
© anyaivanova / Shutterstock.com

Mikael Kubista
TATAA Biocenter, Sweden

Prime time for qPCR – raising the quality bar

Quantitative Real-Time Polymerase Chain Reaction, better known as qPCR, is the most sensitive and specific technique we have for the detection of nucleic acids. Even though it has been around for more than 30 years and is preferred in research applications, it has still to win broad acceptance in routine. Main hurdles are the lack of guidelines, standards, quality controls, and even proper methods to evaluate the diagnostic results. This is now rapidly changing.

The polymerase chain reaction (PCR) was conceived in 1983 by Kary Mullis and soon after its first publication in 1986 it was refined into real-time PCR, which today is better known as quantitative real-time PCR (qPCR). qPCR reaches the ultimate sensitivity of detecting a single target molecule if it is present in the reaction container, its specificity is sufficient to distinguish targets that differ in a single base position only, it has virtually an infinite dynamic range, and replicate qPCR measurements are impressively reproducible. Still the technique finds limited use in routine diagnostics and the results delivered are often hard to compare between laboratories. So where is the problem? There are several problems.

One problem is that qPCR workers sometimes have the impression, possibly from companies' representatives, that qPCR is a very simple method to use, which may lead to misuse. Necessary tests and validations to secure the performance of the particular assays used and analyses performed may be neglected and relevant information describing the method is left out in reports and publications. The situation grew more serious with, surprisingly, higher negligence in reporting in the most prestigious journals[1]. Eventually, this led a group of opinion leaders coordinated by Stephen Bustin to compile and publish the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines[2]. The MIQE guidelines are a checklist of parameters related to the validation of the performance of qPCR measurements that shall be reported when data are submitted for publication in scientific journals. The acceptance of MIQE is broad; many leading journals request that submitted manuscripts reporting qPCR data adhere to the guidelines to be considered for publication, and companies produce instruments, reagents and tools that simplify the compliance with MIQE. Biometric studies indicate that the introduction of MIQE had a pronounced effect on the transparency and hopefully also quality of published results supported by qPCR[1].

The second problem is preanalytics. Detailed studies using tools for the optimisation of experimental designs show that the confounding variation in a qPCR based analysis rarely, if ever (if performed correctly), is dominated by the actual qPCR measurement[3]. Rather, most of the variation is introduced during the preanalytical steps: sampling, transport, storage, extraction, and in case of RNA analysis, also during the reverse transcription. This was studied extensively by the consortium SPIDIA supported by the European Framework 7 programme[4]. Sampling may dramatically influence genes' expression by

the abrupt change of the cells' environment and milieu. For example, when blood is collected in EDTA tubes the cells remain live, but the Mg^{2+} is chelated, which has major impact on many biological reactions. For expression profiling it is much better to collect blood into a media that immediately lyses the cells and stabilises the RNA. Sampling may also degrade the RNA by exposing it to nucleases or damaging it by added chemicals such as formalin used when preserving tissues. The degradation affects transcripts differently leading to distorted profiles[5]. There are some assay design tricks to get data for

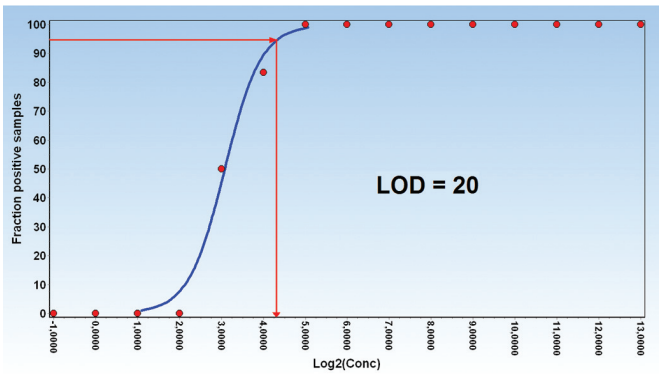


Figure 1: Limit of detection. Graph showing the fraction of replicate samples that shows positive reads at different concentrations (log scale). The fitted data are read out at the relevant confidence level (95%) to give the LOD of the test. Analysis performed with GenEx[22].

more reliable comparison, but it is generally hard to analyse severely degraded samples. Partial degradation can be tested for with electrophoresis, while more extensive degradation is better measured using the long-short qPCR assay strategy[6]. Reverse transcription is less reproducible than qPCR and its yield varies close to 200-fold depending on conditions[7,8]. Sample handling may also introduce variation and even compromise the test material. Recent SPIDIA proficiency ring-trial revealed that about one third of European routine laboratories have issues extracting high quality RNA from blood[9], and were offered training at the TATAA Biocenter[10]. Partly encouraged by the results from SPIDIA the European committee for standardisation has taken the initiative to draft ISO guidelines for the preanalytical process in molecular diagnostics[11]. These are expected to come into force by the end of 2014.

In analytical and clinical chemistry it is routine to measure analytes in complex matrices and estimate their concentrations by means of standard curves. Much of the testing is overseen by regulatory bodies that are different depending on the context and compliance with guidelines requested. A key organisation developing guidelines for molecular testing is the Laboratory and Clinical Standards Institute[12]. Their guidelines EP5 'Evaluation of Precision Performance of Quantitative Measurement', EP6 'Evaluation of Precision Performance of Clinical Chemistry Devices', EP15 'User Verification of Performance for Precision and Trueness', and EP17 'Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures' are potentially most relevant for qPCR analysis of nucleic acids. A practical problem with implementing the CLSI guidelines to qPCR data, however, is that the guidelines primarily consider data measured in linear scale, i.e., where the measured response is proportional to the amount of analyte, and the equations and examples provided are for linear scale data. In qPCR we record the Cq values, which are proportional to the negative log base two of the concentrations/number of molecules ($-\log_2 N$) and is efficiently logarithmic scale measurements. This has several implications for the analysis and interpretation of the data. For example, when measurements are performed on a negative sample, a method such as absorption records a signal proportional to the analyte concentration and for a negative sample it produces a read of the background signal. From repeated measurements the standard deviation (SD) of the background can be calculated and used as a basis to estimate the lowest concentration of the analyte that can be reliably detected. This concentration is known as the Limit of Detection (LOD). SD of the background signal is also used to estimate the lowest analyte concentration that can reliably be quantified, which is known as the Limit of Quantification (LOQ). However, negative qPCR samples do not produce reads; the amplification response curve is not expected to cross the threshold line. Since no Cq values are obtained SD cannot be calculated. Hence, it is not possible to estimate LOD and LOQ by the standard procedures. Instead LOD has to be estimated differently, by analysis of replicate standard curves[13]. Working at 95 per cent confidence LOD can be defined as the analyte concentration that produces at least 95 per cent positive replicates. Under error-free conditions, when only sampling (Poisson) noise contributes to variation, LOD at 95 per cent confidence is three molecules[14]. For real samples LOD is also affected by noise contributed by sampling, extraction, RT, and qPCR, and can be substantially higher. When experimentally determining LOD, standard samples' concentrations should be around the expected LOD, concentration increments shall be small (often 2-fold is used), and the number of replicates should be large (a rough estimate can be obtained with only some six replicates, but precise estimates requires 20 or more). From a plot of the fraction of positive replicates vs. $\log_2 N$ LOD is determined (see **Figure 1**).

LOQ can also be determined from replicate standard curves. This time, though, SD is calculated for the responses of the replicate samples at the different concentrations. SD can be calculated in either log scale (on Cq values) or linear scale[15]. It is expected to increase with decreasing concentration due to sampling ambiguity, which produces an SD of 0.25 cycles at an average number of 35 molecules per analysed aliquot (see **Figure 2**), but also to other errors, such as losses due to

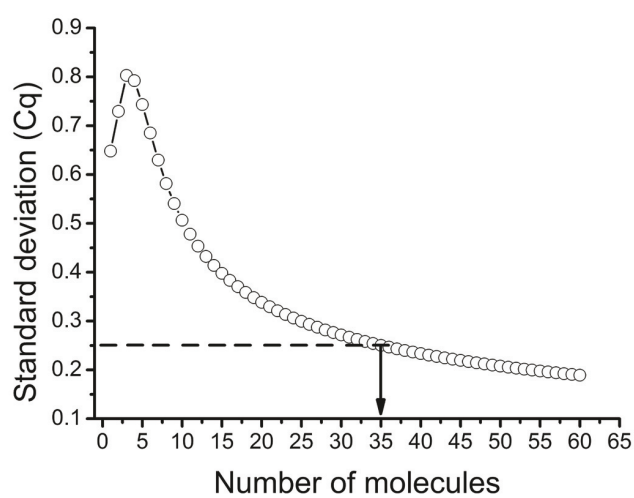


Figure 2: Minimal standard deviation. SD contribution from sampling ambiguity (Poisson noise) on Cq[14].

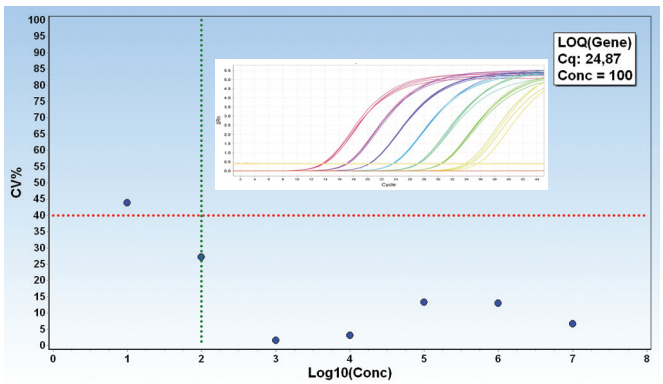


Figure 3: Limit of quantification. Graph showing SD of replicate samples as function of concentration (log scale). LOQ is the highest concentration below the stipulated CV threshold (here 35%). Analysis performed with GenEx[22].

adsorption to surfaces, decreasing reaction yields at lower concentrations, less efficient reactions etc. Calculating SD on data in logarithmic scale, such as the Cq values, has the advantage that data are normally distributed and readily converted to confidence intervals (e.g., 67 per cent of the data expected to be within the mean +/- 1 SD and 95 per cent within the mean +/- 2 SD). However, for easy comparison with other techniques SD of qPCR replicates is often recalculated into linear scale and expressed in percentage as the relative standard deviation, also known as the coefficient of variation (CV = 100xSD/mean), following the outline in the handbook of the National Institute of Standards and Technology[16]. There is no general guidance about the threshold value of LOQ. What is reasonable varies from case to case and depends on the complexity of the samples and the required precision in the follow-up decision-making based on the measured results. For many projects at TATAA, we specify LOQ as the concentration at which CV ≤ 35 per cent. The precision of the LOQ estimate depends on the number of replicates and the concentration increments. A useful estimate can usually be obtained from the regular standard curves, which then have to be performed in at least triplicates and usually are based on 10-fold dilutions (see later on). LOQ is then the lowest concentration for which CV is below the stated threshold (see **Figure 3**). Also, LOQ cannot be lower than LOD. If more precise estimate of LOQ is required higher number of replicates and smaller concentration increments around the expected LOQ can be sued. Notably, SD can also increase towards higher concentrations. This is usually not due to the PCR, but can be an artifact of base-line subtraction. More often, though, increased variation at high concentration is due to saturation of the extraction kit used, limiting amounts of some critical reagent, or contamination. If SD of replicates at high concentration also exceeds threshold, both values are reported and referred to as the lower limit of quantification (LLOQ) and the upper limit of quantification (ULOQ). The stated precision of the test is then only valid within these limits.

The qPCR standard curve is based on the linear relation between the input amount of double-stranded template and the measured Cq value:

$$N_x = N_0(1+E)^x$$

N_0 is the number of double-stranded template molecules present initially in the test-tube, 0<E<1 is the PCR efficiency reflecting the percentage of molecules that is copied every cycle, and x is the total number of cycles. Note, if the template is single-stranded, such as the single-stranded cDNA typically produced by reverse transcription, or a single-stranded vector or virus, one cycle has to be subtracted from x, since the first cycle does not copy the number of template molecules; rather it produces the complimentary copy[17]. The standard curve in qPCR is used for two quite different purposes and depending on the objective it should be designed differently. One objective is to test or validate a newly designed or acquired assay. When testing the performance of a new assay, purified, well-defined template such as a cDNA library, plasmid, or synthetic DNA, and optimum conditions should be used. Wide dynamic range should be covered and replicates (preferably tetraplicates[12]) should be performed. The data are inspected in a standard curve plot, showing the best linear fit and the Working-Hotelling confidence band indicating the area within which the fitted straight line is expected with stipulated confidence, and in a residuals plot showing the deviations of the measured data points from the best fit (see **Figure 4**). The residuals plot is inspected for outliers, which can be done with e.g., the Grubbs' test. One can also test the validity of the assumed linear model with runs test.

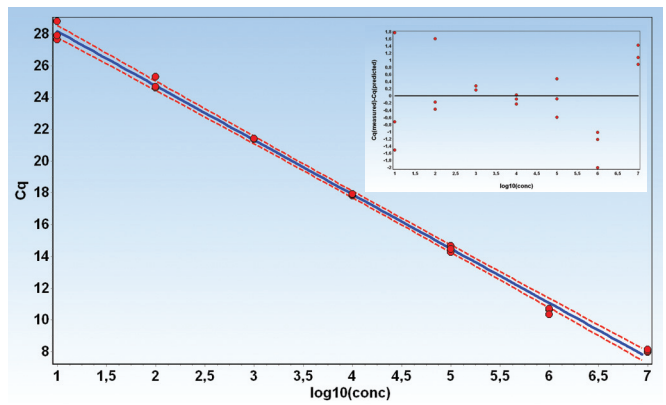


Figure 4: qPCR standard curve. Cq versus the log of the initial number of template molecules in standard samples. Data are fitted to a straight line. Working-Hotelling confidence band is indicated with dashed red lines. Inset shows residuals plot. Analysis performed with GenEx[22].

From the fit the PCR efficiency is estimated as well as its confidence interval[17,18]. For the example data in **Figure 4**, PCR efficiency is estimated at 96 per cent, with the 95 per cent confidence interval: 92 per cent < 96 per cent < 99 per cent. The efficiency is high and estimated with high precision as reflected by the narrow confidence interval. The confidence interval is narrow because a rather large number (21) of standards was included and a wide concentration range was covered. Today, when powerful assay design tools are available, PCR efficiencies for normal RNA targets should reach an efficiency of 90 per cent or more, and most professional assay suppliers do deliver assays of high quality. This is important, particularly if the assay shall be used for analysis of complex matrices, since well performing assays are more robust and therefore less prone to inhibition. A common mistake is to perform multiple independent estimates of PCR efficiency, each based on rather a few standard samples, and use it to correct for day-to-day variations or drift in the performance of the reaction. Such independent standard curves will produce different estimates of the PCR efficiency, but this variation does not reflect true changes in the PCR efficiency; rather they vary

because each standard curve has its particular random noise that has significant impact on the efficiency estimate when only few standards have been used. Estimating the confidence interval of the PCR efficiency is therefore pertinent to reliable analysis.

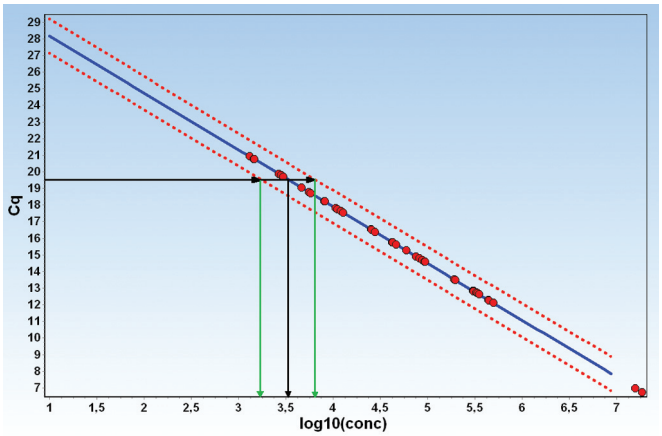


Figure 5: Calibration. Calibration with standard curve. At the measured Cq the standard curve and the Working-Hotelling band are intersected to read out the estimated concentration and confidence interval. Analysis performed with GenEx[22].

The other objective to construct a standard curve is to use it as calibrator to estimate the concentrations of the analyte in field samples. For this purpose the standard samples should be prepared quite differently. If DNA is analysed the standards should of course be based on DNA, but if RNA is being quantified the standards should be based on RNA to also account for the performance of the reverse transcription reaction. As already said, single-stranded standards should be used when quantifying single-stranded targets. Linear standards should be used for quantification of linear targets. Many naturally occurring DNA molecules, such as mitochondrial DNA, chloroplast DNA, many bacterial chromosomes, virus and plasmids are circular, and in their natural state they are supercoiled. The degree of supercoiling, which vary with the state of the cells and may also be affected by the extraction procedure, has a pronounced effect on the priming efficiency of the native template. To avoid bias due to supercoiling circular templates shall be linearised. The template length may affect PCR efficiency, for example, because of flanking sequences in the native molecule folding back on the template sequences making it less available for the primers. Also, local supercoiling in eukaryotic genomic DNA may compromise efficiency. The effect of length can be tested by excising a fragment containing the template using restriction enzymes. Some native targets are in tight complexes with other biomolecules, such as for example eukaryotic DNA being tightly bound by histone proteins in chromatin, or having a protective shield, such as viruses and bacteria. These complexes and interactions may have pronounced effects on the accessibility to the native target and must be mimicked by the standards. Also, the sample matrix usually influences the measurement. Complex matrixes such as blood, faeces, lipid containing tissues, sewer and soil samples etc., may substantially inhibit PCR as well as the reverse transcription, if RNA is measured, and this inhibition must be accounted for in the quantitative analysis. This shall be done using standards constructed in as similar matrix as possible. Recommended procedure is to produce the standard samples by mixing a negative sample and a positive sample, both in representative matrix, at different ratio to cover a relevant concentration range.

When designing standard curves for calibration it is advisable to establish the linear range of the standard curve. The is done by fitting the data also to a second and third order polynomials and test if the coefficients for the higher order polynomials are significant compared to the coefficient of the first order polynomial, which is the slope of the standard curve[12]. If they are significant at a stated confidence (typically 95 per cent) the data deviate from linearity. The data in **Figure 4** (page 65) do not pass the linearity test. Inspection of the residual plot reveals that all three replicates at the highest concentration are larger than predicted by the linear fit, indicating deviation from linearity. This is at the highest concentration in the studied range. We have found this quite frequently, particularly when broad range is covered. The most concentrated samples then have very low Cq values and base-line subtraction, particularly on some instruments, may be prone to systematic error leading to deviation from linearity. Of course, the dynamic range may also be limited by reproducibility, but this is more often at low concentration as influences rather the LOQ than the linear range.

Once the dynamic range of the standard curve has been established, it can be used to predict the concentrations of field samples. Prediction is simple: from the measured Cq value the intercept is subtracted and the difference is divided by the slope, which gives the concentration in logarithmic scale. The precision of the estimate is obtained by calculating the confidence band of the fit also taking into account the imprecision of the measured Cq of the field sample[18,19]. This is illustrated graphically in **Figure 5**: the standard curve is entered from the left at the level of the measured Cq; the best estimate is read out from the standard curve and the confidence interval is obtained by reading out the confidence band. The confidence interval is essentially symmetric around the best concentration estimate on the x-axis, which, is in logarithmic scale. As consequence, when converted to linear scale, the confidence interval around the best concentration estimate is asymmetric (see **Table 1**).

Table 1: Concentration estimates. Concentrations of field samples including confidence intervals estimated in logarithmic and linear scales. Note, while the confidence interval in logarithmic scale is essentially symmetric around the mean, it is asymmetric in linear scale. Analysis performed with GenEx[22].

	CI(min)	Log10(Conc)	CI(max)	Repeats	CI(min)	Conc.	CI(max)
Test 1	4,493449054	4,640334087	4,787219121	3	31149,3547	43685,17576	61265,943
Test 2	3,812430978	3,949469288	4,086507599	3	6492,78434	8901,624868	12204,152
Test 3	2,995439089	3,149469288	3,303499488	3	989,553066	1410,812466	2011,4048
Test 4	4,276880022	4,418166	4,559451977	3	18918,2091	26191,83943	36262,019
Test 5	4,803431446	4,961893162	5,120354877	3	63596,241	91599,51236	131933,44
Test 6	3,30722829	3,451539934	3,595851578	3	2028,74886	2828,394173	3943,2252
Test 7	4,701455427	4,855681225	5,009907023	3	50286,9652	71726,76188	102307,39
Test 8	5,341533245	5,528032017	5,714530789	3	219549,9	337312,1748	518239,83
Test 9	3,939995251	4,077118497	4,214241742	3	8709,54067	11943,13927	16377,279
Test 10	5,113762361	5,287349922	5,460937483	3	129945,834	193798,2815	289026,38
Test 11	3,585513978	3,724377936	3,863241895	3	3850,47207	5301,245738	7298,6392
Test 12	5,301050399	5,485157473	5,669264548	3	200009,396	305602,9016	466943,73
Test 13	6,943287825	7,244962589	7,546637353	3	8775822,4	17577721,9	35207676
Test 14	5,468271867	5,662502175	5,856732483	3	293948,919	459729,291	719005,95

The last problem is about the calibration of standards. Even though we know how to produce standards and use them for calibration, we also need means to calibrate the standards. This has historically been tricky. The standards we use in the research or diagnostic laboratory are known as secondary standards and should be related to a primary standard that is common to all the laboratories to ascertain they report comparable values. To calibrate standard either a reference standards or a reference method is needed. This has not been available for qPCR or rather for molecular diagnostics, and organisations such as the World Health Organisation has made consensus standards available[20]. A consensus standard is produced by sharing a sample among leading laboratories that analyse it and report a test value. The average of the test values is then assumed to be the concentration of the standard. There are many issues with this approach, but recently a new possibility emerged. Digital PCR is a technique to determine the number of DNA target molecules in a sample without comparing with a standard. The sample, which should be rather diluted, is distributed across a very large number of reaction containers. If the number of containers is much larger than the number of target molecules, most containers will not contain any target DNA and most of the rest will contain a single target. All containers are analysed by PCR and counting the number of positive reactions determines, after small correction for Poisson variation, the number of target molecules that were present in the samples[21]. dPCR offers a most wanted reference method to calibrate standards for qPCR use, and was recently introduced by NIST in their development of Standard Reference Materials (SRMs) for molecular diagnostics[16].

In conclusion, the research community is rapidly becoming educated about qPCR through the MIQE guidelines that have become broadly accepted by the community, enforced by prestigious journals, and embraced by the leading instrument/reagents manufactures and service providers[1]. The preanalytics has been studied by SPIDIA and guidelines are expected through ISO[4,11]. Software tools for statistical analysis of qPCR data compliant with the guidelines are emerging[22]. To say everything is in place would be an exaggeration, but qPCR is taking major leaps forward to become accepted in the realm of the routine.



Mikael Kubista studied chemistry at the University of Göteborg, Sweden, and obtained B.Sc. in chemistry in 1984. Mikael then worked at Astra Hässle (today part of AstraZeneca), studying the K+/H+-ATPase inhibitor omeprazole, which became the then most sold pharmaceutical drug under the trade names of Losec (Prilosec in US) and Nexium, and is used to treat ulcers. He returned to academia joining Chalmers University of Technology in Gothenburg and in 1986 received the Technology Licentiate in Chemistry and in 1988 his PhD in physical chemistry on studies of nucleic acid interactions with polarized light spectroscopy. His first postdoc at La Trobe University, Melbourne, Australia, focused on transcriptional foot-printing, and his second postdoc at Yale University, New Haven, USA studied chromatin and epigenetic modulation of nucleosomes. Returning to Gothenburg in 1991, Mikael started his own research group studying DNA-ligand interactions and elucidated some critical details about the RecA catalysed strand exchange process, which led to the establishment of the current model of DNA strand exchange in homologous recombination. They also discovered a novel mechanism of transcriptional activation of oncogenes, which led to the development of a new class of anticancer drugs that target specific quadruplex DNA structures. As part of his team he developed methods for multidimensional data analysis based on which MultiD Analyses AB was founded, and invented the light-up probes for nucleic acid detection in homogeneous solution, which led to the foundation of LightUp Technologies AB as Europe's first company focusing on quantitative real-time PCR (qPCR) based diagnostics. In 2001 Mikael set up the TATAA Biocenter as a centre of excellence in qPCR and gene expression analysis with locations in Gothenburg, Sweden, Prague, Czech Republic, and Saarbrücken, Germany. TATAA Biocenter is the largest provider of qPCR training globally, and Europe's largest provider of qPCR services. It was the first laboratory in Europe to obtain flexible ISO 17025 certification and was presented the Frost & Sullivan Award for Customer Value Leadership as Best-in-Class Services for Analysing Genetic Material in 2013. Mikael also co-authored the MIQE guidelines for RT-qPCR analysis, which receives an average of 25 citations per week, and is a member of the CEN/ISO group drafting guidelines for the pre-analytical process in molecular diagnostics.

References

1. Stephen A Bustin et al. The need for transparency and good practices in the qPCR literature. Nature Methods 11:1063-1067 (November 2013)
2. Stephen Bustin, Jeremy Garson, Jan Hellemans, Jim Huggett, Mikael Kubista, Reinhold Mueller, Tania Nolan, Michael Pfaffl, Gregory Shipley, Jo Vandesompele, Carl Wittwer. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin Chem. 2009 Apr;55(4):611-22.
3. Ales Tichopad, Rob Kitchen, Irmgard Riedmaier, Christiane Becker, Anders Ståhlberg, and Mikael Kubista. Design and Optimization of Reverse-Transcription Quantitative PCR Experiments. Clinical Chemistry 55:10 (2009); doi:10.1373/clinchem.2009.126201
4. www.spidia.eu
5. Joon-Yong Chung, Till Braunschweig, Reginald Williams, Natalie Guerrero, Karl M. Hoffmann, Mijung Kwon, Young K. Song, Steven K. Libutti, and Stephen M. Hewitt. Factors in Tissue Handling and Processing That Impact RNA Obtained From Formalin-fixed, Paraffin-embedded Tissue. Journal of Histochemistry & Cytochemistry 56 (11), 1033-1042 (2008).
6. http://www.tataa.com/products-page/quality-control/
7. Anders Ståhlberg, Joakim Håkansson, Xiaojie Xian, Henrik Semb, and Mikael Kubista. Properties of the Reverse Transcription Reaction in mRNA Quantification. Clinical Chemistry 50:3, 509–515 (2004)
8. Anders Ståhlberg, Mikael Kubista, and Michael Pfaffl. Comparison of Reverse Transcriptases in Gene Expression Analysis. Clinical Chemistry 50, No. 9, 2004
9. M. Pazzagli, F. Malentacchi, L. Simi, C. Orlando, R. Wyrich, K. Günther, C.C. Hartmann,P. Verderio, S. Pizzamiglio, C.M. Ciniselli, A. Tichopad, M. Kubista, S. Gelmini. SPIDIA-RNA: First external quality assessment for the pre-analytical phase of blood samples used for RNA based analyses. Methods 59, 20-31 (2013)
10. http://www.tataa.com/courses/
11. http://www.cen.eu/
12. http://clsi.org/
13. M. Burns, H. Valdivia. Modelling the limit of detection in real-time quantitative PCR. European Food Research and Technology. April 2008, Volume 226, Issue 6, pp 1513-1524
14. Anders Ståhlberg, Mikael Kubista. The workflow of single cell profiling using qPCR. Expert Rev. Mol. Diagn. 14(3), (2014)
15. In contrary to the calculation of confidence interval, calculation of SD does not assume normal distribution. However, the interpretation does. SD is defined as the average distance of the measured values to the average value. This holds for any data. If data are collected from a normal distribution then 967% is expected to be with the mean +/- 1 SD, and 95% with the mean +/- 2 SD.
16. http://www.nist.gov
17. M. Kubista, J. M. Andrade, M. Bengtsson, A. Forootan, J. Jonák, K. Lind, R. Sindelka, R. Sjöback, B. Sjögreen, L. Strömbom, A. Ståhlberg & N. Zoric. The Real-time Polymerase Chain Reaction. Molecular Aspects of Medicine 27, 95-125 (2006).
18. I. M. Mackay, S. A. Bustin, J. M. Andrade, M. Kubista and T. P Sloots. Quantification of Micro-Organisms: Not Human, Not Simple, Not Quick. Chapter 5 in Real-Time PCR in Microbiology: From Diagnosis to Characterization. Publisher: Caister Academic Press. Editor: Ian M. Mackay Sir Albert Sakzewski Virus Research Centre, Queensland, Australia. ISBN: 978-1-904455-18-9 (2007).
19. Paolo Verderio, Sara Pizzamiglio, Fabio Gallo and Simon C Ramsden. FCI: an R-based algorithm for evaluating uncertainty of absolute real-time PCR quantification. BMC Bioinformatics 9:13 (2008)
20. http://www.who.int/
21. Monya Baker. Digital PCR hits its stride. Nature methods, vol.9 no.6, June 2012, p.541-544
22. http://www.multid.se