

UC San Diego

UC San Diego Previously Published Works

Title

Detection of GPCR mRNA Expression in Primary Cells Via qPCR, Microarrays, and RNA-Sequencing.

Permalink

<https://escholarship.org/uc/item/4069p1vc>

Authors

Sriram, Krishna
Salmerón, Cristina
Di Nardo, Anna
et al.

Publication Date

2021

DOI

10.1007/978-1-0716-1221-7_2

Peer reviewed



Published in final edited form as:

Methods Mol Biol. 2021 ; 2268: 21–42. doi:10.1007/978-1-0716-1221-7_2.

Detection of GPCR mRNA Expression in Primary Cells Via qPCR, Microarrays, and RNA-Sequencing

Krishna Sriram,

Department of Pharmacology, University of California San Diego, La Jolla, CA, USA

Cristina Salmerón,

Department of Pharmacology, University of California San Diego, La Jolla, CA, USA

Anna Di Nardo,

Department of Dermatology, University of California San Diego, La Jolla, CA, USA

Paul A. Insel

Department of Pharmacology, University of California San Diego, La Jolla, CA, USA

Department of Medicine, University of California San Diego, La Jolla, CA, USA

Abstract

A workflow is described for assaying the expression of G protein-coupled receptors (GPCRs) in cultured cells, using a combination of methods that assess GPCR mRNAs. Beginning from the isolation of cDNA and preparation of mRNA, we provide protocols for designing and testing qPCR primers, assaying mRNA expression using qPCR and high-throughput analysis of GPCR mRNA expression via TaqMan qPCR-based, GPCR-selective arrays. We also provide a workflow for analysis of expression from RNA-sequencing (RNA-seq) assays, which can be queried to yield expression of GPCRs and related genes in samples of interest, as well as to test changes in expression between groups, such as in cells treated with drugs or from healthy and diseased subjects. We place priority on optimized protocols that distinguish signal from noise, as GPCR mRNAs are typically present in low abundance, necessitating techniques that maximize sensitivity while minimizing noise. These methods may also be applicable for assessing the expression of members of families of other low abundance genes via high-throughput analyses of mRNAs, followed by independent confirmation and validation of results via qPCR.

Keywords

qPCR; RNA-seq; Arrays/microarrays; Screening; Omics

1 Introduction

G protein-coupled receptors (GPCRs) include approximately 800 proteins, forming the largest family of cell surface receptors in the human genome. Based on their membrane association, selectivity of expression, and broad range of cellular processes that they

regulate, GPCRs are the largest family of proteins targeted by approved drugs [1]. Approximately 360 of the ~800 human GPCRs are endoGPCRs, i.e., expressed in tissues and activated by endogenously produced ligands [2]. This subset represents the majority of current drug targets and drug discovery efforts. The remaining GPCRs are involved in taste, vision, and olfaction, although certain of those receptors are also expressed in tissues other than the primary sensory organs [3]. Most studies of GPCRs focus on the endoGPCRs, although efforts are also directed at chemosensing GPCRs. Of the ~360 endoGPCRs, ~100 are primary targets for approved drugs and ~100 are currently orphans (i.e., without known endogenous agonists [1, 2]). Hence, studies of GPCRs are an active area of research, with various approaches used to define their role in health and disease in cultured cells, as well as in cells and tissues from experimental animals and human subjects.

Due to their low abundance at the protein level, combined with the general paucity of well-validated antibodies [3, 4], identification and quantification of GPCR proteins is challenging. Detection of GPCRs thus typically involves the use of methods for quantification of mRNA, followed by validation of these GPCRomics data with quantitative polymerase chain reaction (qPCR) analyses and, importantly, signaling and functional assays to verify that the receptors detected in cells are physiological receptors. These techniques may be applied to primary cells isolated from humans or experimental animals or by using cell lines. We recently published a report in which we compared three omics methods for high-throughput detection of GPCR mRNA expression in primary cells: qPCR-based GPCR microarrays, Affymetrix arrays, and RNA-seq [4]. A key conclusion of our study was that both qPCR-based GPCR microarrays and RNA-seq provide useful, mutually consistent data for detection of GPCRs, whereas Affymetrix arrays do not. Hence, in the subsequent sections, we discuss protocols for the detection of GPCR mRNA abundance and briefly review approaches for analysis of data generated by use of qPCR, qPCR-based GPCR microarrays, and RNA-seq.

GPCR expression data can provide valuable information, especially for drug discovery efforts, by identifying novel roles for GPCRs in disease [5]. In addition, such data are available in the public domain via servers, such as NCBI GEO (Gene Expression Omnibus), or large database studies such as TCGA (The Cancer Genome Atlas) [3]. These findings are mineable by the broader scientific community, thereby facilitating research efforts beyond the individual laboratories that generated the data [3, 4]. Such data have applications in exploring and elucidating roles for GPCRs, including in novel settings such as in cancers [3] and infectious disease, for example, in the COVID-19 pandemic [6].

Figure 1 shows a general schematic of the steps involved in analysis of GPCR mRNA expression. Topics in solid boxes are discussed below while topics in boxes with dashed lines are outside the scope of this chapter. In the protocols described here, a primary goal is to minimize the amount of noise in the data, which can be introduced by artifacts such as sample contamination, improper reagents, and poorly optimized protocols, thereby compromising one's ability to detect GPCR mRNA. This is of particular importance for GPCRs, due to their relatively low magnitude of expression.

2 Materials

2.1 Cell Lines

1. Cells in culture (e.g., pancreatic cancer-associated fibroblasts [4] and see notes for cell numbers and plate layouts).

2.2 Equipment

1. Thermocycler for reading 96-well qPCR plates, compatible with SYBR green chemistry.
2. Thermocycler for reading 384-well qPCR plates, compatible with TaqMan chemistry.
3. Thermocycler for standard end point PCR.
4. Tabletop microcentrifuge able to maintain up to $10,000 \times g$.
5. Nanodrop.
6. UVP imager for imaging of gels.
7. A desktop computer running current versions of any standard operating system, for data analysis.
8. 1000, 200, and 20 μL pipettes.

2.3 Reagents and Kits

1. Qiagen RNeasy Mini Kit.
2. Molecular biology grade ethanol, 200 proof.
3. β -Mercaptoethanol (β -ME) or 2 M dithiothreitol (DTT).
4. RNase-free DNase-1 kit.
5. Ultrapure molecular biology-grade water.
6. 70% Ethanol (200 proof molecular biology-grade ethanol, diluted to 70% with ultrapure nuclease-free water).
7. iScript cDNA synthesis kit.
8. SYBR Green qPCR master mix.
9. qPCR primers (see Note 1).
10. TaqMan Universal PCR Master Mix.
11. TaqMan PCR array.
12. SYBR Safe DNA Gel Stain.
13. RNase away.

2.4 Consumables

1. Nuclease-free 1.7 and 0.6 mL microcentrifuge tubes.
2. Filtered pipette tips, RNase/DNase free for 1000, 200, and 20 μ L pipettes.
3. PCR strip tubes with caps.
4. 96-Well plates for qPCR with adhesive seal.
5. Suitable wipes for cleaning of surfaces.

2.5 Software Tools

1. Manufacturer-supplied software for qPCR thermocyclers.
2. FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, Babraham bioinformatics).
3. BBDDUK [U.S. Department of Energy (DOE) Joint Genome Institute (JGI)].
4. Kallisto [Lior Pachter Lab, <https://pachterlab.github.io/kallisto/> [7]].
5. Reference transcriptome for relevant species, obtained from Ensembl (ensembl.org/info/data/ftp/index.html).
6. R, preferably with R Studio (<https://www.r-project.org/>;<https://rstudio.com/>).
7. R package tximport [<https://bioconductor.org/packages/release/bioc/html/tximport.html>, [8]].
8. R package edgeR [<https://bioconductor.org/packages/release/bioc/html/edgeR.html>, [9]].
9. R package BioMart [<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>, [10]].

3 Methods

3.1 mRNA Isolation from Cultured Cells

1. Harvest cells, typically using 350 μ L of RLT Lysis buffer (included with Qiagen kit) for a semi-confluent cell culture 10 cm dish, or smaller (see Note 2). For a highly confluent dish, or in larger dishes or flasks, increase the volume of lysis buffer to 700 μ L (see Notes 3 and 4).
2. Use a cell scraper to ensure removal of all cells in lysis buffer and collect lysate in an RNase-free 1.7 mL microcentrifuge tube.
3. Vortex the contents vigorously for ~20 s and allow to rest at room temperature ~ 5 min, to allow for complete lysis.
4. Prepare RNase-free DNase-1. Dissolve the bottle of lyophilized enzyme in 550 μ L of nuclease-free water. Store 10 μ L aliquots at -20 $^{\circ}$ C; avoid repeated freeze-thawing. Aliquots are usable for ~3 months.

5. Add 1 volume of 70% ethanol to the lysate from **step 2** and mix well by pipetting. Do not centrifuge (see Note 5).
6. Transfer up to 700 μL of the sample, including any precipitate, to a RNeasy Mini spin column (pink color) placed in a 2 mL collection tube (supplied with Qiagen RNeasy mini kit). Close the lid and centrifuge for 20 s at $8000 \times g$ (see Note 6). Discard the flow-through.
7. Add 350 μL Buffer RW1 (included with Qiagen RNeasy mini kit) to the RNeasy spin column. Close the lid, and centrifuge for 20 s at $8000 \times g$. Discard the flow-through.
8. Add 10 μL DNase stock solution from **step 4** to 70 μL Buffer RDD (included with DNase-1 kit).
9. Mix by gently inverting the tube, and centrifuge briefly to collect residual liquid from sides of tube. Add to RNeasy spin column membrane, and incubate for 15 min at 20–30 $^{\circ}\text{C}$.
10. Add 350 μL Buffer RW1 to the RNeasy column and centrifuge for 20 s at $8000 \times g$. Discard flow-through.
11. Add 500 μL Buffer RPE (included with Qiagen RNeasy mini kit; see Note 7) to the RNeasy column. Centrifuge for 20 s at $8000 \times g$ and discard flow-through.
12. Add 500 μL Buffer RPE to RNeasy column and centrifuge for 2 min at $8000 \times g$.
13. Place the RNeasy column in new 2 mL tube and centrifuge at full speed for 1 min.
14. Place RNeasy column in new 1.7 mL tube. Add 34 μL RNase-free water, close lid and wait ~1 min to allow water to saturate filter membrane. Centrifuge for 1 min at $8000 \times g$.
15. Test RNA quantity and quality via a nanodrop (see Note 8). RNA concentrations in $\text{ng}/\mu\text{L}$ are necessary for the cDNA synthesis steps, so as to estimate the amount of RNA sample to load per cDNA synthesis reaction.

3.2 cDNA Synthesis

1. In PCR strip tubes, prepare 20 μL reactions for cDNA synthesis, by adding 4 μL of 5X iScript Reaction mix, 1 μL iScript Reverse Transcriptase, Input RNA template (20–1000 ng), molecular biology-grade, nuclease-free water to make up volume to 20 μL (see Note 9).
2. Perform PCR in thermocycler, using the following reaction parameters: priming 5 min at 25 $^{\circ}\text{C}$ reverse transcription (RT) 20 min at 46 $^{\circ}\text{C}$; RT inactivation 1 min at 95 $^{\circ}\text{C}$; hold at 4 $^{\circ}\text{C}$.

3.3 qPCR Primer Design (Bioinformatics Protocol)

Due to their low magnitude of expression, signal-to-noise ratios in qPCR detection of GPCRs can be poor, i.e., these assays are heavily compromised by primer-dimer formation. The steps below for primer design are intended to minimize this phenomenon.

1. Query the gene of interest at NCBI (<https://www.ncbi.nlm.nih.gov/>) using the option “gene” in the drop-down menu.
2. In search results, select the gene for the species of interest.
3. Inspect the figure showing the various refseq-annotated transcripts; transcripts with the prefix “NM” correspond to well-validated transcripts, for which evidence exists of their expression in cells as mRNA.
4. Unless the aim is to assess individual splice variants, when examining GPCR mRNA expression as part of a screening exercise, it is preferable to focus on regions of the gene common to all well-validated transcripts. This is typically straight-forward for GPCRs, as they usually possess few, if any, introns.
5. In the drop-down menu lower down the page titled “NCBI Reference Sequences (RefSeq),” click on one of the validated refseq transcripts. This opens a new page that has specific details about the transcript. In the top of the page, click on “Graphics.”
6. This opens a graphical view of the transcript highlighting regions of interest, including exons, protein-coding regions, etc. The coordinates at the top of the graphical view are critical for the next step. Based on regions that are present in all validated transcripts from the previous view, identify a region of the transcript over which primers should be identified. For example, if a transcript is 5000 bp long, you may wish to design primers for a region in the middle of the transcript, perhaps from 1000 to 3500, a region conserved in all variants/transcripts of the gene.
7. Under the menu “Analyze this sequence,” select “Pick Primers.” This opens the Primer-blast primer design tool.
8. In “Range,” for forward primer, enter the first/start coordinate (1000 in the example above), which is the beginning of the region of interest from the graphical view discussed above. Leave the second box in Forward Primer blank.
9. Similarly, for the reverse primer, leave the first box blank, and enter the coordinate corresponding to the end of the region of interest (3500 in the example above) in the second box.
10. Under PCR product size, change “Max” to 200. If difficulty is encountered in identifying suitable primers, this can be increased to 250–300.
11. Do not select the primer to be either exon-junction spanning or to include an intron. In general, this places excessive constraints on primer design which force a compromise in identifying primer pairs with minimal primer-dimer formation, which is critical when studying GPCRs. In addition, the inclusion of DNase-1

digestion during the RNA isolation removes the need for intron-including primer pairs.

12. Check the box “Allow Splice Variants.”
13. In “Advanced Parameters,” in “Secondary Structure Alignment Methods,” select “Use Thermodynamic Oligo Alignment.”
14. Run the tool with all other parameters as default.
15. The tool may produce a prompt that the region identified corresponds to multiple transcripts. Select “All” and proceed.
16. After a few moments, the tool will produce a graphical view that illustrates regions where the primer pairs bind to the transcript of interest, as well as various properties of the primers. An example of this graphical output is shown in Fig. 2.
17. Prioritize primer pairs with GC content closest to 50% and with lowest self-complementarity and Self 3′ complementarity. Typically, Primer-Blast produces primer pairs with both complementarity scores of 0, or < 2.
18. Using the oligo-analyzer tool from IDT (<https://www.idtdna.com/calc/analyzer/>), analyze the primer pairs for their self-dimer formation and their heterodimer characteristics.
19. For self-dimers, we recommend primers with Gibbs free energy >−9.0 kcal/mol (i.e., values closer to 0).
20. For heterodimers, we suggest primer pairs with Gibbs free energy >−6.0 kcal/mol, ideally >4 kcal/mol.
21. Once a primer pair is selected, test each primer using NCBI nucleotide BLAST(<https://blast.ncbi.nlm.nih.gov/>) (Fig. 2). This is to ensure (a) the resulting primer has a 100% match to the gene of interest and (b) within the species of interest, the extent of a match of the primer sequences to other genes. Preferably, to minimize nonspecific amplification, primers should have a < 85% match to any other regions on coding regions of other genes.
22. Repeat this process varying different parameters above (e.g., amplicon size, region of interest) until suitable primer pairs are obtained, with low Gibbs free energies for dimer formation.

3.4 Primer Validation

For examples of melting curves and amplification curves, see Figs. 3–5. Figure 4 shows an example of a suitable amplification curve, along with examples of amplification curves for a nonoptimal assay.

1. Prepare stock (100 μM) and working (1–10 μM) solutions of the oligonucleotides in nuclease-free water with aliquots stored at −20 °C (see Note 10).
2. Prepare cDNA as above in Subheading 3.2, from 1 μg of species-appropriate universal RNA (see Note 11).

3. Using a 6-point serial dilution, with two-fold dilutions, prepare qPCR reactions beginning with initial input of 100 ng universal cDNA template, with the final cDNA input of 3.125 ng (see Note 12). see Subheading 3.5 for preparation of qPCR reactions.
4. Perform qPCR, using the steps described in Subheading 3.5.
5. Determine the efficiency of the qPCR assay by plotting log of cDNA concentration against the cycle quantification (C_q) values for each point in the standard curve (see Note 13). An efficiency of 100% and a gradient (slope) of -3.323 indicates perfect doubling of product in each cycle; primer efficiency scores between 90% and 110% are typically accepted. The efficiency is represented by the following equation: Efficiency (%) = $(10^{-1/\text{slope}} - 1) \times 100$. An example dilution curve for serial dilutions is shown in Fig. 3; for clarity, curves are shown for eight-fold dilutions, which will be separated by ~3 cycles, for a primer pair with ~100% efficiency (see Note 14).
6. Load the qPCR products from the wells containing 100 ng of template and the PCR control with DNA loading buffer diluted to 1X, in a 2% agarose gel (with SYBR Safe DNA Gel Stain, at 10,000 X stock concentration, diluted in agarose gel to 1X), along with a DNA ladder in a separate lane.
7. Perform gel electrophoresis at 100 V for 45 min. Expected results (visualized in an appropriate imager for DNA gels) are a single band of the predicted size in the cDNA template samples and no background in the PCR control samples.

3.5 Independent Qpcr

1. Prepare 10–50 μL of qPCR reactions in 96-well plates. For every 10 μL of reaction mix, the reaction contains 5 μL of 2X SYBR Green Mastermix, 1 μL of diluted mixture of Forward and Reverse Primers, and 4 μL of diluted cDNA template.
2. Perform the two-step qPCR reaction protocol in the thermocycler according to the following steps: Initial denaturing at 95 °C for 3 min, denaturation at 95 °C for 10 s, and annealing and extension at 60 °C for 30 s.
3. Repeat the former steps for 40 cycles (see Notes 14–20).
4. Perform the melting curve analysis over 55–95 °C. Increase temperature in 0.5 °C increments (see Note 21). Specific instructions for holding/reading at each temperature increment will vary among machines.

3.6 qPCR-Based TaqMan Array

Figure 6 shows the primary result of interest from such an array; the Ct values vs. housekeeping gene, for the highest expressed GPCRs in a sample, thus quantifying their magnitudes of expression.

1. Prepare a reaction mixture by combining 500 μL TaqMan mastermix, 1 μg of cDNA template, and nuclease-free water, to a final volume of 1 mL.

2. Using ports for pipetting reaction mixture into microchannels on the TaqMan array, evenly pipette the reaction mixture into each of the five microchannels.
3. Use the following qPCR reaction protocol: incubation at 50 C for 2 min; incubation at 95 C for 10 min; denaturation at 95 C for 15 s; annealing and extension at 60 C for 60 s.
4. Repeat the protocol steps for 40 cycles (see Note 22).

3.7 RNA-Sequencing Data Analysis for GPCR Expression (Bioinformatics Protocol)

Following preparation of RNA samples and validation via qPCR, preparation of libraries for RNA-sequencing and the subsequent sequencing of these libraries is typically undertaken externally, ideally by core facilities available at most research universities or by commercial vendors. Use of such resources is typically done based on the specialized nature of these protocols and the equipment used, in particular the high costs of sequencers. Specific details of protocols for these steps for RNA sequencing are beyond the scope of this discussion (see Note23). Where applicable, we provide links to additional resources. Many of these tools and alternatives can also be accessed via Galaxy, at <https://usegalaxy.org/>.

Once sequencing is completed, users will typically be provided with a download link, to obtain the raw data files from a server. These files will be in FASTQ/FASTQ.GZ format. Figure 7 provides a schematic describing the steps needed to proceed from these raw RNA-seq data, to analyzed data that provide information regarding magnitudes of GPCR expression and changes in expression between groups, where applicable. We describe these steps in brief; users may find more details in the corresponding notes and in links and resources provided below. The steps below will work in any operating system; all tools are open source.

1. Inspect the files for quality using the FASTQC tool (Babraham Bioinformatics). Specifically, one should verify that the quality of sequencing is adequate (i.e., that one has high confidence that each base in a sequenced read was identified correctly). Quality scores >30 are typically considered adequate (see Fig. 8). If the quality scores are poor, trimming for low quality bases may be needed in subsequent steps. In general, with modern sequencing technology, quality scores are usually satisfactory; we have rarely encountered the need for quality trimming. In some cases, detection of bases at the 5' or 3' end can be of poor quality. If this is the case, trimming of the final few or first few bases may be performed. In addition, FASTQC also provides useful information about whether certain sequences are overrepresented in the data, which may indicate contamination with sequencing adapters or from other sources, such as ribosomal RNA. For additional information on FASTQC output and “good” vs. “bad” FASTQC results, consult <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
2. Perform trimming of reads for quality, or adapter trimming as necessary, using BBDUK. The tool may be downloaded at <https://sourceforge.net/projects/bbmap/>, containing executable files and a database of standard sequencing

adapters, which can be used for adapter trimming. The tool can be run from the command line, on raw FASTQ files, using commands described at <https://jgi.doe.gov/data-and-tools/bbtools/bbtools-user-guide/bbduk-guide/> (see Note 24). This will yield new, trimmed FASTQ files for downstream use. Adapter trimming is especially relevant for paired-end sequencing runs.

3. Using the trimmed FASTQ files, analyze the files via Kallisto [7]. The steps are as follows:
 - a. Download the reference transcriptome for the relevant species from ensembl, via ensembl.org/info/data/ftp/index.html. This will produce a “Fasta” file (see Note25).
 - b. Download Kallisto from <https://pachterlab.github.io/kallisto/download> to desired folder.
 - c. Run Kallisto (calling the Kallisto executable file) on the reference transcriptome Fasta file downloaded above. This generates an index file for the next step (see Note 26).
 - d. Run Kallisto to quantify transcript expression, using the constructed index and trimmed fastq files as input. This will yield an abundance.tsv file, with quantification at the transcript-level of expression in TPM (transcripts per million) and estimated counts.
4. To convert transcript-level data to the gene level, run tximport [8], with the abundance file from Kallisto and a mapping file containing corresponding gene IDs for transcript IDs (see Note 27).
5. The resulting file from tximport is ready to query for GPCR expression. A list of gene IDs corresponding to human GPCRs can be found at <https://insellab.github.io/data>. Expression in TPM for these genes can be queried from the file generated by tximport using a variety of tools as convenient, including R, excel, etc. An example of GPCR expression in TPM from the file produced by tximport is shown in Fig. 9, for the 50 highest expressed GPCRs in a cancer-associated fibroblast sample, analyzed using this approach. Sample commands and syntax for running tximport (including how to generate a mapping file of transcript IDs to gene IDs) on output from Kallisto can be found in Note 28.
6. The gene-level data from tximport also contain estimated counts for each gene. A matrix of these counts for all samples in an experiment can be supplied to the edgeR package [9] within R, to perform differential expression (DE) analysis. Do not use standard statistical methods such as T-tests for RNA-seq data. The result from edgeR is the fold-change between test conditions and the corresponding statistical significance (expressed as a false discovery rate, FDR) for all genes. FDR < 0.05 is typically used to infer statistical significance. Results for GPCRs can be queried from the DE analysis. The edgeR Bioconductor page can be accessed by users for detailed tutorials, <https://bioconductor.org/packages/release/bioc/html/edgeR.html>; the details of how to perform such analysis are beyond the scope of this text. For plotting GPCR expression between different

samples, one can use edgeR to generate gene expression in CPM (Counts per Million).

4 Notes

1. Custom oligos were purchased from IDT. Subheading 3.3 describes the design of oligo sequences.
2. We recommend use of the Qiagen RNeasy Mini Kit. This method, using spin columns with on-column DNase digestion, is widely used for producing RNA as input for omics methods. Alternative kits/isolation methods can also be used, for instance methods using trizol (or equivalents), in particular, in cell types with high levels of RNases, such as immune cell types [11]. As an example, we tested the “Directzol” kit from Zymo Research and found similar results to samples prepared with the Qiagen kits, as assessed by qPCR. This method can also be adapted for use with RNA-later stabilized tissue. Please refer to notes below for additional details.
3. For most adherent mammalian cell types, a single well of a 6-well plate, grown to confluency is typically adequate for the assays described here. Independent qPCR analysis of GPCR mRNA expression requires ~200 ng of total RNA, TaqMan arrays and other qPCR-based arrays typically require ~1000 ng of total RNA; RNA-seq can typically be performed using 200 ng of total RNA.
4. Tissue samples stabilized in RNAlater can be homogenized using tools such as a tissue tearor homogenizer, as per the manufacturer’s instructions. Lysates should be resuspended in RLT buffer, with either 10 μ L β -mercaptoethanol or 20 μ L dithiothreitol per mL of RLT lysis buffer. While keeping samples on ice between steps, the lysates should then be passed through a QIAshredder per the manufacturer’s instructions, to obtain a homogenate from which RNA can be isolated. Lysates resuspended in RLT buffer, stored at -80°C are stable for many months or even years in samples with low RNase content.
5. It is critical to minimize RNA degradation due to presence of RNases in samples, especially in some RNase-rich cell types. As a first step, clean down all surfaces and pipettes with RNasea-way, as well as doing the same for the tabletop centrifuge, gloves, etc. As far as possible, work in a PCR hood or a dedicated bench intended as a clean area for working with RNA.
6. The protocol for RNA isolation via RNeasy spin columnssould be performed at room temperature, i.e., $\sim 20\text{--}25^{\circ}\text{C}$. Priority is on working quickly, to minimize opportunities for sample degradation and/or contamination. Following completion of the isolation procedure, RNA samples should be stored at -80°C . RNase-free samples stored at these low temperatures are stable for several years.
7. Upon first use, the RPE buffer needs to be reconstituted withan appropriate volume of 200-proof molecular-biology grade ethanol. Add the indicated volume of ethanol (typically 4 volumes of ethanol to 1 volume of stock RPE buffer

concentrate) to the bottle of RPE buffer, mix well and check the box on the lid that ethanol was added.

8. The nanodrop should yield a clear peak at 260 nm wavelength. Ideal 260/280 ratios (indicating protein contamination, a particular hazard for RNA degradation) are 2.0–2.1, for samples generated via the Qiagen kit. 260/230 ratios, indicating presence of salt content and/or alcohol, which can inhibit downstream PCR reactions, should typically be >1.8. However, for low RNA concentrations (especially <50 ng/uL), the 260/230 ratio is rarely reliable and will often be <1.0.
9. To produce larger amounts of cDNA for experiments needing larger amounts of cDNA template downstream, this reaction can be scaled-up to 50 μ L.
10. With primers designed by the steps in Subheading 3.3, concentrations of primers between ~0.1 and 1 μ M are suitable to produce amplification while avoiding excess primer-dimer formation. Lower primer concentrations can reduce the likelihood of primer-dimer formation. A stock of primers, containing a 1:1 ratio of forward and reverse primers, can be prepared at ~10 μ M each, and diluted as needed, to yield the final concentrations noted above. Primer stocks and working solutions stored at –20 °C are usable for many years, as long as contamination is avoided.
11. For qPCR reactions downstream, the cDNA product should usually be diluted at least tenfold, as some contents of the reaction (from the iScript cDNA synthesis kit) can inhibit qPCR reactions. cDNA concentration should not be measured using a Nanodrop; these values will not be accurate due to reaction contaminants within the reaction mix. The amount of cDNA template should be estimated on the basis of the amount of input RNA in the cDNA synthesis reaction. cDNA samples can be stored at –20 °C and usable for qPCR for many years. As an alternative to the iScript kit, we have obtained comparable results with the qScript kit and the SuperScript 3 kit.
12. For common species (e.g., human, mouse or rat), universal RNA is commercially available. If it is not available for the species of interest, use lysates from a mixture of different tissues, to obtain a representative pool of RNA expressed in that species.
13. Controls in a qPCR experiment, in particular negative controls, ensure that spurious signals are minimized. This is of particular importance when studying expression of low-abundance targets such as GPCRs. No-template Controls (NTCs) are qPCR reactions with all reagents added, except for the cDNA template. This allows estimation of primer-dimer formation. This control should be used especially when initially validating primers. Once a lack of primer-dimer formation has been verified for a given mixture of forward and reverse primers, this control need not be included in every subsequent qPCR assay. In addition, a Reverse Transcription minus control [RT(–)] contains all components of the cDNA synthesis step, except for the presence of the reverse transcriptase. This

RT(-) template is then assayed via qPCR as normal; amplification indicates the presence of genomic DNA contamination, or potentially unforeseen primer-dimer formation between primers for cDNA synthesis and qPCR. This RT(-) control should be included with every qPCR assay.

14. For primers with ~100% efficiency, every two-fold dilution should yield a delay of amplification to a set threshold, of 1 cycle. Correspondingly, for every eight-fold dilution, curves should be separated by 3 cycles. An example is shown in Fig. 3. The cycle threshold should be set approximately in the middle of the linear portion of the amplification curve, where all amplification curves correspond to phases of the qPCR reaction occurring with high-efficiency amplification.
15. The designed primers may be prepared as a mix of forward and reverse primers, at 1–10 μM each, diluted in nuclease-free molecular biology grade water. This will be further diluted at a 1:10 ratio, giving a final primer concentration of 100 nM–1 μM in the qPCR reaction. Lower concentrations are generally preferable, in order to minimize primer-dimer formation.
16. cDNA template from the earlier cDNA synthesis step (Subheading 3.2) should be diluted, in order to obtain a typical amount (1–10 ng) of input cDNA in each qPCR reaction. Input cDNA in these quantities typically yields amplification of both housekeeping genes and GPCRs within acceptable numbers of qPCR cycles.
17. The 2-step qPCR protocol works best with modern, high efficiency thermocyclers. In older machines, it is often preferable to add an extra annealing step (after the incubation at ~60 °C), often at ~68–72 °C, for 30–60 sec. Similarly, in situations involving large amplicons, e.g., those >200 bp in length, this extra annealing step, yielding a 3-step qPCR protocol, may be preferable.
18. Figure 4 shows a typical amplification curve for a qPCR assay (Blue) and examples of amplification curves (in orange and green) associated with common technical difficulties. An ideal amplification curve should have three distinct phases: an S-curve formed by a baseline region with little signal, a rapidly increasing linear portion and a plateau region as amplification ends and the signal stabilizes.
19. In order to normalize qPCR data, one uses a housekeeping gene. Common choices include ACTB (β -actin), GAPDH (Glyceraldehyde 3-phosphate dehydrogenase) and 18S ribosomal RNA. These housekeeping genes are present on qPCR-based GPCR arrays, including TaqMan arrays. We have found that 18S rRNA is effective for normalizing qPCR data, in particular, if comparing estimates of GPCR expression from qPCR and RNA-seq [4]. Expression of a GPCR relative to 18S rRNA may be computed as their difference in cycle threshold (i.e., C_t). The relative expression of two GPCRs A and B within the same sample is then: $\text{Expression of GPCR A} / \text{Expression of GPCR B} = 2^{-(C_t \text{ for B})} / 2^{-(C_t \text{ for A})} = 2^{-(C_t \text{ for B} - C_t \text{ for A})}$. Similarly, differences

in expression of a GPCR between two samples $A1$ and $A2$ are calculated as: Expression in $A1$ /Expression in $A2 = 2^{-(Ct \text{ for } A2 - Ct \text{ for } A1)}$, i.e., the “ Ct ” method. This method may be used to analyze data from independent qPCR and qPCR-based arrays. Besides this semiquantitative approach using housekeeping genes, spiked-in control genes can be used to generate quantitative expression data.

20. In general, we consider GPCRs to be detected if they show linear amplification curves within 25 cycles of that for 18S [4]. Below this limit, one encounters amplification close to the 40-cycle limit. Data at such low mRNA abundance can be highly inconsistent among technical replicates, yielding non-reproducible, non-rigorous data at these extremely delayed amplification levels.
21. An example of a melting curve is shown in Fig. 5, showing a typical melting curve for a primer behaving as designed, yielding a single peak, implying the presence of a single PCR product. The presence of multiple peaks indicates multiple products, either as nonspecific binding or formation of primer-dimers. Such melting curves can also be observed in certain rare cases where multiple variants are amplified, including with inclusion of small introns or exon skipping of small exons. If an anomalous melting curve is observed, the PCR product should be assessed by gel electrophoresis. Formation of primer-dimers should yield bands of lower molecular weight products than the amplicon of interest. A further application of melting curves is in identification of genomic DNA contamination, via the presence of a peak in RT(–) control qPCR reactions.
22. Similar qPCR-based microarrays are available from other vendors, using SYBR green chemistry. Examples include arrays for “PrimePCR Pathways” from Bio-Rad laboratories and “RT² Profiler PCR Arrays” from Qiagen.
23. The general parameters of a sequencing run to generate RNA-seq data sufficient for obtaining data on GPCR expression are as follows:
 - a. Library type: typically, stranded mRNA. These libraries are most frequently prepared via Illumina Truseq stranded mRNA library kits and protocols.
 - b. Sequencing depth: 25–30 million reads, with 50–75 base single reads is generally sufficient, to identify which GPCRs are expressed in a given sample.
24. For detailed instructions, refer to the guide at <https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/>. For adapter trimming, a list of adapters in the file “*adapters.fa*” is downloadable with the BBDOCK package and contains many standard adapter sequences. Enter the command—
 - a. `bbduk.sh in=raw_data.fastq.gz out=clean_data.fastq.gz ref=adapters.fa ktrim=r k=23 mink=11 hdist=1 tpe tbo.`
 - b. Rename files in the example above, as necessary. Once complete, rerun FASTQC to ensure that adapter-trimming was successful. Refer the

BBDUK guide above for more settings and options for the command above.

- c. Similar commands can also be used for quality trimming or quality filtering of reads, the details are beyond the scope of this text and can be found in the BBDUK user guide.
25. Reference transcriptomes for Kallisto should be obtained from ensembl. Ensembl annotations are preferable to refseq, as they tend to be more complete; these differences are most noticeable for nonhuman species. Reference transcriptomes can be downloaded from <http://uswest.ensembl.org/info/data/ftp/index.html>, by downloading the corresponding FASTA file of cDNAs for the species of interest. Alternately, reference transcriptomes and pre-built indexes for Kallisto for several species can be downloaded from <https://github.com/pachterlab/kallisto-transcriptome-indices/releases>
 26. Once the reference transcriptome has been obtained as a FASTA file, one runs Kallisto by entering in the command line—
 - a. *kallisto index -i name_of_index_file.idx name_of_fasta_file.gz,*
 - b. This will create an index file, with name as specified, with extension “.idx”.
 - c. For additional options for running this command, refer to the Kallisto manual, at <https://pachterlab.github.io/kallisto/manual>.
 27. Once an index file has been generated, this can be used along with the trimmed FASTQ files as input for quantification of transcript expression.
 - a. For paired-end data, enter into the command line—
 - b. *kallisto quant -i name_of_index_file.idx -o output_folder_name pairA_1.fastq pairA_2.fastq,*
 - c. For single read data, enter:


```
kallisto quant -i name_of_index_file.idx -o output_folder_name --single -l 200 -s 20 file 1.fastq.gz file 2.fastq.gz file 3.fastq.gz.
```

where file 1, file 2, and file 3 are the FASTQ files corresponding to a single sample. Depending on how the sequencing was performed, each sample may have only a single file, or be divided among multiple FASTQ files. Since Kallisto processes one sample at a time, one does not supply it with FASTQ files corresponding to multiple samples in a single command. The “l” and “s” parameters correspond to the average and standard deviation of fragment lengths in the sequencing run. The 200 and 20 values indicated above for each correspond to standard values for Illumina libraries; values for a specific experiment can be obtained from bioanalyzer data. Communication with a sequencing center will likely be needed to obtain this information.

- d. For additional options for running these commands, refer to the Kallisto manual, at <https://pachterlab.github.io/kallisto/manual>. In particular, the `--bias` and `--pseudobam` options may be relevant to specific projects.
 - e. The output from Kallisto is stored in the output folder as named by the user, in three files. *The abundance.tsv* file is most relevant to the present exercise; it contains abundance quantifications in TPM for all annotated transcripts, effective length and estimated counts. The other two files are described in the Kallisto manual and not discussed here.
28. tximport is a package in R, which uses as input, quantification data from Kallisto (the abundance.tsv file) and a file containing a map of transcript IDs to gene IDs. To install tximport in R, enter in the R or R Studio command line—,

- a. `if(!requireNamespace("BiocManager", quietly = TRUE),`
- b. `install.packages("BiocManager"),`
- c. `BiocManager::install("tximport").`
- d. To obtain a map of transcript IDs to gene IDs, via Bio-Mart [10], first install BioMart—.
- e. `if(!requireNamespace("BiocManager", quietly = TRUE))`
- f. `install.packages("BiocManager").`
- g. `BiocManager::install("biomaRt").`
- h. And then, to generate the mapping file for transcript IDs to gene IDs—
- i. `library(biomaRt).`
- j. `ensembl = useEnsembl(biomart="ensembl",`
`dataset="hsapiens_gene_ensembl", version=96).`
- k. `t2g <- biomaRt::getBM(attributes = c("ensembl_transcript_id",`
`"ensembl_gene_id", "external_gene_name"), mart = ensembl).`
- l. `t2g <- dplyr::rename(t2g, target_id = ensembl_transcript_id, ens_gene`
`= ensembl_gene_id, ext_gene = external_gene_name).`
- m. `save(t2g, file = "t2g.RData").`

This provides a file named "t2g" with mapping of transcript IDs to both ensembl gene IDs and gene symbols, such as those used by HUGO.

The species name and ensembl version in the commands above can be modified as needed. If so, one uses the external gene symbols for gene level data. Hence, one removes the ensembl IDs and then runs tximport. The commands in R are as follows:

- a. `library(tximport).`
- b. `t2g$ens_gene <- NULL.`

- c. `txi = tximport("abundance.tsv", type = c("kallisto"), countsFromAbundance = c("scaledTPM"), tx2gene = t2g).`
- d. `write.table(txi, file = "gene_abundance ", sep = "\t", col.Names=NA).`

This creates a file, named "gene abundance," containing the data from abundance.tsv for transcript level to gene level, i.e., for each gene symbol the corresponding magnitude of expression in TPM and estimated counts. These can be used for downstream gene-level analyses, including differential expression (DE) analysis.

Acknowledgements

Support was provided by Academic Senate of the University of California, San Diego and NIH grant RO1A1093957.

References

1. Sriram K, Insel PA (2018) G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs? *Mol Pharmacol* 93:251–258 [PubMed: 29298813]
2. Alexander SP, Christopoulos A, Davenport AP, Kelly E, Mathie A, Peters JA, Veale EL, Armstrong JF, Faccenda E, Harding SD, Pawson AJ (2019) The concise guide to PHARMACOLOGY 2019/20: G protein-coupled receptors. *Br J Pharmacol* 176:S21–S141 [PubMed: 31710717]
3. Sriram K, Moyung K, Corriden R, Carter H, Insel PA (2019) GPCRs show widespread differential mRNA expression and frequent mutation and copy number variation in solid tumor. *PLoS Biol* 17:e3000434 [PubMed: 31765370]
4. Sriram K, Wiley SZ, Moyung K, Gorr MW, Salmerón C, Marucut J, French RP, Lowy AM, Insel PA (2019) Detection and quantification of GPCR mRNA: an assessment and implications of data from high-content methods. *ACS Omega* 4:17048–17059 [PubMed: 31646252]
5. Insel PA, Sriram K, Gorr MW, Wiley SZ, Michkov A, Salmerón C, Chinn AM (2019) GPCRomics: an approach to discover GPCR drug targets. *Trends Pharmacol Sci* 40:378–387 [PubMed: 31078319]
6. Sriram K, Insel PA (2020) A hypothesis for pathobiology and treatment of COVID-19: the centrality of ACE1/ACE2 imbalance. *Br J Pharmacol*. 10.1111/bph.15082
7. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527 [PubMed: 27043002]
8. Sonesson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4:1521 [PubMed: 26925227]
9. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140 [PubMed: 19910308]
10. Durinck S, Spellman P, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. *Nat Protoc* 4:1184–1191 [PubMed: 19617889]
11. Gupta SK, Haigh BJ, Griffin FJ, Wheeler TT (2013) The mammalian secreted RNases: mechanisms of action in host defence. *Innate Immun* 19:86–97 [PubMed: 22627784]

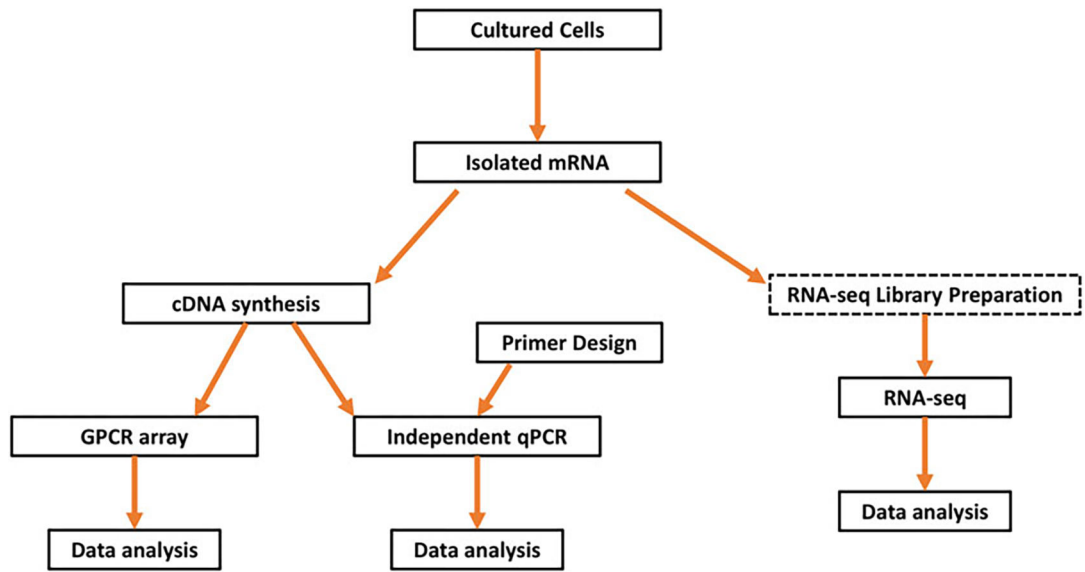


Fig. 1.
The sequence of steps to identify expression of GPCR mRNA. Steps in solid boxes are discussed in this text



Fig. 2. Graphical output from Primer BLAST showing candidate primer pairs (arrows indicate forward and reverse). Red genomic region: protein coding portion of mRNA; Green: the gene of interest; Black: exons for this gene; vertical red lines: demarcating region of interest for primer design

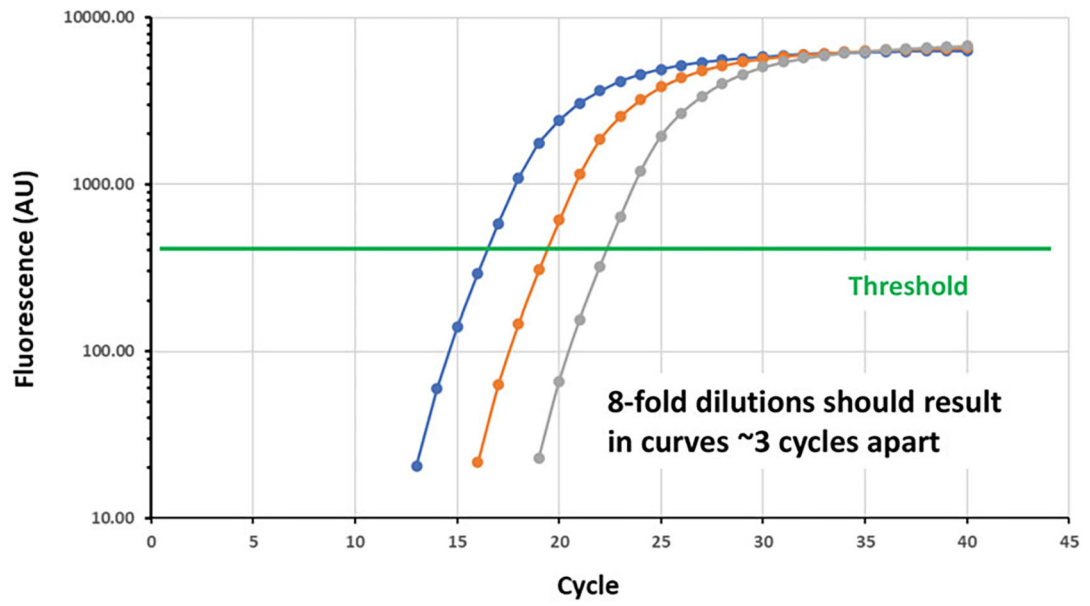


Fig. 3. Amplification curves, corresponding to eightfold dilutions to produce a standard curve for evaluating qPCR primer efficiency. An example of a suitable amplification threshold is also shown

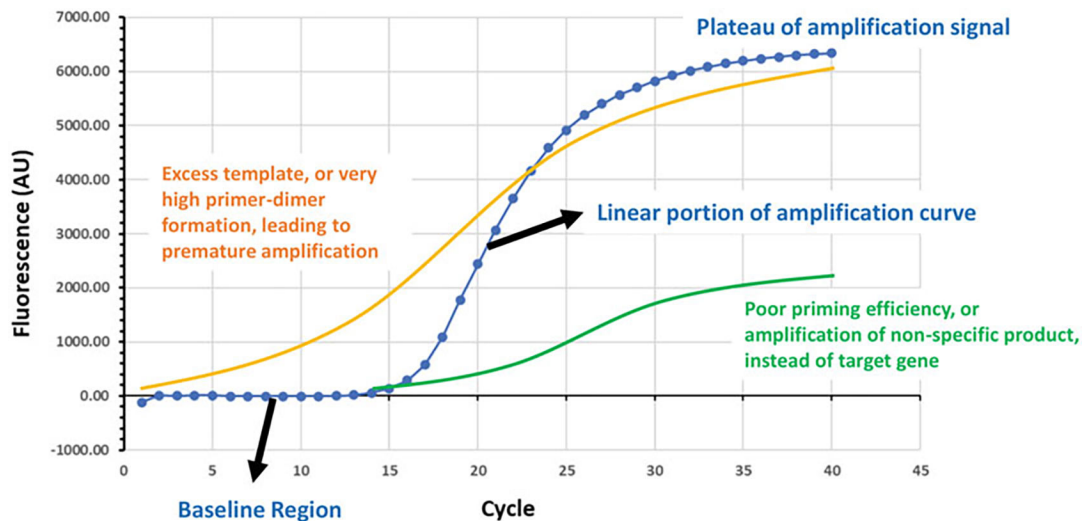


Fig. 4. An example of a typical qPCR amplification curve for an assay performing as designed (Blue), compared to examples (orange, green) of amplification curves associated with commonly encountered technical difficulties

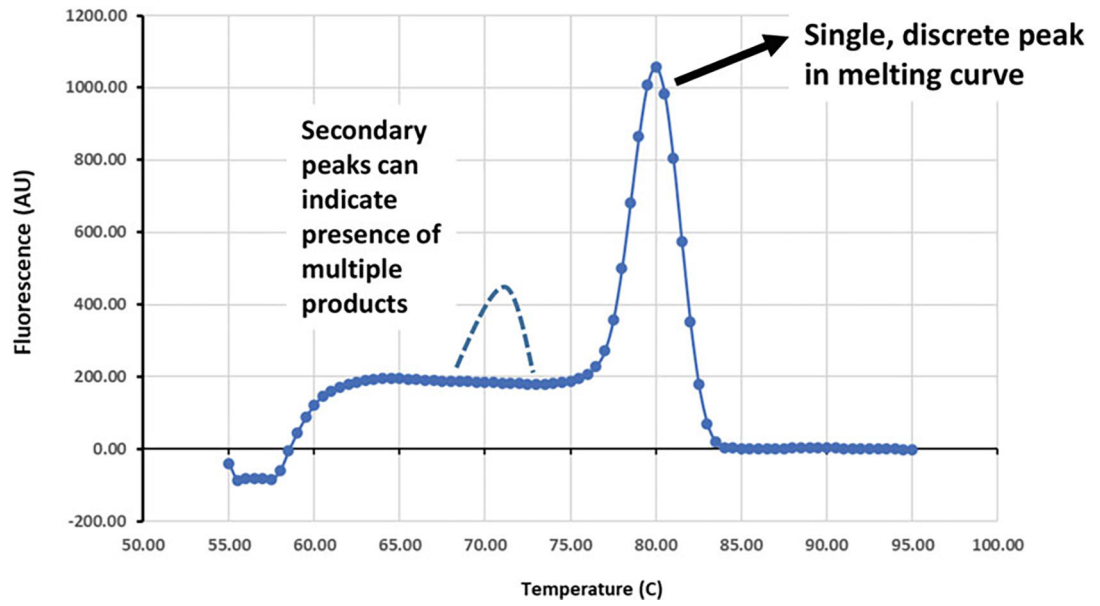


Fig. 5.

An example of a melting curve of qPCR products. A single discrete peak indicates presence of a single product, whereas multiple peaks may indicate presence of primer dimers or other artifacts

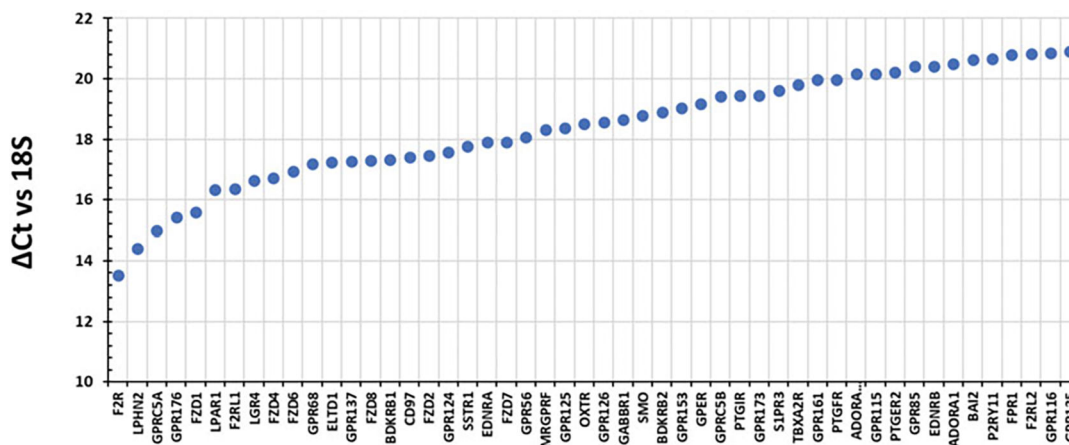


Fig. 6. Sample output data from a TaqMan GPCR array. The key results are the different in cycle threshold (ΔC_t) values compared to 18S rRNA, for all detected GPCRs. The data shown are for the 50 most highly expressed GPCRs in a pancreatic cancer-associated fibroblast sample, from data presented in [4]. For identification of GPCRs expressed in a sample, this is the primary end point of a TaqMan GPCR array experiment

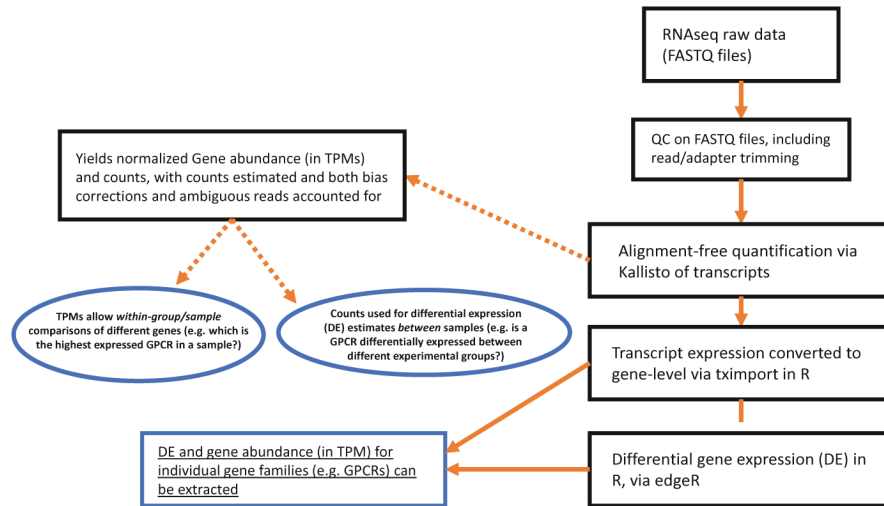


Fig. 7.
A workflow for RNA-seq data analysis

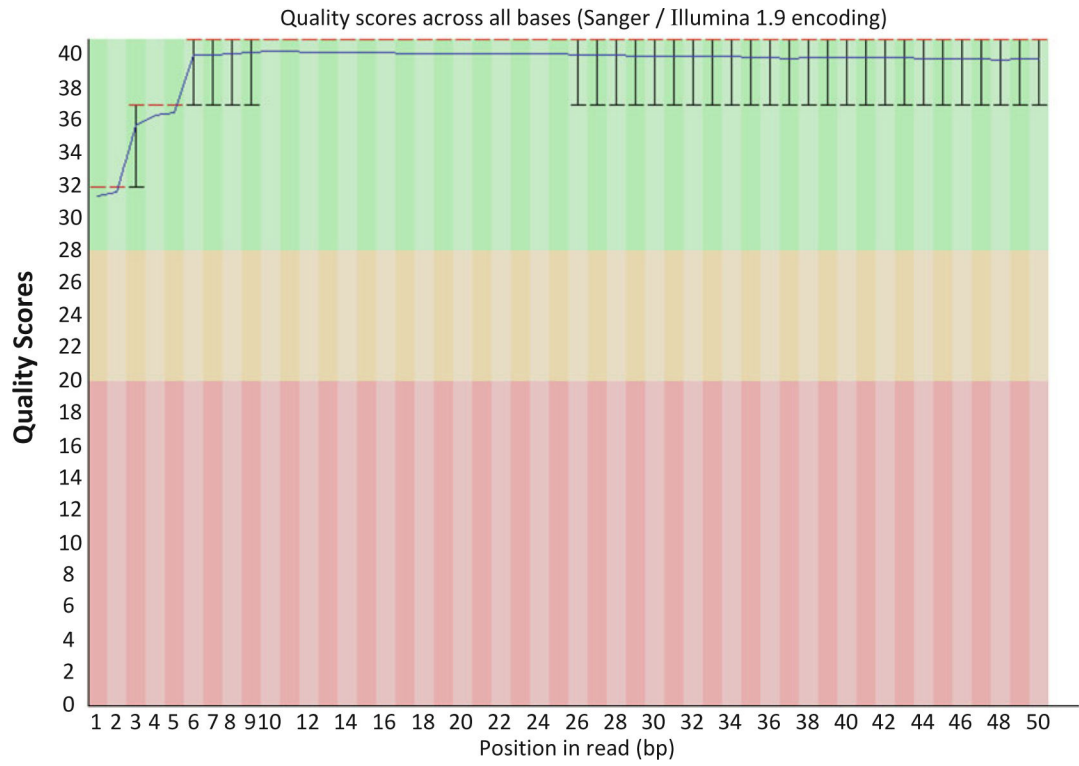


Fig. 8. Example Output, graphing quality scores across the length of reads in the experiment, from the FASTQC software, for RNA-seq raw FASTQ data files. Blue line indicates averages, with ranges highlighted in black

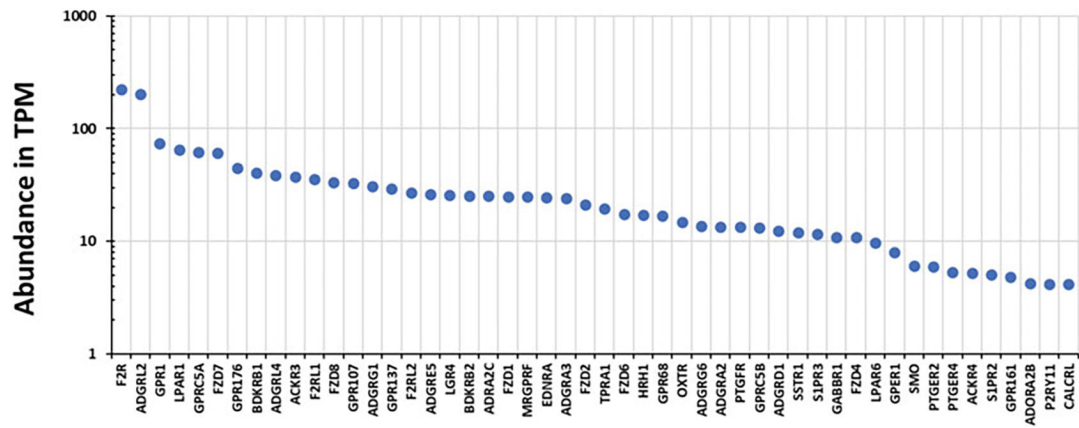


Fig. 9. Expression in TPM for the 50 most highly expressed GPCRs detected via RNA-seq in a pancreatic cancer-associated fibroblast sample, from data presented in [4]. For identification of GPCRs expressed in a sample, this is the primary end point of an RNA-seq experiment