from the paired datasets under various scenarios of differential expression. In this talk, we will report our findings from this study.