

Title: Novel statistical approaches for data analysis in RNA interference high-throughput Screening technology

Abstract:

RNA interference (RNAi) high throughput screening (HTS) allows the identification of genes associated with specific biological phenotypes. This technology has been hailed as the second genomics wave following the first genomics wave of gene expression microarrays and single nucleotide polymorphism (SNP) discovery platforms. HTS experiments generate a huge amount of data. In order to use these data to identify the most effective siRNAs tested, it is critical to adopt and develop appropriate statistical methods for quality control (QC) and hit selection. For QC in HTS assays, J-H Zhang et al (1999) proposed the popular Z-factor and XD Zhang *et al* (2006) proposed strictly standardized mean difference (*SSMD*) and coefficient of variability in difference (*CVD*). For hit selection, mean $\pm k$ SD (or zscore) was initially used. To be more robust to outliers and skewness, Brideau, Gunter and Pikoulis *et al* (2003) proposed B-score method, XD Zhang *et al* (2006) proposed a quartile-based method and suggested using median $\pm k$ MAD or quartile-based methods. Malo and Hanley *et al* (2006) reviewed the methods developed before early 2006. Recently, XD Zhang has developed an *SSMD*-based process for hit selection in RNAi HTS experiments. This process maintains a flexible and balanced control of both false negatives and false positives. Here, XD Zhang first present *SSMD*, β -probability, the *SSMD*-based process for hit selection, and their applications and then compare the *SSMD*-based methods with other methods in RNAi HTS experiments.