



Data Analysis Gene Expression

User Manual v1

(May 2013)

Contents

1. Introduction	3
2. Requirements	4
3. Menu options	4
4. Getting Started	6
4.1. Installation	6
4.2. Import result files and setting parameters	6
4.3. Results	9
4.4. Result charts	10
5. Other utilities	12
5.1. “Efficiency between genes”	12
5.2. Coefficient of variation	13
6. License	14
References	14

1. Introduction

Data Analysis Gene (DAG) Expression software is a simple and free application tool developed in vb.net to simplify the management and analysis of high-throughput gene expression data obtained by real time quantitative PCR (qPCR). The software allows relative quantification using the standard curve method with data from different detection systems (examples: BioMark™ HD System, ABI PRISM® 7700, OpenArray® System). Also, data formatted by the user can be imported and analyzed by the program. The normalization method is based on the principles and formulas described by Vandesompele *et al.* (2002). Single or multiple genes can be used as a normalizer.

Main Features

- Import multiple Fluidigm Real-time .csv results files, SDS 2.1 and above .txt results files, OpenArray® Real-Time .csv results files and user-formatted results files.
- Save project analysis and export results into a .txt file.
- Real-time chart representing standard curves and relative efficiencies between genes, including PCR efficiency and coefficient of determination (R^2).
- Automated analysis of gene expression stability and inter-run coefficient of variation.
- Perform relative standard curve analysis on Ct data and sample normalization with one or more selected controls to obtain normalized data.
- Graphic results visualization via bar-plots.
- Friendly-user and interactive graphical software.


2. Requirements

- Windows XP / 7
- .NET Framework 4


3. Menu options

File


New

This action closes the current file and clears data and creates a new project. A shortcut icon is also available: 

Open

This action opens the previously saved file “Data Analysis Gene Expression (.dag)”. A shortcut icon is also available: 

Save / Save As

This action saves data to .dag file. If the file is new and this is the first time you save it, type a name for the file in the File Name box and then click Save (shortcut icon: ). If you want to save the opened file with another file name click the File menu and then click Save As.

Export

Click the File menu – Export and then click “Text file extension”. This action exports the results data to .txt file.

Import

Click the File menu – Import and then choose the file format to import the data (Fluidigm csv file, SDS txt file, OpenArray csv file or generic file). This action imports results files from the Fluidigm Real-Time PCR analysis software (.csv file), SDS 2.1 and above software (.txt file), OpenArray® Real-Time qPCR analysis software (.csv file) or user-formatted (.csv or .txt files) to the current project.

Note: For User-formatted input data, a sample template (.csv or .txt) may be downloaded by the user. It consists of a sample file that can be filled in by the user with the information required by the software: assay name, sample name and Ct values separated by semicolon. First of all, this file has to be saved in your computer and then you can fill in the data (Fig. 1).

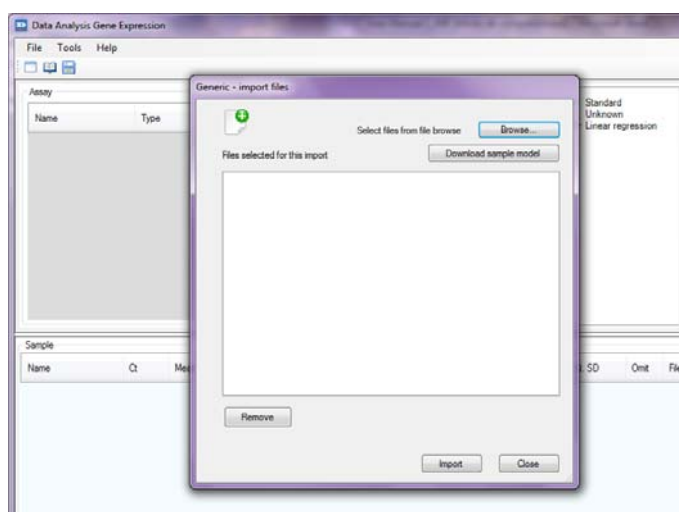




Figure 1. Import generic files

Tools

Charts

This action allows the visualization of normalized data via NQ Plots (NQ Plot:  or NQ Plot by Group: .

Coefficient of variation

This action calculates the inter-run coefficient of variation.

Find control gene-stability

This action calculates the more stable selected controls for normalization (Vandesompele *et al.* 2002).

Options

Decimal Separator: if your data uses the dot or decimal point as separator you must select “English (.)” to import your data. On the other hand, if your data uses a comma as decimal separator, you must select “Other languages (,)”.

Help

DAG Expression Manual

Example Results Data

A demo-experiment (example results data .dag) consisting of a microfluidic dynamic arrayTM IFC (48.48) containing 48 assays (4 reference genes) and 48 samples are available in the Help menu to become familiar with the DAG Expression software.

4. Getting Started

This chapter gives you both a brief overview on installing DAG Expression software and how to analyze your qPCR data.

4.1. Installation

The program doesn't need installation, it is zipped into a file and you can unzip with a standard program to a folder or desktop. In more systems Framework .Net is installed, but if the program gives an error message, you must download and install framework .Net 4 or above (<http://www.microsoft.com/en-us/download/details.aspx?id=17851>).

4.2. Import result files and setting parameters

4.2.1. Set Delimiter on Menu Options if you work with “.” or “,” as decimal separator. Default value: “.”.

4.2.2. Create a new project. Import all necessary output files of different software (Fluidigm csv or SDS txt files, etc) or user-formatted input data to perform the analysis. DAG expression will organize all the data by assay.

4.2.3. Set the assay design (Fig. 2)

Assay Name: This is the name of the assay provided by the user. Normally, it refers to the specific genes of study.

Assay Type: Click the assay name and assign the assay type (selected Control or Target) to the imported assays.

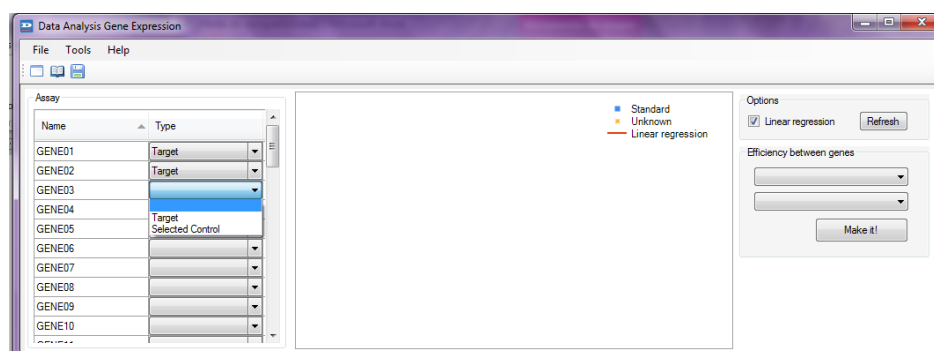


Figure 2. Datagridview of assays

4.2.4. Set the sample design (Fig. 3)

Sample name: This is the name of each sample previously defined by the user and imported with the results files. Samples with the same name are considered by the program as technical replicates.

Please note that before enter the sample type, it is necessary to click into the assay name.

Sample type: Assign the sample type (Standard, Unknown and Non Template Control-NTC) to the samples. This step is very important to construct the relative standard curve, calculate the gene expression stability (geNorm M value) and perform the subsequent analyses.

Note: The sample type is automatically updated from results data (.txt) imported from SDS program.

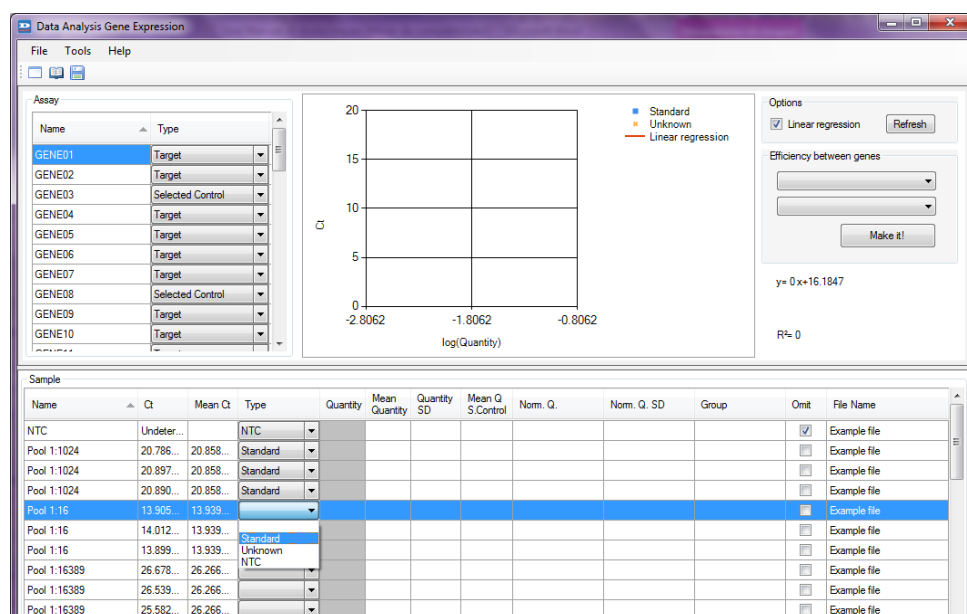


Figure 3. Datagridview of samples

Sample quantity: this column indicates the relative quantities of Standards and Target (Unknown) samples.

Standard quantity: Enter the quantity values of your Standard samples (Fig. 4) to draw the standard curve and extrapolate the quantity of Unknown samples using the linear regression analysis. For relative

standard curve quantification, it is important to know the relative standard dilutions, although the units to express them are irrelevant (example: if four-fold dilutions of a cDNA pool from 10 liver samples are prepared, the units could be the dilution values 0.25, 0.0625, 0.015625, 0.00390625, and so on). Following the recommendations of “The MIQE Guidelines”, the dynamic range should cover at least 3 orders of magnitude (Bustin *et al.* 2009). In this step it is important that the standard curves of the assay and the selected reference controls (see below) consist of the same number of n-fold serial dilutions. Please note that in the demo-experiment the first dilution (1/2) has been omitted due to PCR inhibition.

Note: The standard quantities are automatically updated from results data (.txt) imported from SDS program

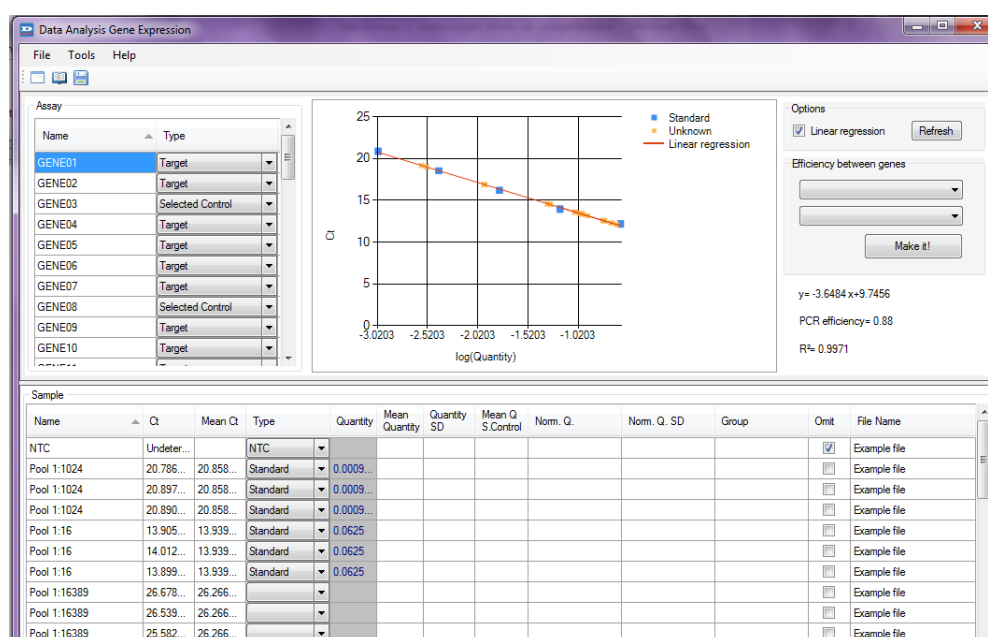


Figure 4. Main window with the quantity column

Note: Assay Name and Sample data columns can be sorted in ascending or descending order by clicking the column heading.

4.2.5. Identify and select the most stable expressed reference genes to perform the analysis. In the Tools, select the “Find control gen-stability” option (Fig. 5). The lowest M value indicates genes with the most stable expression (Vandesompele *et al.*, 2002).

Once the reference genes are selected, the program calculates the arithmetic mean of the mean quantity value of each gene to produce a normalization factor.

Note: To perform this analysis is necessary to select at least three selected controls with the corresponding Unknown samples of each assay. Also the M value will be only calculated if selected control genes have the same samples data (example: if sample 04 has missing data for the selected-control gene 01 but no for the other selected-control genes, the analysis cannot be performed. Sample 04 has to be deleted from all the selected-control gene assays to enable the analysis.)

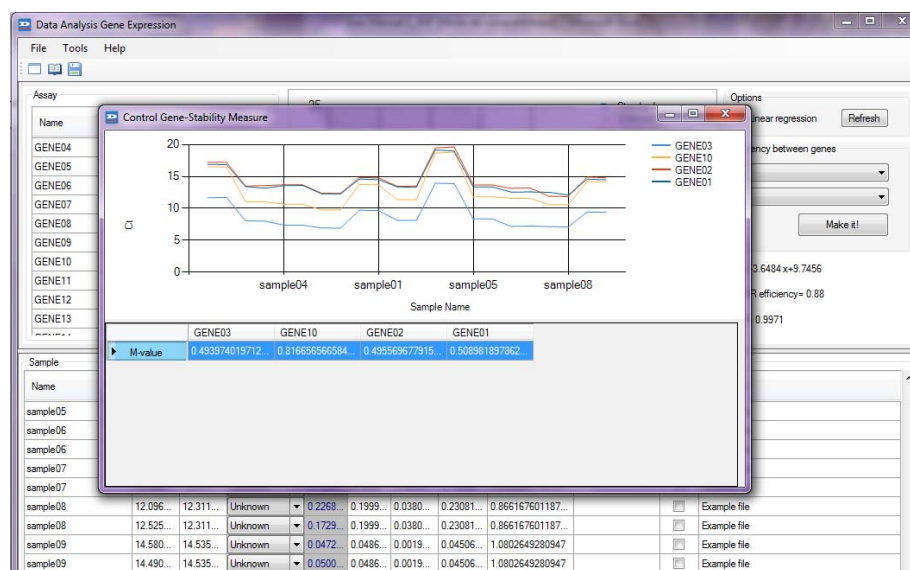


Figure 5. Control Gene Stability analysis

4.3. Results

By clicking in each Assay name, the program will show you a list of samples with results (Fig. 6):

- **Mean:** Average Ct value of sample replicates by assay
- **Mean Quantity:** Average quantity value of sample replicates by assay
- **Quantity SD:** Standard deviation value of quantity values by assay
- **Mean Q S. Control:** Average quantity value of quantity values by selected control assays
- **Norm Q.:** Normalized quantity value
- **Norm. Q. SD:** Standard deviation of Norm Q.

Sample												
Name	Ct	Mean Ct	Type	Quantity	Mean Quantity	Quantity SD	Mean Q S Control	Norm. Q.	Norm. Q. SD	Group	Omit	File Name
sample01	14.585...	14.554...	Unknown	0.0471...	0.0481...	0.0013...	0.04161...	1.156092279145...	0.062010870823...		<input type="checkbox"/>	Example file
sample01	14.522...	14.554...	Unknown	0.0490...	0.0481...	0.0013...	0.04161...	1.156092279145...	0.062010870823...		<input type="checkbox"/>	Example file
sample02	13.156...	13.290...	Unknown	0.1161...	0.1071...	0.0128...	0.10010...	1.070374937556...	0.151968588271...		<input type="checkbox"/>	Example file
sample02	13.425...	13.290...	Unknown	0.0980...	0.1071...	0.0128...	0.10010...	1.070374937556...	0.151968588271...		<input type="checkbox"/>	Example file
sample03	13.362...	13.348...	Unknown	0.1020...	0.1029...	0.0012...	0.09095...	1.131471986464...	0.016953915199...		<input type="checkbox"/>	Example file
sample03	13.335...	13.348...	Unknown	0.1037...	0.1029...	0.0012...	0.09095...	1.131471986464...	0.016953915199...		<input type="checkbox"/>	Example file
sample04	13.572...	13.583...	Unknown	0.0893...	0.0887...	0.0009...	0.12110...	0.732636926317...	0.053863493806...		<input type="checkbox"/>	Example file
sample04	13.595...	13.583...	Unknown	0.0880...	0.0887...	0.0009...	0.12110...	0.732636926317...	0.053863493806...		<input type="checkbox"/>	Example file
sample05	13.352...	13.353...	Unknown	0.1026...	0.1026...	7.0469...	0.08210...	1.250184379653...	0.011004145689...		<input type="checkbox"/>	Example file
sample05	13.353...	13.353...	Unknown	0.1026...	0.1026...	7.0469...	0.08210...	1.250184379653...	0.011004145689...		<input type="checkbox"/>	Example file
sample06	12.272...	12.286...	Unknown	0.2030...	0.2012...	0.0026...	0.20714...	0.971363055029...	0.016837921624...		<input type="checkbox"/>	Example file
sample06	12.301...	12.286...	Unknown	0.1993...	0.2012...	0.0026...	0.20714...	0.971363055029...	0.016837921624...		<input type="checkbox"/>	Example file

Figure 6. Results table

You can omit samples in the analysis clicking on “omit” checkbox (Fig. 7) or leaving the sample empty.

Note: The program recognizes and omits the undetermined Ct values automatically.

Sample												
Name	Ct	Mean Ct	Type	Quantity	Mean Quantity	Quantity SD	Mean Q S Control	Norm. Q.	Norm. Q. SD	Group	Omit	File Name
NTC	Undeter...		NTC								<input checked="" type="checkbox"/>	Example file
Pool 1:1024	20.786...	20.858...	Standard	0.0009...							<input type="checkbox"/>	Example file
Pool 1:1024	20.897...	20.858...	Standard	0.0009...							<input type="checkbox"/>	Example file
Pool 1:1024	20.890...	20.858...	Standard	0.0009...							<input type="checkbox"/>	Example file
Pool 1:16	13.905...	13.939...	Standard	0.0625							<input type="checkbox"/>	Example file
Pool 1:16	14.012...	13.939...	Standard	0.0625							<input type="checkbox"/>	Example file
Pool 1:16	13.899...	13.939...	Standard	0.0625							<input type="checkbox"/>	Example file
Pool 1:16389	26.678...	26.678...									<input checked="" type="checkbox"/>	Example file
Pool 1:16389	26.539...	26.539...									<input checked="" type="checkbox"/>	Example file
Pool 1:16389	25.582...	25.582...									<input checked="" type="checkbox"/>	Example file

Figure 7. Omit checkbox column

4.4. Result charts

4.4.1. Standard Curve charts

Displays Ct vs log Quantity (\log_{10}). The scatter plot represents the current assay selected (Fig. 8).

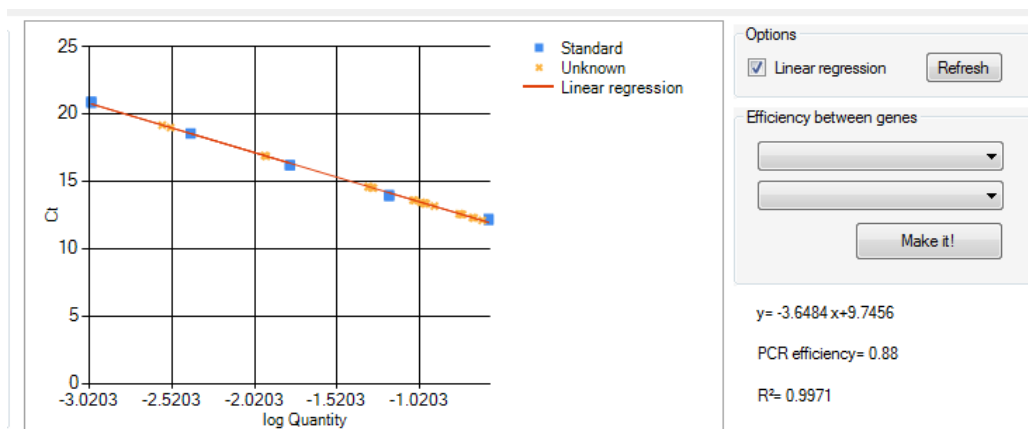


Figure 8. Standard curve

On the top-right of the application you can find options to both refresh the chart and show/hide the regression line. The PCR efficiency calculated using the slope of the linear regression ($E = 10^{(-1/\text{slope})} - 1$) and the R^2 are indicated in the plot.

4.4.2. Results Bar Plots

This utility allows visualizing DAG expression normalized data via bar plots (NQ Plots):

NQ Plot: Displays NQ (normalized quantity) vs target or sample (Fig. 9). NQ data can be displayed as Linear or logarithmic (Log_2 and Log_{10}).

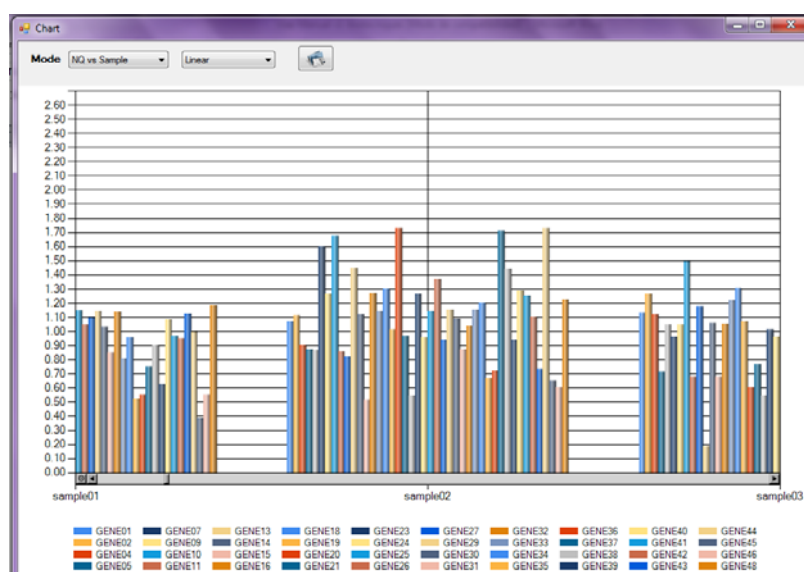


Figure 9. Bar plot displaying NQ vs sample

NQ Plot by Group: In the sample design box (Fig. 3), the user can assign samples to biological groups (examples: genotype 1, genotype 2,... or male vs female) by clicking in the Group box and entering the group name. Groups can be only assigned to Unknown samples. The software plots the arithmetic mean of NQ of each group vs target or group (Fig. 10). NQ data can be displayed as Linear or logarithmic (Log₂ and Log₁₀).

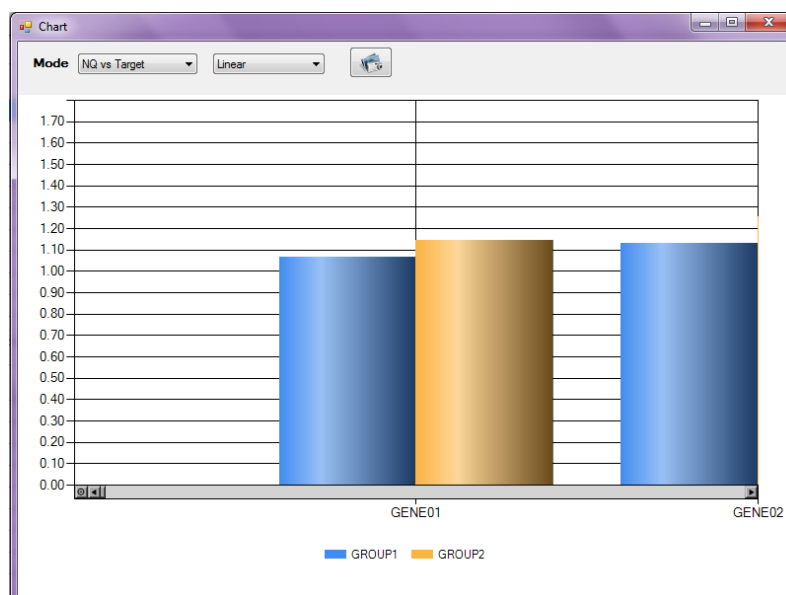


Figure 10. Bar plot by group displaying NQ of each group (1 and 2) vs target

Note: NQ plots can be saved as .jpg, .jpeg, .gif, .bmp, .png by clicking the shortcut icon available in the chart view.



5. Other utilities

5.1. “Efficiency between genes”

This utility allows checking if efficiencies of target and selected control are approximately equal, a prerequisite to use the $\Delta\Delta C_t$ method for quantification (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008). To validate it, the DAG expression program plots the log input amount vs ΔC_t (C_t of target gene $- C_t$ of selected control gene). If the absolute slope of the obtained trend line is <0.1 , the user can use the $\Delta\Delta C_t$ calculation.

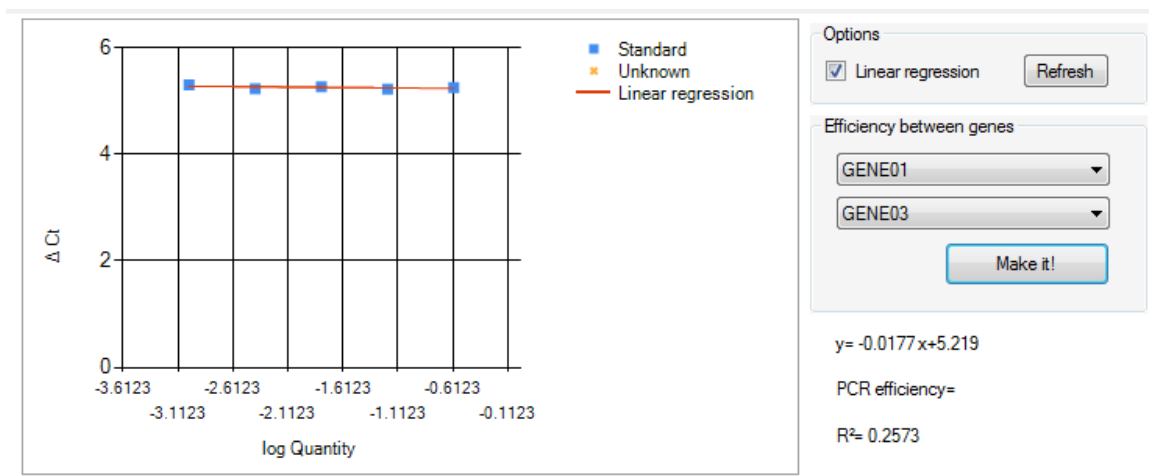


Figure 11. Relative efficiency plot

5.2. Coefficient of variation

This utility calculates the coefficient of variation inter-run (Fig. 12). If the user wants to compare the same assays between different samples distributed in different runs, an identical sample (Unknown) has to be added in all the independent runs performed.

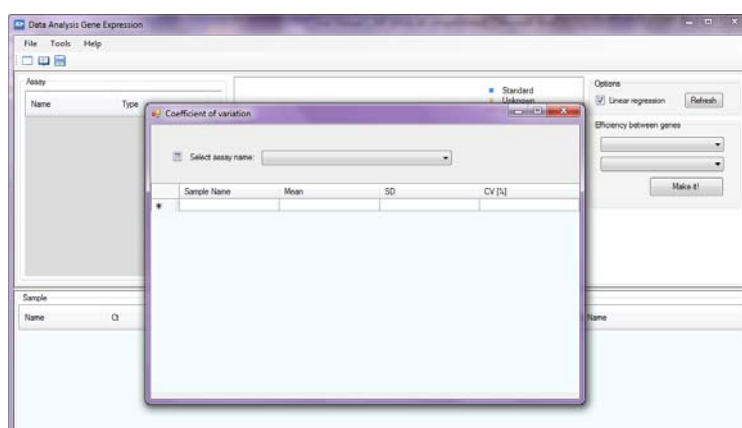


Figure 12. Inter-run coefficient of variation

6. License

Data Analysis Gene Expression is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License](#). This program is free software; you can copy, distribute and transmit the work under the following conditions: **Attribution** - You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). **Noncommercial** - You may not use this work for commercial purposes. **No Derivative Works** - You may not alter, transform, or build upon this work.

References

- Bustin, S.A. *et al.* (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry*, 55, 611-622.
- Livak, K.J., and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta C_t$ method. *Methods*, 25, 402-408.
- Schmittgen, T.D., and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, 3, 1101-1108.
- Vandesompele, J. *et al.* (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3(7), research0034.