Clinical Chemistry 55:10
1816–1823 (2009)

Design and Optimization of Reverse-Transcription Quantitative PCR Experiments

Ales Tichopad,[1,2*] Rob Kitchen,[3] Irmgard Riedmaier,[1] Christiane Becker,[1] Anders Ståhlberg,[2,4] and Mikael Kubista[2,5]

BACKGROUND: Quantitative PCR (qPCR) is a valuable technique for accurately and reliably profiling and quantifying gene expression. Typically, samples obtained from the organism of study have to be processed via several preparative steps before qPCR.

METHOD: We estimated the errors of sample withdrawal and extraction, reverse transcription (RT), and qPCR that are introduced into measurements of mRNA concentrations. We performed hierarchically arranged experiments with 3 animals, 3 samples, 3 RT reactions, and 3 qPCRs and quantified the expression of several genes in solid tissue, blood, cell culture, and single cells.

RESULTS: A nested ANOVA design was used to model the experiments, and relative and absolute errors were calculated with this model for each processing level in the hierarchical design. We found that intersubject differences became easily confounded by sample heterogeneity for single cells and solid tissue. In cell cultures and blood, the noise from the RT and qPCR steps contributed substantially to the overall error because the sampling noise was less pronounced.

CONCLUSIONS: We recommend the use of sample replicates preferentially to any other replicates when working with solid tissue, cell cultures, and single cells, and we recommend the use of RT replicates when working with blood. We show how an optimal sampling plan can be calculated for a limited budget.

Experimental design, first developed by Sir Ronald A. Fisher *(1)*, is a structured, organized method for determining the relationship between the different factors affecting an experimental process and the output of that process. The use of quantitative PCR (qPCR)[6] to study gene expression *(2–6)* requires statistical considerations of all invoked factors: the treatment effect, the intersubject biological variance, and the noise due to sample processing. In addition, the gene-specific effect on the error structure must be considered. A typical qPCR experiment designed to test a hypothesis that a difference in gene expression exists between groups of biological subjects exposed to different treatments involves sampling the biological material, extracting the RNA, reverse transcription (RT) of the RNA into cDNA, and amplification of the cDNA by the qPCR. Too often, experiments are designed and conducted with excessive emphasis on the amplification step while ignoring the preceding steps and their contribution to the measurement error.

The measured difference between any 2 groups has 3 contributions: the treatment effect, intersubject variation, and processing noise (Table 1). Exact definitions for these variance contributions, as well as the models for their calculation, are given in Supplemental Text 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol55/issue10. The relationship between the effect, variation, and the number of subjects studied can be calculated with the power test, which is exemplified in Fig. 2 in the online Data Supplement.

In general, independent errors are additive. Fluctuations in yield due to pipetting errors, uncertainties in instrument readings, and chemical noise in the different processing steps are expected to be independent. Interference due to inhibitors is not independent, however, because any inhibiting substance present in a sample will propagate through the subsequent processing steps of that sample, although the inhibitor will gradually be removed and diluted. The noise observed for a given processing step is only partially attributable

[1] Physiology Weihenstephan, Technical University Munich, Freising, Germany; [2] TATAA Biocenter, Göteborg, Sweden; [3] National e-Science Centre, School of Physics, University of Edinburgh, Edinburgh, UK; [4] Lundberg Laboratory for Cancer, Department of Pathology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden; [5] Institute of Biotechnology AS CR, Prague, Czech Republic.
* Address correspondence to this author at: Technical University Munich, Physiology Weihenstephan, Weihenstephaner Berg 3, 85354 Freising, Germany. Fax +49-8161714204; e-mail ales@tichopad.de.
Received February 26, 2009; accepted June 29, 2009.
Previously published online at DOI: 10.1373/clinchem.2009.126201
[6] Nonstandard abbreviations: qPCR, quantitative PCR; RT, reverse transcription; Cq, cycle of quantification; GOI, gene of interest.

Table 1. Variance in biological experiments.[a]

	Confounding variance		Studied variance
	Intersubject variance	Processing noise	Treatment effect
Source	Different baseline expression	Sampling	The difference between groups induced by treatment
	Different responses to treatment	RT	
		qPCR	
Intervention	Randomize	Replicates	Maximize effect (e.g., dose selection)
	Large N	Normalization to internal standard or spike	
	Paired measures		

[a] The confounding variance consists of the intersubject variance and the processing variance. To maximize the resolution of the effect, the confounding variance must be substantially lower than the effect.

to that step, because it also accommodates the noise from any previous step(s) performed after the compared samples were separated. For example, the qPCR noise as measured by the SD (σ) reflects not only that of the qPCR but also the RT noise if the replicate samples were separated before the RT step. To objectively assess the contribution of individual processing steps to the overall noise requires a model that reflects the additive noise structure.

An experiment that studies gene expression with real-time PCR has a nested (i.e., hierarchical or clustered) structure *(7, 8)*, meaning that the qPCR is run on individual RT products, mRNAs are obtained from samples, samples are collected from subjects (e.g., experimental animals), and subjects are selected from experimental groups (Fig. 1). The nested design is justified because the same qPCR replicate cannot be dispensed from 2 different RT tubes, just as a single sample cannot be obtained from 2 different subjects. The nested design is a refinement of the factorial design that exploits the special character of the variance structure. In a factorial design, each factor is crossed with all

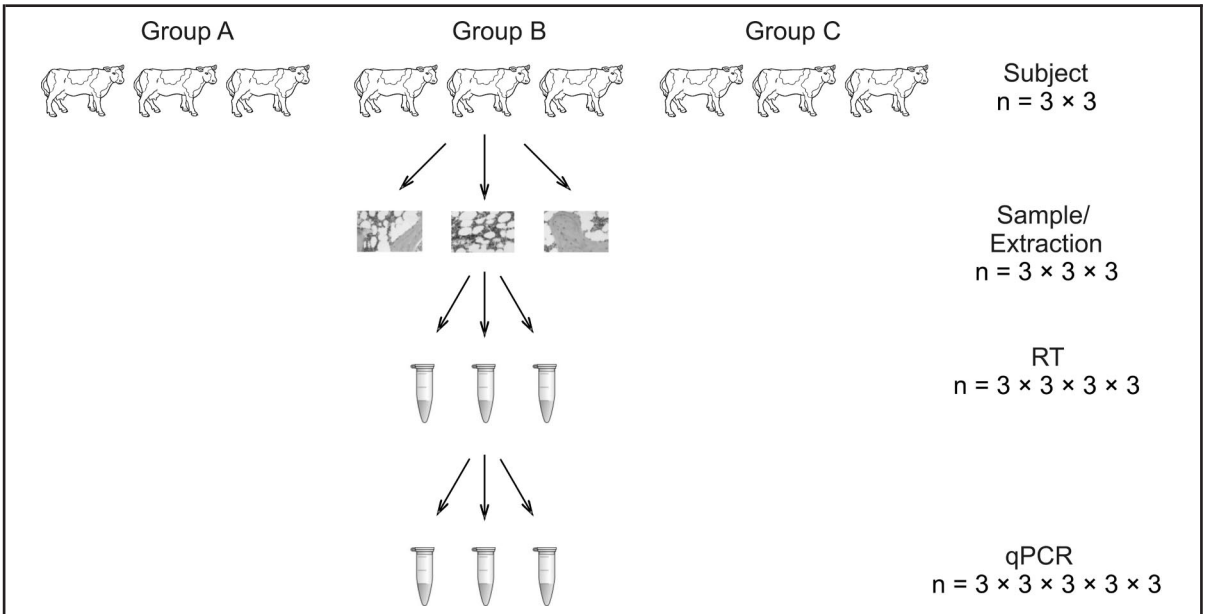

Fig. 1. Comparison of 3 groups with a nested experimental design.

Each group consists of 3 subjects, from which 3 samples are collected and extracted. The extracted and then split into 3 RT reactions, which are then finally into 3 qPCRs. The nested design is $3 \times 3 \times 3 \times 3 \times 3$ and produces a total of 81 Cq values.

others, whereas in the nested design, factors meet in unique combinations; there are no homogeneous replicates in the whole experiment, but rather clusters of replicates representing the integrated effects of the upstream processes.

The linear model of all hierarchical sample-processing effects within a single treatment group can be written as follows:

$$Cq_{ijkl} = \mu + a_i + b_{j(i)} + c_{k(ij)} + d_{l(ijk)}, \quad (1)$$

where Cq_{ijkl} is the individual qPCR response, typically expressed as the number of amplification cycles required to reach a threshold fluorescence level; it incorporates the random effects of the *i*th subject, the *j*th sample, the *k*th RT, and the *l*th qPCR. μ is the total mean of the group expressed in the same unit of measure as Cq_{ijkl}; a_i is the random effect of the *i*th subject; $b_{j(i)}$ is the random effect of the *j*th sample taken from subject *i*; $c_{k(ij)}$ is the random effect of the *k*th RT reaction from sample *j* of subject *i*; and $d_{l(ijk)}$ is the random effect of the *l*th qPCR from the *k*th cDNA of sample *j* from subject *i*.

In this model, all factors are randomly invoked. This feature requires the random selection of subjects, the collection of random samples and extraction of the RNA from each sample, the random dispensing of the RNA into the RT reactions, and the random dispensing of the cDNAs into the qPCR. In most experiments, the noise associated with the collection of a sample also encompasses the extraction noise, because 1 sample is typically acquired for 1 extraction.

The linear equation describing multiple groups with hierarchical replicates follows from Eq. 1:

$$Cq_{gijkl} = \mu_g + a_{i(g)} + b_{j(gi)} + c_{k(gij)} + d_{l(gijk)}, \quad (2)$$

where *g* indexes the experimental groups.

The model defined in Eq. 2 facilitates the comparison of various levels of factor *g* (the treatment effect).

The optimal experimental design requires knowing the sources of error throughout sample processing. To address this question, we performed several RT-qPCR assays arranged in a nested hierarchy for different biological materials.

Materials and Methods

We studied 4 different kinds of samples: liver tissue from slaughtered heifers, blood from heifers, cultures of adherent, growing IPI-2I cells from porcine ileum, and individual astrocytes (single cells) from mouse brain *(9)*. For tissue, blood, and cell culture samples, we reverse-transcribed 500 ng of total RNA and amplified the cDNA by the qPCR, whereas for single cells we reverse-transcribed all of the RNA. Protocols and technical details are provided in Supplemental Texts 2 and 3 in the online Data Supplement.

STATISTICAL ANALYSIS

Variance analysis was performed with the PROC NESTED program in SAS software (version 9.1 for Microsoft Windows) with the model defined in Eq. 1 and according to the procedure outlined in Supplemental Text 1. We estimated variance contributions of the processing steps and the total variance as follows:

$$\hat{\sigma}^2_{Cq} = \hat{\sigma}^2_i + \hat{\sigma}^2_j + \hat{\sigma}^2_k + \hat{\sigma}^2_l. \quad (3)$$

Variance contributions are expressed as:

$$\text{Variance contribution} = 100 \times \hat{\sigma}^2_x / \hat{\sigma}^2_{Cq},$$
$$\text{where } x = i, j, k, \text{ or } l. \quad (4)$$

The SD (σ) is defined in this report with a superscript indicating the sample type and a subscript indicating the processing step. For example, $\hat{\sigma}^{\text{Solid tissue}}_{\text{Sampling}}$ is the estimated SD of the sampling step for solid tissue.

Results

All assays showed linearity over the entire range of Cq (cycle of quantification) values with efficiencies of 85%–100% (see Fig. 2 in the online Data Supplement).

SOLID TISSUE

We collected 3 liver samples from each of 3 slaughtered heifers and extracted the RNA. Each extract was split into 3 aliquots for RT, each of which was split again into 3 qPCR aliquots. These aliquots were analyzed for 4 genes: *ACTB*[7] (actin, beta), *IL1B* (interleukin 1, beta), *CASP3* (caspase 3, apoptosis-related cysteine peptidase), and *FGF7* [fibroblast growth factor 7 (keratinocyte growth factor)] in singleplex format. Hence, we used a nested design (3 subjects × 3 samples × 3 RTs × 3 qPCRs) that yielded 81 Cq values for each of the 4 genes. Estimated SDs ($\hat{\sigma}$'s) for the various processing levels are shown in Table 2. Also shown is the cumulative variation, which is expressed as the SD of measured Cq values ($\hat{\sigma}^{\text{Solid tissue}}_{Cq}$) obtained from different animals (as if they were processed independently without technical replicates). *ACTB*, *IL1B*, and *CASP3* had Cqs of <28 cycles, whereas *FGF7* had low expression (mean Cq, 31.5 cycles). The largest SD was estimated for the

[7] Genes: *ACTB*, actin, beta (*Bos taurus*); *IL1B*, interleukin 1, beta (*B. taurus*); *CASP3*, caspase 3 (*B. taurus*); *FGF7*, fibroblast growth factor 7 (*B. taurus*); *IFNG*, interferon, gamma (*B. taurus*); *BCL2*, B-cell CLL/lymphoma 2 (*Sus scrofa*); *ACTB*, actin, beta (*S. scrofa*); *H3F3A*, histone H3.3A (*S. scrofa*); *IL8*, interleukin 8 (*S. scrofa*); *Rn18s*, 18S RNA (*Mus musculus*).

Table 2. SD estimates for intersubject variation and sample-processing steps. Also shown are values for mean Cq and total noise ($\hat{\sigma}_{Cq}$).

	Sample types and genes												
	Liver				Blood				Cell culture				Single cell
	ACTB	*IL1B*	*CASP3*	*FGF7*	*ACTB*	*IL1B*	*CASP3*	*IFNG*	*ACTB*	*H3F3A*	*IL8*	*BCL2*	18S rRNA
Mean Cq	20.41	26.76	27.25	31.52	16.05	17.6	24.71	32.2	15.87	20.1	23.4	28.5	29.95
SDs													
Intersubject variation	0.00	0.00	0.00	0.00	0.07	0.94	0.00	0.95	0.00	0.00	0.00	0.00	0.00
Processing noise													
Sampling	1.56	1.64	1.20	0.40	0.10	0.00	0.11	0.00	0.37	0.20	0.29	0.20	1.90
RT	0.46	0.39	0.27	0.90	0.21	0.32	0.18	0.24	0.35	0.35	0.31	0.21	0.30
qPCR	0.07	0.12	0.08	0.39	0.18	0.20	0.13	0.40	0.21	0.10	0.09	0.16	0.51
Total noise	1.63	1.70	1.23	1.06	0.31	1.01	0.25	1.06	0.55	0.42	0.44	0.33	1.99

sampling step, with $\hat{\sigma}^{\text{Solid tissue}}_{\text{Sampling}}$ values of 1.20–1.64 cycles for the first 3 genes and a $\hat{\sigma}^{\text{Solid tissue}}_{\text{Sampling}}$ value of 0.40 cycles for *FGF7*. The mean was 1.20 cycles. This value corresponds to a >2-fold variation in sampling yield, which most likely reflects sample heterogeneity. $\hat{\sigma}^{\text{Solid tissue}}_{\text{RT}}$ for the first 3 genes was 0.27–0.46 cycles (mean, 0.39 cycles), and the *FGF7* $\hat{\sigma}^{\text{Solid tissue}}_{\text{RT}}$ was 0.90 cycles. The qPCR step showed the highest reproducibility, with $\hat{\sigma}^{\text{Solid tissue}}_{\text{qPCR}}$ values of 0.07–0.12 cycles (mean, 0.09 cycles) for the first 3 genes and 0.39 cycles for *FGF7*. It was not possible to estimate $\hat{\sigma}^{\text{Solid tissue}}_{\text{Subjects}}$ because its contribution to the overall noise was too small compared with the contributions of the subsequent steps. $\hat{\sigma}^{\text{Solid tissue}}_{\text{Cq}}$ values were 1.06–1.70 cycles. Expressed as variance contributions (Eq. 4), sampling accounted for >90% of the total variance for the first 3 genes (Fig. 2), whereas the contribution of sampling to total variance for the low-expressed *FGF7* gene was <20%.

BLOOD

One blood sample was collected from each of 3 heifers. Lysed cells from the blood samples were split into 3 vials for RNA extractions. Each extract was then split into 3 RT reactions, and each cDNA product was analyzed in 3 qPCR replicates for each of 4 genes [*ACTB*, *CASP3*, *IL1B*, and *IFNG* (interferon, gamma)]. In total, 81 Cq values were measured with the nested design for each of the 4 genes (3 subjects × 1 sample × 3 extractions × 3 RTs × 3 qPCRs). *ACTB*, *CASP3*, and *IL1B* had Cq values in the range of 16–25 cycles, whereas *IFNG* had low expression, with a mean Cq of 32.2 cycles. For the first 3 genes, $\hat{\sigma}^{\text{Blood}}_{\text{qPCR}}$ values were 0.13–0.20 cycles (mean, 0.17 cycles). For *IL1B*, $\hat{\sigma}^{\text{Solid tissue}}_{\text{Sampling}}$ was somewhat higher (0.40 cycles). $\hat{\sigma}^{\text{Blood}}_{\text{RT}}$ values for the 4 genes were similar (0.18–0.32 cycles; mean, 0.24 cycles). The extraction step showed highest reproducibility with $\hat{\sigma}^{\text{Blood}}_{\text{Extraction}}$ values <0.12 cycles. $\hat{\sigma}^{\text{Blood}}_{\text{Subjects}}$ was negligible for *ACTB* and *CASP3* and was approximately 1 cycle for *IL1B* and *IFNG*. For all genes, the total noise was in the range of 0.25–1.06 cycles. Our inspection of the variance contributions revealed $\hat{\sigma}^{\text{Blood}}_{\text{Cq}}$ that no step dominated (Table 2).

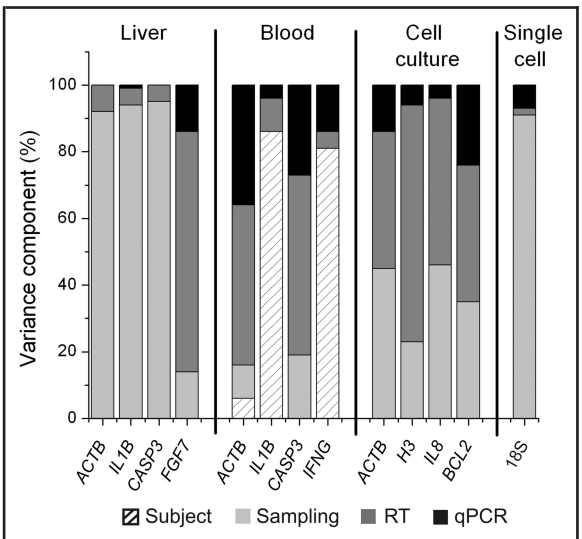

Fig. 2. Estimated confounding variation contributed by the sampling-processing steps.

The contributions to the overall noise are expressed as percentages. A variance contribution of 0% is obtained when subsequent downstream processing steps confound the calculation of variance by their own high variances (i.e., signal-to-noise ratio is too low) or if the real processing noise is close to 0%.

CELL CULTURE

We disseminated 6×10^4 adherent, growing IPI-2I cells from porcine ileum in 10 wells, from which we extracted total RNA. Each RNA sample was reverse-transcribed in 3 reactions, each of which was divided into 3 qPCR replicates for each of 4 genes [*BCL2* (B-cell CLL/lymphoma 2), *ACTB*, *H3F3A* (histone H3.3A), and *IL8* (interleukin 8)]. We obtained a total of 90 readings with the nested design for each of the 4 genes analyzed (1 subject × 10 samples × 1 extraction × 3 RTs × 3 qPCRs). Mean Cq values for the 4 genes were in the range 15 cycles < Cq < 29 cycles, and SD values were 0.090–0.21 cycles for $\hat{\sigma}_{\text{qPCR}}^{\text{Cell culture}}$ (mean, 0.14 cycles), 0.21–0.35 cycles for $\hat{\sigma}_{\text{RT}}^{\text{Cell culture}}$ (mean, 0.31 cycles), and 0.20–0.37 cycles for $\hat{\sigma}_{\text{Sampling}}^{\text{Cell culture}}$ (mean, 0.27 cycles) (Fig. 2). Because each sample was extracted without replicates, $\hat{\sigma}_{\text{Sampling}}^{\text{Cell culture}}$ reflects the combined noise introduced by the growing of the samples and their subsequent extraction. Expressed as noise contributions, the sampling and RT steps each accounted for at least 30% of the variation, and the qPCR step accounted for >25% (Table 2).

SINGLE CELLS

Three astrocytes were collected from each of 3 mice, lysed, and split into 3 replicates for RT. These replicates were then split into another 3 replicates for qPCR assays for 18S rRNA (*Rn18s*, 18S RNA). The experimental design was as follows: 3 subjects × 3 samples × 1 extraction × 3 RTs × 3 qPCRs. The estimated SDs were 0.51 cycles for $\hat{\sigma}_{\text{qPCR}}^{\text{Single cell}}$, 0.30 cycles for $\hat{\sigma}_{\text{RT}}^{\text{Single cell}}$, and 1.90 cycles for $\hat{\sigma}_{\text{Samples}}^{\text{Single cell}}$. The last SD represents the combined SDs of the samples (which are the cells) and the cell lysis/extraction step. The very high $\hat{\sigma}_{\text{Samples}}^{\text{Single cell}}$ value is consistent with previous findings that mRNA concentrations in cells show lognormal variation *(10)*. $\hat{\sigma}_{\text{Subjects}}^{\text{Single cell}}$ was negligible, and $\hat{\sigma}_{\text{Cq}}^{\text{Single cell}}$ was 1.99 cycles. Our inspection of the variance contributions revealed that heterogeneity among the cells totally dominated the overall error, with a contribution of >90%.

Discussion

INDIVIDUAL CELLS AND SOLID TISSUE SHOW THE HIGHEST VARIATION

Confounding biological variation and analytical noise in a qPCR assay limits the ability to observe differential expression of treatment groups. The total variance, including biological variation and experimental noise, is given by Eq. 3 and is denoted as σ_{Cq}^2. The corresponding SD is $\hat{\sigma}_{\text{Cq}}$. For the samples studied, our estimates are 1.99 cycles for $\hat{\sigma}_{\text{Cq}}^{\text{Single cell}}$, 1.40 cycles for $\hat{\sigma}_{\text{Cq}}^{\text{Solid tissue}}$, 0.66 cycles for $\hat{\sigma}_{\text{Cq}}^{\text{Blood}}$, and 0.43 cycles for $\hat{\sigma}_{\text{Cq}}^{\text{Cell culture}}$. The greatest variation is between individual cells and is equivalent to the normal distribution of Cq values. Thus, 68% of the cells in the population are expected to have an expression within the mean ± 1 SD. Given the mean Cq of 18S mRNA in single astrocytes (Table 2), taking into account the dilutions that were performed, and assuming extraction and RT yields of 100% and a PCR assay efficiency of 90% *(11–12)*, we estimate the mean number of copies per cell to be on the order of 10 000. Hence, 16% [(100 − 68)/2] of the cells should have <2500 copies/cell (10 000/8), and 16% should have >40 000 copies/cell. This large observed variation is consistent with previously published findings *(13–16)*.

Our analysis also revealed substantial variation for solid tissue, with a mean $\hat{\sigma}_{\text{Cq}}^{\text{Solid tissue}}$ of 1.40 cycles, which corresponds to variation of 2.6-fold (i.e., $2^{1.4}$). Cq values obtained with samples of solid tissues are commonly assumed to be normally distributed, but to our knowledge, the validity of this assumption has yet to be demonstrated. Because the variance is dominated by the sampling step, whether Cq values are normally distributed may depend on the complexity of the tissue, as well as on the size of the excised sample. It may also be gene dependent. For the low-expressed *FGF7* gene, the majority of the overall noise was introduced at the RT step. The reason for this result is unknown, and with only a single case, we cannot be sure about the significance of the observation. One possibility is that it may be due to the lower abundance of *FGF7* mRNA in the liver tissue, perhaps combined with a more homogeneous distribution of *FGF7*-expressing cells. Because the same amount of total RNA was processed in all samples, the data are inherently normalized to total RNA. Hence, sampling variation reflects heterogeneity in the distribution of the targeted mRNA relative to that of total RNA in the tissue and is not due, for example, to different quantities of material processed.

In the blood samples, 2 genes had $\hat{\sigma}_{\text{Cq}}^{\text{Blood}}$ values of 0.25 and 0.31 cycles, and the other 2 genes had $\hat{\sigma}_{\text{Cq}}^{\text{Blood}}$ values of approximately 1 cycle. This result is because the first 2 genes had only the variance contributed from the RT and qPCR steps, whereas the intersubject variance was substantial for the other 2 genes. This finding suggests that the spread in measured Cq values can have different origins, depending on the gene, even when the same subjects and tissue are analyzed *(17)*. The noise arising from sampling and extraction was consistently small across all of the studied genes, indicating that this step is very reproducible for blood samples.

The lowest overall confounding variation was found for the cell culture samples, which had a mean $\hat{\sigma}_{\text{Cq}}^{\text{Cell culture}}$ value of 0.43 cycles.

USE OF qPCR REPLICATES HAS LITTLE JUSTIFICATION

The total confounding variance, $\hat{\sigma}_{\text{Cq}}$, can be reduced not only by increasing the number of subjects but also

by performing technical replicates to "average out" processing noise. The effect of technical replicates depends on both the noise contributed by that particular step and the noise contributed by subsequent steps. The final step in RT-qPCR process is the qPCR. Many published reports have described the use of experimental protocols that perform only qPCR replicates. On the basis of the variance contributions we have estimated for the 4 studied sample types, we are able to evaluate the importance of qPCR replicates. For 10 of 13 measurements, we find $\hat{\sigma}_{qPCR}$ values of 0.07–0.21 cycles, with a mean of 0.13 cycles. Clearly, $\hat{\sigma}_{qPCR}$ does not depend on sample type, a reasonable finding given that the preceding steps should have removed substances that interfere with the PCR. The mean $\hat{\sigma}_{qPCR}$ of 0.13 cycles is similar to previous findings *(11)*. The 3 higher $\hat{\sigma}_{qPCR}$ values are for the low-expressed genes *FGF7* in liver and *IFNG* in blood, and for 18S rRNA in the individual astrocytes. Low amounts of cDNA template are expected to give data with greater noise because of a Poisson-distributed sampling error and the replication noise introduced during early PCR cycles. A practical threshold of approximately 25 copies per reaction has been suggested. We conclude that a $\hat{\sigma}_{qPCR}$ value of 0.13 cycles is a good estimate for genes that are expressed at reasonable levels and are assayed with a protocol that yields at least some 25 template copies per qPCR.

The noise introduced by the RT appears to be consistent and small for samples of all types ($\hat{\sigma}_{RT}$, 0.18–0.90 cycles; mean, 0.35 cycles). This estimate of $\hat{\sigma}_{RT}$ should be reasonable and suggests that the reproducibility of the RT step is approximately 2 times less than that of the qPCR, as has also previously been reported *(11–12)*.

The noise introduced by sampling depends on the specimen. The mean $\hat{\sigma}_{Sampling}$ was 1.9 cycles for single cells, 1.20 cycles for liver tissue, 0.27 cycles for cell culture, and 0.05 cycles for blood. Compared with the imprecision of the RT and qPCR steps, sampling of single cells and liver tissue shows far greater variation than the RT-qPCR, whereas for cell culture, sampling contributes noise that is comparable to that of the RT. Sampling blood shows negligible variation.

Intersubject variation, which represents the biological variation among inbred heifers and mice in this study, has been difficult to estimate. It is the top level of the nested design, and its contribution is therefore confounded by the contributions from the processing steps that follow. Therefore, this source of the variance cannot be estimated with reasonable precision unless it dominates. For the liver samples and single cells, estimation of intersubject variation failed. Very little intersubject biological variation occurred in the cell culture samples because only a single clone was studied. We were able to estimate $\hat{\sigma}_{Subject}$ only for the blood samples. For *ACTB* and *CASP3*, $\hat{\sigma}^{Blood}_{Subjects}$ was negligible, whereas it was approximately 1 cycle for *IL1B* and *IFNG*. For the latter 2 genes, $\hat{\sigma}^{Blood}_{Subjects}$ made the dominant contribution, accounting for >80% of the total variance. This result indicates that intersubject noise may be gene dependent, although more data are needed to substantiate this conclusion.

qPCR was performed in duplicate in most published studies, although performing such replicates has usually not been justified. According to our estimates of the noise contributions of the sample-processing steps, $\hat{\sigma}_{qPCR}$ is consistently lower than the noise contributions of most of the other steps. The highest estimated contribution of qPCR to the total variance was 36% for *ACTB* in blood; however, this value is abnormal and may be accidental owing to low contributions from the other steps in this particular experiment. The mean variance contribution from the qPCR step was only 11%. Because neither qPCR nor RT shows any dependence on the gene and sample studied and because $\hat{\sigma}_{RT}$ is 2 times higher than $\hat{\sigma}_{qPCR}$, the noise contributed by the qPCR should always be smaller than that contributed by the RT and other steps.

The noise contributed by any step in the processing of a sample can be reduced by performing replicates and by use of mean values in subsequent analyses. The SD of a mean is the SE, which for an isolated processing step is:

$$SE = \hat{\sigma}/\sqrt{N}. \quad (5)$$

Performing a qPCR in duplicate reduces its noise contribution from 0.13 cycles to:

$$SE_{qPCR} = \hat{\sigma}_{qPCR}/\sqrt{2} = \hat{\sigma}_{qPCR} \times 0.7 \cong 0.09 \text{ cycles}. \quad (6)$$

This reduction is not a substantial improvement, compared with the total processing noise.

One can imagine that qPCR replicates might substantially reduce the total variance when the amount of cDNA template is very low *(18)*. This strategy is not valid in practice, however. If the amount of cDNA is indeed limiting, it is preferable to keep all of the material in a single sample and analyze it in a single run than to split it into duplicates that are analyzed separately and averaged. This is because the splitting process further reduces the already low number of cDNA copies per sample and thereby increases the noise. When many genes are assayed, splitting samples and the subsequent reduction in the number of cDNA copies may be unavoidable. The preferred strategy in this situation is first to preamplify the cDNA in a multiplex reaction to increase the number of all cDNAs and then to split the reaction into aliquots for singleplex analyses of the individual genes. This strategy avoids the introduction of severe noise at low template copy numbers. Thus, there is really no situation in which splitting qPCRs

Table 3. Optimization of cost effectiveness.[a]

No. of replicates					
Subject	Sampling	RT	qPCR	$\hat{\sigma}^2_{\overline{Cq}}$	Cost
3	3	1	1	0.08	$401.40
3	1	3	1	0.11	$371.40
3	1	1	3	0.29	$317.40
4	1	1	1	0.24	$418.40
3	1	2	1	0.17	$342.60

[a] Different sampling plans for quantifying *FGF7* expression in liver tissue by qPCR were evaluated with PowerNest software (http://www.powernest.net). For example, the $3 \times 3 \times 1 \times 1$ plan produces a lower variance of the mean Cq ($\hat{\sigma}^2_{\overline{Cq}}$) than the $3 \times 1 \times 3 \times 1$ design, despite the major source of noise being the RT step. This analysis shows that a design that incorporates upstream replicates is superior to one that incorporates replicates at subsequent processing steps.

into technical replicates is justifiable from a data-quality perspective. The only reason to run replicate qPCRs is to insure against a failed reaction so that the data point is not missed. But this consideration is relevant only if no upstream technical replicates are available, because they too provide this redundancy.

The decision of where to produce replicates and how many to use is often a compromise between attaining the maximum accuracy and performing the experiments within a specified budget. If costs are known for the subject, sample, RT, and qPCR steps, the optimal sampling plan can be calculated as follows:

$$\hat{\sigma}^2_{\overline{Cq}} = \hat{\sigma}^2_i/n_i + \hat{\sigma}^2_j/n_i n_j + \hat{\sigma}^2_k/n_i n_j n_k + \hat{\sigma}^2_l/n_i n_j n_k n_l, \quad (7)$$

where $\hat{\sigma}^2_{\overline{Cq}}$ is the variance of the mean Cq and n_x is the number of replicates at step x. Various combinations of n_x can be created, and the one that minimizes $\hat{\sigma}^2_{\overline{Cq}}$ in Eq. 7 and does not exceed the specified budget is chosen as the "cost-optimal" plan (Table 3). General recommendations for experimental setup are given in Table 1 in the online Data Supplement. A software tool can be requested from the corresponding author or downloaded (http://www.powernest.net).

NORMALIZATION WITH CONTROLS DOES NOT NECESSARILY IMPROVE DATA QUALITY

A common approach to reduce confounding variation is to use controls, most commonly endogenous reference genes. Normalizing with a control gene corresponds to calculating the difference between the Cqs of the gene of interest (GOI) and the reference gene (Ref):

$$\Delta Cq = Cq_{Ref} - Cq_{GOI}. \quad (8)$$

Table 4. Effect of normalization.[a]

	Tissues and genes					
	Liver			Blood		
	CASP3	*IL1B*	*IFNG*	*CASP3*	*IL1B*	*FGF7*
Cq	0.47	0.63	0.08	0.09	0.93	1.00
ΔCq	0.15	0.49	0.55	0.21	1.04	0.99

[a] Noise in the data normalized with total RNA and with *ACTB*, as reflected in SDs for Cq and ΔCq values, respectively, for the genes. Averaging technical replicates, we obtained Cq and ΔCq values for each of the 3 subjects and calculated means and associated SDs. There is no systematic improvement in the quality of the data.

The purpose of normalization is to reduce the SD of ΔCq relative to that of Cq_{GOI} (i.e., for successful normalization, $\hat{\sigma}_{\Delta Cq} < \hat{\sigma}_{Cq}$). This is by no means an expected result. The SD of ΔCq is:

$$\hat{\sigma}_{\Delta Cq} = \sqrt{\hat{\sigma}^2_{Ref} + \hat{\sigma}^2_{GOI} - 2\hat{\sigma}^2_{Ref/GOI}}, \quad (9)$$

where $\hat{\sigma}^2_{Ref/GOI}$ is the covariance between the Cqs of the GOI and the reference gene. If the covariance is smaller than the variance of the GOI and the reference gene each, normalization will actually increase the confounding noise. The covariance is a measure of how the preprocessing steps are affecting the GOI and the reference gene relative to each other. If the noise introduced in the preprocessing steps always affects the GOI and the reference gene in the same way (i.e., both are always processed in either high or low yield), then the covariance is high. If the noise affects them independently, however, normalization is meaningless; in fact, it even worsens the quality of the data (Table 4). With respect to the genes we have analyzed, the SD increases with normalization for *IFNG* in liver and for *CASP3* in blood.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures of Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: A. Tichopad, TATAA Biocenter.
Stock Ownership: A. Tichopad, TATAA Biocenter; M. Kubista, TATAA Biocenter.
Honoraria: A. Tichopad, TATAA Biocenter.

Research Funding: A. Tichopad, Technical University Munich; M. Kubista, European Framework projects SmartHEALTH and SPIDIA. This work was supported in part by the European Community Seventh Research Framework Programme project SPIDIA (http://www.spidia.eu) Grant Agreement no. 222916 to the Biotechnology Institute of the Czech Academy of Sciences and the TATAA Biocenter, and the European Community Sixth Research Framework Programme project SmartHEALTH (http://www.smarthealthip.com) Grant Agreement no. 016817 to the TATAA Biocenter.

Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References

1. Fisher RA. The arrangement of field experiments. J Min Agric G Br 1926;33:503–13.
2. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology 1993;11: 1026–30.
3. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonák J, Lind K, Šindelka R, et al. The real-time polymerase chain reaction. Mol Aspects Med 2006;27:95–125.
4. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 2000;25:169–93.
5. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. Nat Protoc 2006; 1:1559–82.
6. Kubista M, Šindelka R, Tichopad A, Bergkvist A, Lindh D, Forootan A. The prime technique. Real-time PCR data analysis. G.I.T. Lab J 2007;9–10: 33–5.
7. Oberg AL, Mahoney DW. Linear mixed effects models. Methods Mol Biol 2007;404:213–34.
8. McLean R, Sanders WL, Stroup WW. A unified approach to mixed linear models. Am Stat 1991; 45:54–64.
9. Pekny M, Eliasson C, Chien CL, Kindblom LG, Liem R, Hamberger A, Betsholtz C. GFAP-deficient astrocytes are capable of stellation in vitro when cocultured with neurons and exhibit a reduced amount of intermediate filaments and an increased cell saturation density. Exp Cell Res 1998;239:332–43.
10. Bengtsson M, Ståhlberg A, Rorsman P, Kubista M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals log-normal distribution of mRNA levels. Genome Res 2005;15:1388–92.
11. Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M. Properties of the reverse transcription reaction in mRNA quantification. Clin Chem 2004;50:509–15.
12. Ståhlberg A, Kubista M, Pfaffl M. Comparison of reverse transcriptases in gene expression analysis. Clin Chem 2004;50:1678–80.
13. Blake WJ, Kærn M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature 2003;422: 633–7.
14. Raj A, van Oudenaarden A. Nurture, nature, or chance: stochastic gene expression and its consequences. Cell 2008;135:216–26.
15. Ross IL, Browne CM, Hume, DA. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. Immunol Cell Biol 1994; 72:177–85.
16. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol 2006;4:e309.
17. Pritchard CC, Hsu L, Delrow J, Nelson PS. Project normal: defining normal variance in mouse gene expression. Proc Natl Acad Sci U S A 2001;23: 13266–71.
18. Bengtsson M, Hemberg M, Rorsman P, Ståhlberg A. Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. BMC Mol Biol 2008;9:63.