Gene expression analysis

Real-time PCR and gene expression specificity

How to optimize your results and overcome specificity challenges in gene expression analysis

Introduction

Real-time PCR for gene expression analysis is typically used to quantify gene-specific RNA transcripts expressed by cells or viruses. Gene expression analysis is a particularly challenging application because results of gene expression experiments may not fall into recognizable patterns, and positive controls are not readily available. In addition, reverse transcription is vulnerable to inhibitors and performance problems. Predictions of gene expression based on protein data are unreliable, and the often-stated quantitative discrimination objective of two-fold requires a great deal of accuracy. Perhaps the most challenging and complex aspect of gene expression analysis is assay specificity. Fortunately, easily accessible solutions to these challenges exist for many species.

Common gene expression research applications using real-time PCR include:

- Cancer research
- Immuno-oncology
- Neurodegenerative and complex disease research
- Validation of results from other technologies (next-generation sequencing (NGS), microarray)
- Pathway analysis
- Biomarker signatures
- Expression quantitative trait loci (eQTL) analysis
- Medium- to high-throughput studies using the Applied Biosystems™ OpenArray™ platform

Additional resources:

- Real-Time PCR Learning Center: Gene expression analysis
- Getting Started Guide: Planning your gene expression experiment
- Infographic: SYBR Green I dye vs. TaqMan Assay chemistry

Specificity challenges for gene expression analysis

Specificity is the degree to which assay results include signal from the target and exclude nontarget contributions. Ideally, every assay would have 100% specificity for the target and no signal from nontarget sequences. In practice, it may be very difficult to completely exclude all possible nontarget contributions. Acceptance of nontarget signal varies with the application. In a gene expression experiment, a 1% nontarget contribution may be acceptable since it would not have a significant impact on the final analysis. On the other hand, signals from nontargets may be far less acceptable for critical research applications like potential pathogen or cancer detection.

Specificity is a top consideration for optimal assay performance in gene expression analysis. To achieve high specificity in an assay for your gene, select or design your assay to:

- Detect all known transcripts of your gene of interest
- Detect a unique splice variant
- Discriminate between closely related members of a gene family

Nontarget PCR products

PCR routinely amplifies nontarget products, because primers can bind to sequences that are not 100% complementary to them. The DNA polymerase can then extend the mismatched primers. This phenomenon, first described many years ago, is called primer mismatch extension. Nontarget PCR products can be divided into two types: nonspecific products and homologs.

Nonspecific PCR products

Nonspecific PCR products are those that do not have a sequence relationship with the target. These products are formed, seemingly randomly, from primers binding to nontarget sequences in samples. If a primer is oriented in a way that can produce a PCR product, the product will start to be amplified, and amplification will continue throughout the PCR process. The nonspecific sequences that primers initially bind to are expected to have many mismatches to the primer sequences, but the primer sequences become part of the nonspecific products and then get amplified with high efficiency. Another potential byproduct of PCR is primer-dimer formation, which results when two primers have sufficient complementarity to anneal to one another and then become a template for DNA polymerase.

Homologs

Homologs are genes with a sequence relationship to the target gene. They are thought to be the result of gene duplication that occurred at some point in the evolutionary past. Over time, the genes diverged in function and sequence. Homologs are very common in genomic DNA and cellular RNA, both within species and across species. Some genes can retain a high degree of homology despite many millions of years of evolution. For example, the Applied Biosystems™ TaqMan™ Assay for 18S ribosomal RNA (rRNA) takes advantage of sequence conservation so that the assay amplifies and detects the 18S rRNA gene from diverse eukaryotic species.

Some real-time PCR assays are designed to include specific homologs while excluding others (e.g., a pan-influenza A assay designed to detect all influenza A viruses and exclude influenza B and C viruses). Gene expression assays are usually intended to measure only target transcripts and exclude homologs. For example, the goal of a gene expression assay could be to measure transcripts from one member of the interleukin family of genes; other members of the interleukin gene family should be excluded from the results.

Homologs pose a particularly significant challenge to gene expression assay specificity, because they are very common in cellular RNA. Homologs may be easily amplified by the target primers and can produce PCR products of similar size and melting temperature as the target, and they can potentially be more abundant than the target.

Transcript abundance and specificity

When the target transcript is abundant in samples, assay specificity requirements are lower; when target transcript abundance is low, assay specificity requirements are much higher. Since the abundance of the target and homolog transcripts will likely not be known ahead of time, gene expression assays should have a high degree of specificity to be prepared for samples with low target and higher homolog abundances.

As an example, say an assay is 10,000 times more specific for the target than the closest homolog. In sample 1, the target and homolog cDNA are both present at 50,000 copies. Assay results from sample 1 would include 50,000 target copies plus the contribution of five homolog copies. The homolog is contributing only 0.01% to the result, which should not have a significant impact.

In sample 2, the target is present at one copy, and the homolog is present at 50,000 copies. In this case, results include one target copy plus five homolog copies. The result is derived primarily (~83%) from the homolog, which means the data are compromised.

qPCR detection chemistries

Given the specificity challenges in gene expression analysis, qPCR detection chemistry is a critical component of experimental design. Both probe-based and dye-based detection methods are widely used for gene expression analysis, including:

- Applied Biosystems™ TaqMan™ Assays (fluorogenic 5′ nuclease assays)
- Invitrogen™ SYBR™ Green I dye–based assays

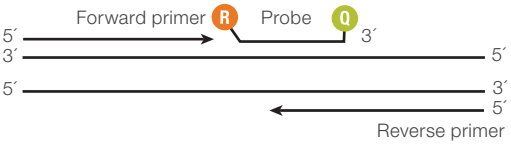
TaqMan Assays use sets of primers and fluorogenic probes to enable detection of specific PCR products as they accumulate during PCR cycles (Figure 1).

SYBR Green I dye binds to double-stranded DNA and enables detection of double-stranded PCR products as they accumulate during PCR cycles. It binds to all double-stranded DNA, including specific and nonspecific reaction products (Figure 1). DNA-bound SYBR Green I dye fluoresces with greater intensity than the free dye.

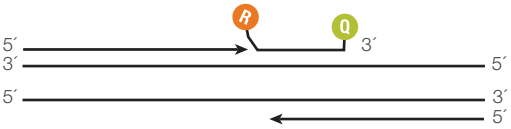
In the following sections, we discuss why TaqMan probe based chemistry is often preferred when specificity is a key concern, and why the detection step can help you achieve maximum assay specificity in your gene expression experiment.

TaqMan probe-based assay

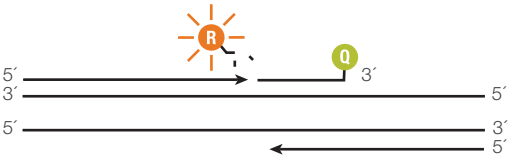
1. Polymerization: A fluorescent reporter (R) dye and a quencher (Q) are attached to the 5′ and 3′ ends of a TaqMan probe, respectively.



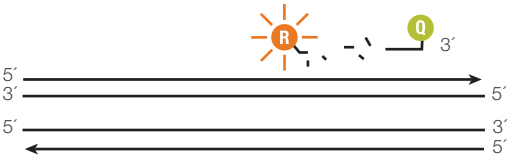
2. Strand displacement: When the probe is intact, the reporter dye's emission is quenched.



3. Cleavage: During each extension cycle, the DNA polymerase cleaves the reporter dye from the probe.



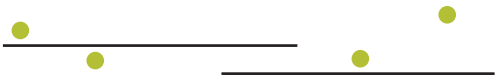
4. Polymerization completed: Once separated from the quencher, the reporter dye emits its characteristic fluorescence.



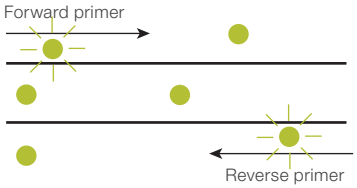
SYBR Green I dye assay

1. Reaction setup: SYBR Green I dye fluoresces when bound to double-stranded DNA.

2. Denaturation: When the DNA is denatured, SYBR Green I dye is released and fluorescence is drastically reduced.



3. Polymerization: In each extension cycle, primers anneal and PCR product is generated.



4. Polymerization completed: When polymerization is complete, SYBR Green I dye binds to the double-stranded product, resulting in a net increase in detected fluorescence.

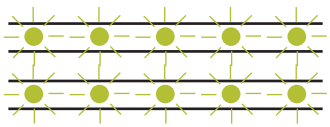


Figure 1. Comparison of TaqMan and SYBR Green I dye chemistries in real-time PCR.

Comparison of TaqMan probe-based and SYBR Green dye–based detection technologies.

	TaqMan probe-based detection	SYBR Green dye–based detection
Chemistry	Uses a fluorogenic probe specific to the target sequence to detect the target as it accumulates during PCR	Uses the dsDNA-binding SYBR Green I dye to detect PCR product as it accumulates during PCR
Specificity	Higher	Lower
Sensitivity (with low number of template copies)	1–10 copies	Variable*
Reproducibility	High	Medium*
Multiplexing	Yes	No
Predesigned assays	Yes	No, user must design and optimize primer sets
Typically requires user design and experimental optimization	No	Yes
Specificity in gene expression analysis	High	Low

* Depends on template quality and primer/design optimization.

Additional resources:

- Infographic: **SYBR Green I dye vs. TaqMan Assay chemistry**
- Real-Time PCR Learning Center: **TaqMan vs. SYBR chemistry for real-time PCR**

TaqMan Assay specificity in gene expression analysis experiments

PCR amplification products were first detected using gel electrophoresis, which enabled at least provisional identification of target amplification based on molecular size estimates of gel bands. The development of real-time PCR required a method to ensure that signals were target-specific without the need for size-based discrimination afforded by gels. The solution was a fluorescently labeled probe, termed the TaqMan probe (Figure 2). The probe is designed to bind to a target sequence located between the two PCR primers. Unlike the primers themselves, a TaqMan probe cannot be extended by DNA polymerase because its 3′-OH is blocked by a quencher moiety.

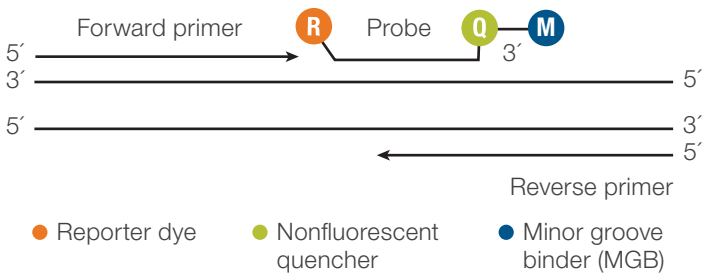


Figure 2. Illustration of a TaqMan probe.

Signal production by TaqMan probes

In an intact TaqMan probe, the fluorescence of the reporter dye is low because of the proximity of the quencher moiety. When the probe is cleaved during an extension cycle by the 5′-nuclease activity of *Taq* DNA polymerase, the reporter and quencher are separated from each other. The reporter fragment then fluoresces strongly, and the signal accumulates with each cycle as target amplification continues.

Multiple conditions must be met for probe cleavage to take place:

- The probe must be bound to a probe binding site with a melting temperature that will allow probe cleavage.
- A PCR primer must be bound a short distance upstream of the TaqMan probe.
- Active DNA synthesis must be occurring from the upstream PCR primer.

When these conditions are met to enable probe cleavage, the signal from a TaqMan probe in real-time PCR has a high degree of specificity. TaqMan probe-based chemistry produces target-specific signals even in the presence of nontarget amplification.

Exclusion of nonspecific products

Nonspecific PCR products contain nontarget sequences between the primers, so they will not harbor probe binding sites. Consequently, nonspecific products are automatically excluded from all TaqMan Assay results.

Exclusion of homologs

When designing a TaqMan probe-based assay, the probe should be positioned such that any sites in homologs that the probe might bind to will contain mismatches. Mismatches in the probe binding site will reduce the melting temperature of the probe bound to the homolog, thus reducing probe cleavage. The degree to which probe cleavage is reduced depends on multiple factors, such as the numbers, types, and locations of the mismatches as well as the probe type. Minor groove binder (MGB) probes are more sensitive to mismatches than non-MGB probes are. MGB probes are particularly useful for applications with high specificity demands like gene expression analysis.

We recommend the general use of TaqMan MGB probes for allelic discrimination assays, especially when conventional TaqMan probes exceed 30 nucleotides. A TaqMan MGB probe contains:

- A nonfluorescent quencher at the 3′ end. The real-time PCR instrument can measure the reporter dye contributions more precisely because the quencher does not fluoresce (Figure 3).
- An MGB at the 3′ end. An MGB increases the melting temperature (T_m) of a probe, allowing the use of shorter probes.

Consequently, TaqMan MGB probes exhibit greater differences in T_m between matched and mismatched probes, which help provide more accurate allelic discrimination.

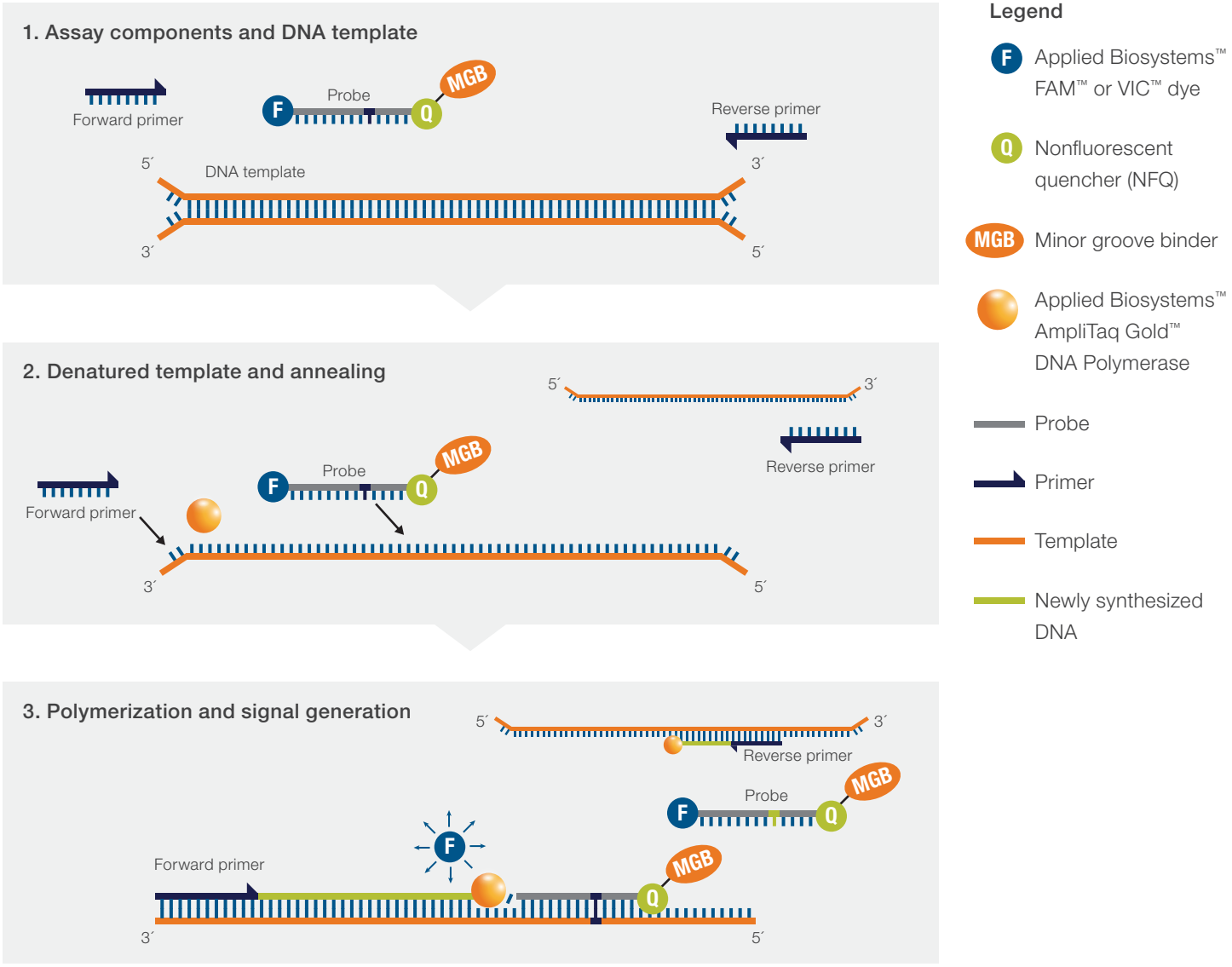


Figure 3. The process in a TaqMan Assay.

Measuring the specificity of TaqMan gene expression assays

Gene expression assay specificity should be measured prior to use to help ensure results will be target-specific with no significant nontarget contribution. As discussed previously, gene expression assay specificity can be expressed as a sensitivity bias in favor of the target vs. the most similar homolog expected to be encountered in samples.

The specificity of a TaqMan gene expression assay is measured using artificial DNA templates for the target and the most similar homologs known to exist in the designated species. An equal number of artificial DNA template molecules for the target and homologs are run in the real-time PCR assay being tested. If a threshold cycle (C_t) is determined for the homolog, the target's C_t is subtracted from it to produce a ΔC_t value. The specificity goal is a minimum ΔC_t of 15, which corresponds to a >30,000-fold specificity difference in favor of the target over the most similar homolog in that species. As an example, if the specificity of an assay is 30,000-fold and the homolog is 300 times more abundant than the target, the homolog contribution would be 1%.

After many specificity measurements, a database was constructed that could predict assay specificity for a given target in the presence of total RNA from a given species. This database is used by the **Custom TaqMan Assay Design Tool**, a web-based tool from Thermo Fisher Scientific that offers a bioinformatics option for a list of species. If the bioinformatics option is chosen by the user, the **Custom TaqMan Assay Design Tool** will not present an assay if none can be found that meets the minimum specificity requirements.

Each **TaqMan gene expression assay** has a unique assay ID number. The gene expression assay ID suffix indicates the assay placement. A small percentage of TaqMan gene expression assays have the suffix H, which indicates that due to extreme similarity between homologs, the minimum assay specificity is lower—at minimum, 1,000-fold lower.

Currently, over 2.8 million TaqMan gene expression assays are available for multiple species. TaqMan gene expression assays are designed, manufactured, tested, and distributed by Thermo Fisher. Search for predesigned TaqMan gene expression assays using the **TaqMan Assay Search Wizard**, or design your own with tools and resources available at the **Assay Design Hub**.

Additional resources:

- Web page: **Gene expression analysis using TaqMan Assays**
- Web page: **Real-time PCR master mixes for gene expression analysis**

Using SYBR Green I dye in gene expression analysis

SYBR Green I dye is often used for gene expression experiments. SYBR Green I dye–based assays typically consist of two PCR primers with no probe. During PCR amplification, SYBR Green I dye binds to double-stranded amplicons, increasing the fluorescence signal. However, SYBR Green I dye may also bind to non-amplicon double-stranded DNA like genomic DNA, which can affect the baseline. As such, SYBR Green I dye–based assays can be more problematic, putting data accuracy at risk.

Considerations when using SYBR Green I dye for gene expression analysis

SYBR Green I dye binds to all double-stranded DNA molecules, including target amplicons, nontarget amplicons, and non-amplicon DNA (e.g., genomic DNA). Regardless of how many different PCR products are being amplified, only one amplification plot will be displayed for SYBR Green I dye fluorescence, which may give the impression that only one PCR product is being amplified. In addition, the shape of the amplification plot cannot be used to identify a target because the shape will usually look the same whether the target, nontarget, or a mixture of PCR products are being amplified.

Given the potential of PCR to amplify nontarget products, there is risk in assuming that only the target is being amplified without any supporting evidence. A TaqMan probe helps provide confidence in specificity, but SYBR Green I assays lack such assurance, so assurance of target-specific signal must be obtained in other ways. While gene expression research applications can be less demanding than some other applications, specificity remains important. Therefore, using a low-specificity detection tool like SYBR Green I dye for this application may cause some concerns.

How to measure SYBR Green I assay specificity

For SYBR Green I assays, measuring specificity involves the use of target-negative samples. Target-negative samples are samples that contain all the genetic complexity of samples that will be used in experiments but lack the target sequence. Target-negative samples must contain all homologs that would be found in experimental samples. Specificity measurements involving only pure target and pure homolog, as done with TaqMan Assays, would not suffice for SYBR Green I–based assays, because pure targets and homologs would not represent all products that could potentially form.

The Applied Biosystems™ MycoSEQ™ Mycoplasma Detection System detects mycoplasmas in cell culture and illustrates SYBR Green I assay specificity measurement for pathogen detection. In this case, the target-negative samples are uninfected cells. If no amplification is observed using the SYBR Green I assay for mycoplasmas with 30,000 copies of DNA from uninfected cells, specificity could be described as >30,000-fold in favor of the target versus one copy of cellular DNA.

Obtaining target-negative samples for gene expression analysis is likely to be more difficult. Genetic research techniques exist to delete specific genes in cells or organisms, but this takes time and expertise. In addition, deleting certain genes may severely impact the health of cells or organisms.

Alternatives to specificity measurements for SYBR Green I assays using target-negative samples

Creating target-negative samples to measure the specificity of SYBR Green I dye–based gene expression assays can be difficult and time-consuming. Consequently, very few researchers create them. Instead, users of SYBR Green I assays sometimes utilize other techniques to improve confidence in the specificity of their results.

1. Primer BLAST analysis

The Basic Local Alignment Search Tool (BLAST™) is a nucleic acid analysis program offered on the National Center for Biotechnology Information website. BLAST analysis can identify similarities between input sequences and sequences found in databases. BLAST analysis is widely considered to be a helpful bioinformatics tool for designing PCR primers. However, BLAST analysis does not perform a specificity measurement. Examples from researchers using BLAST analysis on PCR primers have demonstrated that primers can amplify nontarget sequences not predicted by BLAST analysis. Therefore, BLAST analysis results should not be interpreted as a guarantee that a PCR assay will have the necessary degree of specificity.

2. Melt curve analysis for SYBR Green I assays

A dissociation or melt protocol is often used post-PCR for SYBR Green I dye–based assays. In the melt protocol, the temperature of the block is gradually increased, causing double-stranded DNA to dissociate into single-stranded DNA. Conversion to single-stranded DNA results in a decrease in the SYBR Green I fluorescence signal. Real-time PCR software plots the melt data as a negative first derivative, producing one or more melt peaks. The purpose of melt analysis for SYBR Green I assays is to help ascertain whether the target alone was amplified.

Melt curve analysis is not relevant for TaqMan Assays. It is done for SYBR Green I assays to help compensate for the lack of specificity provided by the TaqMan probe.

Analysis of melt curves requires some fundamental understanding of PCR amplification. For a SYBR Green I melt peak to be consistent with amplification of a single PCR species, there must be one narrow, symmetrical peak that aligns with melt peaks of other samples in the same assay. If melt peak data do not conform to these requirements, the data from those sample wells should be considered compromised in specificity.

Does an apparently anomaly-free melt peak mean with certainty that the target alone was amplified? The answer is no. Here are some reasons why:

- Unlike DNA bands in gel electrophoresis, the sizes of which can be predicted, the melting temperature of a PCR amplicon cannot be predicted with accuracy in a SYBR Green I assay. Melt curve analysis of SYBR Green I assays is not a target identification method in and of itself. Other methods must be used to identify a melt peak in a SYBR Green I assay.
- Melt curve analysis of SYBR Green I assays is low in resolution. Examples from users demonstrate that multiple amplification products can still form one melt peak. For example, running a gel of the PCR products from a SYBR Green I assay can yield multiple bands despite the appearance of one melt peak (Figure 4).

Thermo Fisher supports high-resolution melt (HRM) analysis, but HRM requires the use of a different type of DNA-binding dye that binds to DNA very differently from SYBR Green I dye.

3. Gel electrophoresis of SYBR Green I dye–based PCR reactions

Acknowledging the limitations of melt curve analysis for SYBR Green I assays, some users will perform gel electrophoresis analysis of their SYBR Green I dye–based reactions and evaluate specificity based on the molecular sizes of the PCR products. However, there are many limitations to this method.

- Gel electrophoresis analysis provides limited endpoint information and does not provide a specificity measurement.
- Agarose gels are more often used than acrylamide gels to analyze SYBR Green I–based PCR products, but most amplicons in real-time PCR assays are very short, and agarose gels have poorer resolution, so small amplicons may appear fuzzy or diffuse in the gel. They may also be difficult to distinguish from primer-dimers.
- Agarose gels are often stained with ethidium bromide, which has a 10-fold lower signal-to-noise ratio than that of SYBR Green I dye. Nontarget DNA molecules may not be visible in ethidium bromide–stained gels, but those molecules could be significant contributors in the PCR reactions.
- A collection of DNA molecules of varying sizes can produce “smears” in gels. It may be difficult to see or assess the impact of such smears in a SYBR Green I–based experiment.
- Homologs can generate amplicons similar in size to the target and can be indistinguishable from the target in an agarose gel.

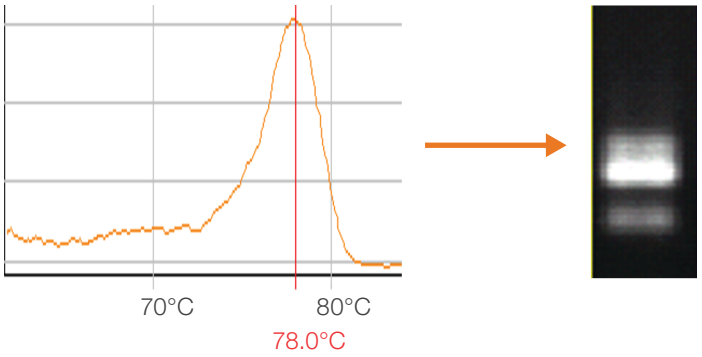


Figure 4. Melt curve analysis of a SYBR Green I–based gene expression assay intended to amplify one specific gene target. Gel electrophoresis of the real-time PCR reaction yielded three distinct bands, indicating at least three different PCR amplification products.

4. Sequencing of DNA products from SYBR Green I dye–based PCR

Another technique used to identify melt peaks from SYBR Green I assays is DNA sequencing, but sequencing also has significant limitations.

- Ideally, PCR primers should be designed with mismatches to homolog sequences to minimize homolog amplification. However, these primer binding site mismatches will never be observed in amplicons, even when they are derived from homologs, because the PCR primers are incorporated into the amplicons. Mismatches would have to exist between the primer binding sites in order to detect, by sequencing, the origin of the amplicon.
- Depending on the type of sequencing, minority sequences may go unrecognized in a mixed population of PCR products, yet all of those products contribute to results in SYBR Green I–based assays.
- Sequencing can add significant cost, which makes sequencing every sample impractical. Consequently, sequencing only a small number of reactions offers very limited evidence of specificity.

5. Comparing gene expression to protein data

Some researchers may feel that if gene expression results follow those of protein expression, this is proof that the gene expression results are accurate. For example, if the researcher has evidence that protein expression increased from control to treated states, and gene expression data also show an increase, the researcher may interpret the correlation as strong evidence that the gene expression experiment was successful. However, this is not always true, for multiple reasons.

- If gene expression is divided into two categories, increase or decrease, randomness will result in a 50% chance of meeting expectations. This high probability of random correlation means such a correlation has no value.
- Eukaryotic gene expression involves many regulatory mechanisms, such as splicing, RNA transport, protein processing, and more. Transcript concentration and protein concentration may not correlate. For example, a decrease in protein from control to the treated state could be the result of increased microRNA and not a lower transcript concentration.
- Most protein-encoding eukaryotic genes produce many transcripts, which vary due to differential splicing. Each gene expression assay for a target gene usually detects only a subset of those transcripts. Typically, only one assay per target gene is used in gene expression experiments, so important transcript information may be missing. For example, the transcript likely to have the greatest impact on gene expression is the one at the highest concentration, but the assay being used may not detect it. This problem can be mitigated by using multiple gene expression assays designed for different locations on the same gene (e.g., exon–exon junction sites), so that a higher percentage of the transcripts produced from that gene can be detected.

Additional resources:

- White paper: **Design and optimization of SYBR Green assays for qPCR measurement of relative gene expression**
- Web page: **Real-time PCR master mixes for gene expression analysis**
- Web page: **SYBR Green real-time PCR master mixes**

Conclusion

Given the specificity challenges inherent to real-time PCR gene expression analysis, the choice of detection chemistry is a critical factor to help ensure accurate results. SYBR Green I dye–based gene expression analysis, while popular, offers less specificity than analysis with TaqMan probes. This is because SYBR Green I dye binds to all double-stranded DNA, leading to potential detection of nonspecific products. While techniques such as melt curve analysis and gel electrophoresis provide some insights, they have significant limitations and do not guarantee specificity. In contrast, TaqMan assays, which use fluorogenic probes specific to the target sequences, offer significantly higher specificity. This probe-based detection method helps ensure that only the intended PCR products are detected, enhancing the reliability and reproducibility of gene expression data.

Learn more at
thermofisher.com/gene-expression-specificity

WTP-9566476 0225