Chapter 4
Quantitative Real-Time PCR Analysis of Degradome Gene Expression

Caroline J. Pennington, Robert K. Nuttall, Clara Sampieri-Ramirez, Matthew Wallard, Simon Pilgrim, and Dylan R. Edwards

Abstract Dissection of the contribution of proteases and inhibitors in the complex molecular events involved in cancer initiation, growth, and spread requires as a starting point detailed knowledge of the degradome genes that are expressed and dysregulated in cancer. This information identifies candidate genes for functional investigations and also reveals potential markers of disease progression and severity. Quantitative real-time polymerase chain reaction (qRT-PCR) analysis provides optimal sensitivity and specificity for analysis of RNA from human tumors and nonneoplastic tissues. In this chapter, we outline basic qRT-PCR technologies, and approaches for normalization and analysis of expression data. Degradation of RNA is a major problem for microarray analyses, but we demonstrate that TaqMan® qRT-PCR is a remarkably robust technique that can provide reliable information on archival specimens that would not be appropriate for other transcriptomic analyses. We also highlight the utility of low-density TaqMan arrays for degradome expression analysis.

Introduction

Quantitative real-time polymerase chain reaction (qRT-PCR) is one of the principal platform technologies of the genomic age that overcomes some of the main challenges associated with characterization and accurate quantification of protease expression in tissues and cell lines (Bustin et al. 2005). Unlike earlier methods of quantifying gene expression, such as Northern blotting, which requires 5–30 μg of RNA, qRT-PCR can accurately detect as little as 100 copies of target sequence in a 5 ng pool of reverse-transcribed complementary DNA (cDNA), equivalent to about 1 copy per cell (Nuttall et al. 2003). The technique can, therefore, be adapted when clinical tissue or cell samples are limited. As qRT-PCR simultaneously detects and quantifies the presence

C.J. Pennington
School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ UK,
e-mail: c.j.pennington@uea.ac.uk

D. Edwards et al. (eds.), *The Cancer Degradome*.

of a specific region of DNA during the early, efficient phase of PCR, results obtained are more accurate and reproducible than those obtained from semiquantitative PCR techniques. The relative ease of methodology has resulted in high throughput profiling studies of large gene families in extensive collections of samples (Wall and Edwards 2002, Morimoto et al. 2004, Porter et al. 2004, Overbergh et al. 2003).

Technologies

qRT-PCR follows the general pattern of polymerase chain reaction, that is, the exponential amplification of target DNA, but with the added advantage of quantification after each round of amplification; this is the "real-time" aspect of the process. The accumulation of data at each cycle of the PCR greatly increases the sensitivity of the reaction and negates the need for post-PCR image processing involved in semiquantitative end-point strategies such as competitive PCR and "primer-dropping" PCR (Wall and Edwards 2002). Quantification of up to 384 samples can be determined in the time it takes to run a 40-cycle PCR, which, depending on the instrument, can be as little as 35 min.

Quantification Strategies

Four quantification strategies are currently available for qRT-PCR amplification: SYBR® green, Taqman® probes (Applied Biosystems, Foster City, CA), and Molecular Beacons and Scorpions® (DxS Ltd., Manchester, UK). Each of these chemistries relies on the detection of a fluorescent signal. SYBR uses a fluorescent molecule in the reaction mix that emits little fluorescence when in solution but a strong signal when incorporated in double-strand DNA (dsDNA) during primer extension. Taqman probes, Molecular Beacons, and Scorpions depend on Förster resonance energy transfer (FRET), in which a fluorogenic dye molecule and a quencher are coupled on a gene-specific oligonucleotide. In both cases, the increase in fluorescence signal is proportional to the amount of product produced during each PCR cycle. Individual samples are quantified relative to each other by determining the cycle at which the signal rises above background fluorescence, termed the cycle threshold or C_t. The lower the C_t, the earlier the signal is detectable above the threshold, the more target is present. The ability to quantify the amount of template in a sample remains accurate over a wide (at least 6 log) dynamic range (Nuttall et al. 2003).

SYBR

SYBR intercalating dyes are perhaps the simplest and certainly the cheapest method of quantifying gene expression using qRT-PCR. Nonlabeled target-specific primers are generated following qRT-PCR primer design protocols (see below), and these are

combined with a PCR reaction mix that contains SYBR dye and a passive reference dye. The passive reference dye provides an internal reference to which the SYBR green signal can be normalized during data analysis, which is necessary to correct for fluorescent fluctuations caused by changes in concentration or volume. However, despite the apparent cost-effectiveness of using SYBR reactions one major disadvantage is the ability of SYBR dye to bind to any dsDNA including unspecific products and primer dimers, thus overestimating the amount of target in a sample. The presence of nonspecific amplification can be recognized by detecting amplification products in a "no template" control and by performing a post-run melting curve analysis to display dissociation curves for each target gene. If nonspecific products are amplified it will be necessary to carry out reaction optimization or redesign primers to overcome the problem. The need to run melt curves and conduct PCR optimization adds to the complexity of the analysis increasing the time and potential cost of the reaction. Another drawback of SYBR is that because multiple dye molecules bind to each product, longer amplicons will incorporate more dye molecules, resulting in a higher signal. More efficient reactions will also bind more SYBR molecules than would less-efficient reactions. Both these problems can be overcome by standardizing amplicon size and optimizing PCR efficiencies.

Probe-Based Chemistry

A major advantage of probe-based detection systems over SYBR is the added specificity of using a third gene-specific oligonucleotide or probe in the reaction. In this case, a dual-labeled fluorescent probe is positioned between the forward and reverse primers. These oligonucleotides are combined in a reaction mix that unlike the SYBR does not contain additional signaling dyes. A signal will only be generated if the probe itself hybridizes to its complementary target and fluoresces. Probes can be labeled with dyes of different wavelength emission spectra, so multiplexing qRT-PCR reactions is possible. It should be noted, however, that lack of detection of nonspecific products does not mean that the reaction is completely specific. Undetectable amplification of nontarget products will affect the efficiency of a PCR reaction and, consequently, the relative fold differences between individual samples. Assessing the efficiency of the reaction by analyzing a standard curve should indicate if this is a problem but generally the risk of amplifying and detecting additional products will be significantly reduced compared to SYBR reactions. An additional drawback of using probe-based technology in qRT-PCR is the cost of the individual probes that are required for each target sequence.

Three different probe chemistries are commonly available: Taqman probes, Molecular Beacons, and Scorpion probes. During PCR, a Taqman fluorogenic probe, consisting of an oligonucleotide labeled with a reporter and a quencher dye in close proximity, anneals specifically to the target sequence between the forward and reverse primers. When the probe is cleaved by the 5′ nuclease activity of the DNA polymerase during primer extension, the reporter dye is separated from the quencher dye, FRET no longer occurs, and a sequence-specific signal is generated. With each subsequent

cycle, additional reporter dye molecules are cleaved from their respective probes and the increasing fluorescence intensity is monitored. Minor groove binding (MGB) Taqman probes are a modification of Taqman probes. Because of the minor groove binding moiety, probes can be shorter than other Taqman probes; this is an advantage when designing probes for single nucleotide polymorphism (SNP) and allelic discrimination analysis or at GC-rich or other sequence regions for which larger probes are unsuitable. MGB probes have added target specificity but are more expensive than non-MGB Taqman probes.

Molecular Beacons, like Taqman probes, rely on FRET for detection and quantification of target sequences. Unlike Taqman probes, Molecular Beacons form stem-loop structures when not hybridized to target DNA with a reporter dye on one arm and quencher in close proximity on the other arm. When annealed to its complementary target strand, a conformational transition occurs, the stem structure opens, and the entire length of the probe anneals to the target. The reporter and quencher are spread apart and florescence is emitted. Unlike Taqman probes, Molecular Beacons are not hydrolyzed and can be denatured from the target to reform the stem-loop configuration and can then be reused in the next cycle of PCR.

Scorpion probes are described as unimolecular in that the probe is linked, via a nonamplifiable linker to one of the gene-specific primers. Following annealing and extension of the primer, the stem-loop of the attached probe disassociates, stretches out, and anneals to the target, again separating the reporter and quencher dyes and preventing FRET occurring.

The higher initial costs of probe-based assays is partially offset since careful design can result in little requirement of reaction optimization, but there is no doubt that probe-based assays are expensive. One possible compromise between high specificity and cost considerations is the recent introduction of the Roche Universal Probe Library (Roche Applied Science, Burgess Hill, UK). This is a library of short, 8–9 nucleotide (nt) probes which offer transcriptome-wide coverage for specific organisms. Locked nucleic acid (LNA) chemistry means these probes are highly specific for 8–9 nt complementary strands. Added specificity is achieved by designing 5′ and 3′ flanking primers. A full set of 165 Universal Probe Library probes will effectively provide a probe for every gene in a number of specified organisms (including human, mouse, rat). Although initially expensive the Universal Probe Library allows a researcher to design and buy only relatively cheap primers to match an existing "in-house" probe. This makes preliminary studies into potential target genes a much more financially feasible option.

Methodological Aspects of TaqMan qRT-PCR

Primer Design

As with standard end-point PCR, primer design is vital to the success of qRT-PCR. Several design programs are available, for example, Primer Express (Applied Biosystems, Warrington, UK), Primer 3 (Rozen and Skaletsky 2000), or HUSAR (DKFZ,

Heidelberg, Germany). All should allow the strict criteria for qRT-PCR amplicon design to be selected; for example, Taqman probes should have

- not more than 2 Gs or Cs in the last five bases.
- a gauss/call second content of 30–80%.
- an amplicon size range of 50–150 bp.
- a maximum amplicon melting temperature of 85°C.
- a primer length of 9–40 bp.
- a primer melting temperature of 58–60°C with a difference of less than 2°C between primers.
- a probe melting temperature that is 10°C higher than that of the primers.
- more call second than gauss in the probes sequence.
- no gauss on the 5′ end of the probe.
- no self-binding complementarity.

It is important that at least one primer, but preferably the probe, crosses an exon junction. The primers thus created would amplify cDNA but not genomic DNA. This is important since DNase treatment of RNA is rarely 100% efficient and genomic DNA amplification will clearly lead to erroneous results. Primer and probe sets should be checked for sequence specificity by BLAST analysis and by sequencing of the PCR products.

Roche provides an online design center for use with the Universal Probe Library to find the best primer set to flank one of the universal probes. A sequence or accession number can be input and a list of potential primers and corresponding probes are returned in order of suitability. Likewise, PrimerBank provides predesigned and validated primer sequences for SYBR reactions via their Web site (Wang and Seed 2003).

Designing primers and probes for a gene of interest is relatively quick and allows the targeting of specific splice variants that may be of interest and offers more control of the regions targeted. However, time can be saved by purchasing ready-made primer/probe sets or SYBR primer pairs from companies specializing in their manufacture, for example, Qiagen (Crawley, UK) and Applied Biosystems (Warrington, UK) both have a genome-wide stock of fully validated primers and probes. It should be noted, however, that "off-the-shelf" primer/probes are more expensive than custom-designed sets. Whichever option is chosen for primer design, a useful validation of downstream analysis is to use more than one primer/probe set. Nolan et al. (2006a) suggest a 5′:3′ assay for glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in which RNA integrity can be assessed. This assay uses primer sets designed for the 5′, mid, and 3′ region of GAPDH mRNA. Since these authors use oligo(dT) priming for reverse transcription, the ratio of amplicon in this assay indicates whether the reverse transcription reaction represents full-length cDNAs and thus intact starting RNA. This assay has the added benefit of providing a validation that a given gene is correctly amplified and quantified by a chosen primer set. Although this is an expensive option, if broad scale profiling is to be undertaken, it is a useful tool for genes that will be focused on more specifically. This is also an option for validating gene expression of rare transcripts where the sensitivity of the reaction may be borderline.

RNA Quality and Integrity

One of the prime benefits of using qRT-PCR to quantify gene expression is the sensitivity of the technique when using very small amounts of precious clinical specimens or cells, laser-captured tissue, or even formalin-fixed paraffin-embedded archived material. Each of these strategies imposes challenges on the extraction of high quality, intact RNA that is free of genomic DNA, nucleases, and other PCR contaminants. It is easy to argue that the quality of RNA is perhaps the most important determinant of the reproducibility and biological relevance of qRT-PCR data, and so it is important to recognize and remove samples that have degraded to such an extent that their use could result in a misinterpretation of results. Conversely, the meaningful statistical analysis of many tissue-based studies depends on the number of samples included. It is, therefore, equally important to maximize sample numbers and not reject samples that have some degradation but are still of adequate quality for qPCR.

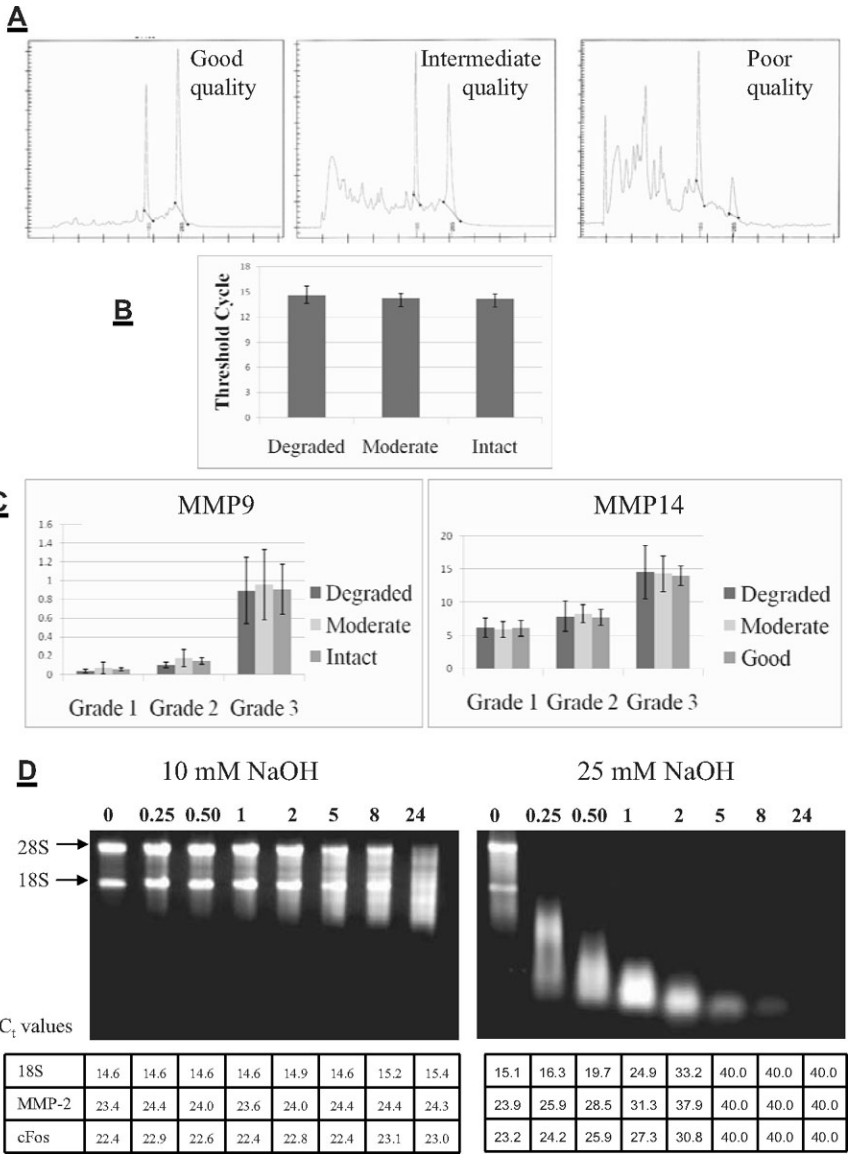
Since most of the actual RNA isolation procedure takes place in a strong denaturant that renders RNases inactive, it is typically before isolation when RNA integrity is most at risk. At this stage, the quality of RNA can be maximized by careful and prompt treatment of starting material; for example, samples should be snap frozen or stored in buffers such as RNAlater® (Ambion, Warrington, UK) immediately following collection and then stored correctly before RNA extraction (Schoor et al. 2003). During tissue homogenization for RNA isolation, it is vital that the denaturant be in contact with the cellular contents from the very moment that the cells are disrupted. This can be problematic when tissues are difficult to break down (e.g., bone) or when samples are numerous, making rapid processing difficult, consequently some optimization of technique and a pilot quality check may be necessary before large numbers of extractions are undertaken. Additionally, where possible, tissue sample collection and extraction protocols should be highly standardized since different protocols will have implications on the type of RNA, the amount, and quality of RNA extracted.

Generally, column-based kits for RNA isolation, for example, SV total RNA isolation kit (Promega Southampton, UK) or High Pure RNA isolation kits (Roche Burgess Hill, UK), are preferable to the more traditional phenol/chloroform extraction methods as these greatly reduce the potential for contamination of extracted RNA with salts, alcohols, and proteins, although it should be noted that small RNAs can remain in silica-based columns during the extraction procedure, and so kits specifically designed for the extraction of microRNAs, for example, mirVana™ (Ambion, Warrington, UK), should be used. The quality and quantity of RNA is initially assessed by measuring UV absorbance in a spectrophotometer. Nucleic acids absorb light at a wavelength of 260 nm, whereas organic compounds like phenol/trizol/rnazol, sugars, and alcohols absorb light at 230 nm and proteins at 280 nm (Sambrook et al. 2001). 260/280 and 260/230 ratios greater than 1.8 are generally indicative of good quality, uncontaminated RNA. We use the NanoDrop® ND-1000 (Nanodrop Technologies, Wilmington, DE), which is highly accurate and requires the use of just 1 μl of RNA that can be used without dilution or the need for cuvettes.

A 260/280 ratio of less than 1.8 can indicate protein contamination or possibly overdried RNA pellets that are not fully in solution. Freezing overnight at −80°C or briefly heating the RNA to 65°C can help rectify the latter problem. Contamination of RNA with residual extraction substances is more of a problem and can have wide ranging consequences for qRT-PCR results. A significant reduction in the sensitivity and kinetics of PCR assays is caused by inhibitory components; also different reactions may not be inhibited to the same degree and the effects can be compounded in absolute quantification where an external calibration curve is used to calculate the number of transcripts in the test samples. Various methods can be used to assess the presence of PCR inhibitors in samples with low 260:280/230 ratios. The PCR efficiency can be tested by the standard curve method, in which a single sample is serially diluted and the slope of the curve calculated. This is probably adequate if samples are not limited. Alternatively, the SPUD assay (Nolan et al. 2006b) can be used in which a potentially contaminated RNA sample is spiked with an uncontaminated high-quality control RNA that yields a defined C_t in an uninhibited reaction. PCR inhibitors will result in a higher than expected C_t value when the control RNA is amplified in the presence of contaminants in the test RNA.

A major problem that can be encountered when handling RNA for transcriptomic studies is RNA degradation. Degraded RNA can reduce the yield of reverse-transcribed cDNA and subsequent qPCR resulting in inaccurate representations of gene expression: this is particularly acute if oligo(dT) is used for RT priming. Traditionally, RNA quality has been evaluated by observation of ethidium-bromide stained bands of ribosomal RNA (rRNA) on nondenaturing agarose gels; this is still a good and cheap way of visualizing RNA quality and quantity although alone this would not be adequate for qRT-PCR. More recently, the introduction of the Agilent™ 2100 Bioanalyser (Agilent Technologies UK Ltd., Stockport, UK) and the BioRad Experion microfluidic capillary electrophoresis (BioRad, Hemel Hempstead, UK) systems have provided a more informative, qualitative, and quantitative assessment of RNA by calculating more precisely total RNA concentration based on the 28S:18S rRNA ratio. Additionally, an RNA integrity number (RIN) can be calculated and used to plot samples on a scale of quality and suitability for qRT-PCR (Schroeder et al. 2006).

However, our experience has been that some degradation of RNA for TaqMan qRT-PCR can be tolerated, allowing analysis of samples that would be unsuitable for microarray studies. The rationale for this is that amplicons for qRT-PCR are typically short (70–150 bp), and with random hexamer priming of RT reactions (see below), even partially degraded transcripts are successfully and reproducibly quantified. It is, however, vitally important to recognize and remove samples that have degraded to such an extent that their use could result in a misinterpretation of results. As an example, RNA samples from a bank of 169 archived urothelial carcinomas (UCCs) (Wallard et al. 2006) were assessed by Agilent 2100 Bioanalyser, and found to fall into approximately three equal groups representing "Poor," "Intermediate," and "Good" quality based on 28S:18S rRNA ratios (Fig. 4.1a). All samples were analyzed by TaqMan qRT-PCR for 18S rRNA and 30 protease genes. Although Agilent traces suggested that RNA samples classified as "Poor" quality



18S	14.6	14.6	14.6	14.6	14.9	14.6	15.2	15.4
MMP-2	23.4	24.4	24.0	23.6	24.0	24.4	24.4	24.3
cFos	22.4	22.9	22.6	22.4	22.8	22.4	23.1	23.0

15.1	16.3	19.7	24.9	33.2	40.0	40.0	40.0
23.9	25.9	28.5	31.3	37.9	40.0	40.0	40.0
23.2	24.2	25.9	27.3	30.8	40.0	40.0	40.0

Fig. 4.1 TaqMan® quantitative real-time polymerase chain reaction (*qRT-PCR*) allows quantitative analysis of partially degraded *RNAs*. (**a**) Agilent™ *RNA* traces: Representative traces obtained for samples from 169 urothelial carcinoma (*UCC*) samples representing intact (*left*), intermediate quality (*middle*) and poor quality *RNA* (*right*). (**b**) Comparison of threshold cycles (C_t) with Agilent total *RNA* analysis. The graph shows the median C_t value ± one standard deviation. (**c**) Relative expression of matrix metalloproteinase (*MMP*) transcripts in relation increasing tumor grades, grouped by Agilent total *RNA* trace quality. Data show the median expression ± one standard deviation for *MMP-9* (*left*) and *MMP-14* (*right*). (**d**) Agarose gel analysis of *RNA* degraded over time on incubation with 10 mM or 25 mM of NaOH and corresponding C_t values for 18S, *MMP-2*, and cFos are shown in the columns underneath

had undergone degradation, the majority of the 18S C_t values for such samples did not differ significantly from those for samples showing an intermediate or good trace (Fig. 4.1**a** and **b**). Conversely, the few samples with high C_t values for 18S *always* corresponded to poor quality or low yield Agilent traces. Exclusion of these 18 out of the initial 169 samples based on high 18S C_t (operationally we exclude any samples that are more than 1 C_t different from the median 18S C_t of the sample set) demonstrated that the remaining 151 samples showed similar patterns of expression of all degradome genes analyzed, regardless of their Agilent quality level. Figure 4.1**c** shows quantification of matrix metalloproteinase (MMP)-9 and MMP-14 in specimens of increasing histological grade, demonstrating that the patterns of expression in relation to tumor grade are evident, regardless of the RNA quality as assessed by Agilent. Agilent analysis would have excluded approximately two-thirds of the specimens, thus reducing the power of the study. These data show that small amplicon size – which is a necessary criterion of the qRT-PCR system – allows the use of moderately degraded samples that might be rejected if inclusion criteria relied on Agilent trace alone.

We extended this analysis by evaluating RNA that had been subjected to alkali degradation following isolation. Figure 4.1**d** shows that 18S and 28S bands were present on agarose gels following up to 24 h incubation with 10 mM NaOH, but these bands were no longer present after 15 min of incubation with 25 mM NaOH. C_t values for 18S rRNA remained stable, within 1 C_t of the control, at all time points during incubation with 10 mM NaOH. After 15 min of incubation with 25 mM NaOH, C_t values had increased by more than 1 C_t and continued to rise until there was no amplification of 18S 5 h after incubation with 25 mM NaOH ($C_t = 40$). Analysis of mRNA expression for MMP-2, a relatively stable mRNA with a long half-life (Overall et al. 1991), showed a similar pattern with C_t values rising by 2 as RNA degradation became detectable by raised 18S C_t and on agarose gels. Expression of cFos mRNA, an immediate early gene with a short half-life, also remained stable alongside TaqMan 18S values and amplification diminished in parallel with 18S as the RNA degraded. This again demonstrates that some RNA degradation is tolerated by qRT-PCR as long as amplicons are kept short and expression is normalized or analyzed alongside an endogenous control gene, a conclusion also reached by (Hamalainen et al. 2001).

Reverse Transcription

Once RNA has been extracted, checked for quality, and accurately quantified, it is necessary to reverse transcribe the RNA to generate cDNA. This can be performed as a one-tube, single-combined reverse transcription and PCR method or as a two-tube method with an initial reverse transcription followed by multiple qRT-PCRs using aliquots of cDNA. This section will focus solely on the more commonly used latter method.

Surprisingly, the relatively small stage of reverse transcribing RNA into cDNA can be an important contributor to variability and lack of reproducibility observed

in qRT-PCR (Stahlberg et al. 2004a, 2004b; Bustin et al. 2005). A major source of that variability can be introduced by poor pipetting practice. Also, interassay variability can be introduced because the efficiency of reverse transcription reactions depends somewhat on the relative abundance of transcripts (Karrer et al. 1995, Curry et al. 2002, Bustin and Nolan 2004). Rare transcript templates will not reverse transcribe as efficiently as more abundant transcripts and may be disproportionately affected by background nucleic acid contamination. This problem can be exacerbated by the reverse transcription priming strategy. Three priming strategies are in common use; oligo(dT), random hexamers, and gene-specific priming (GSP). The effectiveness of each of these may vary depending on the concentration of the target genes, the quality and configuration of RNA, and the number of genes the researcher may wish to profile.

Oligo(dT) primers used in reverse transcription target the poly A tail of mRNA and are used in ~40% of reported assays using qRT-PCR (Bustin et al. 2005). Since rRNA will represent 75–80% of total RNA extracted and this abundance can affect the reverse transcription efficiency of rare mRNA species, priming specifically to target mRNA has some advantage. However, oligo(dT) priming is less efficient at generating cDNA from RNAs with significant secondary structures that block elongation of the cDNA strand. Oligo(dT)-primed cDNA will not include targets without a poly A tail, for example, histones, viral RNAs, or rRNAs that may be required as endogenous controls or quality control genes (see later). As discussed in the previous section, oligo(dT) priming is also only possible for good quality, intact RNA since cDNA synthesis of fragmented RNA will fail to reach the qRT-PCR amplicon site if this is located toward the 5′ end of a long mRNA, resulting in false negatives at qRT-PCR.

An alternative to oligo(dT) priming used in ~30% of qRT-PCR assays is priming by random hexamers (Bustin et al. 2005). These are 6 bp oligonucleotides of a varying sequence that prime at multiple origins along all the RNAs in a sample. The drawback of using this priming strategy is that rare mRNA targets may not be primed proportionately due to competition for priming sites by the more abundant rRNA molecules. This may have implications for the accuracy of qRT-PCR quantification. Random priming has also been reported to overestimate mRNA copy numbers compared to a 22 base gene-specific primer (Zhang and Byrne 1999).

Gene-specific priming uses a unique antisense primer or the reverse primer of the subsequent qRT-PCR to target and reverse transcribe only a gene of interest. Because of this specificity, this strategy has been considered sensitive when analyzing rare transcripts. However, if a broader profiling of gene expression is required this strategy becomes both time consuming and financially prohibitive since a reverse transcription reaction would be necessary for every gene to be analyzed.

More recently, the use of random pentadecamer (15-mer) priming has been reported to be 40% more efficient than using random hexamers (Stangegaard et al. 2006). Also, Abgene (Epsom, UK) recommend a 3:1 mix of random hexamers to oligo(dT). We have recently compared qRT-PCR results of rare (MMP-8) and more abundant (MMP-1) protease genes and 18S rRNA primed with either oligo (dT), random hexamers, GSPs for each target, a random pentadecamer or a 3:1 mix

of random hexamers:oligo(dT). Gene-specific primers did not increase the sensitivity of the reverse transcription reaction for the rare MMP-8 transcript. There was no significant difference in quantification of targets using any of the priming strategies except that random hexamers allowed for the use of rRNA in qRT-PCR and was cheaper and quicker than using GSPs for each target. For ease and flexibility, random hexamers are probably the most versatile solution to reverse transcription priming, and are the basis of all of our published work.

A third source of variation at the reverse transcription stage of qRT-PCR can be introduced by varying reverse transcriptase enzymes used in reverse transcription reactions. This is particularly relevant if samples are reverse transcribed in separate laboratories and gene expression data are compared. Reverse transcriptases vary in efficiency and it is prudent to standardize protocols between laboratories or indeed experiments within the same laboratory that will be compared at a later date. Finally, it is important to include negative controls in any reverse transcription reaction. In this case, this should include a sample that has been through the reverse transcription reaction except that reverse transcriptase has been omitted and a reverse transcription in which no template RNA has been added. These controls will detect genomic DNA contamination of RNA and contamination of nucleic acids by other reagents or pipettes, tubes, and so on used during the preparation of the reaction. These controls should be included in subsequent qRT-PCR reactions.

Validation and Normalization

Ideally, an internal control used to normalize between samples should be constitutively expressed in all cell types at similar levels to the target gene and should remain constant, independent of disease status or experimental conditions. Historically, "housekeeping" genes, including GAPDH and β-actin, have been used as internal references, but their use has been largely discredited as expression of these genes can alter with varying cell culture conditions, hypoxia, in malignancy, and following treatment with tumor promoters (Hamalainen et al. 2001, Zhong and Simons 1999, Bhatia et al. 1994, Goldsworthy et al. 1993). In contrast, the expression of rRNA has been found to be relatively stable and has become a commonly used endogenous control in qPCR assays (Zhong and Simons 1999, Schmittgen and Zakrajsek 2000, Bhatia et al. 1994) Since rRNA and mRNA are generated by distinct polymerases, their levels are less likely to vary under conditions which would affect the expression of mRNA (Paule and White 2000). These features suggest that rRNA may be an appropriate gene for intersample standardization. However, concern has been expressed regarding possible imbalances between rRNA and mRNA fractions of different samples (Solanas et al. 2001), and the general unsuitability of rRNA as a normalizer for genes expressed at very different levels. Despite these hesitations, we believe that 18S rRNA is the most reliable method of assessing the quality of samples used in qRT-PCR. This is based on the fact that rRNA constitutes 75–80% of total RNA, so if the same amount of RNA of equal quality is used for the reverse transcription and if the same amount of cDNA is

used in qRT-PCR, amplification should be identical in all samples. This is operationally equivalent, therefore, to loading equal amounts of total RNA, as determined spectrophotometrically, on a gel for Northern blotting. Based on the analysis shown in Fig. 4.1 on the effects of RNA degradation on 18S rRNA C_t values and our experience, accumulated from analysis of thousands of clinical samples from diverse tissue origins, exclusion of samples that show more than 1 C_t variation from the median 18S C_t value is essential: when such samples are not removed these can lead to a distortion of expression profiles. We believe this level of quality control is central to all qPCR analysis and that 18S rRNA amplification efficiency serves as a quality control and possibly a potential normalizing agent.

The consensus regarding subsequent normalization of qRT-PCR data has been reached following years of publications advocating the use of one or another reference gene followed by contradictions and further propositions. The recent introduction of GeNorm^{TM} (Vandesompele et al. 2002) provides an Excel-based program for determining the most stable reference genes from a panel of potential endogenous control genes. Vandesompele et al. show that the common practice of using a single normalizing gene can lead to erroneous normalization. They suggest that an ideal, universal control gene probably does not exist and that normalizing to more than one endogenous control gene may be a more robust strategy. To determine the most appropriate normalizing genes, GeNorm analyzes a panel of 6 or 12 genes (ACTB, B2M, GAPD, HMBS, HPRT1, RPL13A, RPL32, RPS18, SDHA, TBP, UBC, YWHAZ) to find the optimal genes to use as endogenous controls in each unique experimental system. Primer Design (Southampton, UK) provides a panel of primers for use with SYBR primer or fluorescently labeled probe sets which can be used in conjunction with GeNorm, although other potential normalizes can be added to this panel if appropriate. GeNorm has become the gold standard for determining the number of identity of the most stable normalizing genes for qRT-PCR analysis and overcomes many of the uncertainties that existed before its introduction.

Standard Curve Versus Δ/Δ C_t Post-Run Analysis

qRT-PCR data can be analyzed using an absolute or relative standard curves method (Nuttall et al. 2004) or a comparative C_t method. Full details of both methods and examples are found in User Bulletin #2 produced by Applied Biosystems (Warrington, UK). It is ideal to include standard curves on every plate so that unknown samples that may fall within the less-sensitive region of amplification are recognized (as detailed earlier). Standard curves also show samples have been pipetted accurately, that the probe and primers are amplifying product in a meaningful way, and that the PCR is efficient (no contaminants etc.). The slope of the standard curve should be close to -3.2, which shows the PCR is 100% efficient. A slope with a value varying from this will indicate that each PCR cycle does not represent a doubling of product. The comparative method of data analysis is commonly used, but the strict rules regarding the efficiency of compared reaction and validation assays are not always adhered to. In brief, the absolute values of the slope of the log input amount versus the ΔC_t should be less than 0.1. See User Bulletin #2 for further clarification.

Degradome Expression in Cancer Cell Lines and Human Tumors

TaqMan qRT-PCR analyses of the MMPs and tissue inhibitor of metalloproteinases (TIMPs) have been reported in human cancer cell lines and gliomas (Nuttall et al. 2003), for the ADAMTS subfamily in breast cancers (Porter et al. 2004, 2006), for broader collections of serine and metalloproteinase and inhibitors in prostate cancer (Riddick et al. 2005), urothelial carcinoma (Wallard et al. 2006), and in comparisons of peripheral blood mononuclear cells and microglia (Nuttall et al. 2007). The developmental profiles of murine MMP and TIMP genes have also been presented (Young et al. 2002, Nuttall et al. 2004). One of the major issues in display of TaqMan data is the comparison of expression for different genes across the same sample set, since numerical values generated by the standard curve method discussed above are specific for particular primer–probe combinations, making intergene comparisons difficult. This can be overcome by the $\Delta/\Delta C_t$ method by reporting C_t differences between the genes of interest and 18S rRNA (*see*, for example, Jones et al. 2006). However, using the standard curve method and calibrating against synthetic RNA templates for a collection of MMP and TIMP genes (Young et al. 2002, Nuttall et al. 2003), we found that raw C_t values could be used for approximate intergene comparisons because no two genes showed more than a 5 C_t difference for equivalent absolute levels. Consequently, we developed a "tile-pattern" display that groups genes by 5 C_t ranges of C_t values, represented as very high expression ($C_t \leq 25$), high expression ($C_t = 26$–30), moderate ($C_t = 31$–35), low ($C_t = 36$–39), and not detected ($C_t = 40$). This has proved to be a valuable way to display comparative expression across degradome gene families, as shown for the analysis of MMPs and TIMPs in a panel of human mammary tissues as shown in Fig. 4.2. Another useful display is the box-and-whisker plot of $\Delta/\Delta C_t$ differences relative to 18S rRNA, though it needs to be emphasized that unless amplification efficiency is the same for each gene probe, comparison of levels of expression between different genes is only an approximation. We have also used heatmap displays of C_t data, which is useful in conjunction with hierarchical cluster analysis (Eisen et al. 1998).

We have recently developed a 384-gene TaqMan low-density array (TLDA) that includes the entire human metalloproteinase and serine proteinase families, along with their inhibitors and additional control genes. This TLDA has been used to analyze expression in a small collection of breast cancers and normal mammary tissue, revealing excellent fidelity and sensitivity compared to conventional TaqMan qRT-PCR. In addition to the confirmation of several MMP genes that are dysregulated (including MMP-1, −3, −10, −11, −12; ADAMTS-14), this analysis revealed for the first time several other genes whose expression is elevated in tumors compared to normal breast (*see* Chap. 30 by Pennington et al., this volume). These genes are being evaluated further using in silico transcriptomic data mining of resources such as Oncomine and In Silico Transcriptomics (*see* Chap. 31 by Iljin et al., this volume).

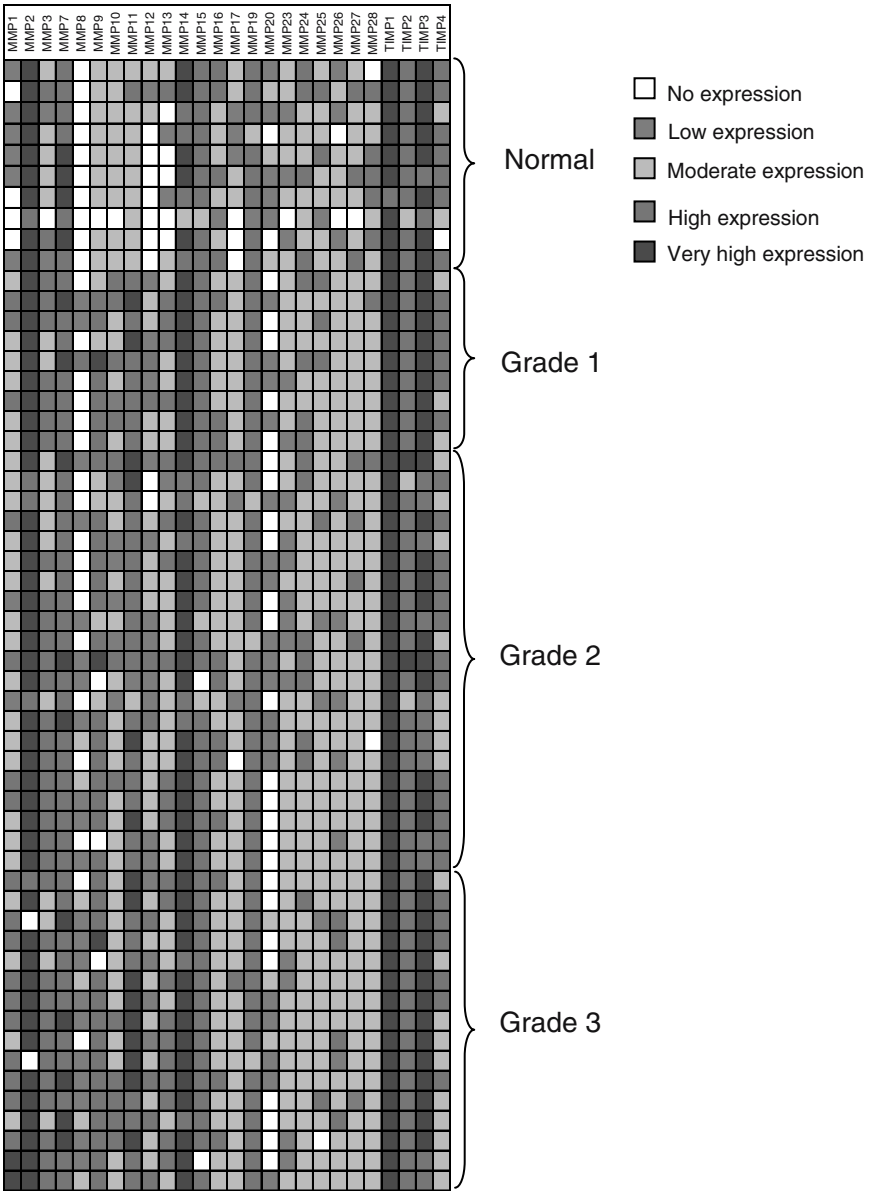


Fig. 4.2 Comparative TaqMan® expression profiles of matrix metalloproteinase (*MMP*) and tissue inhibitor of metalloproteinase (*TIMP*) gene families in normal and malignant human mammary tissues. Data are displayed as the expression levels of the indicated genes in human mammary tissue samples, which have been reported in detail elsewhere (Porter et al. 2004). The tissues are grouped as normal and histopathological grades 1, 2, and 3, respectively. Expression levels (raw C_t levels) are shown in the associated key. (*See also* Color Insert I)

Conclusions and Perspectives

qRT-PCR continues to have value in the analysis of the cancer degradome both as a tool for validation of microarray findings and for primary data generation, the latter, in particular, through the development of low-density arrays that provide comprehensive coverage of gene families. Cost is certainly a factor, but the sensitivity provided by TaqMan and the robustness of the technique, which can tolerate levels of RNA degradation that exclude samples from use on array platforms, make it applicable to analysis of archival material and laser-captured tissue specimens, without a requirement for RNA amplification (Pedersen et al. 2005). The universal probe library strategy also increases its versatility. The specificity of the technique makes it ideal for single nucleotide polymorphism analysis or quantification of alternatively spliced mRNA variants, which are beginning to be studied in detail for degradome genes. We have also applied the technique for parallel quantification of heterogeneous nuclear RNA (hnRNA) and mature mRNA in the same total RNA preparations, the hnRNA providing a surrogate of gene transcription that was shown to correlate well with transcription rates determined in nuclear run-on assays (C. Pennington, unpublished data).

Perhaps one of the major areas where TaqMan analysis will become increasingly valuable is in clinical diagnostics using expression signatures that identify patients at high- and low risk of recurrence, and in prediction of response to therapy (*see* also Chap. 30 by Pennington et al., this volume). The sensitivity, reliability, and speed of TaqMan make it attractive for analysis of small gene subsets, such as the Oncotype 21-gene and the 70-gene Amsterdam "MammaPrint®" signatures (Sotiriou and Piccart 2007). As will be discussed in Chap. 30, it is possible that these signatures could be refined further, potentially by including genes such as MMP-8 and ADAMTS-15 that have been shown to have prognostic value in breast cancer (Porter et al. 2006, Gutierrez-Fernandez et al., submitted), but which likely are expressed at levels that are too low for reliable quantification by microarray analysis.

References

Bhatia P., Taylor W. R., et al. (1994). Comparison of glyceraldehyde-3-phosphate dehydrogenase and 28S-ribosomal RNA gene expression as RNA loading controls for northern blot analysis of cell lines of varying malignant potential. Anal Biochem 216(1): 223–6.

Bustin S. A., and Nolan T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J Biomol Tech 15(3): 155–66.

Bustin S. A., Benes V., et al. (2005). Quantitative real-time RT-PCR—a perspective. J Mol Endocrinol 34(3): 597–601.

Curry J., McHale et al. (2002). Low efficiency of the Moloney murine leukemia virus reverse transcriptase during reverse transcription of rare t(8;21) fusion gene transcripts. Biotechniques 32(4): 768, 770, 772, 754–5.

Eisen M. B., Spellman P. T., et al. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95(25): 14863–8.

Goldsworthy S. M., Goldsworthy T. L., et al. (1993). Variation in expression of genes used for normalization of Northern blots after induction of cell proliferation. Cell Prolif 26(6): 511–8.
Hamalainen H. K., Tubman J. C. et al. (2001). Identification and validation of endogenous reference genes for expression profiling of T helper cell differentiation by quantitative real-time RT-PCR. Anal Biochem 299(1): 63–70.
Jones G. C., Corps A. N., et al. (2006). Expression profiling of metalloproteinases and tissue inhibitors of metalloproteinases in normal and degenerate human achilles tendon. Arthritis Rheum 54(3): 832–42.
Karrer E. E., Lincoln J. E., et al. (1995). In situ isolation of mRNA from individual plant cells: creation of cell-specific cDNA libraries. Proc Natl Acad Sci USA 92(9): 3814–8.
Morimoto A. M., Tan N., et al. (2004). Gene expression profiling of human colon xenograft tumors following treatment with SU11248, a multitargeted tyrosine kinase inhibitor. Oncogene 23(8): 1618–26.
Nolan T., Hands R. E., et al. (2006a). Quantification of mRNA using real-time RT-PCR. Nat Protoc 1(3): 1559–82.
Nolan T., Hands R. E., et al. (2006b). SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. Anal Biochem 351(2): 308–10.
Nuttall R. K., Pennington C. J., et al. (2003). Elevated membrane-type matrix metalloproteinases in gliomas revealed by profiling proteases and inhibitors in human cancer cells. Mol Cancer Res 1(5): 333–45.
Nuttall R. K., Sampieri C. L., et al. (2004). Expression analysis of the entire MMP and TIMP gene families during mouse tissue development. FEBS Letters 563: 129–134.
Nuttall R. K., Silva C., et al. (2007). Metalloproteinases are enriched in microglia compared with leukocytes and they regulate cytokine levels in activated microglia. Glia 55(5): 516–26.
Overall C. M., Wrana J. L., et al. (1991). Transcriptional and post-transcriptional regulation of 72-kDa gelatinase/type IV collagenase by transforming growth factor-beta 1 in human fibroblasts. Comparisons with collagenase and tissue inhibitor of matrix metalloproteinase gene expression. J Biol Chem 266(21): 14064–71.
Overbergh L., Giulietti A., et al. (2003). The use of real-time reverse transcriptase PCR for the quantification of cytokine gene expression. J Biomol Tech 14(1): 33–43.
Paule M. R., and White R. J. (2000). Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res 28(6): 1283–98.
Pedersen T. X., Pennington C. J., et al. (2005). Extracellular protease mRNAs are predominantly expressed in the stromal areas of microdissected mouse breast carcinomas. Carcinogenesis 26(7): 1233–40.
Porter S., Scott S. D., et al. (2004). Dysregulated expression of adamalysin-thrombospondin genes in human breast carcinoma. Clin Cancer Res 10(7): 2429–40.
Porter S., Span P. N., et al. (2006). ADAMTS8 and ADAMTS15 expression predicts survival in human breast carcinoma. Int J Cancer 118(5): 1241–7.
Riddick A. C. P., Shukla C. J., et al. (2005). Identification of degradome components associated with prostate cancer prostate cancer progression by expression analysis of human prostatic tissues. Br. J. Cancer 92: 2171–2180.
Rozen S., and Skaletsky H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365–86.
Sambrook J., MacCullum P., Russell D. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.
Schmittgen T. D., and Zakrajsek B. A. (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods 46(1–2): 69–81.
Schoor O., Weinschenk T., et al. (2003). Moderate degradation does not preclude microarray analysis of small amounts of RNA. Biotechniques 35(6): 1192–6, 1198–201.
Schroeder A., Mueller O., et al. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 7: 3.

Solanas M., Moral R., et al. (2001). Unsuitability of using ribosomal RNA as loading control for Northern blot analyses related to the imbalance between messenger and ribosomal RNA content in rat mammary tumors. Anal Biochem 288(1): 99–102.

Sotiriou C., and Piccart M. J. (2007). Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? Nat Rev Cancer 7(7): 545–53.

Stahlberg A., Hakansson J., et al. (2004a). Properties of the reverse transcription reaction in mRNA quantification. Clin Chem 50(3): 509–15.

Stahlberg A., Kubista M., et al. (2004b). Comparison of reverse transcriptases in gene expression analysis. Clin Chem 50(9): 1678–80.

Stangegaard M., Dufva I. H., et al. (2006). Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA. Biotechniques 40(5): 649–57.

Vandesompele J., De Preter K., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3(7): RESEARCH0034.

Wall S. J., and Edwards D. R. (2002). Quantitative reverse transcription-polymerase chain reaction (RT-PCR): a comparison of primer-dropping, competitive, and real-time RT-PCRs. Anal Biochem 300(2): 269–73.

Wallard M. J., Pennington C. J., et al. (2006). Comprehensive profiling and localisation of the matrix metalloproteinases in urothelial carcinoma. Br J Cancer 94(4): 569–77.

Wang X., and Seed B. (2003). A PCR primer bank for quantitative gene expression analysis. Nucleic Acids Res 31(24): e154.

Young D. A., Phillips B. W., et al. (2002). Identification of an initiator-like element essential for the expression of the tissue inhibitor of metalloproteinases-4 (Timp-4) gene. Biochem J 364 (Pt. 1): 89–99.

Zhang J., and Byrne C. D. (1999). Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR. Biochem J 337 (Pt. 2): 231–41.

Zhong H., and Simons J. W. (1999). Direct comparison of GAPDH, beta-actin, cyclophilin, and 28S rRNA as internal standards for quantifying RNA levels under hypoxia. Biochem Biophys Res Commun 259(3): 523–6.

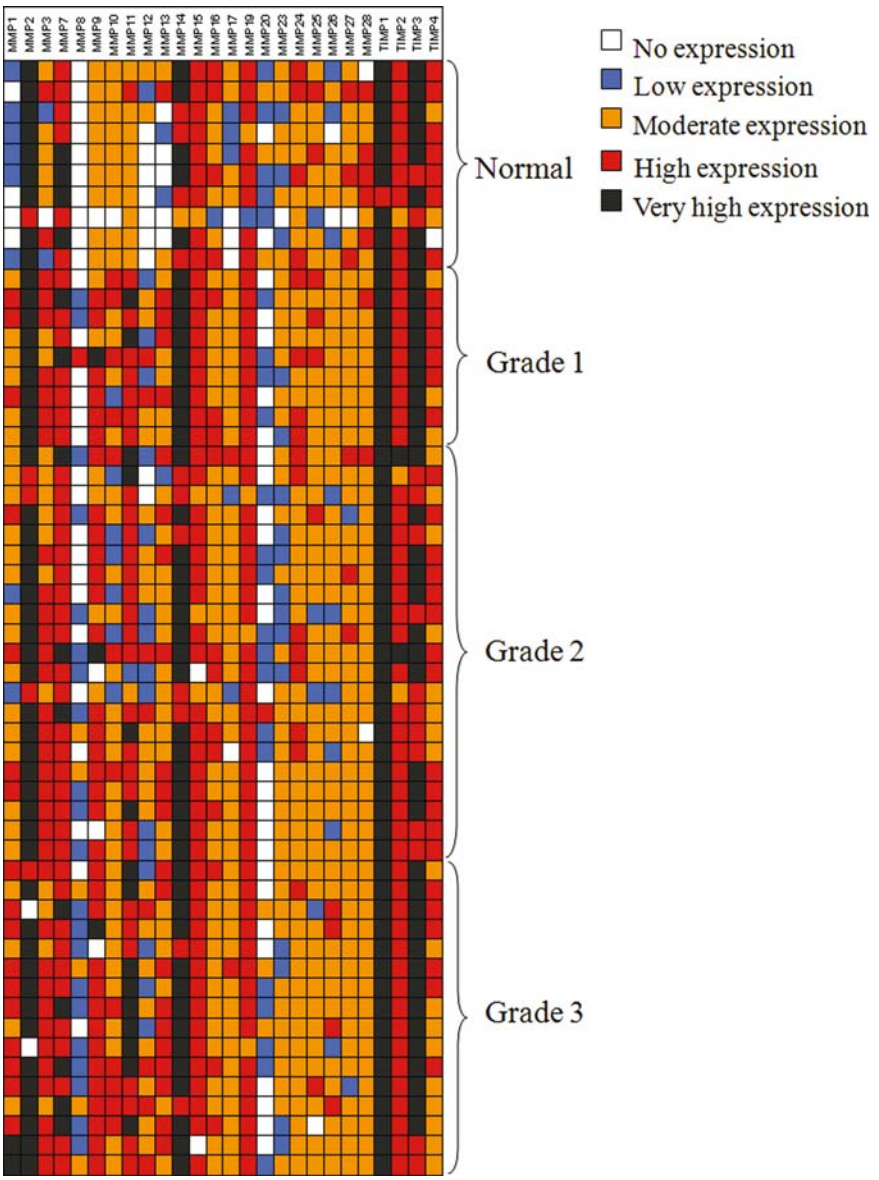


Fig. 4.2 Comparative TaqMan® expression profiles of matrix metalloproteinase (*MMP*) and tissue inhibitor of metalloproteinase (*TIMP*) gene families in normal and malignant human mammary tissues. Data are displayed as the expression levels of the indicated genes in human mammary tissue samples, which have been reported in detail elsewhere (Porter et al. 2004). The tissues are grouped as normal and histopathological grades 1, 2, and 3, respectively. Expression levels (raw C_t levels) are shown in the associated key.