GenEx User Guide

Version 1.0

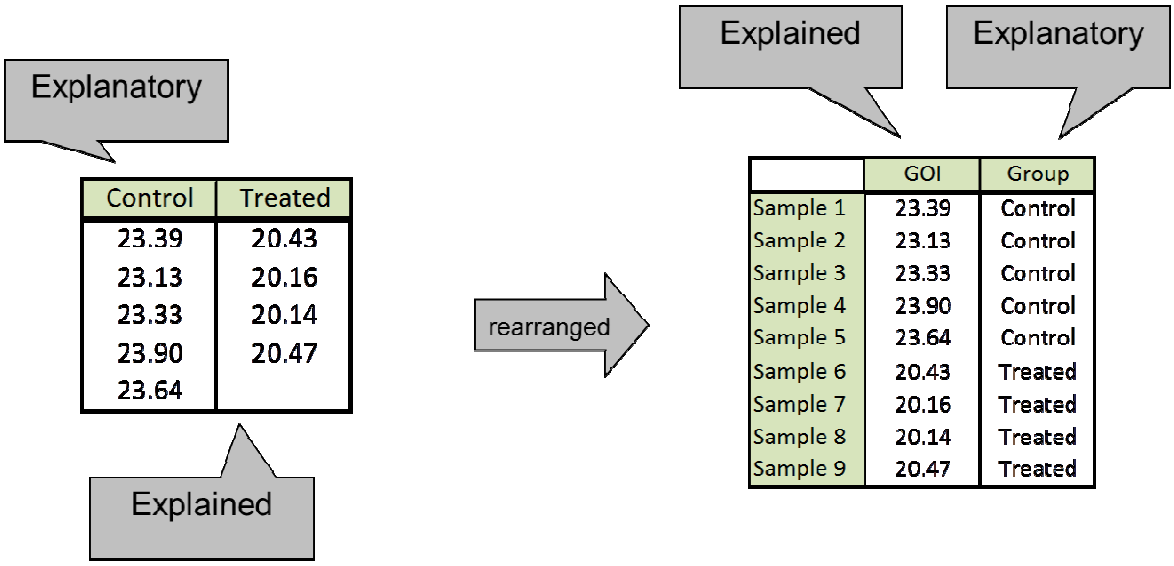
www.multid.se

Introduction

As the qPCR field advances, the design of experiments and the analysis of data are becoming more important and more challenging. Studies are now designed using multiple markers, nested levels, exploring or confirming the effect of multiple factors, occasionally in paired designs, etc. These analyses require more information on how the experiment was set up to handle references, standards, and controls appropriately, and correctly account for the variance and covariance in the measured data. Proper handling of such data requires software that support the planning and design of experiments, and data analysis. GenEx is the main qPCR software on the market today.

Data arrangement

When groups are compared, data are classically arranged with the measured Cq (explained variables) in columns headed with the experimental group label (explanatory variables). This arrangement provides easy overview of the data (Figure 1). However, it is not practical for advanced studies that may include more than one nominal factor or covariate (variable of metric character, e.g. time, age, dose etc.), multiple markers, replicate measurement, repeated measurements (same subject sampled repeatedly), multiplate measurements, etc. A more flexible approach is to arrange the data with samples as rows and all variables in columns (Figure 1). The format is readily generalized to any number of markers and additional columns and rows can be added that specify the experimental design, indexing samples, markers, plates etc. These are referred to as classification columns and classification rows and have labels starting with #. In the example shown in Figure 2, #Repeat indexes qPCR technical replicates (samples with the same index are replicates on the qPCR level). These are expected to be highly similar and shall be averaged during data pre-processing.



mn's heads (left)

	MG	RG	#Repeat	#Treatment	#Pairing
S1	22.22481...	16.71773...	1	1	1
S2	17.93896...	16.59702...	1	1	2
S3	19.76137...	16.60889...	2	1	3
S4	17.23068...	16.23785...	2	1	4
S5	16.57810...	16.89222...	3	1	5
S6	18.25507...	16.54029...	3	1	6
S7	29.47699...	16.86087...	1	2	1
S8	18.68488...	16.08867...	1	2	2
S9	22.91167...	16.67638...	2	2	3
S10	17.80890...	16.15741...	2	2	4
S11	16.80731...	16.05589...	3	2	5
S12	19.77758...	16.19985...	3	2	6
#Gene	0	1			

Figure 2. Example of data arrangement in GenEx. First column list the samples. 2nd and 3rd columns are measured Cq values. 4th column indexes technical replicates, 5th column indexes treatment groups and 6th column indexes paired samples. Bottom row identifies maker and reference genes.

#Treatment indexes treatment groups that eventually shall be compared using a statistical test. Finally, the study is paired, meaning that each subject received both treatments and a sample was collected after each treatment. Paired study designs are more powerful, because the pairing reduces confounding variation. This elevates the power of the test and the experiment requires fewer subjects. It can be, for example, samples collected from all subjects before treatment and a second set collected after treatment. It can also be positive and negative samples collected from the same subject or genetically similar individuals such as siblings, identical twins or clones. A special type of paired study design is repeated samplings, used in more than two subsequent measurements. In general, the word "paired" is replaced by "repeated". Specialized statistical procedures are available to analyze repeated samplings.

Data import

The experimental design is defined in part by deciding on the experimental factors and covariates involved in the experiment and in part when the samples and assays are mixed while dispensed into the qPCR containers. This information is critical for proper analysis and mining of the measured data. This has been realized by several of the leading qPCR instrument and assay providers. Roche LC480 software for RealTime Ready custom and focus panels, for example, names all genes and samples, indicates reference genes, and specifies technical and biological replicates at various levels. This information is transferred to GenEx and arranged appropriately for downstream analysis using a wizard. A similar high level user friendly solution is provided by Exiqon, who offers a customized version of GenEx with a powerful wizard to read their miRCURY LNA™ Universal RT microRNA PCR platform (www.exiqon.com/qpcr-software). On the BIOMARK microfluidic platform from Fluidigm, technical and biological replicates are indicated by the naming of assays and samples, and appropriate classification columns are created automatically. Data generated on other qPCR instruments can also be read by GenEx, including the Stratagene MX300X from Agilent, Realplex from Eppendorf, CFX96/384 from Bio-Rad, Eco from Illumina, and the many different qPCR platforms from Life Technologies. These efforts from the instrument

manufacturers to transfer experimental design information automatically or at least semi-automatically into GenEx substantially simplify the pre-processing needed to prepare qPCR data for statistical analysis.

Data pre-processing

For most studies performed today the ΔΔCq method is not sufficient to analyze qPCR data. Not that we are calculating differently, but the studies have become larger and the experiments more complex. In fact, for most studies performed today, it is not even possible to write a closed form expression to calculate the resulting expression response. Rather the measured data must be processed sequentially to account for the various aspects of the experiment. In particular, it is essential to correctly define the statistical unit (often referred to as a subject when organisms are used). Each unit should be associated with a single value for each variable to use common statistical methods. This value must, however, frequently be assembled from various measurements, i.e. responses of target and reference genes, estimated amplification efficiencies, etc. In order to integrate the process into a logical workflow, GenEx provides an intuitive wizard with the following sequential operations:

1. **Interplate calibration.** Many studies cannot be fitted in a single experimental run or for practical reasons have to be extended over time. qPCR instruments perform base-line correction and set threshold separately for each run, which introduces a bias between the Cqs measured in different runs. This bias can be compensated for by performing a common amplification in all plates, where the same sample is analyzed for a given assay. This sample is called Inter-Plate Calibrator, IPC. Any variation in the measured Cqs of the IPC among runs reflects systematic variation due to instrument factors and should be compensated for. It is sufficient to run a single IPC for each channel in the instrument if a common threshold and base-line correction is used. It is not recommended to perform separate inter-plate calibrations for each target. Since every correction adds confounding variation to the data, unimportant corrections shall be avoided, as they may impair data quality rather than improving it. For the same reason, it is a good strategy to use a robust sample for IPC and analyze it in replicates. In multiplate experiments, the runs and the inter-plate calibrators shall be indexed in classification columns.

2. **Efficiency correction.** If PCR efficiency has been estimated, the measured Cq values can be corrected to account for suboptimal amplification. Typically, PCR efficiencies are estimated from serial dilutions run separately. The PCR efficiencies may then be listed in a classification row for automatic correction in GenEx.

3. **Normalize using spiking.** PCR efficiency depends on the sample matrix. Usually it is assumed that the sample matrix and thus the PCR efficiency is constant. But occasionally there are variations, which can be tested for using an exogenous spike added to the samples. Differential expression of the spike between the test and a standard sample reflects the sample's specific inhibition and can to some degree be accounted for.

4. **Normalize to sample amount.** Measured Cq values depend on the sample input. This can be the sample volume processed, amount of RNA used for reverse transcription, or cell count. If sample input vary, data may have to be normalized. The sample input shall be indicated in a classification column.

5. **Average qPCR replicates.** If qPCR replicates are available they shall be indexed in a classification column and their Cq values shall be averaged.

6. **Correct for genomic DNA background.** When quantifying RNA levels using RT-qPCR, the assays may also amplify genomic copies of the target if the DNase treatment used is insufficient. The amount of genomic background can be assessed by measuring either NoRT controls or by using the ValidPrime approach. The contribution to Cq from the genomic background can be calculated and the Cqs corrected.

7. **Normalization with reference genes.** In expression studies normalization to endogenous controls, such as stably expressed reference genes, is popular. In GenEx, you can normalize to any number of reference genes; you can even normalize sets of reporter genes to sets of reference genes to match the genes' properties such as expression levels, stabilities, distribution in tissues, etc. It is also possible to normalize to the mean expression of all the genes (global normalization). Optionally, reference genes can be indexed in a classification row for automatic processing. Normalization to the expression of reference genes corresponds to calculating ΔCq in the classical approach.

8. **Average technical replicates.** If additional technical replicates are available, such as RT, extraction, and sampling replicates, they shall be indexed in classification columns and averaged.

9. **Normalize with Reference Sample(s)**. In some paired designs, systematic variation can be reduced by normalizing to the paired sample during pre-processing.

10. **Relative quantities**. An arbitrary reference level is selected (which corresponds to ΔΔCq in the classical approach) and data are converted to linear scale ($2^{-\Delta\Delta Cq}$ in the classical approach). The reference level can be the most expressed sample, the least expressed sample, mean expression of all the samples, mean expression of a group of samples, or percentage (sum of the expression in all samples set to 1). It is also possible to convert the ΔCq values to an arbitrary linear scale ($2^{-\Delta Cq}$).

11. **Convert to log scale**. For statistical analysis with parametric methods, the data shall be converted to logarithmic scale. Available options are $\log_2$, $\log_{10}$, ln, and log(X+1).

All the steps in the workflow are not needed, since some cancel the effect of others. The appropriate steps depend on the experimental design, the controls and references that are available, and the analysis that will be performed.

In addition to the pre-processing work flow, GenEx has correction for missing data. GenEx recognizes two types of missing data, random missing (failed experiments) and non-random missing (off-scale data). There is built in handling of random missing data among technical replicates, which are replaced based on available information, in the course of the pre-processing. This is very useful, since the missing information can be ignored and is automatically accounted for. There are also several means to handle non-random missing or off scale data that are due to too low target amounts, which may bias the biological effect and invalidate the statistical inference in the majority of the statistical tests employed. Outliers in the data can be tested for based on standard deviation and the Grubb's outlier test.

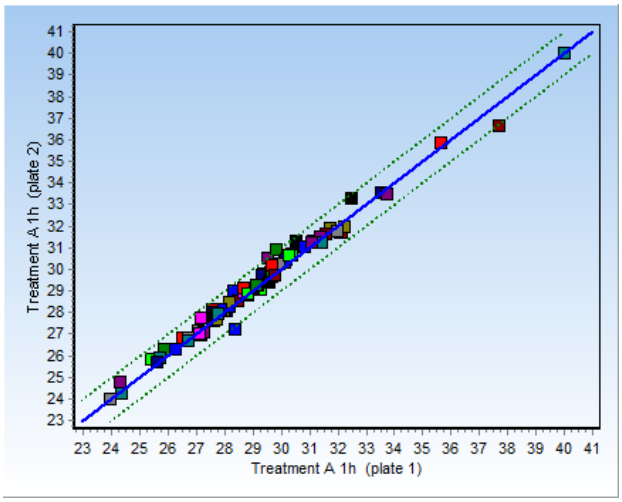
The pre-processing of data is logged and stored in a log file.

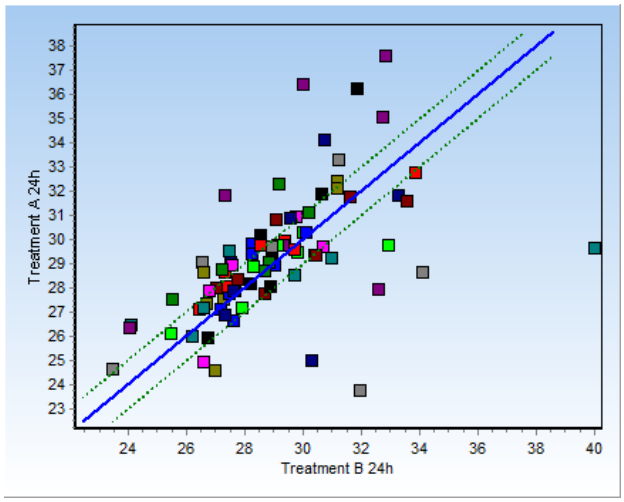
Screening by correlation

Several companies, including Roche, Exiqon, Life Technologies, Lonza, Qiagen and TATAA Biocenter, offer pre-plated assays for smooth expression profiling and screening purposes. Data from those plates are readily read into GenEx. Rarely are all assays relevant for every study and a strategy is to analyze a few representative samples of each kind in a pilot study to identify differentially expressed genes to be used in a larger downstream study. This is readily done using the GenEx scatter plot (Figure 3). Replicated measurements can be compared to test the reproducibility between plates (top left), or screen for differentially expressed genes under two conditions (top right, bottom). Correlations between genes' expressions can be quantified by calculating the Pearson or Spearman correlation coefficients. This is typically applied to larger number of samples and has, for example, been used to reveal correlations between genes expressed in individual cells (Stahlberg et al. (2010) *Nucl. Acids Res.*, 1–12, doi:10.1093/nar/gkq1182).

Preparing the data for analysis

Groups for comparison are created using the GenEx Data Manager. Treatment groups or treatment factors in multifactorial studies, such as studies of the effect of gender or covariates such as age, time, or drug load, can be indexed in classification columns and used to assign subjects into groups automatically. The groups are assigned colors and symbols for plotting. A neat feature is that colors and symbols can be set independently, which makes it possible to assign subjects to multiple groups and identify these in plots by the shape, size, and color of the symbol. Even shades of colors can be used creatively to indicate various levels of covariates at ordinary scale (e.g. darker shade indicates higher drug load).





Over-exp...				Under-ex...			
	Treatm...	Treatme...	Diff		Treat...	Treat...	Diff
gene 95	36.4	30	6.41	gene 24	29.64	40	-10.36
gene 37	37.61	32.84	4.77	gene 55	23.73	31.93	-8.21
gene 83	31.81	27.34	4.47	gene 12	28.65	34.09	-5.43
gene 42	36.27	31.87	4.39	gene 78	24.97	30.32	-5.35
gene 61	34.13	30.75	3.38	gene 8	27.92	32.58	-4.66
gene 32	32.32	29.17	3.15	gene 93	29.78	32.94	-3.15
gene 29	29.07	26.54	2.52	gene 85	24.59	27	-2.4
gene 9	26.46	24.13	2.34	gene 81	31.61	33.59	-1.98
gene 65	35.05	32.73	2.31	gene 84	29.25	30.96	-1.72
gene 49	26.33	24.05	2.28	gene 70	24.93	26.61	-1.68
gene 71	33.3	31.21	2.08	gene 34	31.82	33.3	-1.48
gene 96	29.55	27.48	2.06	gene 51	28.54	29.7	-1.17
gene 39	28.64	26.61	2.04	gene 63	29.36	30.47	-1.11
gene 16	27.55	25.54	2.01	gene 44	29.32	30.41	-1.08
gene 94	30.86	29.09	1.76	gene 60	32.8	33.87	-1.07
gene 73	30.2	28.54	1.67				
gene 2	29.84	28.27	1.56				
gene 59	28.75	27.21	1.53				
gene 5	29.05	27.52	1.53				
gene 54	28.95	27.57	1.38				
gene 17	28.66	27.31	1.35				
gene 18	30.9	29.57	1.33				
gene 53	32.41	31.16	1.26				

Figure 3. Scatter plot comparing replicate measurements performed in separate plates (left) and two different conditions (right). Differentially expressed genes are tabulated (bottom)

In the data manager, subjects and genes can be removed temporarily from analysis to compare results based on analyses of subsets of data. Data can also be mean centered (subtraction of the mean value) or autoscaled (subtraction of mean followed by division with the standard deviation) to change the weights of the genes/samples in analyses. This is particularly useful in expression profiling analysis, where genes having different expression levels can be assigned equal weights. For analyses that apply models based on measured data, such as the standard curve, reverse calibration, neural networks, self organized map, potential curves etc., samples (and genes) can be assigned either training or test. Training data are used to create the model, which is applied to classify the test samples. The various analyses available in GenEx are listed in Table 1.

Table 1. Analyses available in different versions of GenEx.

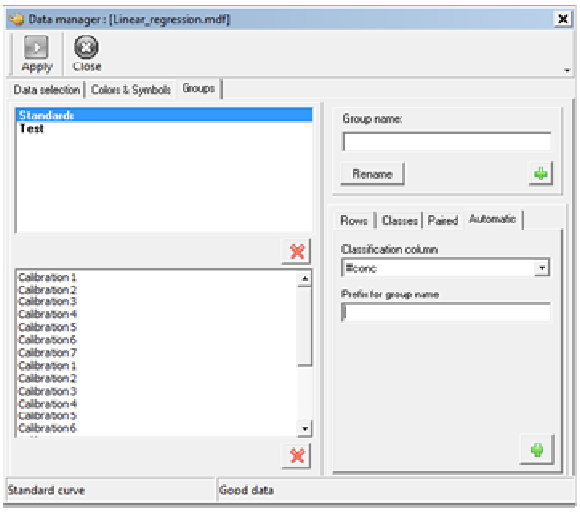
GenEx version comparison	Standard	Pro	Enterprise
Pre-processing of data			
Interplate calibration	x	x	x
PCR efficiency correction	x	x	x
Normalize to sample amount	x	x	x
Normalize to reference genes/samples	x	x	x
Normalize to spike	x	x	x
Missing data handling and primer dimer correction	x	x	x
Relative quantities and fold changes	x	x	x
Finding optimal reference genes			
geNorm	x	x	x
NormFinder	x	x	x
Plots			
Scatterplots	x	x	x
Line plots	x	x	x
Bar plots	x	x	x
Box and whiskers plot	x	x	x
Principal component analysis			
PCA		x	x
P-curve			x
Cluster analysis			
Hierarchical clustering/dendogram	x	x	x
Heatmap analysis		x	x
Networks			
Self-organizing map (SOM)		x	x
Artificial neural networks (ANN)			x
Support vector machine (SVM)			x
Regression analysis			
Standard curve	x	x	x
Reverse calibration		x	x
Limit of detection (LOD)		x	x
Partial least square (PLS)			x
Three-way analysis			
Trilinear decomposition			x
Statistics			
Descriptive statistics	x	x	x
Parametric t-test	x	x	x
Non-parametric tests	x	x	x
One-way ANOVA	x	x	x
Two-way ANOVA		x	x
Nested ANOVA		x	x
Correlation			
Spearman rank correlation coefficient	x	x	x
Pearson correlation coefficient	x	x	x
Experimental design			
Sample size		x	x
Experimental design optimization			x

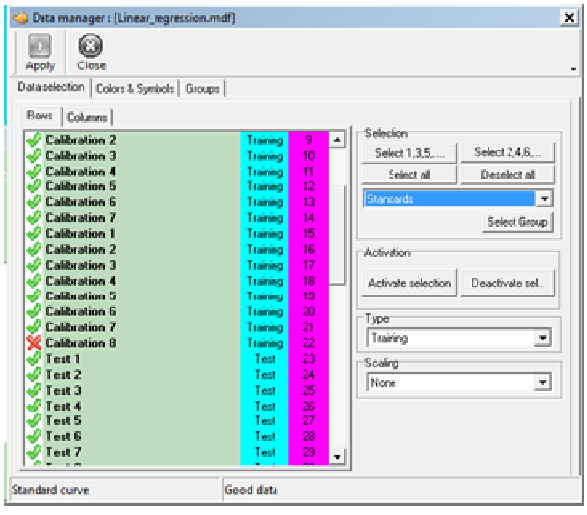
Standard curve and reverse calibration

Amounts of pathogens in field samples can be quantified using qPCR by comparing the measured Cqs of the field samples with those of standard samples by means of a standard curve. Representative data arrangement is shown in Figure 4. In addition to the measured data, the concentration of the standards is given in a classification column. Additional classification columns can be used to index replicates and to identify the standard and test samples. Any technical replicates shall be averaged during pre-processing and the averaged Cq shall be considered a single data point. Independently prepared standards are treated as different data points, while replicate measurements of field samples are averaged and used as a single more precise estimate. Groups are created in the GenEx Data Manager and assigned either test or training status. Samples can also be reversibly removed from analysis (Figure 5). A confidence level is set for the analysis.

	A	B	C	D	E
1		Gene	#conc	#repeats	#test&train
2	Calibration 1	27.63	1	1	1
3	Calibration 2	25.31	2	2	1
4	Calibration 3	21.42	3	3	1
5	Calibration 4	17.82	4	4	1
6	Calibration 5	14.65	5	5	1
7	Calibration 6	10.63	6	6	1
8	Calibration 7	7.95	7	7	1
9	Calibration 1	28.79	1	8	1
10	Calibration 2	24.61	2	9	1
11	Calibration 3	21.38	3	10	1
12	Calibration 4	17.87	4	11	1
13	Calibration 5	14.27	5	12	1
14	Calibration 6	10.7	6	13	1
15	Calibration 7	8.02	7	14	1
16	Calibration 1	27.91	1	15	1
17	Calibration 2	24.68	2	16	1
18	Calibration 3	21.42	3	17	1
19	Calibration 4	17.91	4	18	1
20	Calibration 5	14.45	5	19	1
21	Calibration 6	10.35	6	20	1
22	Calibration 7	8.14	7	21	1
23	Test 1	15.76	0	22	2
24	Test 2	18.22	0	23	2
25	Test 3	20.74	0	24	2

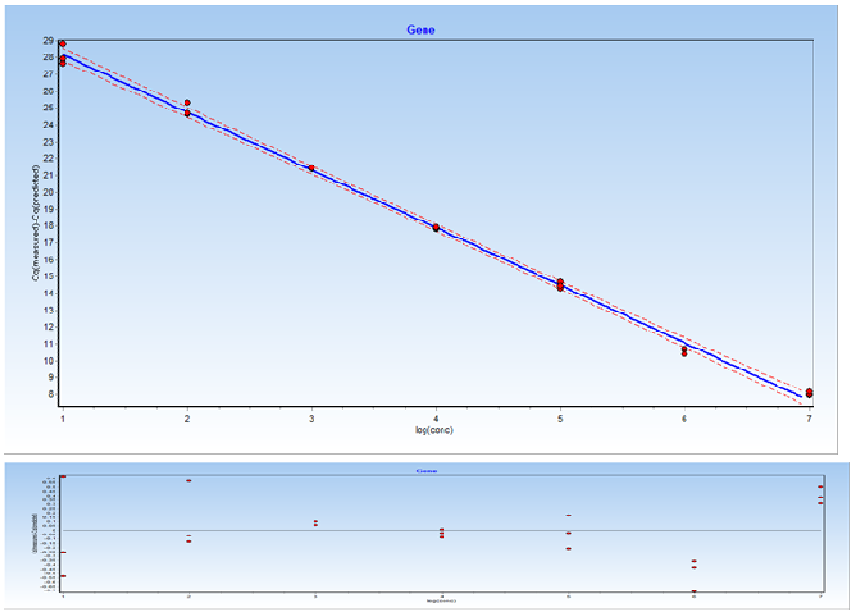
s indicate
les.





ation in
y be reversibly

The standard curve is the best straight line fit of the Cqs measured for the standard samples to their concentration in logarithmic scale (Figure 6). It is calculated using linear regression and defines the intercept, which is the Cq expected for a sample containing a single template molecule, and the slope. From the slope, the PCR efficiency is estimated. GenEx also calculates the uncertainties in the estimates of the slope and the intercept, which are reflected by the dashed lines in the plot as the Working-Hotelling area, and the confidence interval for the PCR efficiency (Figure 6). It is essential to calculate the confidence information since it reflects the precision of the estimated efficiency. In this example, the precision of the estimated efficiency is quite high, because a large number (21) of standards was used and a wide concentration range was covered (6 logs). In the literature, we frequently see standard curves based on a substantially lower number of standards. The PCR efficiencies estimated from such standard curves are highly uncertain and any corrections made are unreliable.



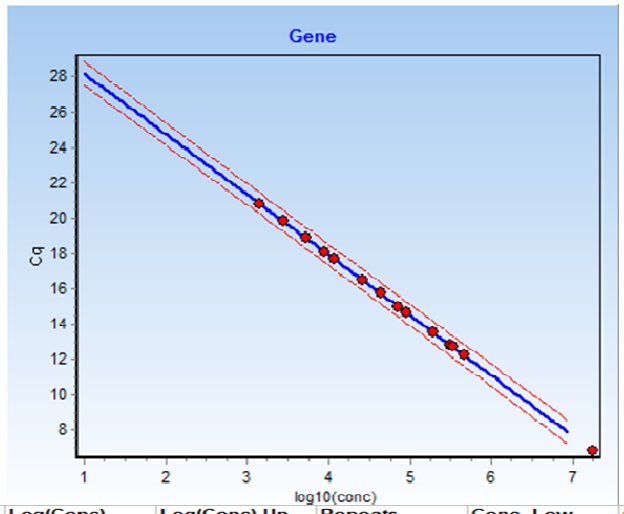
Slope:	-3.503664972...	<	-3.420833333...	<	-3.338001693...
Intercept:	31.213375170...	<	31.583809523...	<	31.954243876...
Efficiency:	0.9283672932...	<	0.9603175643...	<	0.9922678353...

ed red lines).
ed standard curve.
cluding confidence

The residual plot shows the deviations of the standard samples' measured Cq and their predicted Cq by the standard curve (Figure 6). If the straight line standard curve is adequate to model the data, residuals should fluctuate randomly. If the model is inadequate, runs of positive and negative residuals will be seen. GenEx performs a statistical test for the number of runs, and if they are too few, warns that the linear standard curve may be an inadequate model. Outliers are readily identified visually in the residual plot and GenEx further uses the Grubb's test to support the removal of outliers. In general, no more than one outlier should be removed from a standard curve (EP6-A, Vol 23, no 16 (2003). Approved guideline NCCLS). If multiple outliers are indicated, the approach used is likely to be unstable and should be

overseen. When replicates are available, the residual plot also reveals if noise increases at low concentrations.

A reliable standard curve is critical for accurate estimation of the concentrations of field samples, which in GenEx are referred to as test samples. The estimates improve if the field samples are available in replicates that can be averaged to reduce confounding variation. Concentrations of the unknown samples are estimated by entering the standard curve at the measured Cq and reading out the log of the concentration on the x-axis (Figure 7). The Working-Hotelling area, which reflects the prediction uncertainty, is wider than before because of the additional error contribution from the measured Cq. GenEx calculates confidence intervals for the estimated concentrations. The confidence intervals are symmetric around the mean in logarithmic scale, while they are asymmetric around the mean in linear scale (Figure 7). The uncertainty in the estimates is larger than what most people think. Even though the standard curve in the example is based on 7 concentrations of standard covering 6 logs, each measured in triplicates for a total of 21 readings, and the assay has 96% efficiency, the uncertainty in the estimated concentrations of the unknowns is substantial. For example, for “Test 1”, estimated concentration is 46700 copies, with the 95% confidence interval: 31000 – 61000 copies! With less accurate standard curve, the precision in the estimated concentration would be even worse.



	Log(Conc) Low	Log(Conc)	Log(Conc) Up	Repeats	Conc. Low	Conc.	Conc. Up
Test 1	4.4934490538...	4.6403340873...	4.7872191208...	3	31149.354679...	43685.175758...	61265.942768...
Test 2	3.8124309780...	3.9494692883...	4.0865075985...	3	6492.7843427...	8901.6248683...	12204.151734...
Test 3	2.9954390885...	3.1494692883...	3.30349948811...	3	989.55306551...	1410.8124655...	2011.4048274...
Test 4	4.2768800219...	4.4181659996...	4.5594519773...	3	18918.209126...	26191.839432...	36262.018684...
Test 5	4.8034314462...	4.9618931616...	5.1203548770...	3	63596.240975...	91599.512357...	131933.43718...
Test 6	3.3072282896...	3.4516399338...	3.5958515781...	3	2028.7488638...	2828.3941731...	3943.2251776...
Test 7	4.7014554271...	4.8556812249...	5.0099070228...	3	50286.965231...	71726.761884...	102307.39410...
Test 8	5.3415332445...	5.5280320167...	5.7145307888...	3	219549.90047...	337312.17479...	518239.83075...
Test 9	3.9399952514...	4.0771184966...	4.2142417417...	3	8709.5406689...	11943.139268...	16377.278779...
Test 10	5.11376236059...	5.2873499216...	5.4609374828...	3	129945.83394...	193798.28152...	289026.37955...
Test 11	3.5855139776...	3.7243779363...	3.8632418949...	3	3850.4720715...	5301.2457380...	7298.6391935...
Test 12	5.301050398549	5.4851574734...	5.6692645483...	3	200009.39609...	305602.90163...	466943.73021...
Test 13	6.9432878254...	7.2449625891...	7.5466373528...	3	8775822.3962...	17577721.901...	35207675.507...
Test 14	5.4682718668...	5.6625021750...	5.8567324832...	3	293948.91864...	459729.29095...	719005.94817...

mple’s
ıtrations including

Limit of detection

The limit of detection (LOD) is "the lowest amount of analyte in a sample that can be detected with (stated) probability, although perhaps not quantified as an exact value", with analyte here referring to the targeted nucleic acid (World Health Organization (1995) document WHO-BS/95.173, and EP17-A, Vol 24, no 10 (2004). Approved guideline NCCLS). For classical tests, when a signal is measured against a background, LOD is estimated from the standard deviation of the blank readout at the standard curve intercept. This approach is, however, not applicable to qPCR, which, due to its real-time readout, gives no reading for a negative sample (Cq for a blank sample is formally infinity). Instead, for an analytical process that involves qPCR, LOD can be estimated from multiple standard curves (Burns et al., *European Food Research and Technology* Volume 226, Number 6, 1513-1524, DOI: 10.1007/s00217-007-0683-z). A minimum of six is recommended and concentrations around the expected LOD should be assessed. The measured data are transferred to binary format indicating positive and negative PCR's and the fraction of positive calls at each concentration is calculated. LOD is the concentration at which replicates are positive at the stated rates (e.g. 95 %). GenEx fits the measured positive rates at different concentrations to estimate the LOD (Figure 8).

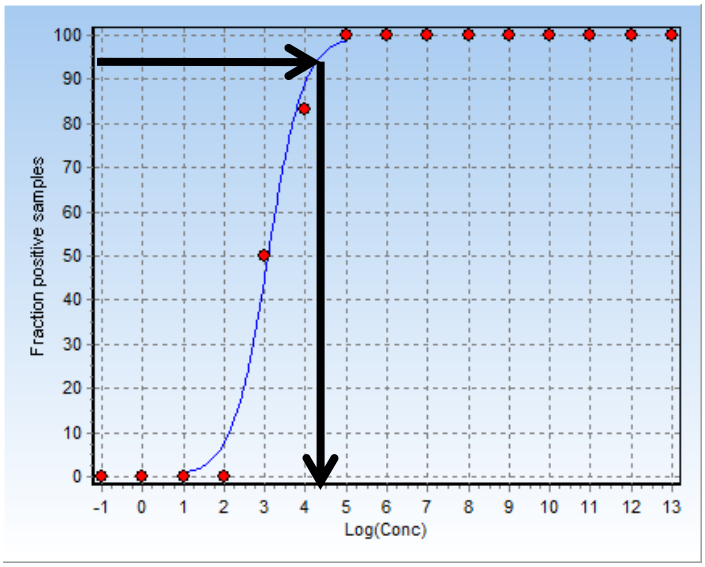


Figure 8. Estimation of LOD by fitting the rate of positive calls as function of concentration.

Selecting reference genes

With qPCR, the amount of target molecules in a sample is measured rather than their concentration. A large sample is expected to contain more target molecules than a small one and to compensate for the effect of size, normalization must be applied. There are several options as to normalize. A popular option in gene expression analysis is to normalize with reference genes, since this should not only compensate for variation in sample amount, but also for variations in extraction yield, reverse transcription efficiency, and RNA quality. In the early days of PCR, genes needed for basic housekeeping functions were thought to have stable expression and could serve as references. Experience has shown this is not always true,

and before a gene is used as a reference, this assumption should be validated. Criteria for a good reference gene is that it has a stable expression among samples and that its expression is invariant of the treatment applied. Stability of expression is reflected by the standard deviation (SD) of biological replicates. However, we cannot just take a set of samples, measure the Cqs and calculate the SD's, because we do not know how to normalize the samples for this exercise. Of course, we can use the same amount of RNA in the analyses, but then, how do we know that all samples were extracted and reverse transcribed with the same yield, and that they have the same mRNA/totalRNA ratio and the same RNA quality? Furthermore, if we were to assume that we can evaluate genes' expression stabilities based on samples normalized to the same amount of RNA, then we would have already decided that total RNA is the best norm. The gene selected based on minimum SD measured on samples having the same amount of RNA will be the gene that shows a variation that correlates the most with that of total amount of RNA, and we may then as well normalize to the amount of RNA directly. If we suspect that the total amount of RNA is not the best norm, we have to identify optimum reference genes using different strategy.

NormFinder (Intergroup variation)

File Edit

	A	B	C
1	**Gene Name**	**1**	**2**
2	PPIA	-0.0349	0.0349
3	HPRT	0.0572	-0.0572
4	YWHAZ	-0.0507	0.0507
5	TUBB5	-0.0178	0.0178
6	B2M	0.1122	-0.1122
7	ACTB	0.0029	-0.0029
8	TBP	0.0658	-0.0658
9	GAPDH	0.0086	-0.0086
10	PGK1	-0.0542	0.0542
11	18S rRNA	-0.0471	0.0471
12	GUSB	-0.0099	0.0099
13	ARBP	-0.0321	0.0321

NormFinder (Intragroup variation)

File Edit

	A	B	C
1	**Gene Name**	**1**	**2**
2	PPIA	0.0008	0.0021
3	HPRT	0.0242	0.0062
4	YWHAZ	0.0190	0.0168
5	TUBB5	0.0086	0.0043
6	B2M	0.0139	0.0057
7	ACTB	0.0158	0.0017
8	TBP	0.0154	0.0157
9	GAPDH	0.0035	0.0139
10	PGK1	0.0055	0.0013
11	18S rRNA	0.0184	0.0152
12	GUSB	0.0072	0.0006
13	ARBP	0.0092	0.0048

Figure 9. Estimated intergroup and intragroup variations with Normfinder of genes from the TATAA reference panels in representative brain samples of wild-type and obese mice.

An appropriate approach to select reference genes is a special form of analysis of variance, which in qPCR literature is best known as using the tool Normfinder (Andersen et al. (2004) *Cancer Res.* Aug 1;64(15):5245-50). Normfinder is applied to a panel of candidate reference genes that is analyzed in a set of representative samples. In essence, Normfinder calculates a global average expression of all the genes in all the samples, to which the individual genes are compared. Based on this comparison, SD for each candidate reference gene is estimated. Furthermore, if the samples are from different treatment groups, Normfinder separates the variation into an intragroup and an intergroup contribution. Figure 9 shows an example where reference genes were sought for an obesity study in mice, where wild-type mice and an obese strain were compared. The genes were selected from the TATAA reference gene panel (www.tataa.com), which was measured on seven representative mice from each strain. The intragroup variation estimated is the SD of the genes in the different treatment groups, while the intergroup variation is differential expression and sums to zero for every gene over all the groups. Good reference genes shall have low intergroup variation in all groups and negligible

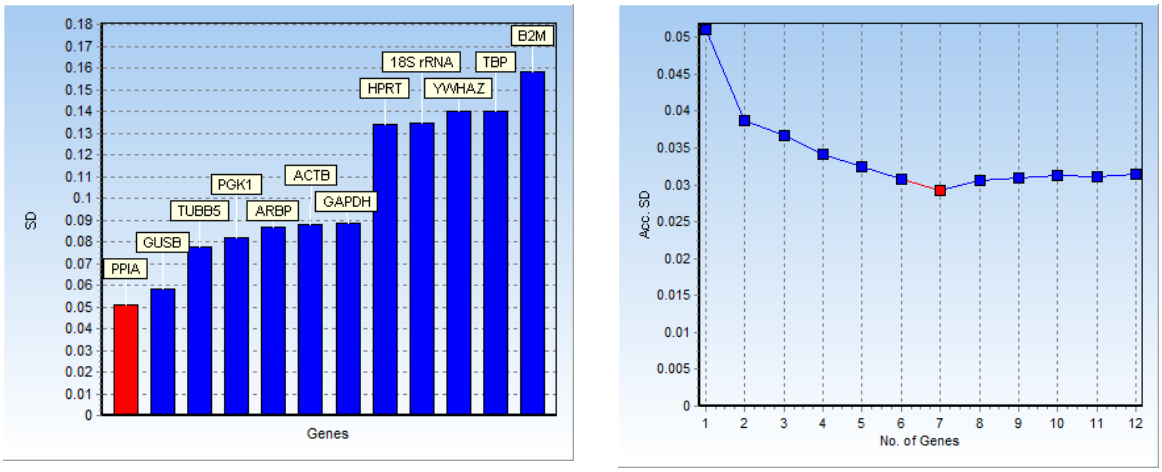


Figure 10. SD (left) and accumulated SD (right) for the reference gene candidates in the TATAA reference gene panel in 14 brain samples from mice estimated using Normfider.

intergroup variation. Good strategy using Normfinder is to inspect the calculated intra and intergroup variations to identify any genes that appear regulated or exceedingly unstable and remove them from the data set using the GenEx Data Manager. Normfinder analysis is then repeated without considering the groups, since the remaining genes are not regulated. This produces more robust result with a single SD estimate for each gene, based on which the genes are ranked (Figure 10). The gene with lowest SD is the optimum reference gene. GenEx also calculates the accumulated SD expected if multiple reference genes are used for normalization. If we use larger number of reference genes, random variation among the genes' expression partially cancel reducing the SD. Comparing the SD contributed from different number of reference genes selected based on stability, a minimum in the accumulated SD plot is obtained, indicating the number of reference genes that give the lowest SD (Figure 10). However, analyzing more genes cost time and money, and one should consider the degree of improvement and the overall noise contributed by the reference genes when making a decision. In the example, the largest improvement is observed when including the 2nd reference gene; including additional reference genes only slightly improves the result. Furthermore, the noise contribution from the best reference gene is only 0.05 cycles and as little as 0.04 cycles when combining the two best reference genes. Considering that the repeatability of a qPCR instrument is rarely less than 0.1 cycle (estimated as SD of technical replicates), using more than one reference gene, and definitely more than two, will in this study not improve the quality of the data appreciably.

Using Normfinder, normalization with reference genes can be compared to normalization with total RNA by adding an extra column in the data sheet with the RNA concentrations per analyzed sample in logarithmic scale (Figure 11). The algorithm is ignorant of the nature of the variables, and will compare their variation. For the data in our example normalization with total RNA is essentially as stable as normalization with PPIA, which is the single optimum reference gene here. In this study the samples analyzed were flash frozen biopsies from mouse brains, from which RNA of very high quality (RIN 8-9) was extracted. Our experience is that for samples with high quality RNA, normalization to total amount of RNA is often as good as normalizing with a single reference gene. In samples of poor RNA integrity, or when expression may have been induced, normalization with reference genes is preferred.

	PGK1	18S rRNA	GUSB	ARBP	log(RNA)	#group
NC2	21.18	7.61	24.88	20.49	2	1
NC3	21.07	7.72	24.83	20.32	2	1
NC4	21.24	7.78	25.03	20.62	2	1
NC5	21.48	7.7	25.12	20.6	2	1
NC6	21.29	7.68	25	20.55	2	1
NC7	21.13	7.7	24.97	20.49	2	1
HFD1	21.38	7.58	24.92	20.7	2	2
HFD2	21.28	7.66	24.97	20.64	2	2
HFD3	21.29	7.63	24.94	20.6	2	2
HFD4	21.41	7.98	25	20.65	2	2
HFD5	21.25	7.8	24.96	20.57	2	2
HFD6	21.27	7.78	24.91	20.51	2	2
HFD7	21.37	7.79	24.99	20.51	2	2

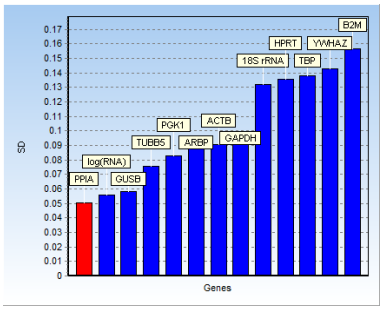


Figure 11. Left: Input data for comparison of normalization with reference genes to normalization with total amount of RNA using Normfider. A column indexing the used total RNA concentrations in logarithmic scale has been added. Right: Output indicating that for these particular samples total RNA normalization is almost as stable as normalization with the single optimum reference gene.

An older method to identify good reference genes that still is being used is geNorm. It uses the same input data as Normfinder, but it does not consider groups; all samples are treated as being from a single population. geNorm sequentially eliminates the gene that shows the highest variation relative to all the other genes based on paired expression values in all the studied samples. The variability is reflected by a so called M-value (Figure 12). Because of the elimination process, geNorm cannot identify an optimum reference gene, and ends up by suggesting a pair of genes that shows high correlation and should be suitable for normalization. The M-value is related to the SD, but as calculated, the M-values for the genes are based on different sample sizes and are therefore not strictly comparable. Furthermore, as the comparison of any individual candidate gene is performed toward a plurality of genes, assumed to resemble most closely the anticipated stable behavior, it is prone to systematic failure where group of co-regulated instable genes may be involved in the analysis. Any such co-regulated complex of instable genes may dominate over the stable genes and hence point at deviant genes as candidates. Usually, the gene rankings by geNorm and Normfinder are similar, which is reassuring. Should the rankings differ, there would be a reason to suspect the selection to include one or more regulated genes, and the result should be interpreted with caution.

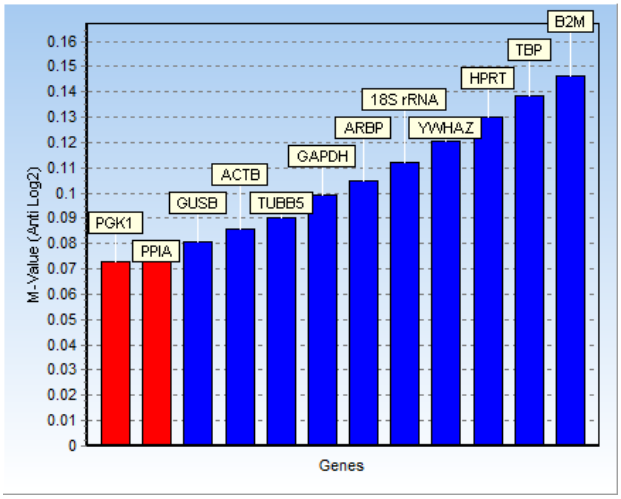


Figure 12. Ranking of reference gene candidates be sequential elimination using geNorm. The final two genes selected cannot be compared further.

Relative quantification

Treatment groups are readily compared visually in bar graphs using descriptive statistics (Figure 13) and statistical comparison is made using ANOVA (one factor, two or more levels) or 2-way ANOVA (two factors, two or more levels each) or, in the case of two groups, with either t-test (paired/unpaired, 1-tail/2-tail) or non-parametric tests (Mann-Whitney, Wilcoxon). The difference between the groups is shown in either linear or logarithmic scale and the confidence interval is indicated (Figure 14). Note that the confidence interval of the differential expression is asymmetric when data are presented in linear scale. When expression of many genes is compared, GenEx offers several means to control for the false discovery rate due to multiple testing, including Bonferroni, Benjamini-Hochberg, Westfall & Young, and Benjamini-Yekutieli (Figure 15).

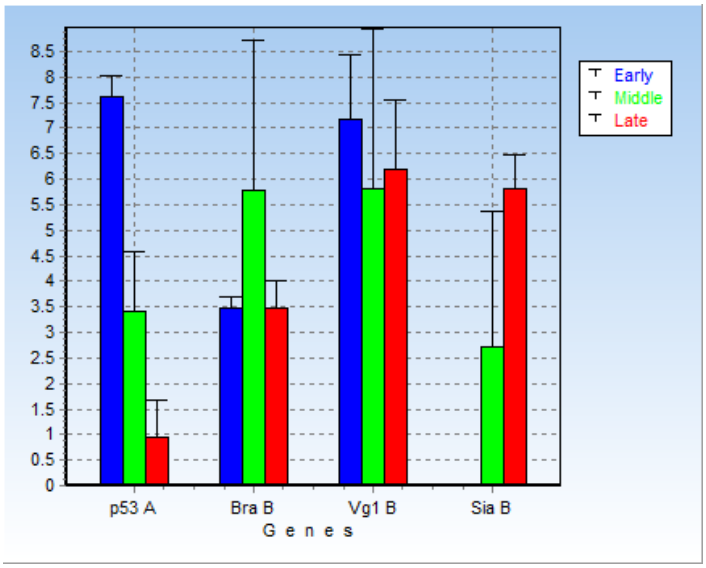


Figure 13. Comparison of the expression of four genes in three groups. Bars indicate mean expression and the error bar either SD, SE or CI.

A	B	C	D	E
	Bra B (Early)	Bra B (Late)	p53 A (Early)	p53 A (Late)
	3.1946979940...	3.1020980521...	8.1487948832...	0
	3.5002778021...	3.9447575230...	7.9635949995...	1.01859936...
	3.7502976451...	3.2039579882...	7.7783951158...	1.48159906...
	3.6113977323...	3.4447178370...	7.4079953484...	1.111199302...
	3.4076778602...	4.6948170520...	7.2227954647...	1.94459877...
	3.3058179242...	3.0743180696...	7.2227954647...	2.03719872...
		2.6946583079...		0
		3.7595576393...		0
KS	0.1122441315...	0.1752407666...	0.2074473675...	0.23978803...
Norm. Dist.	Passed	Passed	Passed	Passed
KS P-Value	>0.1	>0.1	>0.1	>0.1
Count	6	8	6	8
Mean	3.4616944930...	3.4898603086...	7.6240618794...	0.94914940...
STDEV	0.2028269917...	0.6291431816...	0.3957677171...	0.86122924...
df		12		12
SD^2		0.2480368286...		0.49793092...
t		0.1047178816...		17.5153018...
P (2-tail)		0.9183296279...		0.00000001
Confidence level		0.9499999999...		0.94999999...
SEM		0.2373480727...		0.34470251...
t*		2.1788129915...		2.17881299...
Difference (log scale)		-0.028165815...		6.67491247...
CI start/end (log)	-0.5453028800...	0.4889712487...	5.9238701579...	7.42595479...

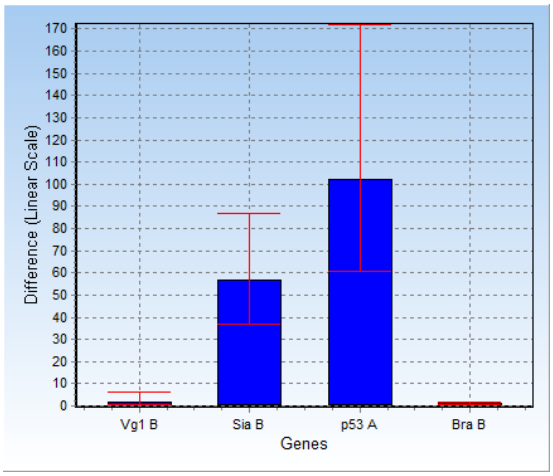


Figure 14. Comparison with t-test of the expression of four genes between two groups. Left: Descriptive statistics includes normality test and p-values. Right: Bar graph showing the differential expression in linear scale. Note, the indicated confidence intervals are asymmetric around the means.

(One) vs (Two)	KS Group 1 (...	KS Group 2 (...	Difference bet...	P Value	P (Benjamini-Hoc...	P (Westfall & Young)	P (Benjamini-Yekutieli)
RG4	Passed	Passed	-1.18826398	0.00005062	0.0011147	0.00257	0.0051251
RG17	Passed	Passed	-1.10118023	0.000118	0.00127812	0.00494	0.00587647
RG8	Passed	Passed	-1.36187953	0.00012155	0.00127812	0.00494	0.00587647
RG7	Passed	Passed	-1.18182055	0.00031428	0.00150089	0.01037	0.00690071
RG2	Passed	Passed	-1.49825382	0.00036413	0.00150089	0.01081	0.00690071
RG10	Passed	Passed	-1.22262582	0.00052804	0.00197747	0.01433	0.00909189
RG11	Passed	Passed	-1.12882321	0.00126256	0.00368418	0.02837	0.01693892
RG19	Passed	Passed	-0.97831093	0.00218714	0.00498312	0.04268	0.02291108
RG1	Failed	Passed	-1.08236183	0.0022424	0.00498312	0.04268	0.02291108
RG16	Passed	Passed	-0.90007345	0.00250654	0.00501308	0.04268	0.02304884
RG12	Passed	Passed	-0.83237989	0.00286956	0.00522101	0.04268	0.02400484
RG5	Passed	Passed	-1.0327149	0.00303341	0.00522101	0.04268	0.02400484
RG14	Passed	Passed	-1.04575987	0.00407095	0.00635184	0.04449	0.02920409
RG3	Failed	Passed	-0.873477	0.00447486	0.00639266	0.04449	0.02939178
RG20	Passed	Passed	-1.05784231	0.00701563	0.00969331	0.05364	0.04456732
RG15	Passed	Passed	-0.61494415	0.00814714	0.01029238	0.05364	0.04732171
RG18	Passed	Passed	-0.76938131	0.00932495	0.01145255	0.05364	0.05265586
RG6	Passed	Passed	-0.73924194	0.01954786	0.0234691	0.06705	0.10790482
RG9	Passed	Passed	-0.77903746	0.04688727	0.04976569	0.09908	0.22880968
RG13	Passed	Passed	-0.48639622	0.20140296	0.20206905	0.20406	0.92906087

Figure 15. Comparison of the expression of multiple genes between two group corrected for the high false discovery rate due to multiple testing. Columns 2 and 3 indicate result of normality test and column 4 the differential expression in log scale. Colum 5 indicates p-values calculated with t-test. Green indicates p-values considered significant based on Bonferroni correction, yellow are p-values below stipulated uncorrected confidence level that are not significant with Bonferroni correction, and red are p-values above confidence threshold. The three last columns indicate p-values calculated using Benjamini-Hochberg, Westfall & Young, and Benjamini-Yekutieli corrections for the high false discovery rate.

Expression profiling

T-test and ANOVA are univariate methods that analyze the expression of every gene separately, effectively assuming that the genes are expressed independently of each other. This is rarely the case; genes' expressions tend to be correlated. This correlation can be exploited in the analyses using multivariate statistical methods. GenEx offers several unsupervised as well as a selection of supervised methods to classify samples and categorize genes based on expression profiles. Unsupervised methods classify samples and genes based on the measured profiles only. They include classical hierarchical clustering combined with heat map, which can be based on various clustering schemes including the Ward's algorithm and several distance measures including the Euclidian distance and the magnitude of the Pearson correlation. While the Euclidian distance clusters genes based on similarities and consider up-regulation and down-regulation being opposite, hence anticorrelated, distance based on the magnitude of the Pearson correlation considers up-regulation and down-regulation to be correlated. The latter is useful to classify, for example, genes that show the same temporal response to treatment independently of the genes being up or down-regulated. The clustering is visualized in a dendrogram, which in GenEx can be mirrored in every node. Mirroring in a node changes the visual appearance of the dendrogram, producing an equivalent mathematical solution. A small Self Organized Map (SOM) can be used to force classification into a defined small number of groups based on expression similarities. SOM can also be used to validate a classification model, based on the distribution of samples/genes in a large map. Principal Component Analysis (PCA) groups samples/genes based on correlated expression in reduced space. Figure 16 shows example of hierarchical clustering, SOM, and PCA of genes expressed during the development of the African claw frog Xenopus

Laevis from the oocyte to tadpole stage (Bergkvist et al. (2010) *Methods* 50:323–335). Hierarchical clustering of sample and genes can be combined showing also the measured intensities in a heat map (Figure 17). The appearance of the heat map can be changed in GenEx within equivalent mathematical solutions by mirroring the dendrograms' in nodes.

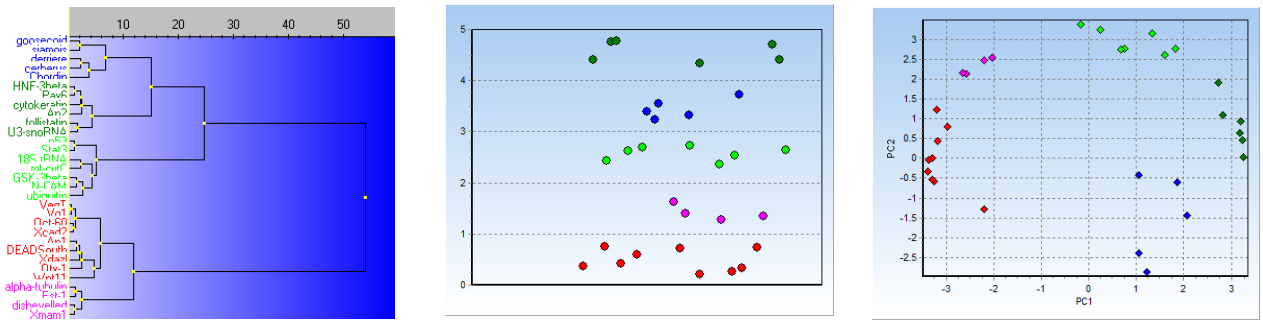


Figure 16. Classification of genes expressed during developmental of Xenopus laevis based on hierarchical clustering (left), SOM (center), and PCA (right). The dendrogram can be mirrored in any of the white nodes.

Supervised methods require a training set of samples with known classification; for example, negative and positive samples, or short, medium and long term survivals. A model is developed based on the training set that can be used to classify new data (in GenEx called test data). The procedure is similar to a regression based on standard curve, but here it is based on multiple genes and the model does not have to be linear. Supervised methods available in GenEx include Partial Least Squares (PLA), which is used to calculate a single standard curve based on the expression of multiple genes to predict concentrations or other measures of test samples. Potential Curves is a variant for prediction of new data based on PCA, and Artificial Neural Network (ANN) and Support Vector Machines (SVM) are multivariate non-linear methods to classify samples. Logistic regression, Probit, Receiver Operating Characteristics (ROC), and Survival Analysis will soon also be available.

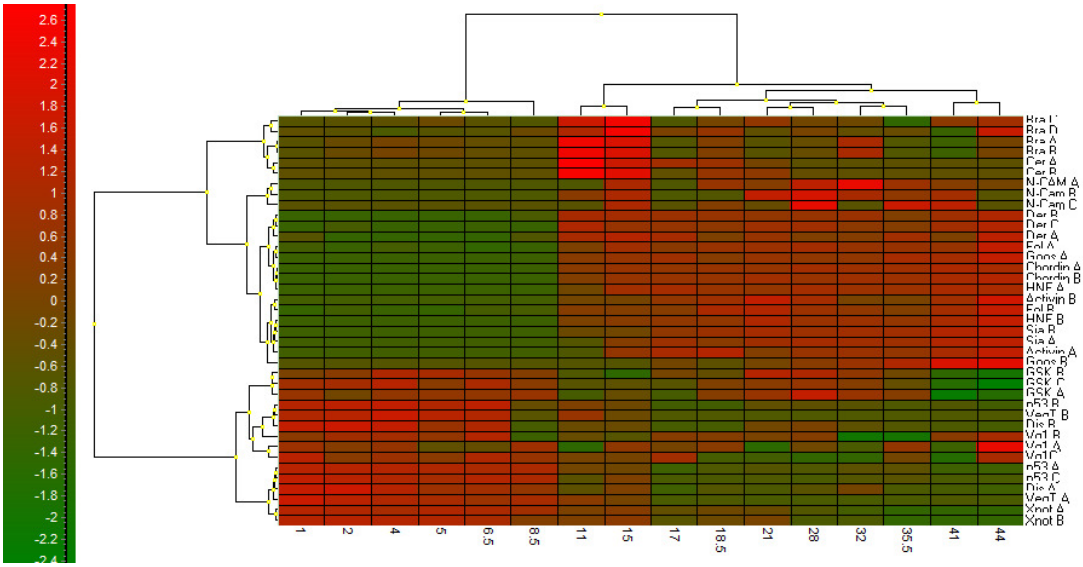


Figure 17. Classification of genes and samples visualized with dendrograms and a heat map indicating expression levels. The heat map appearance can be altered by mirroring in any of the dendrograms' nodes.

Experimental design

Designing experiments is more difficult than analyzing results. In a well designed experiment, confounding variation is minimized and the number of subject is sufficient to obtain conclusive results. A good strategy is to perform a fully nested pilot study before specifying the test protocol for a larger study (Figure 18) (Tichopad et al. (2009) *Clinical Chemistry* 55:101816–1823). Figure 19 shows the result of a nested pilot study wherein three heifers were studied by collecting three blood samples of each that were reverse transcribed in triplicates, and each cDNA was analyzed in triplicate using qPCR. Using a nested ANOVA, the variation arising from the different experimental steps can be estimated and expressed either as standard deviations or as variance contributions (Kitchen et al. (2010) *Methods* 50, 231–236). While β-actine and Caspase-3 show generally low standard deviations in all steps, Interleukin 1-β and Interferon-γ levels varied substantially among the heifers. For liver samples, the picture was different, with the data evidencing large variation in the sampling step. Knowing the costs associated with the different experimental steps, the follow up study can be cost optimized. For example, for genes exhibiting SD = 0.1 cycle for the qPCR and RT steps, SD = 0.2 cycles for the sampling/extraction step, and SD = 1 cycle for the variation among the animals, and assuming a cost of 1 unit for the qPCR, 3 units for the RT, 10 units for sampling/extraction, and 100 unit for each animal, with a total budget of 1000 units, the best we can do is to analyze 8 animals,

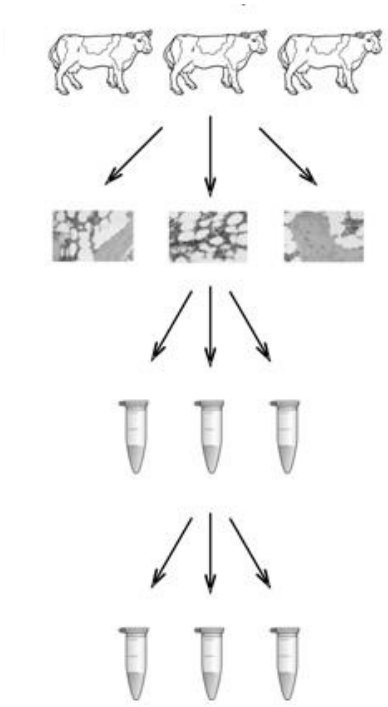


Figure 18. Nested experimental design. Three subjects (heifers) are tested. Three samples are collected from each subject and extracted. The extracted material is reverse transcribed in triplicates. Each cDNA is analyzed in triplicates for qPCR.

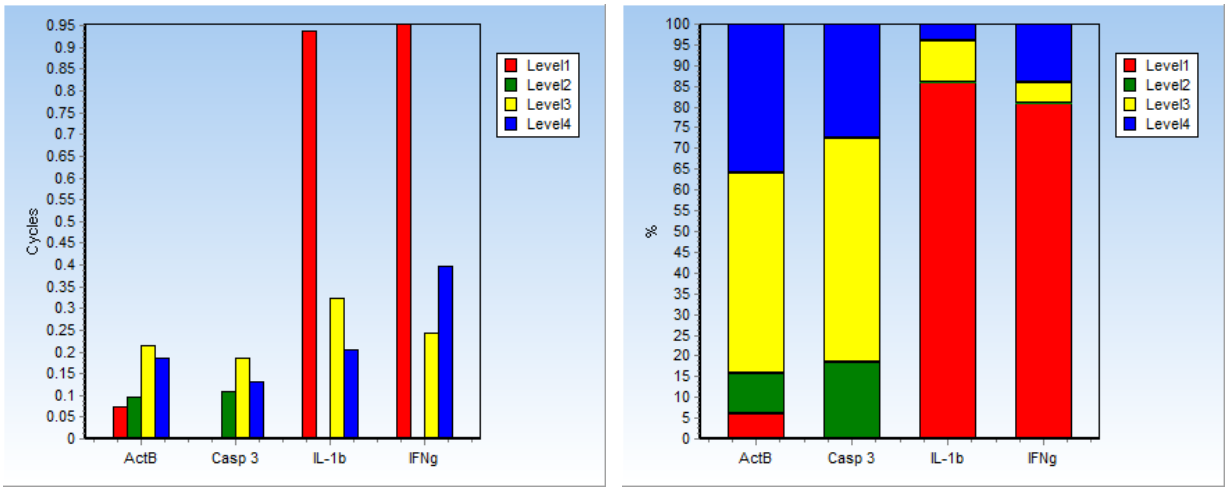


Figure 19. Decomposition of the variance in a 4-level nested study of blood samples collected from heifers. Left: estimated standard deviations (SD) among heifers (red), among replicate blood samplings (green), among reverse transcription replicates (yellow), and among qPCR replicates (blue). Right: Same data presented as variance contributions expressed in percentages.

sample each animal once, perform RT in triplicates and qPCR in duplicates. The total standard error (SE) for this study is expected to be about 0.36 cycles (Figure 20). Using the same tool, the SD among animals analyzed with a single sample collected from each animal, RT performed in triplicates and qPCR in duplicate is estimated to 1.02 cycles. This can be fed into a Power analysis to estimate the number of animals needed to ensure a particular difference with certain confidence and power. If we accept 5 % false positive rate (95 % confidence) and 5 % false negative rate (95% power) we construct a graph showing how many subject are needed to measure a particular difference due to treatment, For example, to measure a 2-fold difference (ΔCq of 1) we require under these criteria 15 animals (Figure 20).

	Level 1	Level 2	Level 3	Level 4	Tot. Var.	SD	Cost
Design Total	8	1	3	2	0.130625	0.3614208073...	1000
Design Subject	1	1	3	2	1.045	1.0222524150...	125

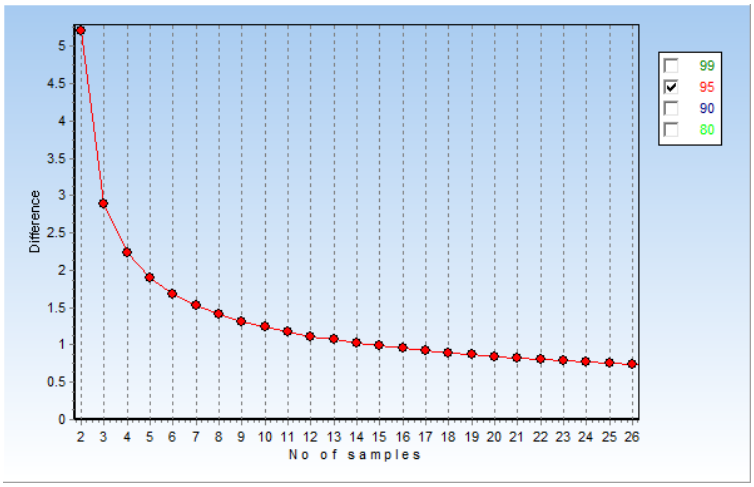


Figure 20. Power analysis assuming SD=1.055 cycles and 5% false positive and 5 % false negative calls. Straight line and arrow indicates that 15 subjects are needed to measure a difference of 1 cycle due to treatment under those conditions.

More information

An extensive help file is accessible in the software that further describes how the various analyses are performed. Help can also be found in the audiovisual tutorials at MultiD's homepage (www.multid.se/tutorials.php) and in the free support forum, where hundreds of GenEx users share experiences and advises (www.qpcrforum.com).