8 The road from qualitative to quantitative assay: What is next?

Michael W. Pfaffl

INTRODUCTION	110
PRE-PCR STEPS	111
INVENTIONS MADE IN "ABSOLUTE" QUANTIFICATION ASSAYS	112
HOW THE RELATIVE QUANTIFICATION STRATEGY CHANGED	114
WHAT ABOUT PCR EFFICIENCY?	115
ASSAY VARIANCE AND HOW TO PERFORM A PROPER NORMALIZATION	116
EXPRESSION PROFILING, qPCR BIOINFORMATICS, AND STATISTICAL ANALYSIS	117
WHAT IS NEXT IN REAL-TIME PCR?	118
GENE EXPRESSION AND MORE – THE SYSTEM BIOLOGY IDEA	120
CONCLUSION	124
REFERENCES	124

The PCR is widely used in many applications throughout the world. It has its secure place in the history of molecular biology as one of the most revolutionary methods ever. The principles of PCR are clear, but how can the reaction procedure be optimized to bring out the best in each assay? What is the status quo and what is next? Where are there areas for improvement?

INTRODUCTION

PCR is defined as a relatively simple heat-stable *Taq* polymerase–based technique, invented by Kary B. Mullis and coworkers,[1,2] who were awarded the Nobel Prize for chemistry in 1993 for this discovery. However, this terrain is contested, and many other scientists were instrumental in making PCR work in all kinds of deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and protein (immuno quantitative PCR [qPCR])–based applications. Reverse transcription (RT) followed by PCR represents a powerful tool for messenger RNA (mRNA) quantification.[3–5] Nowadays, real-time RT–PCR is widely and increasingly used

because of its high sensitivity, good reproducibility, and wide dynamic quantification range.[6,7] Today, quantitative real-time RT–PCR (qRT–PCR) represents the most sensitive method for the detection and quantification of gene expression levels. It has its tremendous advantages in elucidating small changes in mRNA expression levels in samples with low RNA concentrations, from limited tissue samples and in single cell analysis.[8,9] This [what] is a particular requirement of expression profiling, which focuses on the fully quantitative approach for mRNA quantification, rather than simply qualitative analysis.

The enormous potential for scientific and diagnostic assays makes a comprehensive understanding of the underlying principles of RT–qPCR mandatory. As a quantitative method, it suffers from accumulated problems arising during the amplification workflow in (1) the pre-PCR steps (tissue handling, RNA extraction and storage), (2) the RT and PCR steps (RT and PCR enzyme, primer design, detection dye, plastic ware, sealing), and (3) the post-PCR steps (data acquisition, background correction, quantification method, efficiency correction, normalization, statistical testing, data visualization) (summarized by Pfaffl[10]). Importantly, the absolute fidelity of a qRT–PCR assay is associated with its "true" specificity, sensitivity, reproducibility, robustness, and correctness.[11]

This chapter explains the improvements in chemistry, hardware, and software over the last two decades; focuses on considerations of specificity, sensitivity, variability, reproducibility, and data analysis; and presents some new ideas for data analysis.

PRE-PCR STEPS

The so-called pre-PCR steps are important and influence the result of a quantitative assay in a substantial way.[12,13] The process of sampling, tissue handling, and storage, followed by RNA extraction, is important for a reliable and quantitative assay. The scientific community has recognized this over the last years, and the preanalytical steps are now gaining more attention. The development of RNA integrity testing by innovative lab-on-a-chip capillary electrophoresis has made a particularly big step toward quality control. All pre-PCR steps up to the extracted total RNA can now be carefully controlled to preserve the quality and integrity of the RNA material. It is well known that mRNA is sensitive to degradation by post-mortem processes and inadequate sample handling or storage.[14] For a reliable quantification we need high integrity RNA that should be preferentially free of any DNA or inhibitors.[15,16] To prevent any RNA degradation, we recommend the RNA*later*® (Ambion) and PAXgene™ systems (Qiagen), which were recently optimized for high-quality total mRNA and microRNA extraction.[17] The accuracy of gene expression evaluation is recognized to be influenced by the quantity and quality of starting RNA.[18] The RNA purity and integrity are the most determining factors for the overall success of RNA-based quantification. Starting with low-quality RNA may strongly compromise the results of downstream applications that are often labor-intensive, time-consuming, and highly expensive.[18,19] It is

therefore important to use high-quality intact RNA, ideally with RNA integrity numbers higher than five[12] as a starting point in quantitative molecular biological as well as diagnostic applications. In clinical applications with unique and precious limited tissue material – such as samples obtained after surgery, by biopsy, or from single cell studies – a reliable RNA quality analysis is necessary.[20–22]

A second important parameter relating to the pre-PCR step is the RT. It is one of the most variable reaction steps in the entire quantification assay. Even today, after the development of recombinant enzyme types with various new properties, it is the major source of variability. Each reverse transcriptase enzyme has specific reaction conditions that have to be optimized for each application and primer pair. The reaction fidelity suffers from differences in RT efficiencies, resulting in highly variable amounts of synthesized complementary DNA (cDNA) copies.[13] For most quantitative applications, Moloney murine leukemia virus (MMLV) H^- RT is the enzyme of choice,[23,24] as its cDNA synthesis rate can be up to fifty-fold greater than that of avian myeloblastosis virus (AMV).[25,26] Newly available thermostable RT enzymes maintain their activity up to 70°C, thus relieving the amount of secondary RNA structure during RT and permitting increased specificity and efficiency of first primer annealing. Each of the enzymes used to generate cDNA differs significantly with respect to specificity as well as cDNA yield and variety. Consequently, it is important to realize that RT–PCR results are comparable only when the same priming strategy and reaction conditions are used.[13] In addition, by using mFold software,[27] the first primer binding site can be checked for better mRNA accessibility and the RT reaction step can be optimized to prevent any false priming.[22] To circumvent these high inter-assay variations, an internal quality control for cDNA synthesis can be used. These internally grown controls can be artificially, like alien RNA, or naturally occurring reference genes, like glyceraldehyde 3-phosphate dehydrogenase (GAPDH), albumin, actins, tubulins, cyclophilin, microglobulins, or ribosomal subunits (18S or 28S ribosomal RNA [rRNA]).[28,29]

In summary, the vast efforts made at improving the RT step in terms of enzyme development, protocols optimizing the preamplification step, and software improvements, as well as the RNA integrity testing, have resulted in substantial improvements to the standardization and reliability in the pre-PCR setup.

INVENTIONS MADE IN "ABSOLUTE" QUANTIFICATION ASSAYS

The fidelity of a quantification assay is measured by its specificity, low background fluorescence, steep fluorescence increase, high amplification efficiency, and high level plateau. The absolute dynamic range of the detectable fluorescence (maximal plateau minus background fluorescence) should be maximized in a quantitative assay. For single PCR product reactions with well-designed primers, intercalating dyes like SYBR® Green I work perfectly well, with spurious nonspecific background showing up only in very late cycles.[30,31] Among the real-time detection chemistry, SYBR® Green I and probe-based TaqMan® assays produce

comparable quantification ranges and sensitivities, although SYBR® Green I detection is more precise and produces a more linear decay plot than do the TaqMan® probes.[32,33] Nowadays new intercalating and saturated dyes are available (SYBR® GreenER, SYTO® 9, EvaGreen®, LCGreen®, BEBO, BOXTO) that give higher fluorescence readouts and reduce the risk of primer–dimer formation.[34] This [what] has the added advantage that, at least in theory, the sensitivity of the assay should be increased, because C_t acquisition can take place at earlie cycles.

Assay improvements are not solely due to improved dyes and chemistries, however. A whole range of new polymerase types and mixtures has been introduced to the market. In addition to single polymerase reaction mixes, multiple polymerase mixes are now available, such as combinations of the classical *Taq* polymerase and proofreading polymerases. "Hot start PCR" was already a topic in the early days of classical block PCR, when we worked with wax to prevent early reaction start-up at too-low temperatures. Combining PCR components at low temperatures often leads to nonspecific high backgrounds and low product yield. Certain PCR enzymes exhibit significant polymerase activity at the typical reaction setup temperatures lower than 25°C or during the ramping steps. Nonspecific primer annealing and extension at nonrestrictive temperatures produce undesirable products that are amplified throughout the remaining PCR cycles. Today the polymerase is usually activated via antibody blockage,[35,36] through chemical modifications of the enzyme, or by an inert ligand that detaches immediately from the active enzyme center of the polymerase when there is an increase in temperature. The inert ligand has the advantage that the activation step is unnecessary; furthermore it has a "Cold Stop" feature (Eppendorf®; 5-Prime, Germany): When the temperature drops beneath a critical threshold value during the primer-annealing step, the inert ligand "binds" onto the polymerase again and deactivates it. Again, improvements to the enzymes themselves are but one aspect of the improvements made to the qPCR assay. Significant efforts have been made to optimize buffer conditions to simplify the reaction setup. For example, "self-adjusting" magnesium (Mg^{2+}) buffers reduce the need for pipetting during PCR setup, with optimal Mg^{2+} concentration always present in the tube (Eppendorf®; 5-Prime, Germany).

Besides these chemical and enzymatic improvements, hardware, plastic ware, and cycling procedures have been improved significantly. Today the term "rapid cycling" is a synonym for quicker and better results. The heating and cooling performances of the blocks have been improved, allowing the shortening of the "ramping time" between single amplification cycles. Better block surface alloying and thinner tube materials have led to higher temperature uniformity and conduction while cycling. Therefore, the unspecific reaction times have been minimized, resulting in better PCR amplification performance.[37] Additional attention has been paid to seemingly minor items such as tube sealing: Instead of self-adhesive sealing foil that can result in poor seals of the reaction tube at plate borders, new automatic heat-sealing methods that use a glue-free and highly transparent foil guarantee tube-to-tube individual sealed reaction chambers,

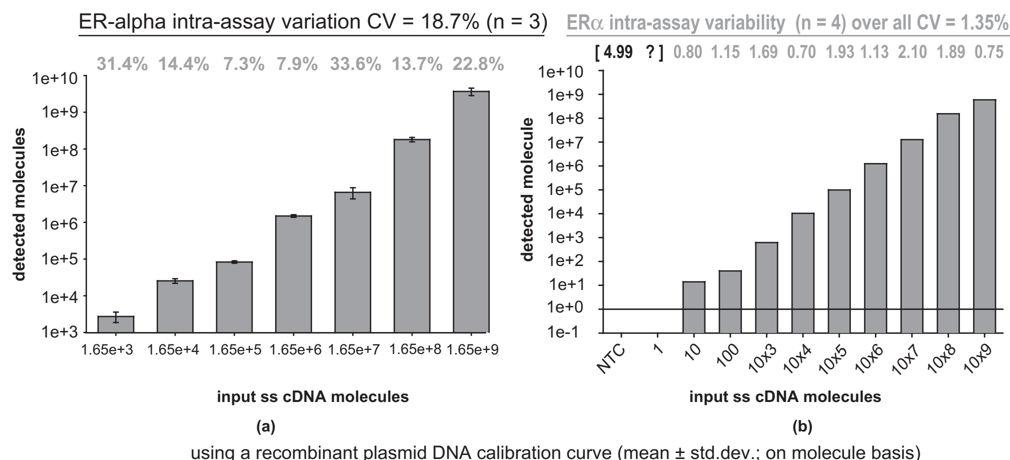

Figure 8–1. Standard curve variability performed with Estrogen Receptor alpha (ERα) single-stranded plasmid deoxyribonucleic acid (ssDNA). **(A)**, Assay variability using 1,650 to 1.65 billion ERα ssDNA start molecules and a classical SYBR® Green I dye ($n = 3$). **(B)**, One to one billion ERα start molecules using a new generation intercalating saturated dye ($n = 4$). Assay variability is indicated as a percentage. CV, coefficient of variation; cDNA, complementary DNA.

thus preventing any evaporation (Abgene, UK; 4-titude, UK; Eppendorf®, Germany).

The improvements made during the last ten years are nicely demonstrated on an estrogen receptor alpha (ERα) assay developed in 1997.[38,39] Both assays shown were run with the same plasmid DNA standard material using different kits and platforms, one in 1997 and the other in 2007 (Figure 8–1, a and b). On the left-hand side, assay variability is plotted using 1.65×10^3 to 1.65×10^9 ERα single-strand DNA (ssDNA) starting molecules and a classical SYBR® Green I dye ($n = 3$). An assay overall variability of 18.7% was derived in 1997. One decade later the standard material was run from 1 molecule to 1×10^9 starting molecules using a new generation intercalating saturated dye. The average variability in four replicates was 1.45%, which is remarkably low. Furthermore, the assay sensitivity was ten molecules per reaction tube.

Summarizing this, we can conclude that chemicals and hardware are made more sensitive and more reproducible while resulting in remarkable reductions in assay variability.

HOW THE RELATIVE QUANTIFICATION STRATEGY CHANGED

Alongside the "absolute" quantification according to a given standard curve, relative quantification has been of particular interest to all areas of physiological science. Relative quantification in qRT–PCR is easier to perform than the absolute assay setup, because a calibration curve is not necessary. It is based on the expression levels of a target gene versus one or more reference genes (sometimes called housekeeping or internal control genes). It is adequate for most purposes

to investigate minor physiological changes in gene expression levels.[40,41] The units used to express relative quantities are irrelevant, and the relative quantities can be compared across multiple real-time RT–PCR experiments.[42] Relative quantification setup determines the changes in steady-state mRNA levels of a gene across multiple samples and expresses it relative to the levels of an internal control RNA. This reference gene can be coamplified in the same tube in a multiplex assay or can be amplified in a separate tube.[43] Therefore, relative quantification does not require standards with known concentrations, and the reference can be any transcript, as long as its sequence is known.[44]

The calculation of the expression changes will be measured by mathematical algorithms that are based on the "delta delta C_t method," established originally by Livak and Schmittgen.[45] Calculations rely on the comparison of the distinct cycle, such as threshold values (C_t) at a constant level of fluorescence or C_t acquisition according to established mathematic algorithms.[46,47] To date, several quantification models that calculate the relative expression ratio have been developed. Relative quantification models with and without efficiency correction use single or multiple reference genes for normalization are available and published (summarized by Pfaffl[10]). According to such ratio calculation models, appropriate software applications were developed, such as LightCycler® Relative Quantification Software (Roche Applied Science),[48] QGene,[49] Relative Expression Software Tool (REST),[50] SoFar,[51] Data Assimilation Research Testbed (DART),[52] qPCR–data analysis and management system (qPCR–DAMS),[53] and Qbase®.[54]

The application of such algorithms that calculate PCR efficiency on a single PCR run basis has been shown to be important for the generation of correct results.[18,47] Therefore, PCR efficiency corrections are being included in new relative quantification software (e.g., REST 2008; http://REST.gene-quantification.info/). It is desirable that the real-time qPCR software applications should calculate automatically the qPCR efficiency and implement it in proven relative quantification modules.[10]

WHAT ABOUT PCR EFFICIENCY?

All qPCR methods, absolute and relative, assume that the target and the sample amplify with similar efficiency,[45] but we know that is not the case! Unfortunately, unknown samples may contain substances that significantly reduce the efficiency of the RT[12] as well as in the PCR.[55] As discussed, sporadic RT and PCR inhibitors or different RNA/cDNA distributions can occur. A dilution series can be run on the unknown samples, and the inhibitory factors often can be diluted out, causing a nonlinear standard curve.[56,57] Individual samples can generate different fluorescence histories in real-time RT–PCR. The shapes of amplification curves differ in the steepness of any fluorescence increase and in the absolute fluorescence levels at plateau depending on background fluorescence levels. The PCR efficiency has a major impact on the overall fidelity as well as accuracy of

the assay, and is critically influenced by PCR components. Efficiency evaluation is an essential marker in gene quantification procedure.[47]

A correction for efficiency, as performed in efficiency-corrected mathematical models, is strongly recommended and results in a more reliable estimation of the "real expression ratio" compared to no efficiency correction.[55] Small efficiency differences between target and reference gene generate false expression ratios, and the researcher over- or underestimates the "real and initial" mRNA amount present in the biological sample (LightCycler® Relative Quantification Software; Roche Applied Science).[48]

To conclude, quantitative efficiency corrections should be included in the automation and calculation procedure in relative quantification models, and are a major goal for the future in real-time PCR cycler and software development.

ASSAY VARIANCE AND HOW TO PERFORM A PROPER NORMALIZATION

It is important to realize that any measured variation in gene expression between subjects is caused by three sources: (1) processing variance that occurs while sampling and during the RT and PCR reactions, which must be minimized by using more replicates and by normalization with internal standards; (2) individual biological variance, which can be minimized by repeated measurements of RT and PCR reactions and by an additional normalization to an untreated control group; and (3) treatment variance.

The processing variance occurs while sampling, during RT, and during the PCR. This variance can be minimized by using multiple replicates and by normalization with internal standards, such as reference genes. The individual biological variance can be minimized by repeated measurements at RT and PCR levels and by an additional normalization to an untreated control group. In contrast, there is the treatment variation, explaining the phenotype or underlying phenomenon under investigation. This variance should be reduced by random sampling and by taking a large number of biological samples.

One major hurdle in real-time PCR gene expression studies is the removal of this experimentally induced nonbiological variation from the true biological variation. As shown before we are on the right path, but there is still some undefined assay variability left. There are several strategies to remove experimentally induced variation, each with its own advantages and considerations.[58] We can reduce reaction noise through normalization by controlling as many of the confounding variables as possible.[29] Although most of these methods cannot completely reduce all variance sources, it has been shown to be very important to control all the sources of variation along the entire PCR workflow.[59] If one does not meticulously try to standardize each step, variation can and will be introduced in the results and cannot be fully eliminated by applying normalization by reference genes.[13]

Although the use of reference genes for normalization of gene expression levels is certainly the "gold standard," some new ideas for normalization have been

recently developed.[58] The quality of normalized quantitative expression data cannot be better than the quality of the normalizer itself. Any variation in the normalizer will obscure real changes and produce artifactual changes.[44] Real-time RT–PCR–specific errors in the quantification of mRNA transcripts are easily compounded with any variation in the amount of starting material between the samples, for example, caused by sample-to-sample variation, variation in RNA integrity, RT efficiency differences, and cDNA sample loading variation.[18,24,25] Normalization of target gene expression levels must be performed to compensate for intra– and inter–RT–PCR variability (sample-to-sample and run-to-run variations). Therefore, data normalization by more than one reference gives much more reliable results.[60] Vandesompele and colleagues recommended using at least three nonregulated references to perform a proper normalization. A set of candidate reference genes has to be performed on all biological samples under investigation, and a reliable test to determine the most stable reference must be performed. This can be done by various software applets available: geNorm,[29] BestKeeper,[60] or Qbase software.[54] It still remains up to the individual investigator to choose appropriate reference gene(s) that are best for normalization in the particular experimental setting. Over the years a panel of optimal references have been reported, which are more or less stable under specific biological treatments. Also the idea of Global Pattern Recognition (GPR) was developed to evaluate expression changes in real-time PCR data.[61] By comparing the expression of each gene to every other gene in the array, a global pattern was established, and significant changes are identified and ranked. GPR makes use of biological replicates to extract significant changes in gene expression, providing an alternative to relative normalization in real-time PCR experiments.

To summarize, the normalization strategy using software applets is prerequisite for accurate quantification of RT–PCR expression profiling, which opens up the possibility of studying the biological relevance of even small mRNA expression differences. The proper normalization process revolutionized the relative quantification in real-time RT–PCR, and guided us to a more reliable result.

EXPRESSION PROFILING, qPCR BIOINFORMATICS, AND STATISTICAL ANALYSIS

In research and in clinical diagnostics, real-time qRT–PCR is the method of choice for expression profiling. Enormous amounts of expression C_t data are created. However, accurate and straightforward mathematical and statistical analysis of qPCR data and management of growing data sets have become the major hurdles to effective implementation.[62] Nowadays up to 96- and 384-well applications are the standard in research, but in the near future high-throughput applications with more than thousand PCR spots will generate huge amounts of data. Various qPCR data sets need to be grouped, standardized, normalized, and documented by intelligent software applications.[54] The main challenge remains the mathematical and statistical analysis of the enormous amount of data gained, as these

functions are not included in the software provided.[49] The so-called bioinformatics and biostatistics on real-time RT–PCR experimental data are highly variable, because various procedures are possible, involving different ways of performing background correction, threshold settings, or expression normalization. The possibilities in performing data analysis are nearly infinite! Many questions arise: Which one is the right analysis method? Can I use my generated data? Which one gives the best results, in terms of significance? Which one gives realistic results, in terms of the biological question? Which statistical test is the right one?

Prior to normalization or statistical testing, real-time qPCR data should be analyzed by automated verification methods, such as Kinetic Outlier Detection (KOD), to detect outliers and samples with dissimilar efficiencies.[63,64]

Later statistical testing in mRNA gene quantification is nowadays mainly performed on the basis of classical standard parametric tests, such as analysis of variance or *t* tests (summarized by Pfaffl[10]). Parametric tests depend on assumptions, such as normality of distributions, the validity of which is unclear.[49,65] When performing relative quantification analysis, where the quantities of interest are derived from expression ratios, assay variances might be high, normal distributions might not be expected, and it is unclear how a parametric test could be applied.[50] Up to now two available software packages support statistical analysis of expression results: QGene[49] and REST.[50] Both work on the basis of Visual Basic applets on the basis of Excel (Microsoft). In QGene, rapid and menu-guided performance of frequently used parametric and nonparametric statistical tests is provided. In REST, permutation or randomization tests are applied as alternatives to more standard parametric tests for analyzing experimental data. Both tests have the advantage of making no distributional assumptions about the data, while remaining as powerful as more standard tests, and are instead based on our knowledge that treatments were randomly allocated.[66]

WHAT IS NEXT IN REAL-TIME PCR?

During the next years, new PCR applications and improvements will be developed, both on the chemical and the hardware sides. Very interesting is the invention of high-throughput applications – even more than 384-well applications[67] – and digital PCR. Digital PCR represents a powerful example of PCR and provides unprecedented opportunities for molecular diagnostics, either on DNA or RNA levels. The technique is to amplify single DNA or RNA templates from highly diluted samples, therefore generating PCR products that are derived from one template. Thus, digital PCR transforms the exponential and analog signals obtained from conventional PCR to linear digital signals, allowing statistical analysis of the PCR product. Digital PCR has been applied in various applications for mutant detection, but offers high impact in future molecular diagnostics.[68,69]

In this section I want to focus on the new data analysis methods and how these models will help us generate more useful information from multiple gene expression data.[70] First we need a powerful concept and, of course, a set of algorithms to

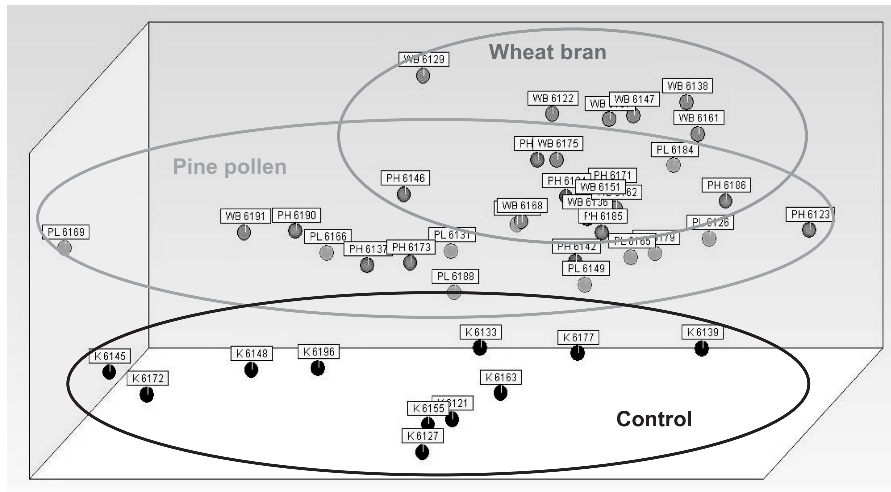

Figure 8–2. Multidimensional regression analysis via a three-dimensional scatter plot of daily intake, daily gain, and area of follicles in lymph node. Three different feeding regimens were investigated: wheat bran, pine pollen, and untreated control group. *See color plates.*

analyze extensive experimental expression data in parallel. Why? Suppose we pretend that our goal is to detect hidden interactions or correlations between genes. We may want to determine whether genes A and B are more influenced by our applied treatment than are genes C or D. A qPCR expression-profiling experiment generates a C_t value for each gene in each sample, thus recording the transcriptional activity of that gene in that particular sample. Although these data provide valuable and accurate information about the transcriptional response of the studied system, an even more powerful experimental design would incorporate an additional third parameter such as treatment time, applied treatment concentration, or type of treatment. Such studies generate so-called three-dimensional data sets (Figure 8–2) that are exceedingly informative and give more insight into the interaction of genes A and B over the parameter C.[71]

To analyze more data sets from an expression-profiling experiment, we need highly sophisticated algorithms, like cluster analysis,[72] which has been long established in the analysis of DNA array experiments, where thousands of data points have to be compared in parallel.[73,74] Gene expression clustering allows open-ended exploration of the data, without getting lost among the thousands of individual genes. Beyond simple visualization, there are also some important computational applications for gene clusters. The goal of clustering is to subdivide a set of items (in our case, genes) in such a way that similar items fall into the same cluster, whereas dissimilar items fall into different clusters.[70] This [what] brings up three questions: (1) How do we decide what is similar – that is, which genes are similarly regulated? (2) How do we use this to cluster the items? (3) How do the different treatments cluster?

The fact that these questions often can be answered independently contributes to the bewildering variety of clustering algorithms. In hierarchical clustering, all information in the data is accounted for, but the data are analyzed sequentially,

which means that not all information is considered at the same time.[70] The distance between two samples in the multidimensional space is typically calculated as the Euclidian shortest distance, by the Ward's algorithm,[75] or by a ranked correlation approach.[76]

mRNA transcripts from different genes often share similar expression patterns. Ma and colleagues[77] developed an approach to reveal related gene expression patterns. The smoothing spline clustering (SSC) algorithm models natural properties of gene expression over time, taking into account natural differences in gene expression.

To summarize, the described three-way dimensional and cluster analysis opens the way to analyze and compare gene expression data in a multidimensional fashion. It creates gene groups, treatment groups, or groups of patients with similar mRNA regulation patterns and will give us much more information than will the classical gene-to-gene comparison.

GENE EXPRESSION AND MORE – THE SYSTEM BIOLOGY IDEA

Cluster analysis of gene expression data by three-dimensional data sets or by SSC is attractive, but we need even more sophisticated approaches. We do not simply wish to compare the gene expression data; what we are really interested in is the comparison between the applied treatment and the biology. This means incorporating a whole range of additional parameters, such as genetic, protein, and metabolic data sets from our samples (Figure 8–3). To visualize this, a nutrition study in forty-five piglets will be presented.[78] Herein the gene expression data (C_t values) from various marker genes (apoptotic, cell-cycle, metabolic, pro- and anti-inflammatory markers), investigated in multiple organs (liver, stomach, jejunum, ileum, colon, lymph node, white blood cells), were implemented and compared with growth parameters (daily intake and daily gain, feed digestibility, feed conversion) as well as morphological data (length and width of villi, size of Peyer plates, various parameters from the lymph node morphology). Even more data sets, such as metabolic and bacterial counts in the gastrointestinal tract (GIT), will be implemented when available. All data were analyzed using GenEx software (http://www.multid.se).[79]

How is it possible to analyze hundreds of data sets that came from different measurement sources? How can we equilibrate all the data to make them comparable?

All data are measured by different analytical methods and therefore have different physical units. How we can bring these different data sets together and generate a complete readout to draw conclusions on treatment efficacy?

To do so, raw data should be autoscaled. Autoscaling is a well-established mathematical conversion that results in data sets of each parameter with the mean value of zero and the standard deviation of one (Figure 8–4). Autoscaling makes the expression data analysis robust.[70] Finally, all 107 data sets – that is, 107 different physiological parameters – from 45 animals underwent a parallel

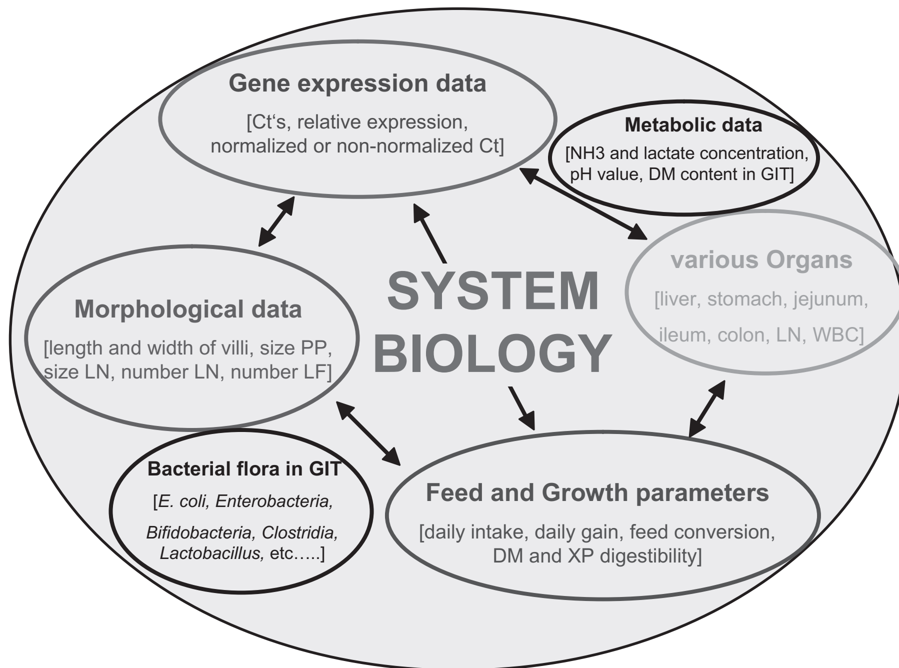

Figure 8–3. Multiple comparison of gene expression data in various organs, feed parameters, growth parameters, and morphological data (metabolic and bacterial) for a piglet feeding study for development of a system biology approach. *See color plates.*

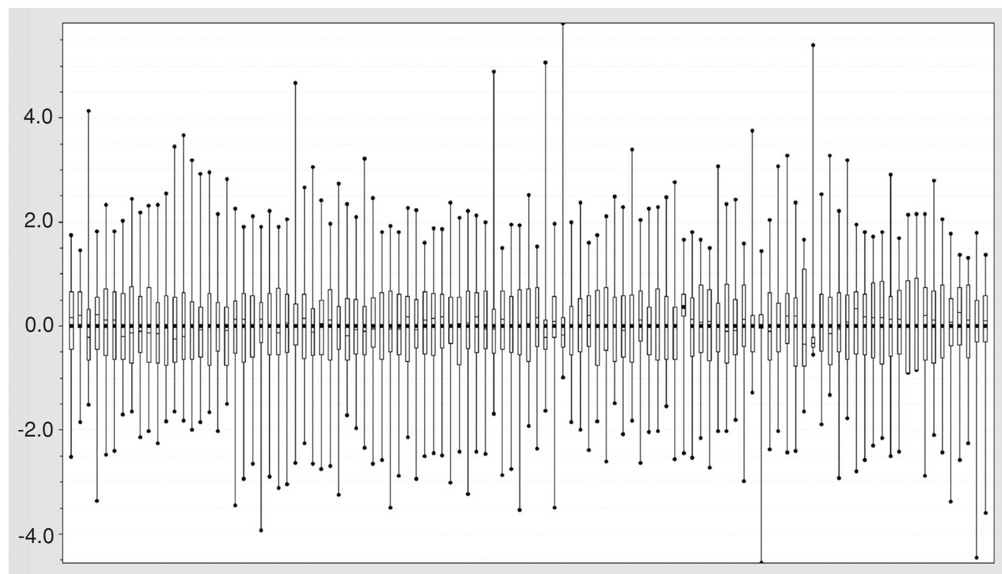

Figure 8–4. Autoscaled data set from 107 parameters in 45 piglets.

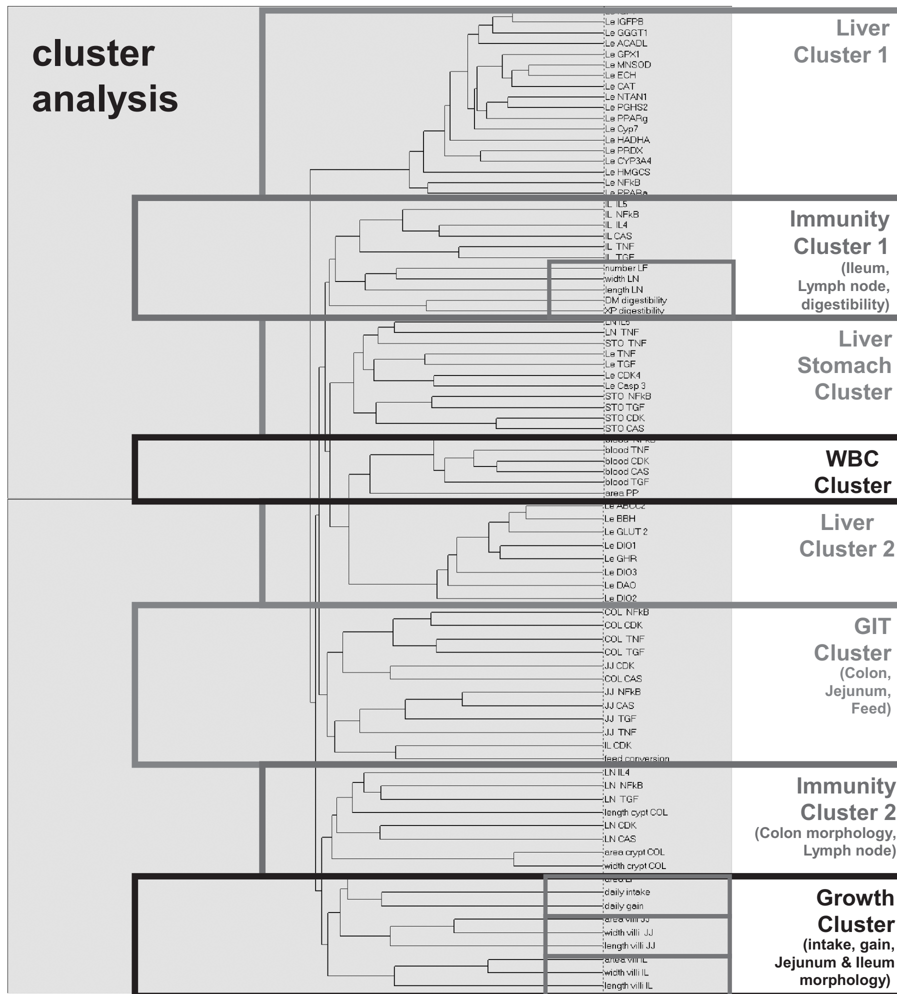

Figure 8–5. Dendrogram as result of cluster analysis. Various data sets cluster in main cluster and subcluster. WBC, white blood cell. *See color plates.*

cluster analysis. Figure 8–5 shows a dendrogram of the applied study. The dendrogram shows various main clusters, and of course subclusters, which correspond to genes expressed in distinct tissues, such as liver (Figure 8–6) or GIT, or belong to a functional group. As an example, many genes expressed in the liver group together, showing that gene expression is not solely regulated gene by gene. Furthermore, there is greater coherence between the individual tissues, and there is further regulation on the tissue level as well.

In immunity cluster 1, the cluster algorithm grouped the following parameters: immunological marker genes expressed in the ileum, lymph node relevant parameters, and feed parameters such as digestibility of dry matter and crude protein content (Figure 8–6). Here a direct conclusion about the overall correlation between gene expression data, morphological appearance in the GIT, and feed properties can be drawn. Furthermore, the growth cluster (Figure 8–7) functioned as a proof of concept. Within all 107 data sets the software conspicuously grouped

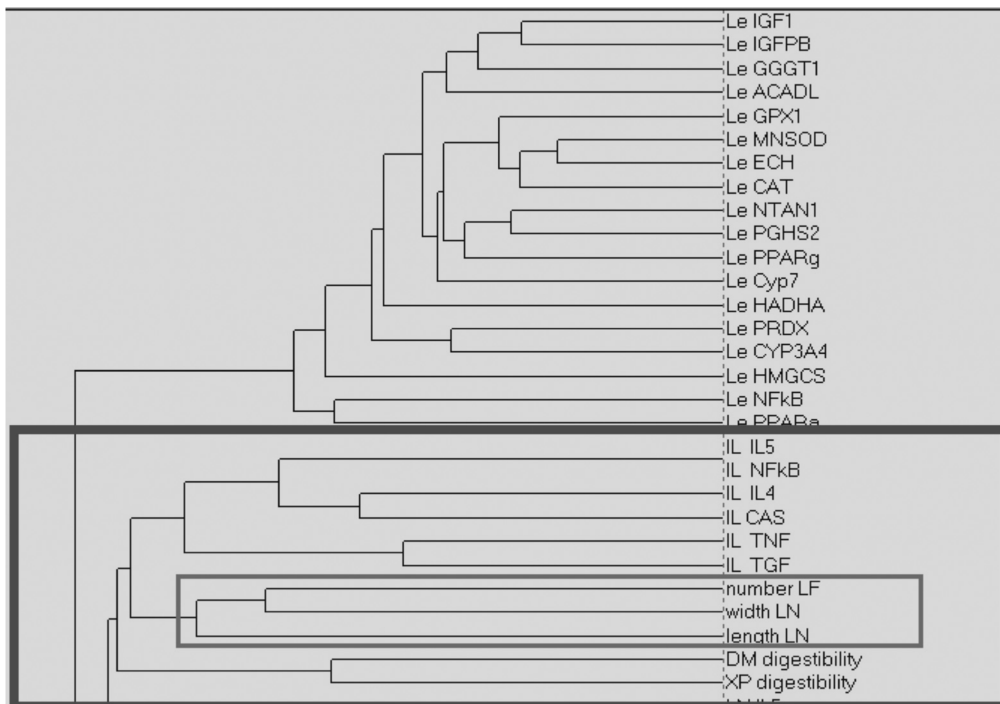

Figure 8–6. Dendrogram of liver and immunity subclusters. All liver genes cluster together in the upper part. Gene expression of immunological marker genes in the ileum, lymph node morphology, dry matter, and crude protein digestibility cluster clearly together (blue frame). *See color plates.*

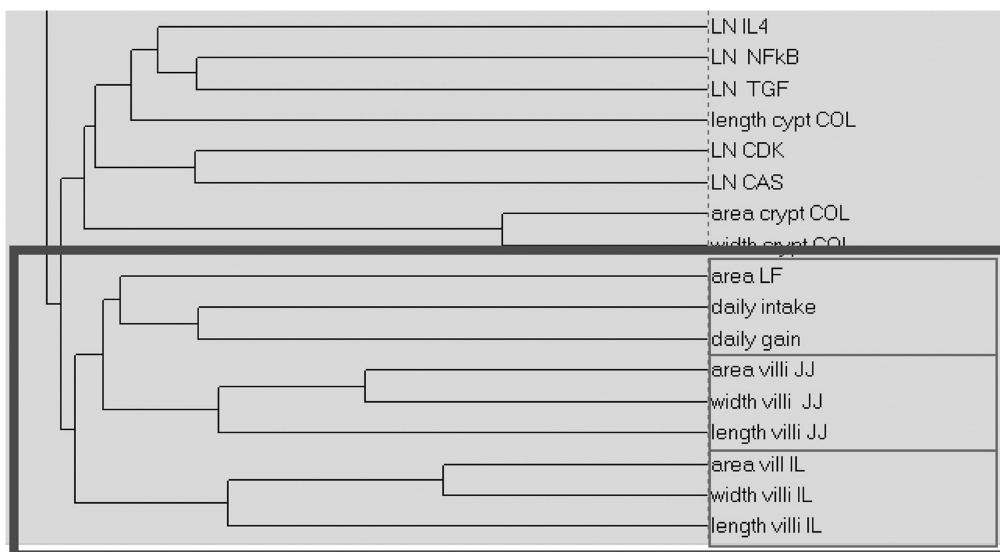

Figure 8–7. Dendrogram of growth cluster and immunity subclusters. *See color plates.*

side by side morphological data from jejunum and ileum (length, width, and area of the villi). This [what] shows us that cluster analysis works and generates no fictitious and meaningless results. Importantly, within the growth cluster morphological data correlate highly with feed intake, daily gain, and the area of the lymph follicles in the ileal mesenterial lymph node.

Compared to cluster analysis, other algorithms are known to detect hidden structures between genes and other parameters. The idea behind the self-organizing map method[80] is to reflect variations in the expression profiles as a collection of cells, each with a representative expression profile, that are arranged to form a map with smooth changes in the profiles. When the expression profiles of the samples are located on the map, similar samples will be found close to each other.[70] In some situations the detailed expression pattern also can have prognostic value. Traditionally expression profiles are measured using microarrays, by which the expression of all genes can be assessed in a single experiment. However, the quality of microarray expression data usually is not good enough for detailed classification and accurate prognosis.[70] Real-time PCR gives much more information, is more sensitive, has a wider dynamic range, and has higher reproducibility.[10,42]

CONCLUSION

During the past two decades, important advances have been introduced, making quantification much more reliable. Improvements have been made in preanalytical steps, detection chemistry, applied dyes, quantification strategy, software application, and instrumentation. These improvements have led to the development of sensitive and stable assays whereby mRNA transcripts can be quantified in high throughput and precisely in a short time. The benefits in terms of increased sensitivity, reduced variability, reduced risk of contamination, increased throughput by automation, and meaningful data interpretation are obvious, even beyond gene expression data.

REFERENCES

1. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51**: 263–273.
2. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, et al. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**(4732): 1350–1354.
3. Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. *Genome Research* **6**: 986–993.
4. Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology* **11**(9): 1026–1030.
5. Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5′-3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **88**(16): 7276–7280.

6. Ding C, Cantor CR (2004) Quantitative analysis of nucleic acids – the last few years of progress. *Journal of Biochemistry and Molecular Biology* **37**(1): 1–10.
7. Pfaffl MW, Hageleit M (2001) Validities of mRNA quantification using recombinant RNA and recombinant DNA external calibration curves in real-time RT-PCR. *Biotechnology Letters* **23**: 275–282.
8. Lockey C, Otto E, Long Z (1998) Real-time fluorescence detection of a single DNA molecule. *BioTechniques* **24**: 744–746.
9. Steuerwald N, Cohen J, Herrera RJ, Brenner CA (1999) Analysis of gene expression in single oocytes and embryos by real-time rapid cycle fluorescence monitored RT-PCR. *Molecular Human Reproduction* **5**: 1034–1039.
10. Pfaffl MW (2004) Quantification strategies in real-time RT-PCR. In: SA Bustin (ed), The Real-Time PCR Encyclopaedia A-Z of Quantitative PCR, pages 87–120. La Jolla, CA: International University Line.
11. Wittwer CT, Garling DJ (1991) Rapid cycle DNA amplification: time and temperature optimization. *BioTechniques* **10**: 76–83.
12. Fleige S, Walf V, Huch S, Prgomet C, Sehm J, Pfaffl MW (2006) Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. *Biotechnology Letters* **28**(19): 1601–1613.
13. Stahlberg A, Hakansson J, Xian X, Semb H, Kubista M (2004) Properties of the reverse transcription reaction in mRNA quantification. *Clinical Chemistry* **50**(3): 509–515.
14. Perez-Novo CA, Claeys C, Speleman F, Cauwenberge PV, Bachert C, Vandesompele J (2005) Impact of RNA quality on reference gene expression stability. *BioTechniques* **39**: 52–56.
15. Mannhalter C, Koizar D, Mitterbauer G (2000) Evaluation of RNA isolation methods and reference genes for RT-PCR analyses of rare target RNA. *Clinical Chemistry and Laboratory Medicine* **38**: 171–177.
16. Swift GH, Peyton MJ, MacDonald RJ (2000) Assessment of RNA quality by semi-quantitative RT-PCR of multiple regions of a long ubiquitous mRNA. *BioTechniques* **28**(3): 524–531.
17. Kruhoffer M, Dyrskjot L, Voss T, Lindberg RL, Wyrich R, Thykjaer T, et al. (2007) Isolation of microarray-grade total RNA, microRNA, and DNA from a single PAXgene blood RNA tube. *The Journal of Molecular Diagnostics* **9**(4): 452–458.
18. Fleige S, Pfaffl MW (2007) RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine* **27**(2–3): 126–139.
19. Raeymarkers L (1993) Quantitative PCR: theoretical consideration with practical implication. *Analytical Biochemistry* **214**: 582–585.
20. Bengtsson M, Ståhlberg A, Rorsman P, Kubista M (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Research* **15**(10): 1388–1392.
21. Freeman TC, Lee K, Richardson PJ (1999) Analysis of gene expression in single cells. *Current Opinions in Biotechnology* **10**(6): 579–582.
22. Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. *Nature Protocols* **1**(3): 1559–1582.
23. Freeman WM, Vrana SL, Vrana KE (1996) Use of elevated reverse transcription reaction temperatures in RT-PCR. *BioTechniques* **20**(5): 782–783.
24. Wong L, Pearson H, Fletcher A, Marquis CP, Mahler S (1998) Comparison of the efficiency of Moloney murine leukaemia virus (M-MuLV) reverse transcriptase, RNase H–M-MuLV reverse transcriptase and avian myeloblastoma leukaemia virus (AMV) reverse transcriptase for the amplification of human immunoglobulin genes. *Biotechnology Techniques* **12**(6): 485–489.
25. Stahlberg A, Kubista M, Pfaffl M (2004) Comparison of reverse transcriptases in gene expression analysis. *Clinical Chemistry* **50**(9): 1678–1680.
26. Stangegaard M, Dufva IH, Dufva M (2006) Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA. *BioTechniques* **40**(5): 649–657.

27. Zuker M (2003) *m*-Fold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**(13): 3406–3415.
28. Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, et al. (1999) Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology* **75**: 291–295.
29. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**(7): RESEARCH0034.
30. Pfaffl MW (2001) Development and validation of an externally standardised quantitative insulin like growth factor-1 (IGF-1) RT-PCR using LightCycler SYBR® Green I technology. In: S Meuer, C Wittwer, and K Nakagawara (eds), Rapid Cycle Real-Time PCR, Methods and Applications, pages 281–291. Heidelberg: Springer Press.
31. Reist M, Pfaffl MW, Morel C, Meylan M, Hirsbrunner G, Blum JW, et al. (2003) Quantitative mRNA analysis of bovine 5-HT receptor subtypes in brain, abomasum, and intestine by real-time PCR. *Journal of Receptors and Signal Transduction* **23**(4): 271–287.
32. Reynisson E, Josefsen MH, Krause M, Hoorfar J (2005) Evaluation of probe chemistries and platforms to improve the detection limit of real-time PCR. *Journal of Microbiological Methods* **66**(2): 206–216.
33. Schmittgen TD, Zakrajsek BA, Mills AG, Gorn V, Singer MJ, Reed MW (2000) Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Analytical Biochemistry* **285**(2): 194–204.
34. Blow N (2007) PCR's next frontier. *Nature Methods* **4**: 869–875.
35. Kellogg DE, Rybalkin I, Chen S, Mukhamedova N, Vlasik T, Siebert PD, et al. (1994) TaqStart antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *BioTechniques* **16**(6): 1134–1137.
36. Sharkey DJ, Scalice ER, Christy KG Jr, Atwood SM, Daiss JL (1994) Antibodies as thermolabile switches: high temperature triggering for the polymerase chain reaction. *Biotechnology* (*NY*) **12**(5): 506–509.
37. Reiter M, Pfaffl MW (2008) Effects of plate position, plate type and sealing systems on real-time PCR results. *Biotechnology & Biotechnological Equipment* **22**: 824–828.
38. Schams D, Kohlenberg S, Amselgruber W, Berisha B, Pfaffl MW, Sinowatz F (2003) Expression and localisation of oestrogen and progesterone receptors in the bovine mammary gland during development, function and involution. *Journal of Endocrinology* **177**: 305–317.
39. Berisha B, Pfaffl MW, Schams D (2002) Expression of steroid receptors in the bovine ovary during estrous cycle and pregnancy. *Endocrine* **17**(3): 207–214.
40. Inderwies T, Pfaffl MW, Meyer HHD, Blum JW, Bruckmaier RM (2003) Detection and quantification of mRNA expression of α- and β-adrenergic receptor subtypes in the bovine mammary gland of dairy cows. *Domestic Animal Endocrinology* **24**: 123–135.
41. Livak KJ (2001) ABI Prism 7700 Sequence detection System User Bulletin #2. Relative quantification of gene expression. Available at: http://docs.appliedbiosystems.com/pebiodocs/04303859.pdf.
42. Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* **25**: 169–193.
43. Wittwer CT, Herrmann MG, Gundry CN, Elenitoba-Johnson KS (2001) Real-time multiplex PCR assays. *Methods* **25**(4): 430–442.
44. Bustin SA (2002) Quantification of mRNA using real-time RT-PCR. Trends and problems. *Journal of Molecular Endocrinology* **29**(1): 23–39.
45. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2'$ [-delta delta C(T)] method. *Methods* **25**(4): 402–408.
46. Tichopad A, Dzidic A, Pfaffl MW (2002) Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency. *Biotechnology Letters* **24**: 2053–2056.

47. Tichopad A, Dilger M, Schwarz G, Pfaffl MW (2003) Standardized determination of real-time PCR efficiency from a single reaction setup. *Nucleic Acids Research* **31**(20): e122.
48. LightCycler Relative Quantification Software. (2001) Version 1.0, Roche Applied Science, Roche Molecular Biochemicals.
49. Muller PY, Janovjak H, Miserez AR, Dobbie Z (2002) Processing of gene expression data generated by quantitative real-time RT-PCR. *BioTechniques* **32**(6): 1372–1378.
50. Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research* **30**(9): e36.
51. Wilhelm J, Pingoud A, Hahn M (2003) SoFAR – Validation of an algorithm for automatic quantification of nucleic acid copy numbers by real-time polymerase chain reaction. *Analytical Biochemistry* **317**(2): 218–225.
52. Peirson SN, Butler JN, Foster RG (2003) Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Research* **31**(14): e73.
53. Jin N, He K, Liu L (2006) qPCR-DAMS: a database tool to analyze, manage, and store both relative and absolute quantitative real-time PCR data. *Physiological Genomics* **25**(3): 525–527.
54. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**(2): R19.
55. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**: e35.
56. Rasmussen R (2001) Quantification on the LightCycler. In: S Meuer, C Wittwer, and K Nakagawara (eds), Rapid Cycle Real-Time PCR, Methods and Applications, pages 21–34. Heidelberg: Springer Press.
57. Morrison TB, Weis JJ, Wittwer CT (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *BioTechniques* **24**(6): 954–962.
58. Huggett J, Dheda K, Bustin S, Zumla A (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity* **6**: 279–284.
59. Bengtsson M, Hemberg M, Rorsman P, Stahlberg A (2008) Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC Molecular Biology* **9**: 63.
60. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnology Letters* 26: 509–515.
61. Akilesh S, Shaffer DJ, Roopenian D (2003) Customized molecular phenotyping by quantitative gene expression and pattern recognition analysis. *Genome Research* **13**(7): 1719–1727.
62. Yuan JS, Reed A, Chen F, Stewart CN Jr (2006) Statistical analysis of real-time PCR data. *BMC Bioinformatics* **7**: 85.
63. Bar T, Stahlberg A, Muszta A, Kubista M (2003) Kinetic Outlier Detection (KOD) in real-time PCR. *Nucleic Acids Research* **31**(17): e105.
64. Burns MJ, Nixon GJ, Foy CA, Harris N (2005) Standardisation of data from real-time quantitative PCR methods – evaluation of outliers and comparison of calibration curves. *BMC Biotechnology* **7**: 31–37.
65. Sheskin D (2000) Handbook of Parametric & Nonparametric Statistical Procedures. Boca Raton, FL: CRC Press.
66. Manly B (1997) Randomization, Bootstrap and Monte Carlo Methods in Biology London: Chapman & Hall.
67. Morrison T, Hurley J, Garcia J, Yoder K, Katz A, Roberts D, et al. (2006) Nanoliter high throughput quantitative PCR. *Nucleic Acids Research* **34**(18): e123.

68. Pohl G, Shih IeM. (2004) Principle and applications of digital PCR. *Expert Review of Molecular Diagnostics* **4**(1): 41–47.
69. Vogelstein B, Kinzler KW (1999) Digital PCR. *Proceedings of the National Academy of Sciences of the United States of America* **96**(16): 9236–9241.
70. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonak J, Lind K, et al. (2006) The real-time polymerase chain reaction. *Molecular Aspects of Medicine* **2–3**: 95–125.
71. Smilde A, Bro R, Geladi P (2004) MultiWay Analysis. Hoboken, NJ: John Wiley & Sons Ltd.
72. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. *Molecular Systems Biology* **3**: 78.
73. D'Haeseleer P (2005) How does gene expression clustering work? *Nature Biotechnology* **23**: 1499–1501.
74. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. *Nature Genetics* **22**: 281–285.
75. Ward JH (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**(301): 236–244.
76. Tichopad A, Pecen L, Pfaffl MW (2006) Distribution-insensitive cluster analysis in SAS on real-time PCR gene expression data of steadily expressed genes. *Computer Methods and Programs in Biomedicine* **82**: 44–50.
77. Ma P, Castillo-Davis CI, Zhong W, Liu JS (2006) A data-driven clustering method for time course gene expression data. *Nucleic Acids Research* **34**: 1261–1269.
78. Schedle K, Pfaffl MW, Plitzner C, Meyer HHD, Windisch W (2008) Effect of insoluble fibre on intestinal morphology and mRNA expression pattern of inflammatory, cell cycle and growth marker genes in a piglet model using cluster analysis. *Archives of Animal Nutrition* (in press October 2008).
79. GenEx software, MultiD, Göteburg, Sweden.
80. Kohonen T (1995) Self-Organizing Maps, Springer Series in Information Sciences. Vol. 30, 3rd ed. Berlin, Heidelberg, New York: Springer.