Real-time polymerase chain reaction – towards a more reliable, accurate and relevant assay

The fluorescence-based quantitative real-time polymerase chain reaction (qPCR) \(^1,2,3\) has become firmly established as the preferred technology for the detection and quantification of nucleic acids in molecular diagnostics, life sciences, agriculture and medicine\(^4\). The combination of conceptual and practical simplicity, large dynamic range of linear quantification, speed, sensitivity and specificity has made it the yardstick for nucleic acid quantification not just in basic research, but has engendered numerous uses ranging from basic research through diagnostic and forensic application to treatment monitoring in a clinical setting\(^6-11\).

Despite its ubiquity, the coming-of-age of this technology has been hampered by significant biological as well as technical issues that frequently combine to obfuscate the interpretation of qPCR data. Biological issues have been discussed elsewhere\(^9,10,12,13\); nevertheless, it is essential to always keep biological relevance in mind when interpreting results, especially when they relate to patient prognosis or drug monitoring. Technical issues are centered on experimental design, data handling, analysis and reporting and whilst peripherally acknowledged, are frequently not adequately addressed\(^4,14,15,16,17\). This insouciant attitude has had major repercussions in the public health domain. A combination of flawed use of qPCR technology, specious data analysis and flawed interpretation resulted in the detection of RNA measles virus in the intestines of children with developmental disorders\(^18\). These data were central to the widespread speculation linking the measles, mumps and rubella (MMR) vaccine with the development of autism. This, in turn, has caused untold distress to thousands of parents, resulted in a dramatic decline in MMR vaccination in a number of countries and was the subject of major class actions in both the UK and the US. At last year’s trial at the vaccine court in Washington DC the reliability of the qPCR data was seriously challenged, with DNA contamination shown to be the most likely cause of most, if not all, positive results\(^19\). Furthermore, the data could never be independently reproduced\(^20,21,22\).

Nevertheless, some lingering doubt continued to remain until recently, when a
paper was published that concludes that there is no evidence for an association between persistent measles virus RNA in the gut and autism. Astonishingly, this publication includes the two main authors of the original paper, and despite publishing evidence that contradicts their own, they have not retracted their original paper. Another example of problems associated with qPCR technology concerns a Science magazine “breakthrough of the year 2005”, which has had to be retracted because of the consequences of poor qPCR assay execution.

Consequently it has been clear for a while that not only are stringent quality control checkpoints at each stage of the experimental workflow indispensable components of a well-designed qPCR experiment, but that these checkpoints must be verifiable. This has resulted in a number of recent initiatives, all aimed at improving the reliability of qPCR-derived data and the transparency of data reporting.

Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE)

Many publications utilising qPCR technology barely provide sufficient information to permit the reader to evaluate the validity of any conclusions derived from the quantitative qPCR data. For example, it is universally accepted that RNA quality assessment is essential for reliable quantification of cellular mRNA using RT-qPCR assays. Yet a brief perusal of 50 BMC open access publications from January to April 2008 reveals that 31 (62%) do not even mention RNA quality, and a further five (10%) describe A<sub>260:280</sub> ratios that are generally accepted as inadequate for accurate quantification. Similarly, a very high percentage of papers continue to normalise gene of interest copy numbers against single reference genes, despite the definitive demonstration that this general approach is invalid and the ready availability of several methods allowing the selection of appropriate sets of reference genes. Other common omissions concern information on sample handling and storage, primers and probes...
JumpStart™ Taq qPCR kits — accurate for all your targets

With a combination of best-in-class JumpStart Taq and highly optimized buffers, Sigma’s qPCR kits are designed for the most demanding applications. The result is the most accurate assay, regardless of the target.

- **Sensitive** – detect even low abundance transcripts with JumpStart Taq
- **Convenient** – ReadyMix™ formulation streamlines pipetting
- **Highly Reproducible** – optimized to minimize inter-assay and intra-assay variability

For more information, go to sigma.com/qpcr

Our Innovation, Your Research — Shaping the Future of Life Science

JumpStart and ReadyMix are trademarks of Sigma-Aldrich Biotechnology LP and Sigma-Aldrich Co.

sigma-aldrich.com
selection, details of the reverse transcription step, efficiency of the PCR reaction, inclusion of controls and data analysis. A summary of the minimum information required to reproduce a rt-qPCR assay is provided in Figure 2.

Since information is knowledge, it is self evident that there is an urgent need to increase the amount of published information. Hence the need for a set of guidelines that can be used by journal reviewers who need to be able to evaluate the reliability of the experimental protocols and ensure the inclusion of all essential information in the final publication. Such guidelines have been discussed and requested for a long time and are finally about to appear in the guise of “Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE)”.

MIQE is modelled on analogous rules drawn up for DNA microarray analysis, proteomics experiments, genome sequence specification and those under discussion for RNAi work and metabolomics, initiatives coordinated under the auspices of MIBBI, Minimum Information for Biological and Biomedical Investigations (www.mibbi.org). MIQE constitutes a set of publication guidelines for researchers, journal reviewers and editors that list the minimum information required to allow potential reproduction as well as unambiguous quality assessment of a qPCR-based experiment. An associated checklist deals with every step of the qPCR assay under the headings “sample, nucleic acids, reverse transcription, target, primers & probes, assay details, PCR cycling and data analysis”. Each heading contains a list of essential or desirable items, e.g. under the heading “primers and probes” the following information needs to be disclosed: “primer sequences, location of any modifications (e.g. LNA), final concentration of primers and optional probe(s), primer purification method, manufacturer of oligonucleotides and probe sequence”.

Systematic adherence to this checklist will provide structured principles for any qPCR experiment which, will provide a significant boost to the quality of data published using this technology. It will enable authors to design and report qPCR experiments with greater inherent value, allow journal reviewers and editors to measure the technical quality of submitted manuscripts against an established yardstick, and of most practical importance, result in the publication of papers that will be much easier to replicate.

Real-time PCR Data Markup Language (RDML)

In the qPCR Stone Age, i.e. around ten years ago, there was a grand choice of two real-time instruments. Characteristically, even then they used incompatible formats (rotor vs. 96-well plate), and quantification schemes (baseline/threshold vs. the second derivative maximum). The former determines the quantification cycle \( C_q \) by drawing a line parallel to the x-axis of the amplification plot and noting the cycle fraction where it crosses the log-linear phase of the amplification plot. The latter method calculates the point of maximal increase of fluorescence within the log-linear phase by determining the second derivative maxima of the amplification curves. The software calculates at which cycle number this point is reached. Both methods remain in common use.

To-day there is a great number of instruments available, all utilising different formats, technologies and software. This proliferation is coupled to the appearance of instruments capable of high throughput: several (Lightcycler 480, BioRad CFX384, Applied Biosystems 7900HT) have 384 well blocks and Biotrove’s OpenArray and Fluidigm’s BioMark technologies permit the analysis of approximately 10,000 reactions simultaneously. Three important consequences of this are that (1) assays are being carried out on a multitude of platforms giving potentially different results, (2) the number of specialised applications is continuously...
increasing, and (3) the number of samples analysed during each run is escalating. Whereas initially most qPCR studies quantified the expression of a handful of target genes in response to, say, drug administration, the introduction of these high throughput systems allows the quantification of numerous genes by qPCR. Consequently qPCR experiments are beginning to match the assessment of complex biological phenomena in the context of high dimensional gene expression profiling.

This proliferation of platforms has resulted in a concomitant increase of instrument-specific software used to generate qPCR data. Since manufacturers do not consider end-user convenience when designing their software, data formats are incompatible. This not only limits data import and export, but data handling and quality control become non-transparent. Every manufacturer saves their run files in proprietary formats and, whilst information can be exported in various file formats (.CSV, .TXT, .XLS), all have different layout and data field terminologies. This creates unnecessary problems for users wanting to exchange data between instrument-specific software packages and analysis tools or for collaborators wanting to share qPCR data between different laboratories. The result is a “Tower of Babel” syndrome, where every instrument speaks a different language, which serves to obscure, confuse and limit data exchange.

A possible solution to this was unveiled in 2005, when a universal XML-based data format initiative for the exchange and publication of qPCR data was proposed at the Freising qPCR Symposium (www.wzw.tum.de/gene-quantification/qpcr2005/pub/). Its aim was to encourage the adoption of a universal data format for real-time PCR data, named RDML (Real-time PCR Data Markup Language). This initiative was followed by the launch of a RDML-website in 2006 and the inauguration of the RDML consortium (www.rdmol.org) in 2008. When implemented, the universal data format will provide sufficient information to allow reviewers and reader to understand the qPCR experimental setup, re-analyse the data and interpret the results. It will be independent of computer hardware, operating system or available software package, and sufficiently flexible to allow future additions of additional information.

The main advantage of a common universal format would be the ease with which raw annotated data could be supplied to manuscript reviewers and readers, collaborators and databases. The main disadvantages are that there is currently no agreed standard for what information should be included, how precisely the information should be handled, whether and how instrument manufacturers would modify their software and whether researchers would be willing to see their raw data re-analysed and re-interpreted. Nevertheless, it does not require much imagination to predict that some implementation of a common data format will become more and more
Data analysis and management programs

Although the analysis of qPCR data has been described as having reached a mature stage of development, the publication of disparate and often contradictory analysis methods suggests that there is little consensus with respect to data analysis and probably even less consistency of data management. Consequently, the coordination and efficient management of experimental data, especially with respect to their statistical evaluation, represent a critical additional, albeit poorly addressed, challenge. Furthermore, the increasing trend towards complex, high throughput qPCR experiments places robust data analysis and management tools at the heart of the drive towards creating a sophisticated data analysis environment yielding reliable and reproducible qPCR data.

The appreciation that PCR efficiency was variable, and hence a key parameter for establishing reliable PCR assays, has resulted in the publication of numerous methods, both linear and non-linear, for calculating optimal (i.e. most accurate) PCR efficiency. Many of the concepts and algorithms developed in these publications lie at the core of the numerous software packages that have been developed for data analysis. One downside of this diversity, incidentally, is that this diversity may lead to variable or even inaccurate results.

Although instrument-specific software has become more flexible and powerful, its development has not kept pace with the major technological advances incorporated into the instruments. All supplied software can extract $C_q$ information from recorded fluorescence measurements, allowing the user to display basic amplification plots, together with threshold lines and melt curves. However, beyond the construction of standard curves, the calculation of sample copy numbers, means and standard deviation for replicates, both their functionality and sophistication are severely limited. Raw data processing is only just beginning to be incorporated by instrument manufacturers: for example BioRad’s CFX software allows limited gene expression analysis by providing options for calculation of differences in a target’s concentration between samples, showing them either as normalised expression ($\Delta C_q$) or relative quantity ($\Delta C_{\Delta q}$).

The need for better experimental organisation and more powerful, statistically reliable data analysis has led to the development of a number of software tools that are designed to standardise, simplify and make qPCR data management and analysis more transparent.

One such tool is PREXCEL-Q, which is unique in that it addresses the labour- and time-intensive set-up and optimisation steps associated with the introduction of a new qPCR assay. It provides a comprehensive set of Excel-based templates that permit rapid calculation of reagent volumes, e.g. for nuclease treatments or RT and qPCR reactions and generates ready-to-use protocol printouts. Especially useful is the feature that allows the researcher to identify appropriate sample dilutions that result in maximum amplification efficiency.

Another category of analysis tool aims to address every step of a qPCR assay by incorporating appropriate algorithms for quantification, statistical tests, error propagation, inclusion of data quality control etc. Following the import of raw $C_q$ data, they may perform quality control and outlier detection, examine the correlation between biological replicates, select the optimal combination of endogenous controls for normalisation based on stability algorithms and compute fold-change and significance results for differential expression analysis.

An early example is the modular Q-Gene, with more recent programs becoming increasingly sophisticated, comprehensive and user-friendly, e.g. qPCR-DAMS, DATAN GenEx and qBasePlus. Use of such software not only speeds up the analysis of raw RT-qPCR data, but also helps improve experimental accuracy by implementing more rigorous analytical workflows. They offer a range of sophisticated multivariate analyses, incorporate algorithms for identifying optimal normalisation genes, e.g. geNorm or NormFinder and use inter-run calibration methods that allow samples analysed in different runs to be compared against each other. For users of Applied Biosystems instruments, Integromics’ StatMiner adds data mining capability to enhance qPCR data analysis by using functional annotations from public databases.

Conclusion

qPCR has been passing through a “cowboy” phase that is characterised by a technological free-for-all in terms of methodology, protocols, data analysis and interpretation and consensus on the amount of information required for publication. However, qPCR has been around for 16 years and is no longer a novel technique. Consequently it is high time for the technology to enter a more consolidated period that will yield verifiable technically reliable, as well as biologically, meaningful data. A combination of appropriate experimental design and an acceptance of the three initiatives discussed above would constitute a significant step towards this goal and will, it is hoped, allow qPCR technology to fulfill its immense promise.

References


