

## DataAssist™ v2.0 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 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  $C_T$  data QC analysis on  $C_T$  data and provide 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 in 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 Version
7900HT Fast Real-Time PCR System	RQ Manager 1.2 (SDS v2.3 & v2.4)* (files must be in the <b>Plate Centric</b> Table Orientation)
7900HT Fast Real-Time PCR System	SDS v2.3 AQ Results
7900HT Fast Real-Time PCR System	SDS v2.2.2 RQ Study Results
7900HT Fast Real-Time PCR System	SDS v2.2.2 AQ Results
7500 & 7500 Fast Real-Time PCR System	SDS v1.4 RQ Study Results
7500 Fast Real-Time PCR System	Software v2.0.1 RQ Study Results
7300 Real-Time PCR System w/ RQ Study Software	SDS v1.4 RQ Study Results
ABI PRISM® 7000 Sequence Detection System	SDS v1.1 & v1.2 RQ Study Results
StepOne™ & StepOnePlus™ Real-Time PCR System	StepOne™ Software v2.0 RQ Study Results

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.

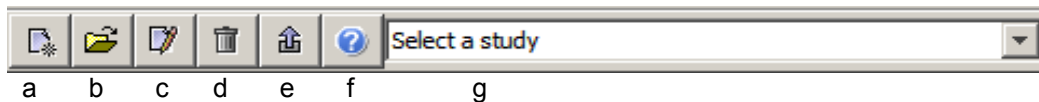
**\*Note: RQ data must be exported from an RQ study file.**

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.

## How to get started with DataAssist™ Software (see below for more details):

1. Go to *File>New Study* (or first icon) to create a new study.
2. Add Result Files exported from supported instrument software (see Table 1 for supported file format)
3. If you are using biological replicates or sample groups, enter the Sample Group names or upload a sample design file (see Appendix A).
4. Input Assay Type (target, control) or upload an assay design file (see Appendix A)
5. Enter Analysis settings.
6. Select Endogenous Control(s)
7. Perform analysis
8. Select Analysis Results tabs to see results
9. Save and export analyzed data

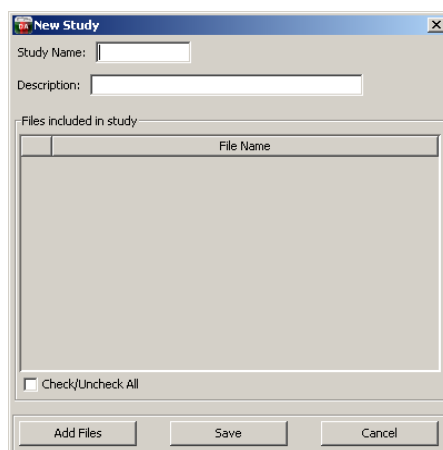
## Helpful Shortcut Icons



- a. Create a new study
- b. Open study from exported  $C_T$  data file (Open a saved study)
- c. Edit a study
- d. Delete (or close) a study
- e. Export and Save analysis results
- f. Help
- g. Toggle between studies

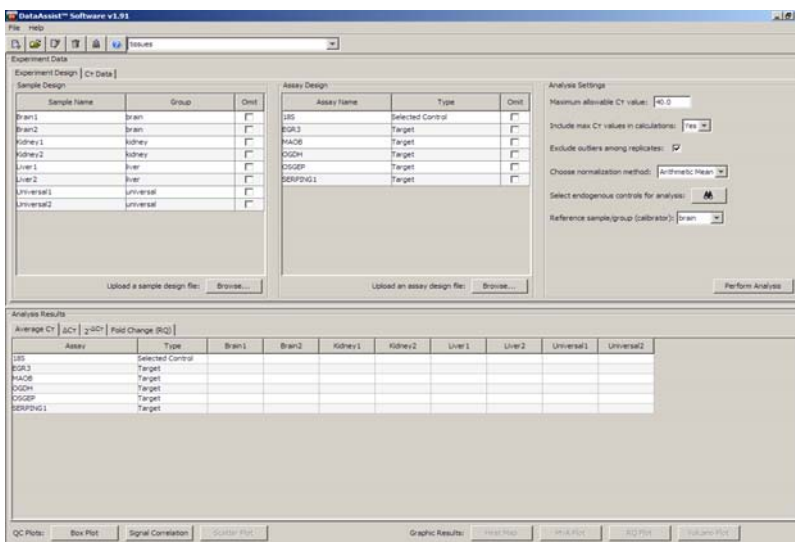
## Create a New Study

1. Create a new study either by clicking on the first icon in tool box or *File > New Study*.
2. Enter a **Study Name** in the box at the top of the window.



3. **Import results files by clicking on the [Add Files](#) button in the New Study window, then click [Save](#).**

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



## Open a Saved Study

Open a study you have previously saved by either clicking on the second icon in tool box or [File > Open Study](#). Browse to find the file containing the results data from the desired study, the *CT Data.csv* file containing the analyzed data, and select to open.

## Sample and Assay Information

- **Sample Name:** The name given to each sample by the user. Wells in a study with the same sample and assay name are considered technical replicates.

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

- (optional) **Sample Group:** The name of the biological replicate group (e.g. normal, disease or time point 1, time point 2).
- **Assay Name:** The detector or target name assigned by user. 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)

- **Experimental Design File:** Sample and Assay Design information can be entered by using experimental design files (see Appendix A). Alternatively, you can enter this information into the analysis software manually.
- **(optional) 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. 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.
- **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

- **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, then 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 ( $CT=40$ ). 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.
- **Choose Normalization Method:** When more than one endogenous control gene is selected for normalization, use either [Arithmetic Mean](#) or [Geometric Mean](#) to calculate the mean  $C_T$  value.
- **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 Stability (M) value (see Appendix B) calculated for those assays. You may select one or more assays to use as an endogenous control(s).
- **Reference sample/group (calibrator):** This is the sample or group to which you want to compare your other samples or groups.
- **Perform Analysis:** Click to start the analysis.

**$C_T$  Data Tab:** 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 only be reflected in the  [\$C\_T\$  Data](#) tab after performing analysis.

DataAssist™ Software v1.999

File Help

new one

Experiment Data

Experiment Design Ct Data

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

## Analysis Results

- **Average C<sub>T</sub>**: Average C<sub>T</sub> value of replicates
- **ΔC<sub>T</sub>**: Normalized C<sub>T</sub> values ± standard deviation (if no biological replicates)
- **2<sup>-ΔC<sub>T</sub></sup>**: Changes ΔC<sub>T</sub> 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 Ct | ΔCt | 2<sup>-ΔCt</sup> | Fold Change (RQ)

Reference: brain

Assay	Type	brain (RQ)	brain (P-Value)	kidney (RQ)	kidney (P-Value)	liver (RQ)	liver (P-Value)	universal (RQ)	universal (P-Value)
18S	Selected Control	1.0	NaN	1.0	NaN	1.0	NaN	1.0	NaN
EGR3	Target	1.0	1.0	0.1894	0.0253	0.0595	0.0217	0.4893	0.0212
MAOB	Target	1.0	1.0	2.5411	0.012	1.5493	0.0376	0.0791	0.0259
OGDH	Target	1.0	1.0	4.6042	0.0415	0.4354	0.0206	1.7752	0.0207
OSGEP	Target	1.0	1.0	2.0319	0.0636	0.8122	0.1093	2.7232	0.0362
SERPING1	Target	1.0	1.0	3.3512	0.0628	50.5616	0.0317	0.4887	0.0088

QC Plots:

## 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 easily identify pertinent sample information. Right clicking on any plot gives you the option to *Copy*, *Save as*, or *Print* the figure.

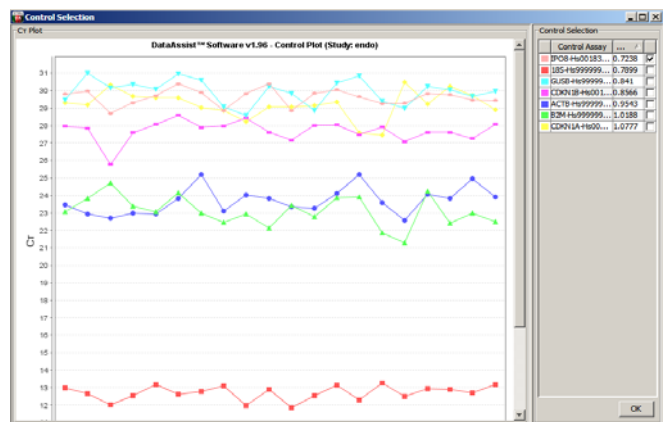
*Note: To view only a subset of assays in any of the **Graphic Results Plots**, select two or more assays. Then only those assays will be shown in the results plots.*

## Endogenous Control Selection:

Displays  $C_T$  values of candidate and selected controls for all samples as well as the calculated stability value (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, select either arithmetic or geometric mean to calculate the normalization factor to be used.

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



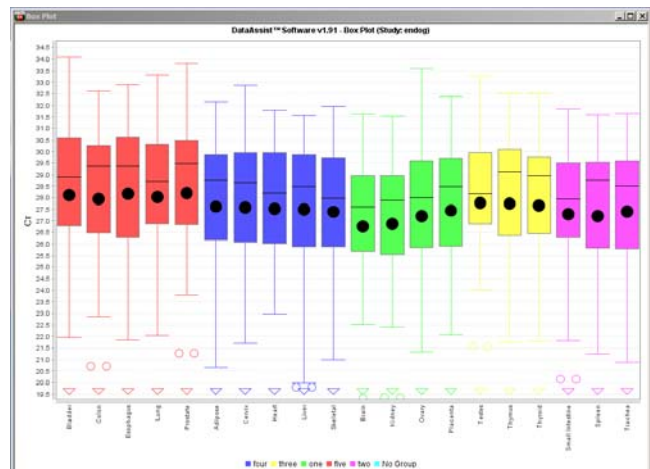
## QC Plots

### Box Plot:

Displays the overall range of  $C_T$  distribution, displayed by **Sample** and sorted by **Group (color)**.

The box contains the middle 50% of the data ( $C_T$  values). The black horizontal line indicates the median  $C_T$  value, and the black dot represents the mean  $C_T$ .

The ends of the vertical lines (“whiskers”) indicate the minimum and maximum  $C_T$  values, unless outliers are present. The points outside the ends of the whiskers are outliers or suspected outliers.



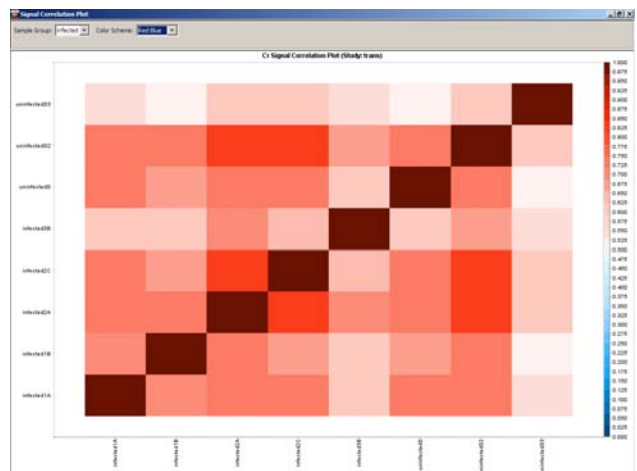
This plot is useful for viewing the variation in the  $C_T$  values among biological replicates.

### 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.

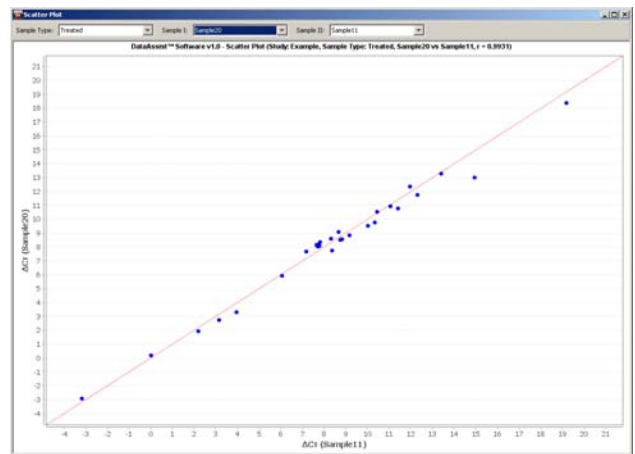
This plot can be displayed in the Red Green or Red Blue color scheme.

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



### Scatter Plot:

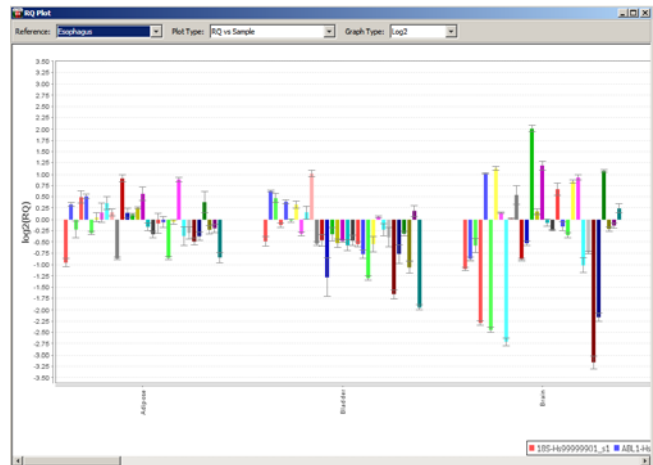
Displays  $\Delta C_T$  correlation between samples within a chosen Group. Pearson's product moment correlation coefficient ( $r$ ) is calculated for each pair of samples in the selected Group and plotted on the Scatter Plot respectively.



### Graphic Results Plots

#### RQ Plot:

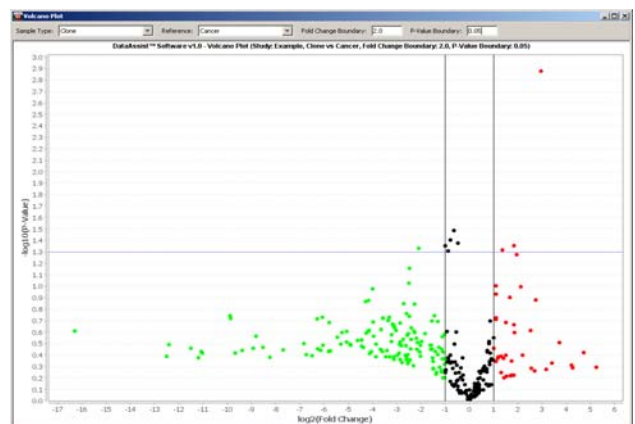
Displays RQ (fold change) vs. Target or RQ vs. Sample. The Graph Types available to view the data are Linear, Log<sub>10</sub>, and Log<sub>2</sub>. If no Group is specified, the standard deviation of the  $\Delta C_T$  is also plotted for each sample on the Log<sub>2</sub> Graph Type.



#### 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 0.01.

**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 hierarchical clustering. Distances between samples and assays are calculated for hierarchical clustering based on the  $\Delta C_T$  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  $\Delta C_T$  and  $\Delta C_T + \text{global control mean}$  are displayed in the mouse-over tool tip text. The **global control mean** is the mean  $\Delta C_T$  value of all selected endogenous controls in the study. This value is added on to the  $\Delta C_T$  to give an idea what the expression level is for each sample and given assay.

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

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

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

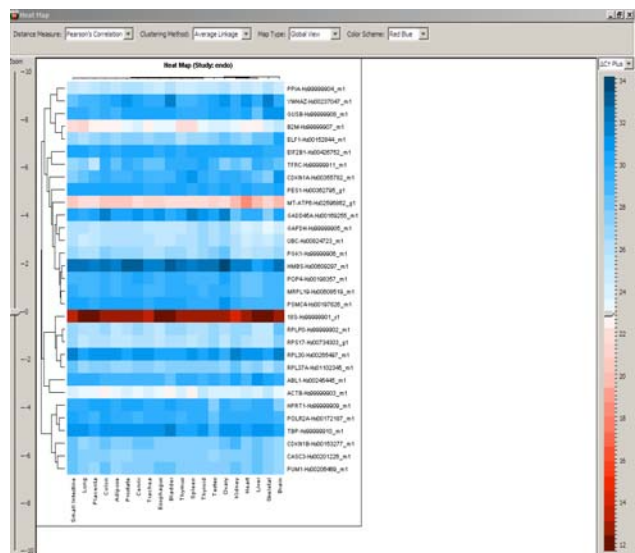
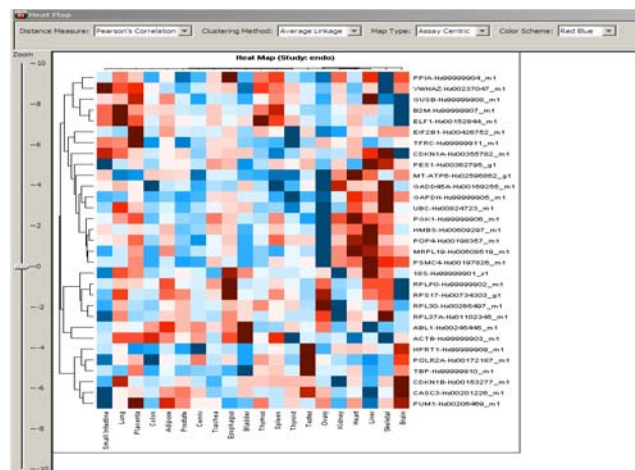
For each **Map Type**, the  $\Delta C_T$  value of the neutral/middle expression level is set differently such that red indicates up-regulated with a  $\Delta C_T$  value below the middle level, and green or blue indicates down-regulated with a  $\Delta C_T$  value above the middle level.

**Assay Centric:** For each assay, the middle expression level is set as the mean of all of the  $\Delta C_T$  values from all samples for that assay. Data can only be compared across a particular row / assay.

**Global View:** The middle expression level is set as the median of all the  $\Delta C_T$  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  $\Delta C_T$  and the  $\Delta C_T + \text{global control mean}$ .


**Sample Centric:** For each sample, the middle expression level is set as the mean of all of the  $\Delta C_T$  values from all assays for that sample. Data can only be compared within a particular column / 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*.





## Exporting and Saving Study Data Files

Save a study and export results data from the study by either clicking on the fifth icon in the tool box () or [File > Export](#). In the **Export** dialog box that is opened, you may select the individual files to be exported and the file export location. The exported files may be saved as a .csv or a .txt file. The results data from the study is saved in the *CT Data.csv* file. This file may be used to open the study at a later time. Other files that can be exported are the C<sub>T</sub> Data, Sample Design, Assay Design, Average C<sub>T</sub>, ΔC<sub>T</sub>, 2<sup>-ΔC<sub>T</sub></sup>, and Fold Change files.

## Appendix A

### Experimental Design Files

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

Start with either Excel<sup>®</sup> 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.

<b>Sample Name</b>	<b>Group</b>
ABC123	untreated
ABC456	untreated
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.

<b>Assay Name</b>	<b>Type</b>
hsa-miR-548d	Target
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<sub>T</sub> 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<sub>T</sub> value. Adjusted C<sub>T</sub> will contain a different C<sub>T</sub> value only if you have set the Maximum allowable C<sub>T</sub> value to lower than the last cycle of your real time run (i.e., lower than C<sub>T</sub> 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,  $\Delta C_T$  values, P-values, etc
- Right click in plots to [Copy](#) figure, [Save As](#) .png or .jpg image files, or [Print](#) figure

## Appendix B

### Analysis Workflow and Calculations in DataAssist™ v2.0 Software

1. The C<sub>T</sub> values for each well are adjusted and included/excluded for analysis based on the following analysis settings:

- **Maximum allowable C<sub>T</sub> value** (Max C<sub>T</sub>): if a C<sub>T</sub>  $\geq$  Max C<sub>T</sub>, it is adjusted to Max C<sub>T</sub>. The undetermined C<sub>T</sub> is also converted to Max C<sub>T</sub>.
- **Include max C<sub>T</sub> values in calculations**: If [Yes](#) is selected, wells with Max C<sub>T</sub> are included in the analysis. If [No](#) is selected, any well with Max C<sub>T</sub> is excluded from analysis.

*Note:* Go to the [C<sub>T</sub> Data](#) tab to see any changes in both the Adjusted C<sub>T</sub> and Omit columns.

2. If analysis setting [Exclude outliers among replicates](#) is checked, a refined Grubbs' outlier test is applied together with a business rule to remove the outlier among technical replicates:

- Find the replicate whose C<sub>T</sub> value has the largest absolute deviation from the mean C<sub>T</sub> value, and calculate the deviation **G** in units of the standard deviation (**SD**):

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

- If the following test is true, and  $(\max C_T - \text{mean } C_T) \geq 0.25$ , then the replicate with  $\max C_T$  is removed as outlier.

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

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

*Note: Go to the **C<sub>T</sub> Data** tab to see outliers in the Omit column.*

3. Specify assay type (at least one Selected Control is required).

Chose normalization method:

- Arithmetic Mean uses the arithmetic mean of  $C_T$  values of the selected controls as the normalization factor (NF), while
- Geometric Mean uses their geometric mean as the NF.

4. Select one or more Endogenous Controls for analysis: Control stability measure (M) is calculated using the geNorm algorithm (<http://genomebiology.com/2002/3/7/research/0034>):

- For control  $i$ , calculate  $\Delta C_{Tij}$  for all samples using another control  $j$  as the normalizer, and calculate the standard deviation ( $SD_{ij}$ ) of the  $\Delta C_{Tij}$  values
- Repeat the above  $SD_{ij}$  calculation for all the other controls,  $j = 1 \dots N-1$  and use the average of all  $SD_{ij}$ 's as the stability measure for control  $i$ .

*Note: A minimum of 2 controls are needed to calculate the stability measure. Since the stability is relative to other controls, the stability 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
  - **$\Delta C_T$**  = Average  $C_T$  – Normalization Factor (NF)
  - $2^{(-\Delta Ct)}$  is also calculated for determining Fold Change
- If sample groups (biological replicates) are specified, for each sample type:
  - Calculate the mean of  $2^{(-\Delta Ct)}$  of the samples in the group
  - Fold Change (RQ) =  $\text{mean } 2^{(-\Delta Ct)} / \text{mean } 2^{(-\Delta Ct_{\text{reference}})}$
  - t-test is performed and a **p-value** is calculated if both the group and the reference have 2 or more samples.
- If no sample group is specified, for each sample:
  - $RQ = 2^{(-\Delta Ct)} / 2^{(-\Delta Ct_{\text{reference}})}$
  - Standard deviation (SD) is calculated for  $C_T$  values of the technical replicates, and is used to calculate the RQ Min and RQ Max:

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

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

6. Pearson's product moment correlation coefficient ( $r$ ) is calculated for  $C_T$  or  $\Delta C_T$  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  $\Delta C_T$  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  $\Delta C_T$  values from all assays, and the distance is defined as  $1 - r$ . For an assay pair, the r is calculated considering all  $\Delta C_T$  values from all samples and the distance is defined as  $1 - r$ .
- Euclidean Distance:  $\text{sqrt}(\sum (\Delta C_T(i) - \Delta C_T(j))^2)$

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