

1 Purpose

Application of scripted computer software is a powerful tool for data analysis of molecular biology data sets. Script design enables rapid and high-throughput statistical testing and data display, facilitating interpretation of experimental results.

2 Script design

2.1 Data Source

Comma Separated Value (.CSV) data structures are a common data format generated by a host of software and hardware platforms. Currently, a large portion of molecular biology data analysis is carried out using basic software packages such as Microsoft Excel or Prism, both of which routinely output .CSV file formats. Generation of data analysis code for interpretation of .CSV files is therefore highly applicable to current molecular biology applications and was thus selected as a data source for analyses.

2.2 Experiment Design & Statistical Analysis

Real-time quantitative PCR (RTqPCR) represents a common molecular biology technique capable of generating large data sets. Multiple RTqPCR experiments were assessed. In each experiment, replicates pertaining to each individual sample were assessed for percentage coefficient of variance (%CV). %CV was averaged across all samples for each experiment, such that an Experimental %CV was defined. Experimental %CV was plotted across experiments thereby displaying experimental reproducibility and accuracies across multiple similar, but not identical, RTqPCR-derived data sets.

3 Script Elements

- The script determines the assay container (96 or 384-well plate) of the experiment, without prior input. Each experiment assayed seven distinct samples. By determining the total number of wells under assessment and knowing the total number of samples, the code determines the number of technical replicates for each sample in addition to the plate type.
- The row formatting of the .CSV data sources was not sequential. The code therefore, once the number of calculated replicates for each sample had been determined, identified the correct row in the data file for each data point and identified the corresponding Ct value for each.
- For each sample, comprised on multiple technical replicates, mean and standard deviation are calculated by the code in order to determine %CV. An Experimental %CV is then calculated and displayed in graphical format
- The number of experiments assessed is not pre-determined, therefore the code appends to lists from which the graphical data is retrieved. Resultantly, indefinite experiments can be sequentially analysed and displayed
- For subsequent data representation (using third party software), an output folder is filled with the mean Ct and %CV values. These are generated in .CSV format for copy/paste applications. The generated .CSV file is appended to such that sequential experiments can easily be added

4 Application

Such data analysis packages are inherently flexible, amendable and time-efficient. Their applications extend to any experimental workflow wherein numerical datasets are created and require statistical testing.