Quantitative real–time PCR for cancer detection: the lymphoma case

Anders Ståhlberg†, Neven Zoric, Pierre Åman and Mikael Kubista

Advances in the biologic sciences and technology are providing molecular targets for diagnosis and treatment of cancer. Lymphoma is a group of cancers with diverse clinical courses. Gene profiling opens new possibilities to classify the disease into subtypes and guide a differentiated treatment. Real-time PCR is characterized by high sensitivity, excellent precision and large dynamic range, and has become the method of choice for quantitative gene expression measurements. For accurate gene expression profiling by real-time PCR, several parameters must be considered and carefully validated. These include the use of reference genes and compensation for PCR inhibition in data normalization. Quantification by real-time PCR may be performed as either absolute measurements using an external standard, or as relative measurements, comparing the expression of a reporter gene with that of a presumed constantly expressed reference gene. Sometimes it is possible to compare expression of reporter genes only, which improves the accuracy of prediction. The amount of biologic material required for real-time PCR analysis is much lower than that required for analysis by traditional methods due to the very high sensitivity of PCR. Fine-needle aspirates and even single cells contain enough material for accurate real-time PCR analysis.

Contents

†*Author for correspondence*
TATAA Biocenter,
Medicinaregatan 7B,
413 90 Gothenburg, Sweden;
Department of Chemistry &
Bioscience, Chalmers University of
Technology, Medicinaregatan 9E,
413 90 Gothenburg, Sweden
Tel.: +46 317 733 916
Fax: +46 317 733 910
anders.stalberg@tataa.com

Most of the human genome sequence is now available and the information can be applied to understand, characterize and treat complex diseases such as cancer. Many cellular functions related to survival, growth and differentiation are reflected in the gene expression pattern, and the ability to quantify transcription levels of specific genes has always been central to research on gene function. For example, normal cell regulation is affected by factors such as viral infections, DNA methylation and sequence alterations. Cancers may develop when these changes affect genes controlling cell division, cell repair, growth control and apoptosis. In the near future it will probably be possible to diagnose an individual's state of health, and also to monitor how individuals respond to medication, treatment and altered living conditions by gene expression profiling.

The expression of virtually all genes in a sample can be roughly assessed by microarray techniques. The expression of selected genes can be measured with very high accuracy by real-time PCR [1–3]. In studies of new systems or in the search for tumor markers, microarray screening is typically used to identify reporter genes, whose expressions are studied in greater detail on a larger number of clinical samples by real-time PCR. The introduction of real-time PCR has significantly improved and simplified quantification of nucleic acids and the technology has become an invaluable tool for many scientists, particularly those in the field of molecular diagnostics [4–7].

Real-time PCR is currently the most sensitive method to determine the amount of a specific DNA in a complex biologic sample [6,8]. In real-time PCR, the amount of product is measured during ongoing amplification using fluorescent reporters. This confers important advantages to classical PCR where only end point quantification is possible. Real-time PCR is characterized by wide dynamic quantification range (seven to nine orders of magnitude), technical sensitivity sufficient to detect a

single molecule and high reproducibility (standard deviation less than 0.2 cycles) [1,2,9]. Another advantage is that no post-PCR processing is required. This drastically reduces the risk of contamination, which otherwise poses serious problems in diagnostics. This review discusses the great potential of quantitative real-time PCR as a method for gene expression analysis and its practical implications in detection and classification of lymphomas.

Quantitative real-time PCR

Quantitative real-time PCR detection involves several steps prior to the actual real-time PCR measurement. These include sample collection, nucleic acid isolation, reverse transcription (RT) and occasionally DNase treatment. For genomic analysis, only the first two steps are needed. The processing of a sample using real-time PCR as a detection method and potential sources of variability are outlined in FIGURE 1. Today, several platforms are available for quantitative real-time PCR. The main differences between these platforms are available excitation and emission wavelengths, speed, and the number of reactions that can be run in parallel [2,10,11]. It is beyond the scope of this review to discuss the steps preceding the RT. However, some crucial issues are mentioned.

Reverse transcription

Before a gene expression measurement can be performed by real-time PCR, the mRNA in the sample must be copied into complementary DNA (cDNA) by RT. The RT step is critical for sensitive and accurate quantification, since the amounts of cDNA produced must correctly reflect input amounts of mRNAs. Several reverse transcriptases are commercially available today. Most

of them are engineered forms of either the Moloney murine leukemia virus (MMLV) or of the Avian myeloblastosis virus (AMV). The RT yields may vary over 100-fold among different target genes due to mRNA secondary and tertiary structures, priming strategy (random hexamers, oligo(dT) or gene-specific primers), and properties of the reverse transcriptase (dynamic range, RNase activity and thermal stability) [12–14]. For quantitative gene expression measurements to be comparable between laboratories, it is important that the same experimental conditions are used. In absolute quantification of mRNA, a known concentration of exogenous mRNA can be added to the sample to compensate for variations in losses during RNA isolation, and variations in RT yields. Still, it is not possible to fully mimic the situation of native mRNA molecules in the cell, making absolute quantification exceedingly difficult.

The RT and real-time PCR reactions can be combined into a one-step RT-PCR reaction. This is particularly convenient when analyzing only one or a few genes, and it reduces the risk of crosscontamination. However, the optimal reaction conditions for RT and real-time PCR are usually different, and one-step reactions are typically somewhat less sensitive than the regular two-step approach [2,14].

In RT-PCR, false-positive signals may arise from genomic DNA in the sample. To avoid this problem, PCR primers for mRNA quantification are usually designed to span over an intron or across exon boundaries [15]. If such a design is not possible the sample must be tested for genomic contamination [14]. This is achieved by running a no-RT control, which is a normal sample but without reverse transcriptase. If genomic background is a problem, the sample may be treated with DNase before RT to digest the DNA.
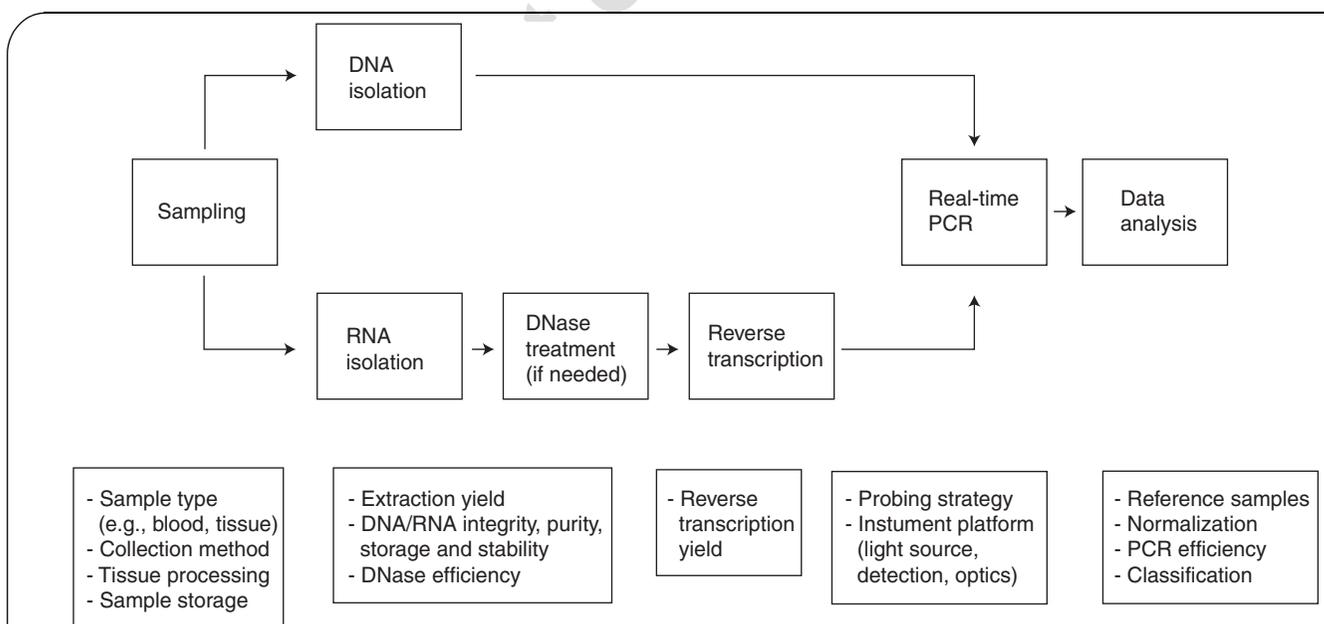

Figure 1. Steps in nucleic acid quantification by real-time PCR. Potential sources of technical variability are indicated.

Probing techniques

Today, fluorescence is exclusively used as the detection method in real-time PCR. The fluorescent reporters can be divided into two categories: sequence-specific and nonspecific labels [10]. The nonspecific labels are DNA-binding dyes such as SYBR® Green I [16,17] and BEBO [18], which become strongly fluorescent when bound to double-stranded DNA. When such dyes are present in the reaction mixture, the fluorescence increases in proportion to the amount of double-stranded products formed. The dyes give signal in presence of any double-stranded DNA including undesired PCR primer-dimer products. By melting curve analysis the shorter primer-dimer products may be distinguished from full-length PCR products [19]. Since dyes do not differentiate between products they cannot be used for multiplex assays. Sequence-specific probes are based on oligonucleotides or analog of oligonucleotides to which one or two fluorescent dyes are coupled. Examples of probes with two dyes include the hydrolysis probes (TaqMan® probes) [20], molecular beacons [21,22], hybridization probes [23] and locked nucleic acid (LNA) probes [24]. The dyes are selected to form a donor–acceptor system, where energy is transferred from the donor to the acceptor dye by fluorescence energy transfer by either the Förster resonance energy transfer (FRET) (long distance) or Dexter (short distance) mechanism. It is also possible to construct the dual-labeled probes with one dye and a quencher molecule that quenches the dye's fluorescence upon contact [25]. Energy transfer and quenching are distance dependent and structural rearrangement of the probe or probe degradation can change the distance between the dyes and, hence, the fluorescence of the system. Probes based on a single dye, whose fluorescence changes upon binding target DNA include the LightUp probes [26], AllGlo™ probes (AlleLogic Biosciences) [101] and SimpleProbes® (IT BioChem) [102]. Chemical modifications and alterations of the oligonucleotide backbone are employed in some probes to improve the binding properties to the target template. This makes it possible to use shorter probes, which is advantageous for detection of targets with short conserved regions such as retroviruses. LightUp probes have a neutral peptide nucleic acid (PNA) backbone that binds to DNA with greater affinity than normal oligonucleotides. The LightUp probes are typically 10–12 bases long compared with normal oligonucleotide probes, which are approximately 25 bases. The LNA probes make use of modified nucleotides to enhance binding affinity. Minor groove-binder (MGB) hydrolysis probes (MGB-TaqMan probes), where a minor groove-binding molecule is attached to the end of the probe [27], is another design to increase probe affinity for target DNA, enabling use of shorter probes. Some other detection systems are based on modified primers to generate signal upon amplification. These include Scorpion primers [28], LUX primers [29], Amplifour primers [30] and the QZyme system [31].

The number of possible analyses made in a single sample is usually limited by the amount of biologic material available. In multiplex real-time PCR, multiple specific primer and probe sets are used to amplify and detect several targets in the same sample.

This saves time, reagents and material. It also eliminates sample-to-sample variation caused by handling and PCR inhibitors, which is particularly important for complex tissue samples [32,33]. The drawback of multiplex PCR systems is that careful optimization is needed to avoid competition for chemicals and reagents between the parallel reactions. Today, multiplexing is mainly used to relate expression of reporter genes to that of an exogenous control gene in diagnostic applications and for single nucleotide polymorphism (SNP) and mutation detection [34].

The different probing technologies have their advantages and limitations. Dyes are substantially cheaper than probes but they do not distinguish between products. Molecular beacons have the highest specificity as the formation of the semistable hairpin structure competes with binding to mismatched targets. This makes the molecular beacons most suitable for SNP and multisite variation (MSV) analysis [35]. However, this requires stringent design, which is hard and may not always be possible for a specific target. TaqMan probes are easier to design and an 80% success rate was recently reported [36]. However, TaqMan probes require two-step PCR to function properly. This is not the optimal condition for the polymerase reaction and short amplicons are necessary to obtain reasonable amplification efficiencies. A good probe, independent of chemistry, should have low background fluorescence, high fluorescence upon target formation (high signal-to-noise ratio) and high target specificity. The dyes' excitation and emission spectra are important parameters to consider when designing multiplex reactions. Spectral overlap in excitation and emission should be minimized to keep crosstalk to a minimum.

Real-time PCR

Through real-time PCR, the amount of any DNA target in a biologic sample can be determined with unprecedented accuracy over an essentially unlimited dynamic range. At a certain point during cycling, enough product accumulates to increase the fluorescence signal significantly above the background noise. This point, typically defined by an arbitrarily set threshold line or by the second derivative maximum of the amplification response curve, is referred to as the threshold cycle (Ct) or crossing point (Cp). A typical real-time PCR amplification curve is shown in FIGURE 2. The Ct value relates to the number of initial template molecules as:

$$N_{Ct} = N_0 \cdot (1 + E)^{Ct} \tag{1}$$

where N_{Ct} is the number of double-stranded DNA molecules after Ct amplification cycles, N_0 is the initial number of double-stranded target molecules, and E is the PCR efficiency. E is assumed to be independent of N_0 over the studied concentration range, and can be estimated from dilution series of mRNA, cDNA or genomic DNA, or from the real-time PCR amplification response curve (FIGURE 3).

Relative gene expression analysis & normalization

Quantitative gene expression analysis is most commonly performed as a relative measurement between two genes. The expression ratio is given by [9,12,37]:

$$\frac{N_A}{N_B} = K_{RS} \frac{\eta_B (1+E_B)^{CT_B - 1}}{\eta_A (1+E_A)^{CT_A - 1}} \quad (2)$$

where N_i is the initial number of mRNA molecules of gene i, K_{RS} is the relative sensitivity of the detection chemistries of the two assays [37], and η_i is the cDNA synthesis yield of gene i, defined as the fraction of mRNA molecules that are transcribed to cDNA in the RT reaction [12]. The exponent Ct-1 accounts for the production of double-stranded DNA in the first PCR cycle from the single-stranded cDNA template generated by the RT reaction. Typically, η is assumed to be independent of both total and target mRNA concentrations. In most applications the expression ratio of two genes is compared in two or more samples (i.e., comparative quantification) [38]. Typically, one is the reporter gene whose expression is expected to be affected by the physiologic change studied and the other is a reference gene whose expression should be constant. Assuming the same RT yields in the samples, K_{RS} and η are eliminated, and the comparative expression ratio of the two samples is given by:

$$\frac{Sample1}{Sample2} = \frac{\left[\frac{N_A}{N_B}\right]_{Sample1}}{\left[\frac{N_A}{N_B}\right]_{Sample2}} = \frac{\left[\frac{(1+E_B)^{Ct_B - 1}}{(1+E_A)^{Ct_A - 1}}\right]_{Sample1}}{\left[\frac{(1+E_B)^{Ct_B - 1}}{(1+E_A)^{Ct_A - 1}}\right]_{Sample2}} \quad (3)$$

Assuming the same PCR efficiencies in the two samples, this expression simplifies to:

$$\frac{Sample1}{Sample2} = \frac{(1+E_B)^{Ct_{B1} - Ct_{B2}}}{(1+E_{BA})^{Ct_{A1} - Ct_{A2}}} \quad (4)$$

Further, also assuming 100% PCR efficiency, it simplifies to:

$$\frac{Sample1}{Sample2} = \frac{2^{Ct_{B1} - Ct_{B2}}}{2^{Ct_{A1} - Ct_{A2}}} = 2^{(Ct_{B1} - Ct_{B2}) - (Ct_{A1} - Ct_{A2})} = 2^{\Delta\Delta Ct} \quad (5)$$

which is commonly referred to as the $\Delta\Delta$Ct method.

Comparing samples requires normalization to compensate for differences in the amount of biologic material analyzed. A number of normalization methods have been suggested based on physical parameters such as volume, mass, size and cell number. Due to the heterogeneity of biologic samples, these methods are usually impractical and unreliable. It is more convenient to normalize to total RNA amount, externally added RNA, ribosomal RNA or internal reference genes. The latter is currently the most popular strategy [1,2]. However, finding appropriate reference genes for data normalization is currently one of the most challenging problems. Extensive evidence indicates that all genes are regulated under some conditions. The expression of many genes is often out of control, particularly in tumor tissue. The field may have to face up to the fact that no universal reference gene with a constant expression in all tissues exists. Due to this uncertainty, any system relying on reference genes should be carefully validated. A number of methods have been proposed to find optimal
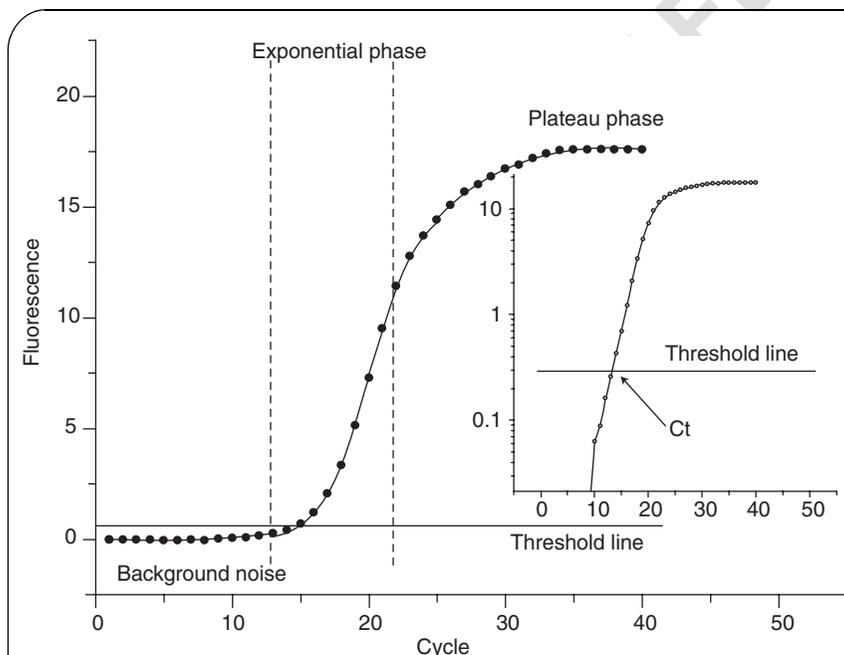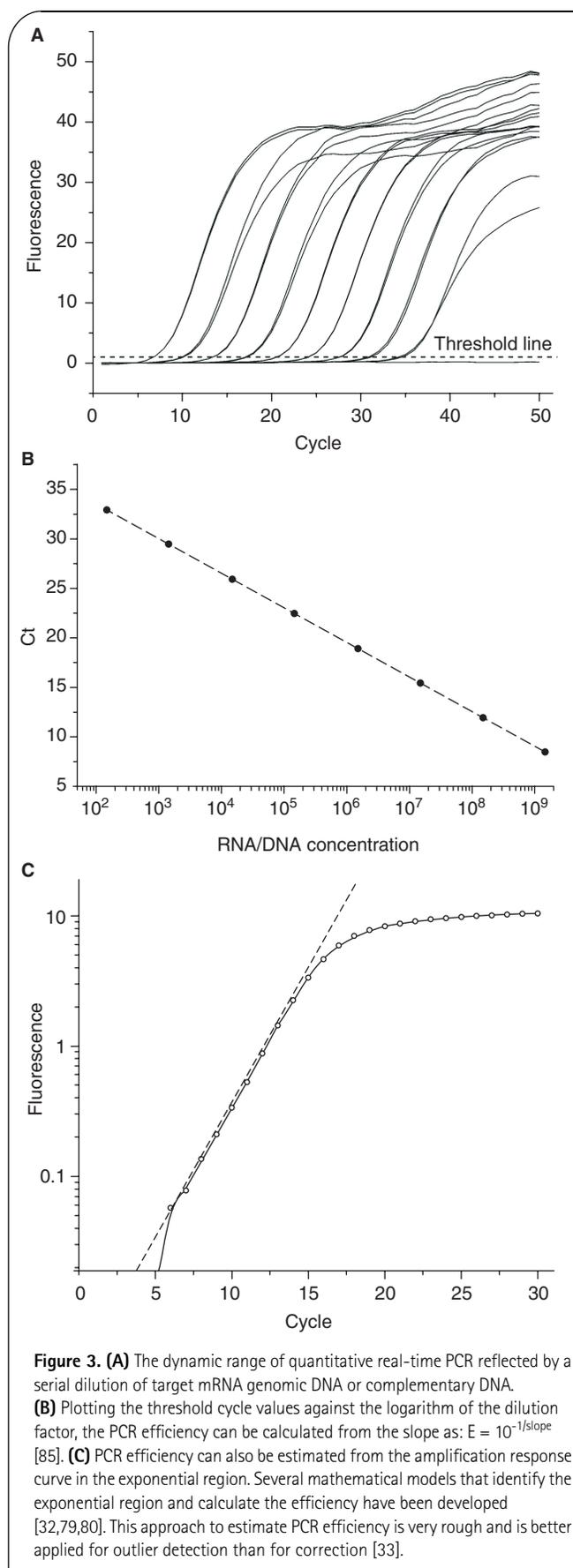

Figure 2. A real–time PCR response curve. The response curve may be divided into three regions. The first part is characterized by background noise, where the target specific florescence is too low to be detected. The second part is the exponential phase where the fluorescence rises significantly above the noise level and is proportional to the amount of product formed. Threshold cycle values should be determined in the exponential phase where extrapolation to determine initial template amount is most reliable. The exponential region is readily identified by plotting the fluorescence signal in logarithmic scale. Eventually, the PCR runs out of chemicals and the response curve enters the plateau phase.

Figure 3. (A) The dynamic range of quantitative real-time PCR reflected by a serial dilution of target mRNA genomic DNA or complementary DNA. **(B)** Plotting the threshold cycle values against the logarithm of the dilution factor, the PCR efficiency can be calculated from the slope as: $E = 10^{-1/\text{slope}}$ [85]. **(C)** PCR efficiency can also be estimated from the amplification response curve in the exponential region. Several mathematical models that identify the exponential region and calculate the efficiency have been developed [32,79,80]. This approach to estimate PCR efficiency is very rough and is better applied for outlier detection than for correction [33].

reference genes for a defined system [39,40]. The GeNorm method developed by Vandesompele and coworkers is currently the most popular [40]. GeNorm assumes that the two genes, whose expression is most correlated are the most appropriate reference genes. The assumption is questionable since regulated genes may still have highly correlated expression due to coregulation, and are thus not suitable as references. This can be partially avoided by choosing genes with various cellular functions. An alternative approach to analyze real-time PCR gene expression data is to use global pattern recognition to identify differently expressed genes [41]. This does not rely on reference genes.

Cancer detection: the lymphoma case

The classification of human lymphomas has steadily evolved since the first description by Thomas Hodgkin in 1831. Beginning with the distinction of Hodgkin's disease from other malignant and nonmalignant conditions, a variety of lymphoma classifications have been advanced on the basis of morphologic and molecular parameters. The most recent classification scheme for lymphoma is the World Health Organization (WHO) classification. It has been introduced to categorize distinct clinicopathologic entities, and also includes the use of new techniques such as microarrays [103].

Gene expression analysis opens for a possibility to classify cancers into more distinct groups for diagnosis and treatment. Microarray studies of lymphoma have revealed new subgroups with distinct gene expression profiles and also identified new potential marker genes [42–44]. Diffuse large B-cell lymphoma (DLBCL), which is one of the most common subtypes of non-Hodgkin's lymphoma [45], has received particular attention due to its highly diverse clinical courses [46–50]. However, in some systems comparison of experiments performed with different microarray technologies has only shown modest agreement in expression pattern, calling for careful validation of microarray data and methodology before any implementation into routine diagnostics [51]. Based on microarray data, Losses and coworkers selected 36 genes whose expression may be used to predict survival in DLBCL patients [52]. Expression of these genes was measured with quantitative real-time PCR in 66 patients. Out of the 36 genes, six (*LMO2*, *BCL6*, *FN1*, *CCND2*, *SCYA3* and *BCL2*) were found to be strong predictors for survival. Other potential lymphoma markers that have recently been validated by quantitative real-time PCR include *CCND1* [53–55], *CDC37* [55], glutathione-S-transferase pi (*GST-pi*) [56] and the cytokines [57].

A number of reference genes have been tested for normalization purposes in classification of lymphoma [54,55,58,59]. However, no gene has yet been identified whose expression is constant in all lymphoma samples. Comparison with reference genes can be avoided by designing tests based on the ratio of two reporter genes, rather than the traditional comparison of one target gene with one reference gene. Comparing reporter genes directly eliminates errors caused by nonideal behavior of reference genes [37,53]. The following

example demonstrates how two target genes can be used as a diagnostic test for B-lymphocyte monoclonality in lymphoma. The vast majority of non-Hodgkin's lymphomas develop from a B-lymphocyte lineage. B-lymphocytes produce immunoglobulins (Igs) with a heavy chain and either a κ or λ light chain. Which light chain will be produced is determined early in the development of each B-lymphocyte. In healthy humans approximately 60% of the B-lymphocytes produce κ chains. Lymphomas, like all malignant tumors, are clonal and arise from one transformed cell. Therefore, in lymphoma tissue, where tumor cells dominate, the κ:λ ratio varies from 60:40. Comparing the expressions of the κ and λ light chain genes can be used as an indicator for B-lymphocyte monoclonality, without relying on any internal reference gene (FIGURE 4) [37]. In a similar approach, the *CCND1:CCND3* ratio can be used to diagnose mantle cell lymphoma [53]. The general strategy of comparing expression of only reporter genes is superior to any strategies relying on reference genes. This approach fails when the reporter genes are highly coregulated, since their relative expression may then not reflect the disease correctly. To avoid making mistakes with correlated genes, reference genes and exogenous controls should be used in the test design phase to verify that no undesired correlation exists between the reporter genes. In applications where reporter genes are coexpressed or only one reporter gene is available, reference genes are needed. For example, the response to imatinib treatment in Philadelphia chromosome-positive chronic myeloid leukemia patients can be monitored by measuring the transcript level of the *BCR–ABL* fusion gene [60,61]. The absolute amount of fusion transcripts is used as an indicator of residual disease. When monitoring changes over time, the *BCR–ABL* transcript levels are usually normalized against *ABL* or *GAPDH*, which are assumed to be expressed at a constant level independent of treatment and other factors. Rituximab treatment of patients with B-cell non-Hodgkin's lymphoma is another therapy where quantitative real-time PCR is used for monitoring [62].

In addition to gene expression profiling, real-time PCR is also useful to detect chromosomal aberrations. Most chromosomal translocations in non-Hodgkin's lymphomas can be detected by genomic PCR. However, due to the very large spread of chromosomal breakpoints, only those in narrow regions can in practice be tested for routine diagnostics. Compared with quantitative gene expression measurements, which require real-time PCR, specific translocations can usually be detected by conventional genomic PCR against zero background. The reader is referred to some recent reviews for an overview of chromosomal translocations in non-Hodgkin's lymphoma and present approaches to test for them [63–66]. Rearrangements of the Ig heavy chain locus are a common target for real-time PCR-based tests [67–70].

Single-cell analysis

The high sensitivity of real-time PCR makes it possible to detect even a single molecule [8,71]. Diagnostics becomes feasible with much lower amounts of biologic material than are required by traditional methods. Fine needle aspirates may replace traditional surgical biopsies in many situations [37]. This reduces patient trauma and also allows many more samples to be collected. Complex tissues contain many cell types, each with a distinct gene expression profile. In a cell population, potentially important variations in gene expression among minority cells may be masked. Analyses of gene expression at the single-cell level therefore open avenues for even more precise understanding of human disease pathogenesis. For single cell analysis, cells are typically collected by flow cytometry, patch-clamp capillaries or laser-based microdissection [71–74]. However, despite the high reproducibility and sensitivity of real-time PCR, robust protocols are needed for single cell analysis before the technique can be used as a standard tool in cancer research. A typical single cell contains only 0.5–1.0 pg mRNA, which accounts to approximately 300,000 mRNA transcripts [75]. This makes single cell real-time PCR analysis sensitive to contamination and statistical uncertainty. For low expressed genes, it may not be possible to split single cell samples into aliquots for multiple gene analysis due to statistical uncertainty (e.g., the Poisson effect becomes important when the number of molecules is <20) [76,77]. Such effects should be accounted for in the experimental design and data treatment, for example, by increasing the number of experimental replicates.

Quality control

One of the most recognized problems in quantitative real-time PCR gene expression analysis is PCR inhibition [32,33].
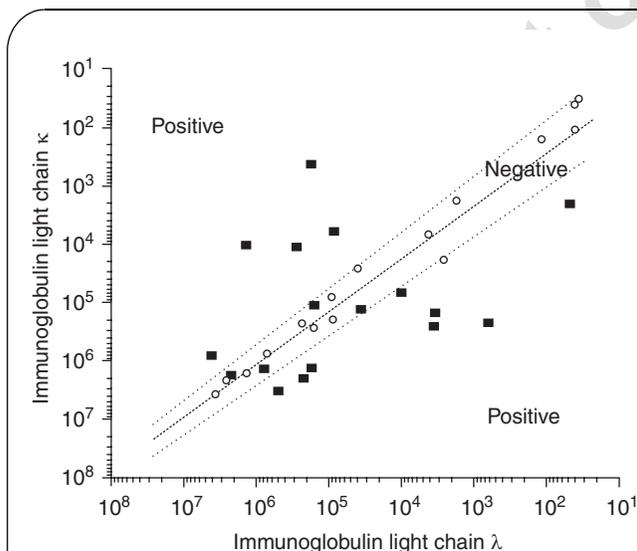

Figure 4. Comparison of k and l immunoglobulin light chain expression. Samples are classified from the k:l expression ratio [37]. A ratio of approximately 60:40 indicates healthy individuals (circles), while deviant ratios indicate lymphoma positive individuals (squares). The dotted lines indicate 95% confidence interval of negative samples.

The inhibiting agents may be extraction chemicals or unpurified components from the biologic sample. There are many ways to purify nucleic acids, including phase separation, filtration, magnetic beads, lysis buffers, anion-exchange columns and ethanol precipitation. Despite the method of isolation employed, there are always material losses, and traces of inhibitors will find their way into the final nucleic acid extract. Heme, heparin and IgG are examples of some common inhibitors in biologic samples that inhibit PCR at very low concentrations [33,78]. In the worst-case scenario, a high degree of inhibition will give false-negative results. Due to the exponential nature of PCR, errors caused by inhibition accumulate with the product. The common procedure to account for any differences in PCR efficiencies between samples is to amplify a reference gene in parallel to the reporter gene and relate their expression levels. This approach assumes that the two assays are inhibited to the same degree. However, this may not always be the case, particularly not for complex tissue samples, and the procedure, although very common, may lead to false classification of samples [33,34]. The problem is even more serious in absolute quantification, where an external calibration curve is used to estimate the number of transcripts in the test samples. The biologic samples are often more complex and contain inhibitors that are not present in the standard samples used to construct the calibration curve. In such cases, the mRNA levels in the test samples will be underestimated. The PCR efficiency in a test sample can be assessed by serial dilution of the sample [37]. It is also possible to obtain a rough estimate of the PCR efficiency from analysis of the amplification response curve. Mathematical algorithms of different degrees of complexity have been developed for this purpose (FIGURE 3) [32,33,79,80]. Finally, amplification of an externally added RNA/DNA standard may be used to test for inhibition, and also to compensate for material losses during sample preparation [81,82].

Expert opinion

The introduction of real-time PCR technology in molecular diagnostics has greatly simplified the quantification of nucleic acids. Very large amounts of data can be generated in a short period of time. Real-time PCR is rapidly becoming a powerful complement to traditional methods such as immunohisto-chemsitry, fluorescence *in situ* hybridization and flow cytometry for cancer detection and classification. Today, methods are even available to analyze nucleic acids from paraffin-fixed tissues, which makes it possible to search for new cancer reporter genes in archival samples [83,84].

Before setting up quantitative real-time PCR-based assays for routine diagnostic purposes, some important parameters should be carefully validated. Most importantly, the reporter genes used for prediction must be statistically validated on a substantial number of patient samples that are characterized and classified by traditional methods. The main technical obstacle in real-time PCR is inhibition caused by unpurified interfering components from the biologic sample. In most cases, this can be accounted and corrected for by proper normalization of all experimental steps including sample collection, nucleic acid extraction, RT and real-time PCR. The authors favor normalizing the entire sample handling process when comparing similar samples. Starting with the same amount of biologic material and using a very rigorous purification process, the authors expect low sample-to-sample variation in RNA purification yield and a uniform low degree of inhibitor contamination. The same amount of total RNA from each sample is then reverse transcribed and the same amount of cDNA is used for real-time PCR analysis. This approach minimizes any variations due to sample handling. To test if any of the studied samples suffer from severe PCR inhibition, PCR efficiencies may be determined by serial dilution [37]. A quality assurance system, such as kinetic outlier detection, can then be implemented to test the data [33]. These strategies may detect any variation in PCR efficiencies. Another approach is to add an external RNA/DNA standard to the sample, which will compensate for variation in nucleic acid isolation yields and RT efficiencies among the samples [83,84]. Experimental designs that do not rely on reference genes, such as the Ig light chain κ:λ ratio [37] and the *CCND1*:*CCND3* ratio [53] mentioned, are always preferred. If that is not possible at least two carefully validated reference genes with fundamentally different cellular functions should be used together as a normalization factor to minimize the risk of coregulation with the reporter genes. Appropriate reference genes may be identified by the GeNorm software [40]. Another important issue when implementing PCR into routine work is to minimize the risk of contamination. An important preventive measure is to perform mastermix preparation, sample loading and post-PCR analysis in separate rooms.

Five-year view

Microarray technologies and real-time PCR have been developed independently and are used today for quite different applications related to gene expression measurements. It is the authors' view that these approaches will become much more synchronized, resulting in highly efficient strategies to identify targets and develop diagnostic, prognostic and theranostic tests based on gene expression. Additionally, on the instrument side these techniques will move closer and possibly merge into a real-time PCR microarray platform. Combining the high-throughput capacity of microarray techniques and the large dynamic range of real-time PCR will make it possible to use both low- and high-expressed genes as reporter genes. Advances in sample enrichment and preparation are other bottlenecks limiting high throughput. During the next 5 years, the authors expect major improvements in sampling techniques. In this manner, a large enough number of lymphoma reporter genes can be measured together on real-time PCR microchips for a variety of diagnostic and prognostic purposes.

Key issues

- All steps from sample collection to data analysis are important sources of variability in real-time PCR experiments that need to be properly controlled for accurate gene expression profiling.
- Real-time PCR measurements are characterized by a large dynamic range, high precision and specificity, and excellent sensitivity.
- Gene expression analysis by microarray techniques and real-time PCR offers new possibilities to classify malignant tumors, such as lymphomas, into more distinct subtypes for diagnosis and treatment.
- Absolute quantification of mRNA is a complex process requiring a known external calibrator with a sequence identical to the target molecule. Gene expression profiling is therefore most frequently performed as a relative measurement comparing different biologic conditions.
- To compensate for the different amount of mRNA among samples, normalization is necessary. Several methods have been proposed, most being based on internal reference genes whose expression is assumed to be constant in all samples.
- More accurate classification can be achieved by comparing expression of target directly, without involving reference genes. The ratio between immunoglobulin light chain κ and λ and between *CCND1* and *CCND3* are two such examples in lymphoma diagnostics.
- The high sensitivity of real-time PCR makes the technique applicable to very small samples, such as fine-needle aspirates. Even a single cell is enough for analysis.
- Biologic samples often contain PCR inhibitors. These affect PCR efficiency and consequently the quantitative gene expression information that is sought. The degree of PCR inhibition can be estimated by analyzing the real-time PCR amplification response curve or by serial dilution of the sample.

References

Papers of special note have been highlighted as:
- of interest
- • of considerable interest

1 Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* 25(2), 169–193 (2000).
•• **First extensive review of quantitative real-time PCR.**

2 Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* 29(1), 23–29 (2002).

3 Dyrskjot L. Classification of bladder cancer by microarray expression profiling: towards a general clinical use of microarrays in cancer diagnostics. *Expert. Rev. Mol. Diagn.* 3(5), 635–647 (2003).

4 Ginzinger DG. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp. Hematol.* 30(6), 503–512 (2002).

5 Bernard PS, Wittwer CT. Real-time PCR technology for cancer diagnostics. *Clin. Chem.* 48(8), 1178–1185 (2002).

6 Klein D. Quantification using real-time PCR technology: application and limitations. *Trends Mol. Med.* 8(6), 257–260 (2002).

7 Walker NJ. Tech. Sight. A technique whose time has come. *Science* 296(5567), 557–559 (2002).

8 Liss B, Roeper J. Correlating function and gene expression of individual basal ganglia neurons. *Trends Neurosci.* 27(8), 475–481 (2004).
• **Discusses the potential of quantitative single-cell analyses.**

9 Pfaffl MW. A new mathematical model for relative quantification in real-time PCR RT-PCR. *Nucleic Acids Res.* 29(9), E45 (2001).
• **Technical study that describes the mathematical model for relative quantification.**

10 Wilhelm J, Pingoud A. Real-time polymerase chain reaction. *Chembiochem* 4(11), 1120–1128 (2003).

11 Kubista M, Zoric N. Real-time PCR Platforms. In: *Encyclopedia of Medical Genomics and Proteomics.* Marcel Dekker Inc., NY, USA (2005) (In Press).

12 Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M. Properties of the reverse transcription reaction in mRNA quantification. *Clin. Chem.* 50(3), 509–515 (2004).

13 Ståhlberg A, Kubista M, Pfaffl M. Comparison of reverse transcriptases in gene expression analysis. *Clin. Chem.* 50(9), 1678–1681 (2004).

14 Peters IR, Helps CR, Day MJ. Real-time RT-PCR: considerations for efficient and sensitive assay design. *J. Immunol. Methods* 286(1–2), 203–217 (2004).

15 Wang X, Seed B. A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res.* 31(24), E154 (2003).

16 Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP. Continuous fluorescence monitoring of rapid cycle DNA amplification. *BioTechniques* 22(1), 130–138 (1997).

17 Zipper H, Brunner H, Bernhagen J, Vitzthum F. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Res.* 32(12), E103 (2004).

18 Bengtsson M, Karlsson JH, Westman G, Kubista M. A new minor groove binding asymmetric cyanine reporter dye for real-time PCR. *Nucleic Acids Res.* 31(8), E45 (2003).

19 Ririe KM, Rasmussen RP, Wittwer CT. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.* 245(2), 154–160 (1997).

20 Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of Thermus aquaticus DNA polymerase. *Proc. Natl Acad. Sci. USA* 88(16), 7276–7280 (1991).

21 Tyagi S, Kramer FR. Molecular beacons: probes that fluorescence upon hybridization. *Nature Biotechnol.* 14(3), 303–308 (1996).

22 Tyagi S, Bratu DP, Kramer FR. Multicolor molecular beacons for allele discrimination. *Nature Biotechnol.* 16(1), 49–53 (1998).

23 Caplin BE, Rasmussen RP, Bernard PS, Wittwer CT. LightCycler™ hybridization probes – the most direct way to monitor PCR amplification and mutation detection. *Biochemica* 1, 5–8 (1999).

24 Braasch DA, Corey DR. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chem. Biol.* 8(1), 1–7 (2001).

25 Lakowicz JR. DNA technology. In: *Principles of Fluorescence Spectroscopy, second edition.* Kluwer Academic/Plenum Publishers, NY, USA, 595–617 (1999).

26 Svanvik N, Ståhlberg A, Sehlstedt U, Sjöback R, Kubista M. Detection of PCR products in real time using light-up probes. *Anal. Biochem.* 287(1), 179–182 (2000).

27 Kutyavin IV, Afonina IA, Mills A *et al.* 3′-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res.* 28(2), 655–661 (2000).

28 Whitcombe D, Theaker J, Guy SP, Brown T, Little S. Detection of PCR products using self-probing amplicons and fluorescence. *Nature Biotechnol.* 17(8), 804–807 (1999).

29 Nazarenko I, Lowe B, Darfler M, Ikonomi P, Schuster D, Rashtchian A. Multiplex quantitative PCR using self-quenched primers labelled with a single fluorophore. *Nucleic Acids Res.* 30(9), E37 (2002).

30 Uehara H, Nardone G, Nazarenko I, Hohman RJ. Detection of telomerase activity utilizing energy transfer primers: comparison with gel- and ELISA-based detection. *Biotechniques* 26(3), 552–558 (1999).

31 BD QZyme™ assays for quantitative PCR. *Clontechniques* 18(4), 2–3 (2003).

32 Tichopad A, Dilger M, Schwarz G, Pfaffl M. Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res.* 31(20), E122 (2003).

33 Bar T, Ståhlberg A, Muszta A, Kubista M. Kinetic outlier detection (KOD) in real-time PCR. *Nucleic Acids Res.* 31(17), E105 (2003).

• **Describes how individual amplification response curves may be used to identify PCR-inhibited samples.**

34 Mhlanga MM, Malmberg L. Using molecular beacons to detect single-nucleotide polymorphisms with real-time PCR. *Methods* 25(4), 463–471 (2001).

35 Bonnet G, Tyagi S, Libchaber A, Kramer FR. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proc. Natl Acad. Sci. USA* 96(5), 6171–6176 (1999).

36 Kubista M. Nucleic acid-based technologies: application amplified. *Pharmacogenomics* 5(6), 767–773 (2004).

37 Ståhlberg A, Åman P, Ridell B, Mostad P, Kubista M. Quantitative real-time PCR method for detection of B-lymphocyte monoclonality by comparison of κ and λ immunoglobulin light chain expression. *Clin. Chem.* 49(1), 51–59 (2003).

• **Describes how immunoglobulin light chains κ and λ can be used as a test for monoclonality in B-cell lymphomas.**

38 Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 30(9), E36 (2002).

39 Szabo A, Perou CM, Karaca M, Perreard L, Quackenbush JF, Bernard PS. Statistical modeling for selecting housekeeper genes. *Genome Biol.* 5(8), R59 (2004).

40 Vandesompele J, de Preter K, Pattyn F *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3(7), research0034 (2002).

•• **Describes GeNorm, a method to identify constantly expressed reference genes.**

41 Akilesh S, Shaffer DJ, Roopenian D. Customized molecular phenotyping by quantitative gene expression and pattern recognition analysis. *Genome Res.* 13(7), 1719–1727 (2003).

42 Husson H, Carideo EG, Neuberg D *et al.* Gene expression profiling of follicular lymphoma and normal germinal center B-cells using cDNA arrays. *Blood* 99(1), 282–289 (2002).

43 Shin MS, Fredrickson TN, Hartley JW, Suzuki T, Agaki K, Morse HC III. High-throughput tagging for identification of genes involved in initiation and progression of mouse splenic marginal zone lymphomas. *Cancer Res.* 64(13), 4419–4427 (2004).

44 Tsukasaki K, Tanosaki S, deVos S *et al.* Identifying progression-associated genes in adult T-cell leukaemia/lymphoma by using oligonucleotide microarrays. *Int. J. Cancer* 109(6), 875–881 (2004).

45 Coiffier B. Diffuse large cell lymphoma. *Curr. Opin. Oncol.* 13(5), 325–334 (2001).

46 Alizadeh AA, Eisen MB, Davis RE *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403(6769), 503–511 (2000).

47 Rosenwald A, Wright G, Chan WC *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma. *N. Engl. J. Med.* 346(25), 1937–1947 (2002).

48 Shipp MA, Ross KN, Tamayo P. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised learning. *Nature Med.* 8(1), 68–74 (2002).

49 Robetorye RS, Bohling SD, Morgan JW, Fillmore GC, Lim MS, Elenitoba-Johnson KSJ. Microarray ananlysis of B-cell lymphoma cell lines with the t(14;18). *J. Mol. Diagn.* 4(3), 123–136 (2002).

50 Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B-cell lymphoma. *Proc. Natl Acad. Sci. USA* 100(17), 9991–9996 (2003).

51 Mah N, Thelin A, Lu T *et al.* A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol. Genomics* 16(3), 361–370 (2004).

52 Lossos IS, Czerwinski DK, Alizadeh AA *et al.* Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N. Engl. J. Med.* 350(18), 1828–1837 (2004).

•• **Validated six genes that may predict survival in diffuse large B-cell lymphoma using real-time PCR.**

53 Jones CD, Darnell KH, Warnke RA, Zehnder JL. Cyclin D1/Cyclin D3 ratio by real-time PCR improves specificity for the diagnosis of mantle cell lymphoma. *J. Mol. Diagn.* 6(2), 84–89 (2004).

54 Howe JG, Crouch J, Cooper D, Smith BR. Real-time quantitative reverse transcription-PCR for cyclin D1 mRNA in blood, marrow, and tissue specimens for diagnosis of mantle cell lymphoma. *Clin. Chem.* 50(1), 80–87 (2004).

55 Katayama Y, Sakai A, Okikawa Y *et al.* Cyclin D1 over expression is not a specific grouping marker, but may collaborate with CDC37 in myeloma cells. *Int. J. Oncol.* 25(3), 79–95 (2004).

56 Bennaceur-Griscelli A, Bosq J, Koscielny S *et al.* High level of glutathione-S-transferase pi expression in mantle cell lymphomas. *Clin. Cancer Res.* 10(9), 3029–3034 (2004).

57 Malec M, Söderqvist M, Sirsjö A *et al.* Real-time polymerase chain reaction determination of cytokine mRNA expression profiles in Hodgkin's lymphoma. *Heamatologica* 89(6), 679–685 (2004).

58 Lossos IS, Czerwinski DK, Wechser MA, Levy R. Optimization of quantitative real-time RT-PCR parameters for the study of lymphoid malignancies. *Leukemia* 17(4), 789–795 (2003).

59 Beillard E, Pallisgaard N, van der Velden VH *et al.* Evaluation of candidate control genes

for diagnosis and residual disease detection in leukemic patients using real-time quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) – a Europe against cancer program. *Leukemia* 17(12), 2474–2486 (2003).

60 Stentoft J, Pallisgaard N, Kjeldsen E, Holm MS, Nielsen JL, Hokland P. Kinetics of BCR–ABL fusion transcript levels in chronic myeloid leukaemia patients treated with STI571 measured by quantitative real-time polymerase chain reaction. *Eur. J. Haematol.* 67(5–6), 302–308 (2001).

61 Merx K, Muller MC, Kreil S *et al.* Early reduction of BCR-ABL mRNA transcript levels predicts cytogenetic response in chronic phase CML patients treated with imatinib after failure of interferon alpha. *Leukemia* 16(9), 1579–1583 (2002).

62 Rambaldi A, Lazzari M, Manzoni C *et al.* Monitoring of minimal residual disease after CHOP and rituximab in previosly untreated patients with follicular lymphoma. *Blood* 99(3), 856–862 (2002).

63 Yao R, Rich SA, Schneider E. Validation of sixteen leukemia and lymphoma cell lines as controls for molecular gene rearrangement assays. *Clin. Chem.* 48(8), 1344–1351 (2002).

64 Arber DA. Molecular diagnostics approach to non-Hodgkin's lymphoma. *J. Mol. Diagn.* 2(4), 178–190 (2000).

65 Vega F, Medeiros J. Chromosomal translocations involved in non-Hodgkin lymphomas. *Arch. Pathol. Lab. Med.* 127(9), 1148–1160 (2003).

66 van del Veldn VHJ, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJM. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* 17(6), 1013–1034 (2003).

67 Yashima A, Maesawa C, Uchiyama M *et al.* Quantitative assessment of contaminating tumor cells in autologous peripheral blood stem cells of B-cell non-Hodgkin lymphomas using immunoglobulin heavy chain gene allele-specific oligonucleotide real-time quantitative-polymerase chain reaction. *Leukemia Res.* 27(10), 925–934 (2003).

68 Donovan JW, Ladetto M, Zou G *et al.* Immunoglobulin heavy-chain consensus probes for real-time PCR quantification of residual disease in acute lymphoblastic leukemia. *Blood* 95(8), 2651–2658 (2000).

69 Uchiyama M, Maesawa C, Yashima A *et al.* Development of consensus fluorogenically labelled probes of the immunoglobulin heavy-chain gene for detecting minimal residual disease in B-cell non-Hodgkin lymphoma. *Cancer Sci.* 94(10), 877–885 (2003).

70 Lee WI, Cabanillas F, Lee MS. Quantitative assessment of disease involvement by follicular lymphoma using real-time polymerase chain reaction measurement of t(14;18)-carrying cells. *Int. J. Hematol.* 79(2), 152–156 (2004).

71 Liss B. Improved quantitative real-time RT-PCR for expression profiling of individual cells. *Nucleic Acids Res.* 30(17), E89 (2002).

72 Liss B, Franz O, Sewing S, Bruns R, Neuhoff H, Roeper J. Tuning pacemaker frequency of individual dopaminergic neurons by Kv4.3L and KCip3.1 transcription. *EMBO J.* 20(20), 5715–5724 (2001).

73 Chiang M, Melton DA. Single-cell transcript analysis pf pancreas development. *Dev. Cell* 4(3), 383–393 (2003).

74 Nutt SL, Vambrie S, Steinlein P *et al.* Independent regulation of the two Pax5 alleles during B-cell development. *Nature Genet.* 21(4), 390–395 (1999).

75 Velculscu VE, Madden SL, Zhang L *et al.* Analyis of human transcriptomes. *Nature Genet.* 23(4), 387–388 (1999).

76 Todd R, Margolin DH. Challenges of single-cell diagnostics: analysis of gene expression. *Trends Mol. Med.* 8(6), 254–257 (2002).

77 Peccoud J, Jacob C. Theoretical uncertainty of measurements using quantitative polymerase chain reaction. *Biophys. J.* 71(1), 101–108 (1996).

78 Al-Soud WA, Jonsson LJ, Rådström P. Identification and characterization of immunoglobulin G in blood as a major inhibitor of diagnostic PCR. *J. Clin. Microbiol.* 38 (1), 345–350 (2000).

79 Ramakers C, Ruijter JM, Deprez RH, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* 339(1), 62–66 (2003).

80 Liu W, Saint DA. Validation of a quantitative method for real time PCR kinetics. *Biochem. Biophys. Res. Commun.* 294(2), 347–353 (2002).

81 Burggraf S, Olgemöller B. Simple technique for internal control of real-time amplification assays. *Clin. Chem.* 50(5), 819–825 (2004).

82 Cronin M, Ghosh K, Sistare F, Quackenbush J, Vilker V, O'Connell C. Universal RNA reference materials for gene expression. *Clin. Chem.* 50(8), 1464–1471 (2004).

• **Meeting report that discusses the design and use of universal reference samples.**

83 Specht K, Richter T, Muller U, Walch A, Werner M, Hoefler H. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. *Am. J. Pathol.* 158(2), 419–429 (2001).

84 Godfrey TE, Kim SH, Chavira M *et al.* Quantitative mRNA expression analysis from formalin-fixed, paraffin-embedded tissues using 5' nuclease quantitative reverse transcription-polymerase chain reaction. *J. Mol. Diagn.* 2(2), 84–91 (2000).

85 Rutledge RG, Cote C. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Res.* 31(16), E93 (2003).

Websites

86 AlleLogic Biosciences Corp. www.allelogic.com (Viewed February 2005)

87 IT BioChem www.idahotech.com/itbiochem/ simpleprobes.html (Viewed February 2005)

88 Lymphoma Information Network www.lymphomainfo.net/nhl/classify.html (Viewed February 2005)

Affiliations

• *Anders Ståhlberg, MSc , Scientist, TATAA Biocenter, Medicinaregatan 7B, 413 90 Gothenburg, Sweden; Department of Chemistry & Bioscience, Chalmers University of Technology, Medicinaregatan 9E, 413 90 Gothenburg, Sweden Tel.: +46 317 733 916 Fax: +46 317 733 910 anders.stalberg@tataa.com*

• *Neven Zoric, MSc, Director, TATAA Biocenter, Medicinaregatan 7B, 413 90 Gothenburg, Sweden Tel.: +46 317 733 916 Fax: +46 317 733 910 neven.zoric@tataa.com*

• *Pierre Åman, PhD, Head of Research Group, Lundberg Laboratory for Cancer Research, Department of Pathology, Gothenburg University, Sahlgrenska University Hospital, Gula stråket 8, 413 45 Gothenburg, Sweden Tel.: +46 313 422 842 Fax: +46 31 828 733 pierre.aman@llcr.med.gUSe*

• *Mikael Kubista, PhD, Research Director, TATAA Biocenter, Medicinaregatan 7B, 413 90 Gothenburg, Sweden; Department of Chemistry & Bioscience, Chalmers University of Technology, Medicinaregatan 9E, 413 90 Gothenburg, Sweden Tel.: +46 317 733 926 Fax: +46 317 733 910 mikael.kubista@tataa.com*