Gene Expression

|  |
| --- |
| * [***Introduction***](https://drupal.unbc.ca/genetics/training/gene-expression#intro)
* [***Reverse Transcription: Converting RNA to cDNA***](https://drupal.unbc.ca/genetics/training/gene-expression#rnatocdna)
* [***Real Time PCR***](https://drupal.unbc.ca/genetics/training/gene-expression#rtpcr)
	+ [***Fluorescence detection chemistries:***](https://drupal.unbc.ca/genetics/training/gene-expression#fluor)
1. [***Fluorescence Resonance Energy Transfer probes***](https://drupal.unbc.ca/genetics/training/gene-expression#FRET)
2. [***SYBR® Green***](https://drupal.unbc.ca/genetics/training/gene-expression#sybr)

 [***Relative Quantification:***](https://drupal.unbc.ca/genetics/training/gene-expression#relative)1. [***Standard Curve Method***](https://drupal.unbc.ca/genetics/training/gene-expression#standard)
2. [***ΔCt Method***](https://drupal.unbc.ca/genetics/training/gene-expression#ct)

 [***Endogenous Control***](https://drupal.unbc.ca/genetics/training/gene-expression#end) [***Preparing a Reaction Plate***](https://drupal.unbc.ca/genetics/training/gene-expression#prep) [***Genome GeXP***](https://drupal.unbc.ca/genetics/training/gene-expression#GeXp) |

**Introduction**

The exciting prospect of gene expression moves above the realm of DNA and into the larger world of proteins and epigenetics. We do this by interrogating the transitional messenger RNA (mRNA) biochemistry that translates DNA sequence into proteins. This directional flow of information, from DNA -> mRNA -> protein, is known as the central dogma of molecular biology.

Quantifying gene expression through mRNA levels is one of the most active and growing fields of research in genetics. Historically, northern blots (or nuclease protection assays) requiring several micrograms of total RNA were used to qualitatively probe levels of mRNA expression using hybridization, radioactive labeling, and scoring of relative band intensities.

UNBC's genetics facility houses two pieces of equipment to study gene expression. Our Applied Biosystem 7300 real-time PCR machine and Beckman Genome GeXP systems allow researchers to simultaneously examine two to forty genes for relative expression.

The following flow chart outlines a path that a researcher would follow and the stages to consider when designing and executing a gene expression study. This manual on gene expression gives an overview with methods required to execute each of the stages. The reader should first read through sections on RNA and DNA extraction where the starting topic of tissue storage and collection is covered.

  

Once tissues have been collected, the process of a gene expression study commences by isolating the nucleotide sequence from the genes of interest.Published gene sequences are deposited into the[***National Center for Biotechnology Information (NCBI) website***](http://www.ncbi.nlm.nih.gov/). You can search to see if your genes of interest have been identified, sequenced, and deposited for retrieval on the NCBI website. Either the genes of interest have been previously sequenced in the organism of study, or a homolog is available in a close relative. A homolog is a gene that has a similar function, location, and sequence identity in different species by way of common ancestry. If your genes of interest are not on the NCBI website you need to obtain sequence reads of your own.

Some sleuthing is required to obtain novel gene sequences by turning to the [***Tree of Life Web Project***](http://www.tolweb.org/tree/)(ToL). If you are fortunate, primers can be borrowed from homologous genes that may have already been sequenced in close relatives of your target organism. In some instances the ToL website doesn't give a complete phylogenetic picture and more research is needed for your particular taxonomic group of interest. However, even distant homologs generally contain enough conserved sequence that primers can be designed to effectively target the gene for DNA sequencing. Use a few test subjects to extract DNA, PCR amplify your target genes, and run a DNA sequence analysis for your genes of interest. Our facility uses the ABI Primer Express software to design effective sequencing, real-time PCR, and reverse transcription primers. The Gene eXpress profiler uses gene sequence data to effectively design chimeric primers for a GeXP experiment.

**Reverse Transcription: Converting RNA to cDNA**

To circumvent the short lifespan of RNA a reverse transcriptase enzyme is used to transcribe total messenger RNA (mRNA) into a more stable working stock of copy DNA (cDNA). The reverse transcriptase enzyme is primed by random haxamers or 15-mer oligo-nucleotides, oligo d(T) nucleotides, a mixture of the two, or gene specific primers. Surprisingly, cDNA can be generated in the absence of these primers, suggesting RNA molecules or nucleotides is enough to prime the enzyme, but with decreased efficiency (Stahlberg et al. 2004). Random hexamers, 15-mer, or target specific primers are required for prokaryote transcription. The oligo d(T) nucleotides are used to target mRNA in nucleated eukaryotes that are polyadenylated at the 3' end with a tail of around 200 adenosine residues before exiting the nucleus (Proudfood et al. 2002).

While most mRNA transcription kits claim 100% efficiency there is tremendous variability in the protocols, kits, genes, primers, and researchers performing the lab work (Stahlberg et al. 2004;[***NARG study 2006***](http://www.abrf.org/ResearchGroups/NucleicAcids/Studies/NARG06_DSG2_8.pdf)). Hence, reverse transcription (RT) is inherently the most difficult stage in the process of quantifying gene expression. A creative solution to the inherent inconsistency in the RT procedure uses an RNA competitor (Zentilin and Giacca 2007). A simple procedure for constructing an RNA (or DNA) competitor is shown in the image below (adopted from Zentlin and Giacca 2007).



The final RNA competitor template is quantified, included in the RT to cDNA phase of the experiment, and co-amplified with the target sequence during real-time PCR or GeXP PCR stages. Such competitor templates are considered advantageous over heterologous competitor molecules because their similarity to the target strand serves to control for differences in amplification efficiencies that may arise in dissimilar strands. A similar procedure is used to generate a shorter competitor template using only three primers. Moreover, the DNA segment that is generated in this procedure can also be incorporated into the PCR procedures using serial dilutions for greater precision and accuracy in the quantification process using competitive PCR procedures. For added utility, restriction enzyme sites or probe specific sequence can also be incorporated into the inserts (Zentilin and Giacca 2007).

**Real-Time PCR**

Quantitative real-time PCR pioneered by ***[Higushi et al. (1993)](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&list_uids=7764001&dopt=Abstract" \o "Higushi et al. pioneering paper." \t "_blank)***, uses a fluorescent marker to signal an increasing amount of amplicon during PCR. The acronym RT-PCR is confusing because it can refer to either reverse transcription or real-time PCR. Reverse transcription is the name of the method used to convert messenger RNA (mRNA) into copy DNA (cDNA) with reverse transcriptase. Reverse transcription is a preliminary step in quantitiative PCR methods, hence it is better to refer to '*real-time quantitative reverse transcription-polymerase chain reaction*', or qRT-PCR for short.

Once mRNA is exctracted and converted to cDNA a real-time PCR reaction is performed with either SYBR green or TaqMan fluorescence. By now it is clear that the preceding steps are problematic and give inaccurate readings on RNA concentrations. In contrast, the real-time PCR method is precise and repeatable. It is strongly recommended that you become familiar with the protocol suggestions in Nolan et al. (2006) and Stahlbert et al (2004) when performing qRT-PCR methods. Moreover, Edwards et al. (2004) is available to UNBC students and faculty through our genetics facility if you require additional information when preparing for your research.

**Fluorescence detection chemistries**

Real-time fluorescent detection chemistries are sub-divided into two groups, a) specific or b) non-specific binding to the amplicon.  I discuss the Fluorescence ([***or Förster***](http://en.wikipedia.org/wiki/Fluorescence_resonance_energy_transfer)) Resonance Energy Transfer (FRET) probes as an example of specific binding probes and the SYBR® Green dye chemistry as an example of the non-specific binding fluorescence probe; for additional examples see Edwards et al. (2004).

**a. Fluorescence Resonance Energy Transfer probes**

FRET probes hybridize to an internal region of the PCR amplicon during the annealing stage. The taqman probe is most routinely used in our facility. The probe is a linear oligonucleotide with a fluorogenic dye attached to the 5' end and a quencher molecule attached to the 3' end. During the extension phase the reporter and quencher are separated by exonuclease activity of the Taq polymerase. Hence, the reporter signal is proportional to the amount of amplicon produced. While the probes are binding specific, like all PCR reactions they often prime and result in the co-amplification of non-specific targets. Other examples of FRET probes vary in their abilities or uses for different purposes and this is an area of active investigation (e.g.Yao et al. 2006).

**b. SYBR® Green**

The SYBR Green dye attaches to any double stranded DNA and its signal increases as more double stranded amplicon is produced. Due to its non-specific annealing, the SYBR Green dye is subject to produce false quantification readings should there be any contamination or non-targeted priming. However, the SYBR Green dye is less costly and it is has highly useful application in melt-curve analysis (described below) and strain identification (see Varga and James 2006).

**Relative Quantification**

Applied Biosystems gives an overview on [***"The essentials of real-time PCR."***](http://docs.appliedbiosystems.com/pebiodocs/04371089.pdf)  The document differentiates between absolute and relative quantification methods. The former method is distinguished from the later method by reference to a standard curve versus reference to another reference sample. This distinction is misleading and should not be used. Absolute quantification is a myth, because all readings are related to some reference material.

Key to understanding real-time PCR  is the exponential growth of amplicon during the PCR cycling process. An assumption is that the PCR reaction is running at 100% efficient, meaning that there is a complete doubling of amplicon during the geometric phase of amplification. Figure 1 shows an image taken from our ABI 7300 RT-PCR system with five samples at different concentrations but run in duplicate. It is immediately clear that the duplicate pairs give consistent results as they pass through the threshold point (Fig. 1a) and this is generally true of RT-PCR reactions. The threshold (Fig. 1a) is selected either manually or set automatically by the software. It is not critical where the threshold lies, so long as it consistently passes through the geometric phase to standardize the critical threshold (Ct) values for comparing samples. The first sample to cross the threshold has the greatest starting template concentration because it had a head-start in amplicon production. This threshold principle is used to extrapolate the relative amounts of starting concentration.



**Fig. 1** Real-time PCR results showing the fluorescence emission intensity (∆Rn) of the reporter dye divided by the fluorescence emission intensity of the passive reference dye (Rn) versus the cycle number. The results are subdivided to show the (a) threshold, (b) baseline, (c) background, (d) geometric phase, (e) linear phase, (f) plateau phase.

Closer examination of the real-time experiment (Fig. 1) shows a bend just prior to the geometric phase and is most exaggerated in the least concentrated samples. The first downward bend may indicate a problem with primer concentrations or primer dimers. These hypotheses require testing, but it is clear that there is a problem with efficiency in the reaction. The second geometric peak in the most dilute samples may be due to primer-dimer template interactions. If you run an excess amount of PCR cycles, a second peak will universally occur as new primer-dimer products increasingly contaminate the sample.

**a. Standard Curve Method**

A standard curve is generated by obtaining relative Ct values from a diluted series of independent template. The standard curve method is described in Applied Biosystems tutorial "[***Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR***](http://docs.appliedbiosystems.com/pebiodocs/04371095.pdf)". The example provided in the AB document refers specifically to an experimental set-up that tests for a treatment effect. Unfortunately, the document uses the term 'calibrator' to refer to an untreated sample in the experiment that is a confusing reference because the untreated sample isn't used for calibration. Samples, including the untreated sample, are calibrated relative to the standard curve (Fig. 2).



**Fig. 2** i) An illustrated series of 10X dilutions setting the standard for a calibration curve. ii) The most concentrated solution reaches the Ct threshold first and is followed in sequence by the more dilute samples. iii) A standard curve is generated by plotting the Ct values against the logarithmic dilution on the x-axis.

Once a standard curve is created any other sample's Ct value can be plotted along the curve and the expression value is given relative to the dilution factor. A PCR reaction with 100% efficiency has a slope of –3.32. If the value is <–3.32, then reactions that are less than 100% efficient. If the value is >–3.32, then there is a problem with sample quality or pipetting error. Applied Biosystems provides a ***[protocol](http://www.appliedbiosystems.com/support/tutorials/pdf/quant_pcr.pdf%22%20%5Co%20%22Applied%20Biosystems%3A%20Creating%20standard%20curves%20with%20genomic%20DNA%20or%20plasmid%20DNA%20templates%20for%20use%20in%20quantitative%20PCR.%22%20%5Ct%20%22_blank)***for creating genomic DNA or plasmid DNA dilution standards.

**b. ΔCt Method**

The mathematical relationship between the exponential amplification of PCR and Ct values is explained in detail by ***[Livack and Schmittgen (2001)](http://www.umich.edu/~caparray/Files/LeivakddCt_paper.pdf%22%20%5Co%20%22Livak%20and%20Schmittgen%20%282001%29%20pdf%22%20%5Ct%20%22_blank)***. The method is used to present a numerical value showing the amount of expression in the targeted gene, normalized to an endogenous control, relative to the amount estimated by the standard curve.

**Endogenous Control**

Relative amounts of expression can be numerically expressed relative to the standard curve or normalized by an endogenous control. An endogenous control is a 'housekeeping' gene that is assumed to express at a constant level in all specimens regardless of the treatment. Commonly employed housekeeping genes include the elongation factor 1-alpha gene, a multi-functional protein that is essential for protein synthesis and cytoskeletal regulation in eukaryotic species, the GAPDH enzyme mentioned in the 3':5' integrity test, β-actin, a protein generally involved in cellular locomotion, and 18s ribosomal RNA (Richards et al. 2003; Filby and Tyler 2007; Huber et al. 2007). To normalize  expression values of genes of interest you must divide by the expression value of the endogenous control. Researchers in other labs are then able to compare normalized expression values relative to an endogenous control. A problem with normalizing against a housekeeping gene, however, is that the assumption of constant expression levels requires testing and often fails, hence it is recommended that several housekeeping genes be targeted for each experimental set-up (Huber et al. 2007).

**Preparing a reaction plate**

Precautions for a standard PCR reaction (Kwok and Higuchi 1989; Mifflin 2003) also apply to real-time PCR. Our facility stresses the importance of using separate enclosures for reagent preparation, specimen processing, and amplification and detection. A [***nucleic acid prep station***](https://products.appliedbiosystems.com/ab/en/US/adirect/ab?cmd=catNavigate2&catID=600646) is also available to automate and enclose the extraction process. Real-time reagents from distributors also include safeguards against false positives such as using a dNTP mix with uracil, instead of thymine, and including the heat activated ***[AmpErase® UNG](https://products.appliedbiosystems.com/ab/en/US/adirect/ab?cmd=catNavigate2&catID=602668" \o "Applied Biosystems: SYBR Green PCR Core Reagents." \t "_blank)*** enzyme that digests template containing uracil prior to the polymerase reaction.

The reaction plates are specially manufactured for optical readings, so it is important that they remain free of light inhibiting contaminants such as a dirty fingerprint.

**Genome GeXP**

Beckman-Coulter's GeXP system is based on the procedures of [***Wang et al. (1998)***](http://circ.ahajournals.org/cgi/reprint/98/22/2422.pdf). The method uses a chimeric set of primers, with the forward part attaching to a gene specific region and a universal tail that is incorporated into the transcript during amplification. Primers are designed to target products ranging in size from 100-400 bp with a minimum 5 bp size separation between products. A dye labeled primer that attaches to the universal tail is added to the amplification process and competitive PCR amplifies the different genes with equal efficiency. Genes are separated by size during capillary electrophoresis and quantified by their relative fluorescent intensity. The flow chart image below shows how the GeXP method works through the cDNA synthesis and PCR phases.

   

Beckman-Coulter has prepared [***a report on an experimental application using the GeXp system***](http://www.beckman.com/Literature/BioResearch/A-10295A.pdf), which offers a complete review of the method and how it works.

**References**

Bustin SA (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* **29**, 23-39.

Edwards K, Logan J, and Saunders N (eds.) (2004) *Real-Time PCR: An Essential Guide.*Horizon Bioscience, Pp. 346.

Filby A, and Charles T (2007) Appropriate 'housekeeping' genes for use in expression profiling the effects of environmental estrogens in fish. *BMC Molecular Biology*, **8**, 1-10

[***Huber DPW, Erickson ML, Leutenegger CM, Bohlmann J, and Seybold SJ (2007) Isolation and extreme sex-specific expression of cytochrome P450 genes in the bark beetle, Ips paraconfusus, following feeding on the phloem of host ponderosa pine, Pinus ponderosa. Insect Molecular Biology, 16, 335-349.***](http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2583.2007.00731.x)

Ingram KK, Oefner P, and Gordon DM (2005) Task-specific expression of the *foraging* gene in harvester ants. *Molecular Ecology,* **14**, 813-818.

Jones LJ, Yue ST, Cheung C, and Singer VL (1998) RNA quantitation by fluorescence-based solution assay: RiboGreen reagent characterization. *Analytical Biochemistry*, **265**, 368-374.

Kwok S, and Higuchi R (1989) Avoiding false positives with PCR. *Nature*, **339**, 237-238.

Livak KJ, and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCtmethod. *Methods*, **25**, 402-408.

Mifflin TE (2003) Setting up a PCR laboratory. In *PCR Primer: A laboratory manual*, 2nd ed. Dieffenbach CW, Dveksler GS (eds.) Pp. 5-14. Cold Spring Harbor laboratory press.

Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, **1**, 1559-1582.

Proudfoot NJ, Furger A, and Dye MJ (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501–512.

Richards JG, Semple JW, Bystriansky JS, and Schulte PM (2003) Na+/K+ ATPase a-isoform switching in gills of rainbow trout (Oncorhynchus mykiss) during salinity transfer. *Journal of Experimental Biology*,**206**, 4475-86.

Stahlberg A, Hakansson J, Xian X,  Semb H, and Kubista M (2004) Properties of the reverse transcription reaction in mRNA quantification. Clinical Chemistry, 50, 509-515.

Vandesompele H, De Paepe A, and Speleman F (2002) Elimination of Primer-Dimer Artifacts and Genomic Coamplification Using a Two-Step SYBR GreenI Real-Time RT-PCR. *Analytical Biochemistry,***303**, 95-98.

Varga A, and James D (2006), Real-time RT-PCR and SYBR Green I melting curve analysis for the identification of Plum pox virus strains C, EA, and W: Effect of amplicon size, melt rate, and dye translocation. *Journal of Virological Methods*, **132**, 146-153.

Wang, Z, L Yue, M White, G Pelletier, and S Nattel. 1998. Differential Distribution of Inward Rectifier Potassium Channel Transcripts in Human Atrium Versus Ventricle. Circulation 98:2422-2428.

Wilfinger WW, Mackey K, and Chomczynski P (2006) Assessing the quantity, purity and integrity of RNA and DNA following nucleic acid purification. Pp. 291-312, *in* Kieleczawa J (ed.) DNA Sequencing II: optimizing preparation and cleanup. Jones and Bartlett Publishers.

Yao Y, Nellaker C, and Karlsson H (2006) Evaluation of minor groove binding probe and Taqman probe PCR assays: Influence of mismatches and template complexity on quantification. *Molecular and Cellular Probes*, **20** 311–316.

Zentilin L, and  Giacca M (2007) Competive PCR for precise nucleic acid quantification. *Nature Protocols*, **2**, 2092-2104.