DataAssist™ v3.0 Software User Instructions

Introduction

DataAssist™ Software is a simple, yet powerful data analysis tool for sample comparison when using the comparative C_T ($\Delta\Delta C_T$) method (Livak and Schmittgen, 2008) for calculating relative quantitation of gene expression. It contains a filtering procedure for outlier removal, various normalization methods based on single or multiple genes, and provides relative quantification analysis of gene expression through a combination of statistical analysis and interactive visualization.

What DataAssist™ Software does

The main steps in the analysis performed by DataAssist™ Software are:

- Read export files (.txt or .csv) of analyzed results from supported instruments and software
- Perform QC analysis on C_T data and associated plots
- Perform sample normalization for each assay
- Perform QC analysis on normalized data
- Perform relative quantification for sample comparison; perform t-test for biological group comparisons; and produce graphics to visualize test results.

Required Files and Formats

DataAssist™ Software is designed to work with .txt or .csv **results files** exported from Applied Biosystems Real-Time PCR instruments, and is able to process multiple Relative Quantitation (RQ) study files simultaneously. Thus, very large studies across a large number of plates or microfluidic cards can be analyzed as one data set.

DataAssist™ Software is compatible with analyzed real time **results** (C_T) data exported from the following versions of Applied Biosystems' software.

Table 1

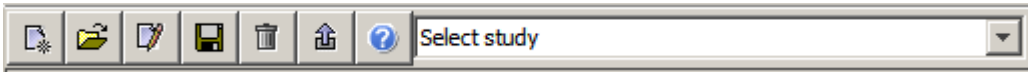
Applied Biosystems Instrument	Software Versions	Instrument File extension
Viia™ 7 Real-Time PCR System	Viia™ 7 Software v1.0 & v1.1 RQ	.edm
7900HT Fast Real-Time PCR System	SDS v2.2.2, 2.3 & 2.4 AQ Results	.sds
7900HT Fast Real-Time PCR System	SDS v2.2.2, 2.3 & 2.4 RQ Study Results* *(files must be in the Plate Centric Table Orientation)*	.sdm
ABI PRISM® 7000 Sequence Detection System	SDS v1.1 & v1.2 RQ Study Results	.sdm
7300 Real-Time PCR System	SDS v1.4 RQ Study Results	.sdm
7500 & 7500 Fast Real-Time PCR System	SDS v1.4 RQ Study Results	.sdm
7500 Fast Real-Time PCR System	Software v2.0 RQ Study Results	.edm
StepOne™ & StepOne Plus™ Real-Time PCR System	StepOne™ Software v2.2 RQ Study Results	.edm
OpenArray® Real-time PCR Instrument	OpenArray® Real-time PCR Software v1.0	Summary results.csv

For all data from Applied Biosystems' Real-Time PCR instruments, make sure that you have set appropriate baselines and thresholds for your experiments before exporting the data. For more information on properly analyzing results, please refer to the instrument user manual.

> ***Note: RQ data must be Results exported from an RQ study file, except StepOne Software v2.0 & OpenArray.***

The results files are imported during the set up of a new experimental study. The Sample Name, Assay Name, Assay Type, and C_T values are automatically entered into the study from the results file.

Helpful Shortcut Icons



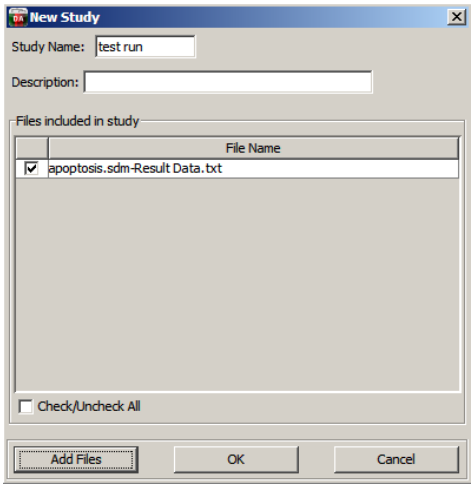
a b c d e f g h

a. Create a new study
b. Open saved study
c. Edit a study
d. Save a study
e. Delete (or close) a study
f. Export analysis results
g. Help
h. Toggle between studies

Create a New Study

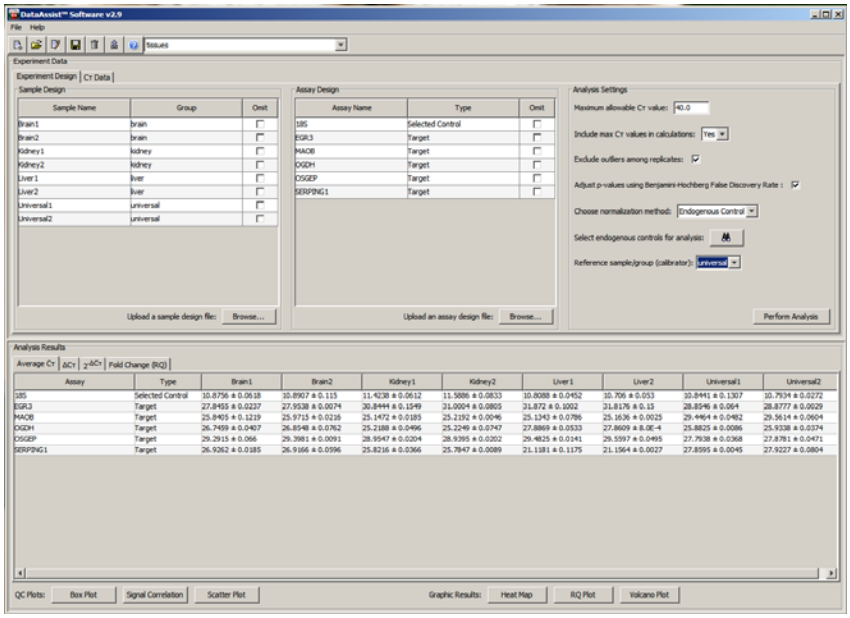
Create a new study either by clicking on the  icon in tool box or ***File > New…*** .

Enter a **Study Name** in the box at the top of the window.



Import results files (see Table 1 for supported file format) by clicking on the Add **Files** button in the New Study window, then click **OK**.

Samples will now appear in the *Sample Design* box and Assay information in the *Assay Design* box.



Exporting and Saving Study & Data Files

Exporting Data Files

Export results data from the study by either clicking on the  icon or navigating to ***File > Export***. In the **Export** dialog box that is opened, you may select individual files (csv or txt), or one file (.xls) with each table in a separate worksheet, to be exported. Tables that can be exported are the C_T Data, Sample Design, Assay Design, Average C_T, ΔC_T, $2^{-\Delta CT}$, and Fold Change files.

Open a Saved Study

Open a study you have previously saved by either clicking on the  icon in tool box or ***File > Open…***. Browse to find the file containing the results data from the desired study, the *.das* file containing the analyzed data, and select to open.

Experiment Data

Experiment Design

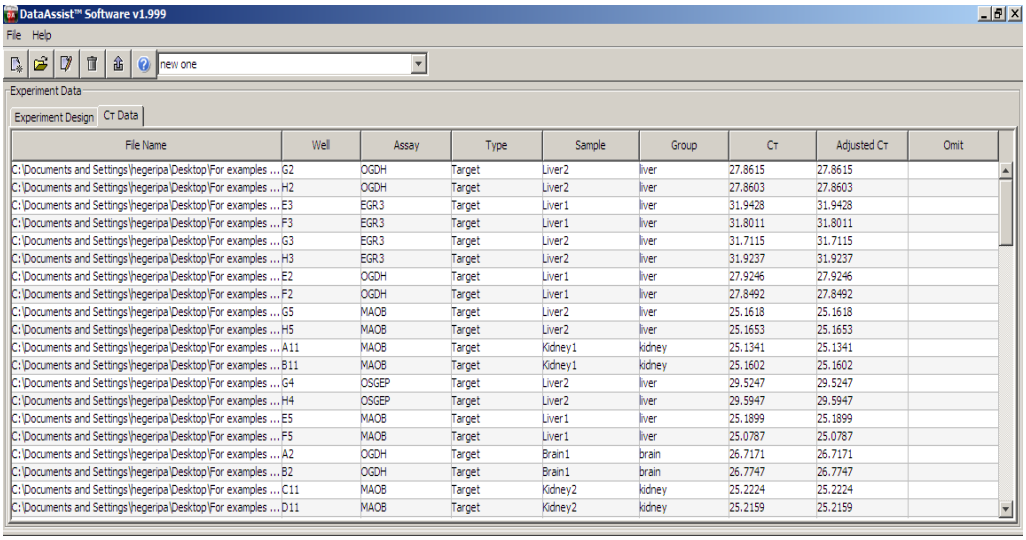
Sample Design

- **Sample Name**: The name given to each sample by the user is pre-defined in and imported with the results files. Data points in a study with the same sample and assay name are considered technical replicates.
- **Group**: The name of the biological replicate group (e.g. normal, disease or time point 1, time point 2).
    - Enter or import Sample Design information: To assign samples to biological groups, such as normal & disease or time point 1, time point 2, either click in the **Group** box and manually enter group name, or upload a sample design file **(see Appendix A)**. Click on **Browse**... to find and select your sample design file. If your samples are solely technical replicates, not biological replicates, no group assignment is necessary.

Note: Samples labeled as NTC (no template control) are not used in calculations.

C_T Data

The **C_T Data** tab displays a table with the following information from the real-time data files: imported file path, well number, assay name, assay type, sample name, group name, C_T value, Adjusted C_T value and reason for omission of sample well from calculation. Changes made to **Sample** or **Assay Design** information will be reflected in the **C_T Data** tab only after performing analysis.



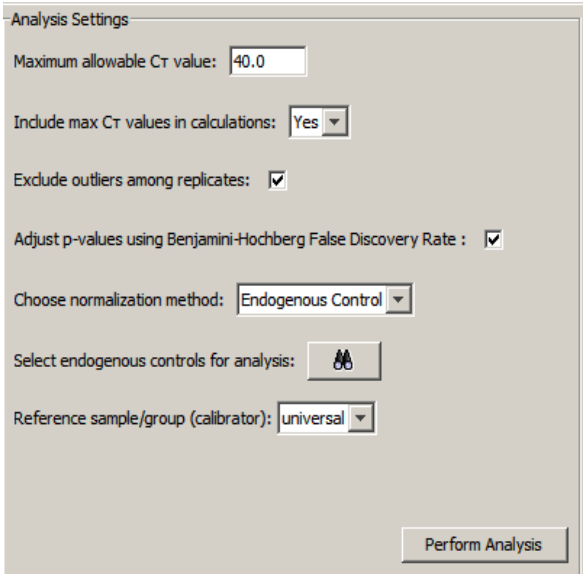
File Name	Well	Assay	Type	Sample	Group	CT	Adjusted CT	Omit
C:\Documents and Settings\hegeripa\Desktop\For examples ...	G2	OGDH	Target	Liver2	liver	27.8615	27.8615	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	H2	OGDH	Target	Liver2	liver	27.8603	27.8603	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	E3	EGR3	Target	Liver1	liver	31.9428	31.9428	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	F3	EGR3	Target	Liver1	liver	31.8011	31.8011	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	G3	EGR3	Target	Liver2	liver	31.7115	31.7115	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	H3	EGR3	Target	Liver2	liver	31.9237	31.9237	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	E2	OGDH	Target	Liver1	liver	27.9246	27.9246	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	F2	OGDH	Target	Liver1	liver	27.8492	27.8492	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	G5	MAOB	Target	Liver2	liver	25.1618	25.1618	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	H5	MAOB	Target	Liver2	liver	25.1653	25.1653	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	A11	MAOB	Target	Kidney1	kidney	25.1341	25.1341	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	B11	MAOB	Target	Kidney1	kidney	25.1602	25.1602	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	G4	OSGEP	Target	Liver2	liver	29.5247	29.5247	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	H4	OSGEP	Target	Liver2	liver	29.5947	29.5947	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	E5	MAOB	Target	Liver1	liver	25.1899	25.1899	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	F5	MAOB	Target	Liver1	liver	25.0787	25.0787	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	A2	OGDH	Target	Brain1	brain	26.7171	26.7171	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	B2	OGDH	Target	Brain1	brain	26.7747	26.7747	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	C11	MAOB	Target	Kidney2	kidney	25.2224	25.2224	
C:\Documents and Settings\hegeripa\Desktop\For examples ...	D11	MAOB	Target	Kidney2	kidney	25.2159	25.2159	

Assay Design

- **Assay Name:** The detector or target name assigned by user, imported in with the results files. This is most often the Assay ID, gene name or the name of the assay provided by the user.
- **Assay Type**: The task assigned by user if required (e.g. target, endogenous control), imported in with the results files.
- **Assay Design File:** Assay Design information can be entered by using an assay design file (see Appendix A). Alternatively, you can enter this information into the analysis software using the drop down menu provided.
- **Enter or import Assay Design information:** To assign assay types (Target, Selected Control or Candidate Control), either click in the **Type** box and change the assignment or upload an assay design file. Click on **Browse...** to find and select your assay design file.

Analysis Settings

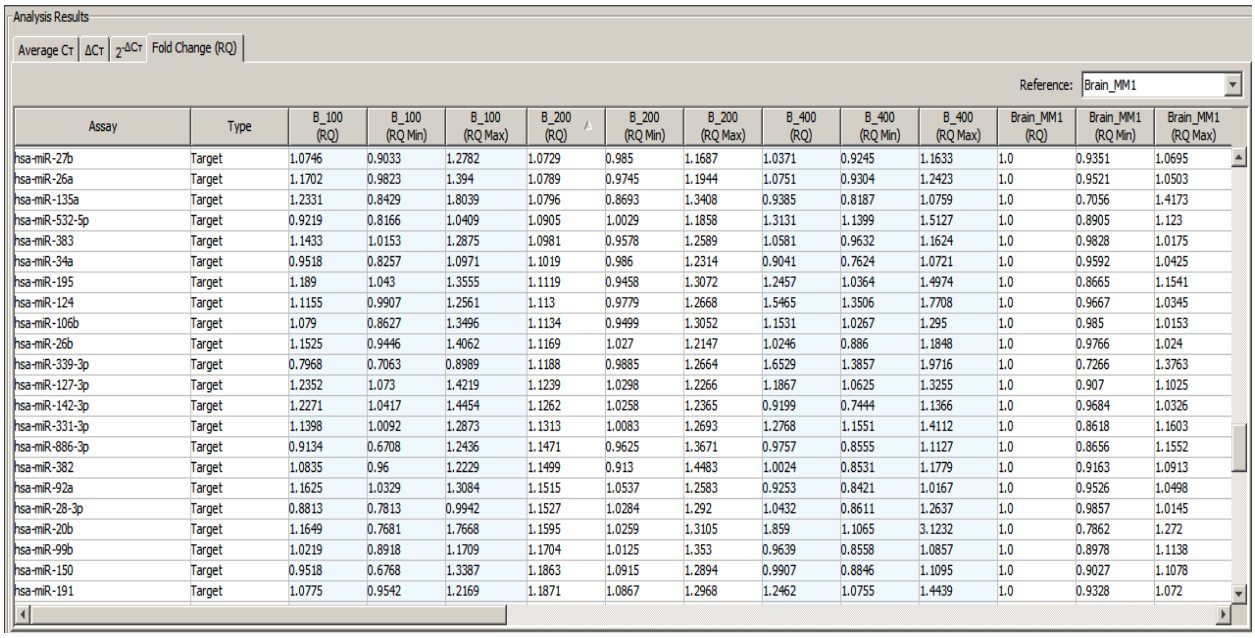
The Analysis Settings section enables specific choices to be made regarding how the calculations are done.



- **Maximum Allowable C_T Value:** This is used as a detection threshold or C_T cut-off value. Any value above the maximum allowable is changed to the maximum allowable value. This change is reflected in the **Adjusted C_T** column in the **C_T Data** tab.
- **Include Max C_T Values in Calculations:** If Yes is selected, wells with Max C_T (C_T cut-off) are included in the analysis. If you set the C_T cut-off at C_T 32, everything above C_T 32 will be converted to 32, including wells with a C_T of 40 (not detected). A fold change will be calculated for each well, even those which displayed no amplification (C_T=40). If No is selected, all wells with a C_T greater than the Max C_T are excluded from analysis.
- **Exclude Outliers Among Replicates:** Outliers within technical replicates will be excluded from data analysis calculations (see Appendix B). This change is reflected in the **Adjusted C_T** column in the **C_T Data** tab as (outlier replicate).
- **Adjust p-values using Benjamin-Hochberg False Discovery Rate (FDR):** If you choose this option, multiple testing corrections will be done to adjust p-values, to correct for occurrence of false positives, using the Benjamini-Hochberg False Discovery Rate method. If you have selected the option to use the FDR method, then all displayed p-values will be the adjusted p-values.
- **Choose Normalization Method:** This allows for two different methods of normalization: use of one or more **endogenous control(s)** or **global normalization** (Mestdagh et al 2009). When more than one endogenous control gene is selected for normalization, the software calculates the mean of the chosen endogenous control genes to use as a normalizer (normalization factor), on a per sample basis. Global normalization first finds the common assays among all samples. The median C_T of those assays is used as the normalizer, on a per sample basis.
- **Select Endogenous Controls For Analysis***:* Click on the binoculars icon to see a plot displaying C_T values of all samples for all assays that are labeled as ***Candidate Control*** or ***Selected Control*** in the Assay Design table. The box to the right lists each candidate/selected control gene and the Score value, based on standard deviation (see Appendix B) calculated for those assays. The lower the score, the more stable the control. You may select one or more assays to use as an endogenous control(s).
- **Reference sample/group (calibrator):** This is the control sample or group to which you want to compare your other samples or groups. When running the t-test, this makes up control group against which the test group is compared.
- **Perform Analysis**: Click to start the analysis.

Analysis Results

- **Average C_T:** Average C_T value of replicates
- **ΔC_T:** Normalized C_T values $\pm$ standard deviation (only applied to technical replicates)
- **$2^{-\Delta CT}$:** Changes ΔC_T values to linear values
- **Fold Change (RQ):** Displays fold change (RQ), RQ Min and RQ Max for each sample. For biological groups, fold change and P-value will be displayed.


Analysis Results

Average C_T | ΔC_T | $2^{-\Delta C_T}$ | Fold Change (RQ)

Reference: Brain_MM1

Assay	Type	B_100 (RQ)	B_100 (RQ Min)	B_100 (RQ Max)	B_200 (RQ)	B_200 (RQ Min)	B_200 (RQ Max)	B_400 (RQ)	B_400 (RQ Min)	B_400 (RQ Max)	Brain_MM1 (RQ)	Brain_MM1 (RQ Min)	Brain_MM1 (RQ Max)
hsa-miR-27b	Target	1.0746	0.9033	1.2782	1.0729	0.985	1.1687	1.0371	0.9245	1.1633	1.0	0.9351	1.0695
hsa-miR-26a	Target	1.1702	0.9823	1.394	1.0789	0.9745	1.1944	1.0751	0.9304	1.2423	1.0	0.9521	1.0503
hsa-miR-135a	Target	1.2331	0.8429	1.8039	1.0796	0.8693	1.3408	0.9385	0.8187	1.0759	1.0	0.7056	1.4173
hsa-miR-532-5p	Target	0.9219	0.8166	1.0409	1.0905	1.0029	1.1858	1.3131	1.1399	1.5127	1.0	0.8905	1.123
hsa-miR-383	Target	1.1433	1.0153	1.2875	1.0981	0.9578	1.2589	1.0581	0.9632	1.1624	1.0	0.9828	1.0175
hsa-miR-34a	Target	0.9518	0.8257	1.0971	1.1019	0.986	1.2314	0.9041	0.7624	1.0721	1.0	0.9592	1.0425
hsa-miR-195	Target	1.189	1.043	1.3555	1.1119	0.9458	1.3072	1.2457	1.0364	1.4974	1.0	0.8665	1.1541
hsa-miR-124	Target	1.1155	0.9907	1.2561	1.113	0.9779	1.2668	1.5465	1.3506	1.7708	1.0	0.9667	1.0345
hsa-miR-106b	Target	1.079	0.8627	1.3496	1.1134	0.9499	1.3052	1.1531	1.0267	1.295	1.0	0.985	1.0153
hsa-miR-26b	Target	1.1525	0.9446	1.4062	1.1169	1.027	1.2147	1.0246	0.886	1.1848	1.0	0.9766	1.024
hsa-miR-339-3p	Target	0.7968	0.7063	0.8989	1.1188	0.9885	1.2664	1.6529	1.3857	1.9716	1.0	0.7266	1.3763
hsa-miR-127-3p	Target	1.2352	1.073	1.4219	1.1239	1.0298	1.2266	1.1867	1.0625	1.3255	1.0	0.907	1.1025
hsa-miR-142-3p	Target	1.2271	1.0417	1.4454	1.1262	1.0258	1.2365	0.9199	0.7444	1.1366	1.0	0.9684	1.0326
hsa-miR-331-3p	Target	1.1398	1.0092	1.2873	1.1313	1.0083	1.2693	1.2768	1.1551	1.4112	1.0	0.8618	1.1603
hsa-miR-886-3p	Target	0.9134	0.6708	1.2436	1.1471	0.9625	1.3671	0.9757	0.8555	1.1127	1.0	0.8656	1.1552
hsa-miR-382	Target	1.0835	0.96	1.2229	1.1499	0.913	1.4483	1.0024	0.8531	1.1779	1.0	0.9163	1.0913
hsa-miR-92a	Target	1.1625	1.0329	1.3084	1.1515	1.0537	1.2583	0.9253	0.8421	1.0167	1.0	0.9526	1.0498
hsa-miR-28-3p	Target	0.8813	0.7813	0.9942	1.1527	1.0284	1.292	1.0432	0.8611	1.2637	1.0	0.9857	1.0145
hsa-miR-20b	Target	1.1649	0.7681	1.7668	1.1595	1.0259	1.3105	1.859	1.1065	3.1232	1.0	0.7862	1.272
hsa-miR-99b	Target	1.0219	0.8918	1.1709	1.1704	1.0125	1.353	0.9639	0.8558	1.0857	1.0	0.8978	1.1138
hsa-miR-150	Target	0.9518	0.6768	1.3387	1.1863	1.0915	1.2894	0.9907	0.8846	1.1095	1.0	0.9027	1.1078
hsa-miR-191	Target	1.0775	0.9542	1.2169	1.1871	1.0867	1.2968	1.2462	1.0755	1.4439	1.0	0.9328	1.072

QC Plots and Graphic Results Plots

QC Plots help to visualize sample and group correlations for a quick quality check of data. **Graphic Results Plots** are helpful to visualize your analyzed data.

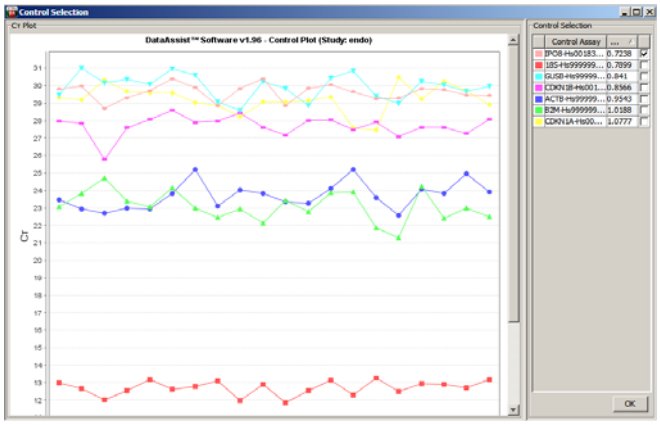
All plots have a mouse over functionality to provide quick access to sample information. Right clicking on any plot gives you the option to ***Copy***, ***Save as*** or ***Print*** the figure.

*Note: To view only a subset of assays in any of the **Graphic Results Plots**, select two or more assays. Then only those assays will be shown in the results plots.*

Endogenous Control Selection:

Displays C_T values of candidate and selected controls for all samples as well as a calculated score (see Appendix B). Checking the box to the right of an assay will assign it as a selected control. One or more controls may be selected for data normalization. If you choose more than one gene for normalization, the mean C_T value of the controls will be used for normalization.

Note: *A minimum of 2 controls are needed to calculate the score. Since the score is relative to other controls, the score will be the same if you only have two controls.*



QC Plots

Box Plot:

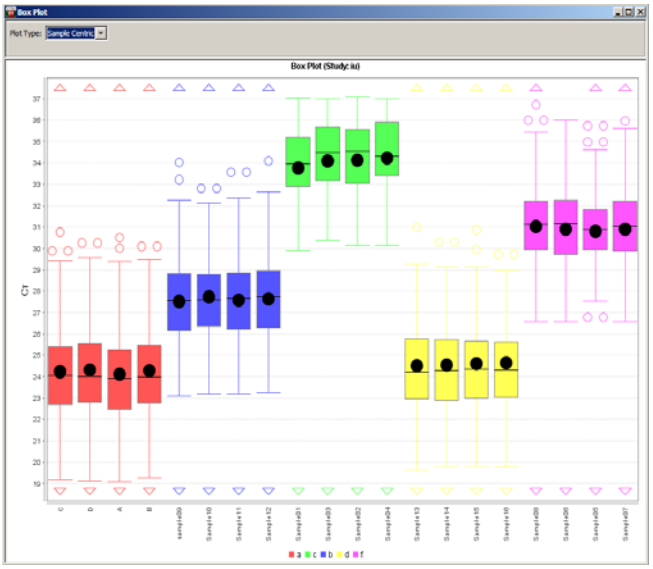
Displays the overall range of C_T distribution, shown by Sample (sorted and colored by Group), or by Assay.

The box contains the middle 50% of the data (C_T values). The black horizontal line and the black dot denote the median C_T value and mean C_T respectively.

The ends of the vertical lines (“whiskers”) indicate the minimum and maximum C_T values, unless these values exceed 1.5 x IQR. The IQR is the inter quartile range: the distance between Q1 and Q3. The points outside the ends of the whiskers are outliers or suspected outliers.

Note: *The outliers represented in a box plot are not the same as outliers calculated by the modified Grubb's calculation that is used to exclude outliers from the data for further calculation.*

This plot is useful for viewing the variation in the C_T values among biological replicates or assays.

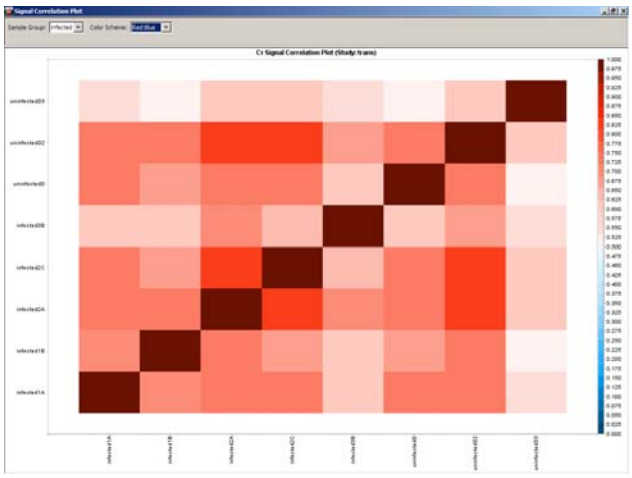


Signal correlation:

Displays C_T (signal) correlation between samples in a chosen Group. Pearson's product moment correlation coefficient (r) is calculated for each pair of samples in the selected Group and plotted on the Signal Correlation Plot.

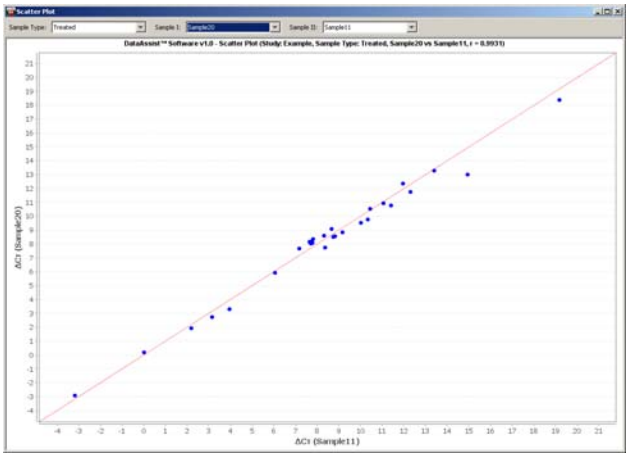
This plot can be displayed in a Red/Green or Red/ Blue color scheme.

Note*: Signal correlation plot is for biological replicates and is not drawn if no* Group *is entered in Sample Design table.*



Scatter Plot:

Displays ΔC_T correlation between samples within a chosen Group. Pearson's product moment correlation coefficient (r) is calculated for each pair of samples in the selected Group and plotted on the Scatter Plot respectively.



Graphic Results Plots

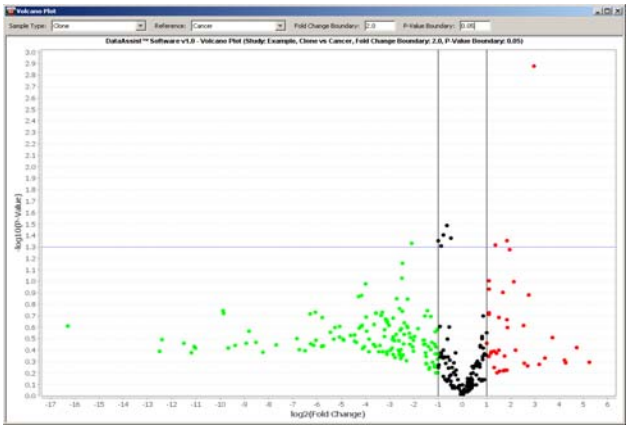
RQ Plot:

Displays RQ (log fold change) vs. Target or RQ vs. Sample. The Graph Types available to view the data are Linear, Log_{10}, and Log_2. If no Group is specified, the standard deviation of the ΔC_T is also plotted for each sample on the Log_2 Graph Type.

Volcano plot:

Displays P-values vs. Fold Change of Groups based on input Fold Change Boundary and P-values. Default is a Fold Change Boundary of 2 (2-fold change) and a P-value of 0.05.

***Note**: Volcano plot is not drawn if no Group is entered in Sample Design table, as no p-values are calculated in this instance.*



Cluster Analysis & Heat Map:

Graphically displays results of unsupervised hierarchical clustering. Distances between samples and assays are calculated for hierarchical clustering based on the ΔC_T values using either Pearson's Correlation or Euclidean Distance.

This plot can be displayed in the Red/Green or Red/Blue color scheme, and has a zoom in / zoom out feature on the left of the plot. Sample names are colored by group, if groups have been designated.

Both ΔC_T and ΔC_T + global control mean (global median if global normalization was used for normalization) are displayed in the mouse-over tool tip. The global control mean is the mean C_T value of all selected endogenous controls in the study. The global median is the median value used if global normalization was use. This value (global control mean or global median) is added on to the ΔC_T to better approximate the original C_T (a rough estimate of expression level) calculated for each sample and given assay prior to normalization.

Distance Measure: *Pearson's Correlation* (default) or *Euclidean Distance*

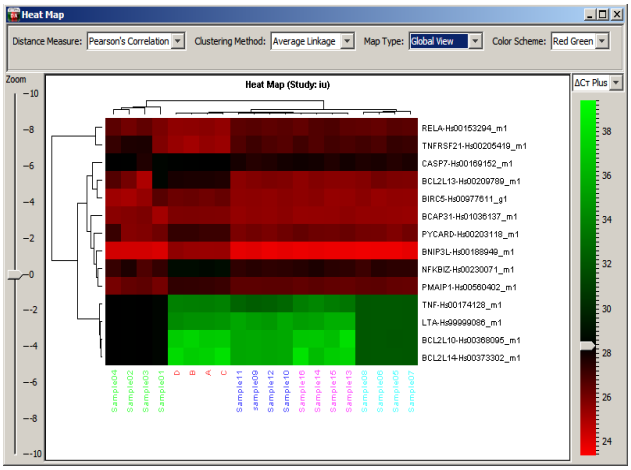
Clustering Method: *Average Linkage* (default), *Complete Linkage* or *Single Linkage*

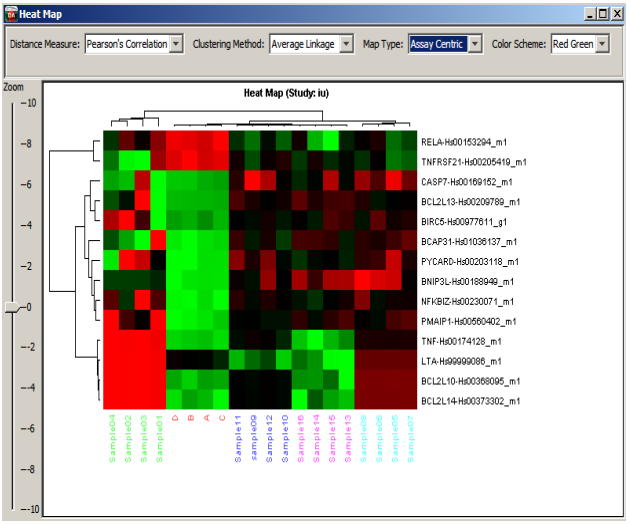
Map Type: *Assay Centric*, *Global View* (default), or *Sample Centric*

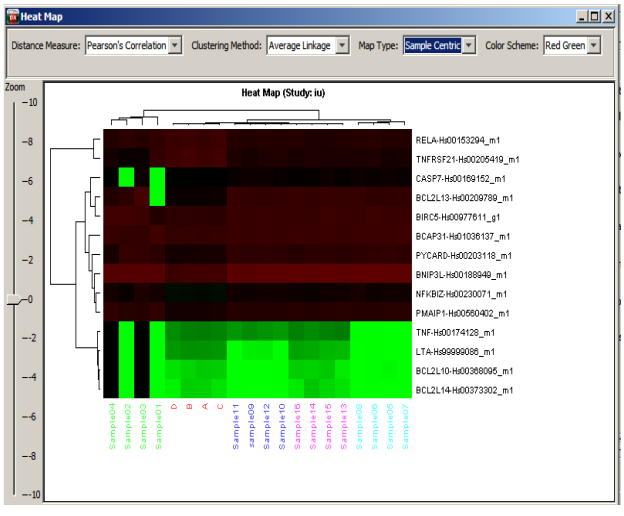
For each **Map Type**, the ΔC_T value of the neutral/middle expression level (mean or median) is set such that red indicates an increase with a ΔC_T value below the middle level, and green or blue indicates a decrease, with a ΔC_T value above the middle level.

> *Global View*: The middle expression level is set as the median of all the ΔC_T values in the study by default, and can be adjusted using the scale on the right side of the plot. The scale for the plot can be changed between the ΔC_T and the ΔC_T Plus global control mean or global median. Any data point in the plot can be compared relative to all others.
>
> *Assay Centric*: For each assay, the middle expression level is set as the median of all of the ΔC_T values from all samples for that assay. Data points for a given assay can only be compared relative to other data points for that assay.

Sample Centric: For each sample, the middle expression level is set as the median of all of the ΔC_T values from all assays for that sample. Data points for a given sample can only be compared relative to other data points for that sample.







All plots shown here are shown with Pearson's Correlation and Average Linkage. Map types are shown in this order: *Global View* ,*Assay Centric,* then *Sample Centric*.

Appendix A

Experimental Design Files

Experimental Design files can be used to import sample and assay information.

Start with either Excel® or WordPad (or equivalent) to create a design file. If using Excel, save the files as a .txt or .csv file. Design files must have 2 columns, and it is important to maintain the file structure (spacing, columns).

Sample Design File:

The sample design file should have two data columns (**Sample Name** and **Group**) separated by a comma or a tab. The rows of this file represent the samples used in an experiment. The first column must be named "Sample Name" or "SampleName" and it will contain the same sample names used in the real-time PCR results files. The second column must be named "Group" or "Type". Use the "Group" column to designate biological groups (i.e. treated and untreated). Enter in the groups you would like to be identified. Save the file as .txt or .csv.

	Sample Name	**Group**
	ABC123	untreated
	ABC456	untreated
EXAMPLE	ABC789	untreated
	DEF123	treated
	DEF456	treated
	DEF789	treated

Assay Design File:

The assay design file should have two data columns (**Assay Name** or **assayName** and **Type**) separated by a comma or a tab, the Type column can be **Target, GEX, Candidate Control, Selected Control, or Endogenous control**.

The rows of this file represent the assays / genes. The first column is "Assay Name" and will contain the same names as your detectors or targets from the SDS output. The second column is "Type" and will contain the task information real-time PCR results, i.e. "target", "endogenous control", "unknown", or "GEX". If you would like to make changes to this file, please be sure that when you save the changes to the text document you maintain the file structure (spacing, columns, etc). Save the file as.txt or .csv.

	Assay Name	**Type**
	hsa-miR-548d	Target
EXAMPLE	hsa-miR-123	Target
	U6 snRNA	endogenous control
	sn2343	Candidate Control

SOFTWARE TIPS

Sample Design Table:

- To copy a group assignment to all samples, click on that group name, then right click and select *Copy To All*. This will assign that group to all samples. Choosing *Clear All* will delete all group assignments from all samples.
- To omit one or more samples from analysis, highlight one or more samples, then right click and select *Omit Sample from Study*. You may also check the *Omit* box to the right of the Group.
- Select two or more samples to view this subset of samples only in the results plots.

Assay Design Table:

- To copy an assay type assignment to all assays, click on that assay type, then right click and select *Copy To All.*
- To omit one or more assays from analysis, highlight one or more assays, then right click and select *Omit Assay from Study*. You may also check the *Omit* box to the right of the Type.
- To view a subset of assays, select two or more assays; then only those assays will be shown in the results plots.

C_T Data tab:

- This table contains information imported from the real-time data results files: imported file path, well number, assay name, sample name and C_T value. Adjusted C_T will contain a different C_T value only if you have set the Maximum allowable C_T value to lower than the last cycle of your real time run (i.e., lower than C_T 40). Omit column contains information if well has been omitted.

QC Plots and Graphic Results:

- To zoom in, click and drag the mouse over desired region in the plot
- To zoom out, click and drag the mouse up.
- To update the Volcano Plot graphic after changing the P-value or Fold Change Boundary, hit enter once you've entered the desired number.
- Mouse over data points to get sample related information such as sample name, assay name, ΔC_T values, P-values, etc
- Right click in plots to *Copy* figure, *Save As* .png or .jpg image files, or *Print* figure
- To view only a subset of data in any of the **Graphic Results Plots**, select two or more assays from the **Assay Design** table, or two or more samples from the **Sample Design** table. Select the Results plot button and only that subset of assays and/or samples will be displayed in the results plots.
- From the **Analysis Results** table, you may sort the data and select a subset of assays to have plotted on the Heat Map, RQ Plot or Volcano Plot.

Appendix B

Analysis Workflow and Calculations in DataAssist™ Software

1. The C_T values for each well are adjusted and included/excluded for analysis based on the following analysis settings:
   - **Maximum allowable C_T value** (Max C_T): if a $C_T \geq$ Max C_T, it is adjusted to Max C_T. The undetermined C_T is also converted to Max C_T.
   - **Include max C_T values in calculations**: If *Yes* is selected, wells with Max C_T are included in the analysis. If *No* is selected, any well with Max C_T is excluded from analysis.

   Note: Go to the **C_T Data** tab to see any changes in both the Adjusted C_T and Omit columns.
2. If analysis setting *Exclude outliers among replicates* is checked, a refined Grubbs' outlier test is applied together with a heuristic rule to remove the outlier among technical replicates:
   - Find the replicate whose C_T value has the largest absolute deviation from the mean C_T value, and calculate the deviation **G** in units of the standard deviation (**SD**):

$$\mathbf{G = (max\ C_T - mean\ C_T) / SD}$$

- If the following test is true, and **(max C_T – mean C_T)** ≥ 0.25, then the replicate with **max C_T** is removed as outlier.

$$G > \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{t^2_{(\alpha/(2N),N-2)}}{N-2+t^2_{(\alpha/(2N),N-2)}}}$$

$t_{(\alpha/(2N),N-2)}$ is the critical value of the *t*-distribution with (*N*-2) degrees of freedom and a significance level of α / (*2N*), α= 0.05 is used.

Note: Go to the C_T Data tab to see outliers in the Omit column.

3. Chose normalization method:
   - **Endogenous Control:** Choose one or more genes to calculate a normalization factor. The Normalization Factor is the mean of the selected endogenous control(s), which is used to normalize the Ct value of each sample.
   - **Global Normalization:** The software first finds the common assays among all samples. The median C_T of those assays is used as the normalizer, on a per sample basis.
4. Select one or more Endogenous Controls for analysis:

   The score of each candidate or selected control is the average pairwise variation of that gene with all other chosen candidate or selected control genes. It is calculated as shown below (Vandesompele et al 2002).

   Note: *The score is only calculated once. There is no iterative exclusion of the highest value. If you would like to exclude the gene with the highest value and recalculate, you may change the assay type of that gene from candidate control to target and the software will recalculate the stability value for the remaining genes.*

   - For control i, calculate ΔC_{Tij} for all samples using another control j as the normalizer, and calculate the standard deviation (SD_{ij}) of the ΔC_{Tij} values
   - Repeat the above SDij calculation for all other candidate controls, j = 1…N–1 and use the average of all SDij's as the stability score for control i.

     Note: A minimum of 2 controls are needed to calculate a score. Since the score is relative to other controls, it will be the same if you only have two controls.
5. Once controls are selected, click the **Perform Analysis** button. The results are calculated as following:
   - For each sample:
     - **Average C_T** = mean of the technical replicates
     - **ΔC_T** = Average C_T – Normalization Factor (NF) (endogenous control)
     - $\mathbf{2^{(-\Delta Ct)}}$ is also calculated for determining Fold Change
   - If sample groups (biological replicates) are specified, for each sample group:
     - Calculate the geometric mean of $\mathbf{2^{(-\Delta Ct)}}$ of the samples in the group
     - Fold Change (RQ) = geometric mean $2^{(-\Delta Ct)}$ / geometric mean $2^{(-\Delta Ctreference)}$
     - A two-sample, two-tailed Student's t-test comparing the ΔC_T values of the two groups is performed and a **p-value** is calculated if both groups have 2 or more samples. The p-value is adjusted if the analysis setting *Adjust p-values using Benjamini-Hochberg False Discovery Rate* is checked. Benjamini-Hochberg FDR method is based on R package multtest at http://cran.r-project.org/web/packages/multtest/index.html.
   - If no sample group is specified, for each sample:
     - RQ = $\mathbf{2^{(-\Delta Ct)} / 2^{(-\Delta Ctreference)}}$
     - Standard deviation (SD) is calculated for C_T values of the technical replicates, and is used to calculate the RQ Min and RQ Max:

RQ Min = $2^{(-\Delta Ct - SD)} / 2^{(-\Delta Ctreference)}$

RQ Max = $2^{(-\Delta Ct + SD)} / 2^{(-\Delta Ctreference)}$

6. Pearson's product moment correlation coefficient (r) is calculated for C_T or ΔC_T values of sample pairs, and plotted on the Signal Correlation Plot and Scatter Plot respectively:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

7. Unsupervised hierarchical clustering is performed and then displayed as a Heat Map. Distances between samples and assays are calculated for hierarchical clustering based on the ΔC_T values using one of the following:

   - Pearson's Correlation: For a sample pair, the Pearson's product moment correlation coefficient (r) is calculated considering all ΔC_T values from all assays, and the distance is defined as 1 – r. For an assay pair, the r is calculated considering all ΔC_T values from all samples and the distance is defined as 1 – r.
   - Euclidean Distance: sqrt(∑(ΔC_T (i) - ΔC_T (j))^2)

     For a sample pair, the calculation is done across all assays for sample i and sample j

     For an assay pair, the calculation is done across all samples for assay i and assay j

Bibliography

Livak, K.J. and Schmittgen, T.D. Analyzing real-time PCR data by the comparative CT method. Nature Protocols 3, 1101-1108 (2008).

Mestdagh P.,Van Vlierberghe P., De Weer A., Muth D., Westermann F., Spelemean F., Vandesompele J. A novel and universal method for microRNA RT-qPCR data normalization. Genome Biology 10,R64 (2009).

Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N, De Paepe A., Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3**,** research0034 (2002).