

High-Throughput and Complex Gene Expression Validation Using the Universal ProbeLibrary and the LightCycler® 480 System

King-Hwa Ling^{1,3,4}, Chelsea A. Hewitt¹, Sarah A. Kinkel^{1,3}, Gordon K. Smyth^{2,3}, and Hamish S. Scott^{1,3,4*}

¹Division of Molecular Medicine, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia;

²Division of Bioinformatics, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia; ³Department of Medical Biology, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, Victoria, Australia; ⁴Division of Molecular Pathology, Institute of Medical and Veterinary Science, Adelaide, South Australia, Australia

*Corresponding author: hamish.scott@imvs.sa.gov.au



King-Hwa Ling

Introduction

During the past decade, a considerable number of transcriptome profiling analyses have been reported. High-throughput approaches such as tagging-based serial analysis of gene expression (SAGE) and hybridization-based microarray technologies have been widely performed to profile gene expression [1]. However, these techniques are either laborious or costly, or may involve advanced technologies which limit the sampling depth, thus diminishing the sensitivity of analysis. Validation of gene expression profiles can be rapidly performed using an independent technique such as quantitative polymerase chain reaction (qPCR) to eliminate false results. In most laboratories, qPCR analysis has so far been performed on a small scale. This limits reliable downstream analysis of molecular interactions and pathway predictions, which are becoming important in the emerging field of systems biology. qPCR study design involves consideration of template quality and quantity, various levels of normalization, PCR efficiency (specificity and sensitivity), and data analysis. These factors are vital to ensure a reliable final quantification. The qPCR design becomes difficult when these factors are considered for a study involving multiple tissues, ages, genes, groups (e.g., treatment and control groups) and biological subjects. In this article, we discuss how the Universal ProbeLibrary and LightCycler® 480 System provided flexibility in high-throughput confirmatory qPCR analysis of datasets generated from SAGE and microarray platforms in three complex studies (Table 1).

Materials and Methods

RNA and complementary DNA preparation

Total RNA was extracted from biological triplicates with additional DNase I treatment. The concentration and quality of isolated total RNA were determined using the 2100 Bioanalyzer (Agilent Technologies). Total RNA with an RNA Integrity Number greater than 8.0 was considered for subsequent complementary DNA (cDNA) synthesis [2]. First-strand cDNA was synthesized from 1-5 µg of total RNA using random hexamer priming.

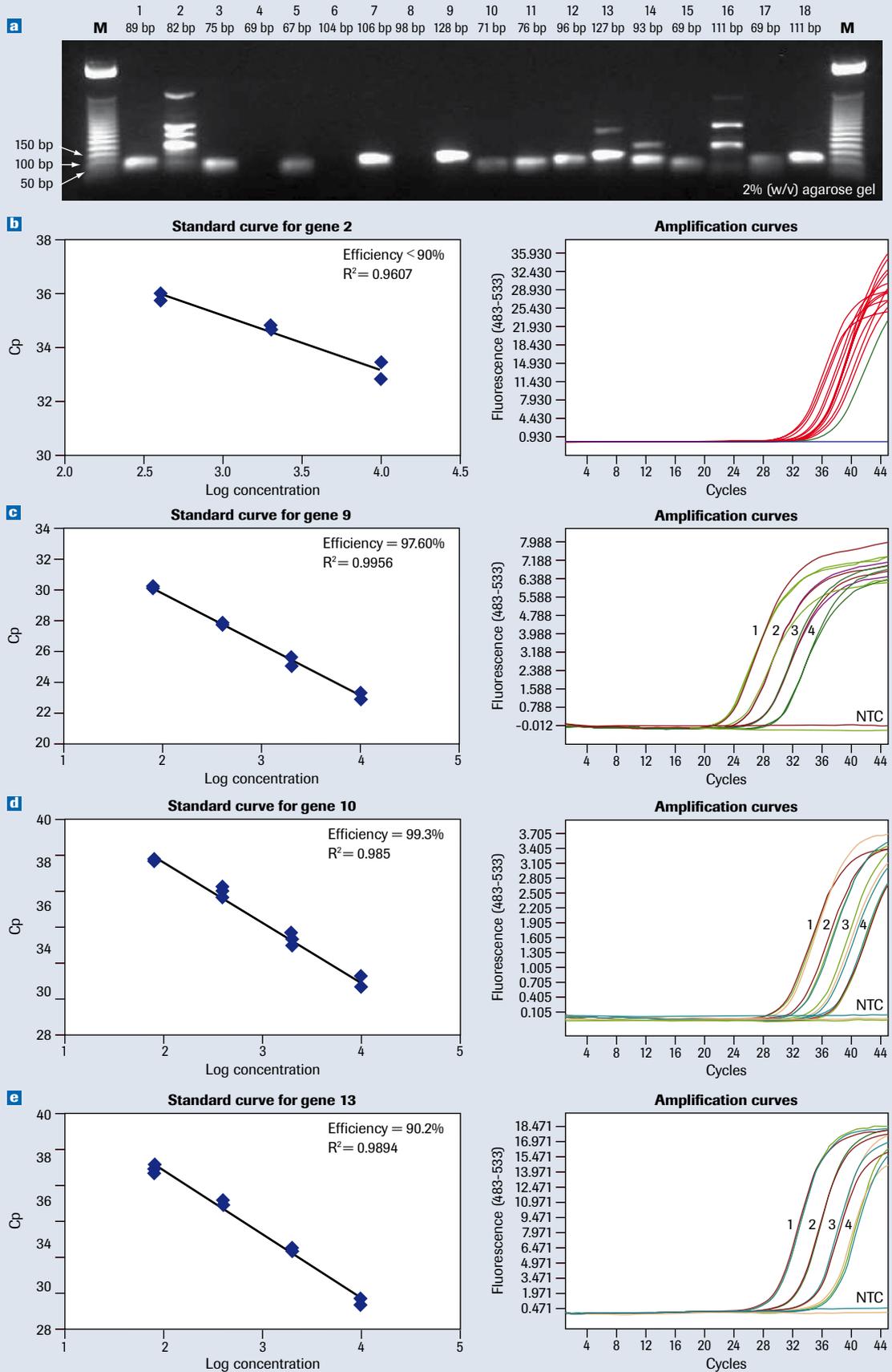
Assay design and qPCR

All primers were designed using the ProbeFinder qPCR assay design software, which is freely accessible at www.universalprobelibrary.com. The following parameters were applied to all primers designed; length = 18-27 bp, GC% = 30-70, T_m = 59-60°C and amplicon size = 50-200 bp. Prior to qPCR, all primer sets were screened using PCR and conventional agarose gel electrophoresis. Only primer sets that generated specific amplicons were used for subsequent qPCR. All qPCR reactions were prepared in a total of 10 µl in a 384-well plate format to the final concentration of 1x LightCycler® 480 Probes Master, 200 nM forward and reverse primers, and 100nM of Universal ProbeLibrary probe. qPCRs were performed using the LightCycler® 480 System with a pre-denaturing step of 95°C for 10 minutes and 45 cycles of 95°C (10 seconds), 60°C (30 seconds), and 72°C (10 seconds) followed by a cooling step at 40°C for 1 minute.

Data analysis

The crossing-point (Cp) value (cycle number in a log-linear region) from each signal was calculated based on the second derivative maximum method performed by the LightCycler® 480 quantification software, which eliminates user interactivity during threshold selection and baseline subtraction. A set of serially diluted cDNAs was used to construct a 4 data-point standard curve for every PCR system. A total of two to three housekeeping genes (Hprt1, Psmb2, Pgk1, or Hsmb) were used as internal controls in order to compensate for variation in inter-reverse transcriptase-PCR (RT-PCR) set up, RNA integrity, RT efficiency, and cDNA sample quantity. The following criteria were adopted to define a successful qPCR assay; PCR efficiency between 90-110%, R-squared > 0.985, and a minimum of two successful housekeeping genes in each run. Based on a successful standard curve, the quantity of starting RNA of both target and reference transcripts was calculated as a linear function of logarithmic concentration and Cp. All estimated starting quantity of transcripts was in arbitrary units. The starting quantity of each target transcript was normalized to the starting quantity of each refer-

Figure 1: (a) Agarose gel screening of selected qPCR assays. The size of each expected amplicon is presented in each lane. Assays 1, 3, 5, 7, 9–12, 15, 17, and 18 generated specific amplicon whereas assays 2, 13, and 14 generated both specific and unspecific amplicons. Assays 4, 6, 8, and 16 are considered failed. M denotes 50 bp DNA Step Ladder (Promega). Panels (b–e) show the standard and amplification curves for assay 2 (b) (specific amplicon and unspecific dominant amplicons), assay 9 (c) (specific amplicon only/highly expressed gene), assay 10 (d) (specific amplicon only/lowly expressed gene) and assay 13 (e) (specific dominant amplicon and unspecific amplicon). Number 1–4 on amplification curves denotes the cDNA stock used (synthesized from 5 µg of either whole brain or testis total RNA) as starting material with 50, 250, 1,250, and 6,250 dilution factors, respectively. NTC denotes no template control.



ence transcript generated from the same biological sample. Generated ratios from each of the target genes from various stages of development or subjects were then analyzed using linear modelling and empirical Bayesian moderated t-statistic provided by limma software package for the R computing environment (www.r-project.org) [3].

Results and Discussion

A total of 252 assays were designed. Of these, 235 (93%) assays generated specific amplicons observed by agarose gel electrophoresis analysis whereas 17 (7%) assays generated unspecific or no amplicons (Table 2). Successful pre-screened assays were used in qPCR analysis to validate differentially expressed transcripts (DETs). Based on previously defined criteria (see methods), 161 (69%) assays were successfully performed using the Universal ProbeLibrary and LightCycler® 480 System under the same amplification conditions (without any optimization). Despite generating a specific amplicon, 74 (31%) assays required further optimization. In most cases, failed assays were due to a low level of transcript expression, which leads to PCR inefficiency (<90% or >110%) and non-reproducible results (R-squared < 0.985). Figure 1 shows the different outcomes of agarose gel analysis during pre-screening assays. Most assays which generated specific dominant amplicons have high PCR efficiency and reproducibility.

By considering only the successful assays, 70 (54%) DET profiles were validated under the stringent analysis in the SAGE brain study. Most of the non-validated DETs in the SAGE study were generated mainly from short tags libraries, which were constructed from only one biological subject; 25 (78%) DET profiles were validated in the microarray brain study. The

microarray datasets were generated from experiments that were performed on different biological subjects (triplicates). Consequently, the validation rate was higher (~80%) and therefore demonstrated better sensitivity compared with the SAGE technique. Depending on the study design or platform, the validation of expression profiles is vital to avoid false-positive results in downstream analysis before any pertinent conclusions or specific hypotheses can be deduced.

In this study, we demonstrated that both the Universal ProbeLibrary and LightCycler® 480 System provide a flexible platform for a successful high-throughput qPCR analysis. We consider there to be four main factors that contribute to a successful and feasible high-throughput qPCR analysis; 1) fast assay design and execution, 2) simultaneous analysis of multiple genes or samples, 3) proper data normalization and handling, and 4) manageable data analysis.

First, quantification of a large number of target genes involves a considerable amount of time in primer design. In addition, it is difficult and time consuming to design intron-spanning primers by manual navigation through various target gene sequences or target genes that are comprised of multiple transcript variants. In our context, the ProbeFinder software provided a batch assay design feature that automatically designed primers and matched a relevant Universal ProbeLibrary probe for up to 10 input sequences or transcript IDs at one time. Furthermore, both common and differentiation assays based on transcript variants can be performed easily. Since the UPL-based qPCR assay requires only standard primer synthesis, the time and cost required from primer synthesis to actual qPCR analysis are reduced compared with pre-validated assays or modified oligonucleotides.

Table 1: Type of gene expression profiling study.

Platform	Gene expression profiling studies	Group ¹	Developmental stage ²	No. of tissues ³	Subjects/group ⁴	No. of datasets ⁵	No. of DETs ⁶
SAGE	Spatio-temporal analysis of developing cerebral cortex	1 (C57BL/6)	4 (E15.5, E17.5, PN1.5, adult)	1-3 (Cx, Hipp & Cer)	1-3	14	1,031
Microarray	Spatio-temporal analysis of postnatally developing mouse brain	2 (C57BL/6 & Ts1Cje)	4 (PN1, PN30, PN60 and PN84)	3 (Cx, Hipp & Cer)	3	72	345
	Aire and autoimmunity	2 (C57BL/6 & KO)	1 (adult)	1 (mTECs)	2	4	1,532

¹ Refers to the mouse strains used in the study. Ts1Cje is a mouse model for Down syndrome whereas KO is an Aire-/- (autoimmune regulator gene) knockout mouse.

² E denotes 'embryonic development day' and PN denotes 'postnatal development day'.

³ Cx denotes cerebral cortex, Hipp denotes hippocampus and Cer denotes cerebellum of the brain, mTECs denotes the 'medullary thymic epithelial cells'.

⁴ Refers to the number of biological replications per group per timepoint per tissue.

⁵ Refers to either the number of library constructed and sequenced for SAGE or the number of microarray experiment performed.

⁶ DETs denotes 'differentially expressed transcripts'. Fisher's exact test and Vencio et al., 2004 [6] approach were used to analyze SAGE datasets with adjusted $p < 0.05$ was considered significant in differential analysis. In microarray analysis, probe set with \log_2 -intensity > 5 , differential analysis with adjusted $p < 0.05$ or t -statistics > 4 or t -statistic < -4 was considered significant. In both studies, the Benjamini & Hochberg, 1995 [7] method for multiple testing correction was carried out to control false discovery rate.

Table 2: Analysis of qPCR assay designed.

Project	No. of gene selected for validation	Agarose gel analysis		No. of successful pre-screening	qPCR analysis		No. of successful assays	No. of DETs validated
		Specific amplicons ¹	Failed PCR ²		Good PCR efficiency ³	R-squared ⁴		
SAGE – brain	195	185	10	185	131	132	129	70 (54%)
Microarray – brain	57	50	7	50	33	39	32	25 (78%)
Microarray – Aire and autoimmunity	87	Work in progress						

¹Specific amplicon is defined as a specific band detected as one or as dominant band among unspecified product in agarose gel analysis.

²Failed PCR is defined as non-detectable specific band in agarose gel analysis.

³Good PCR efficiency is defined as efficiency between 90–110%.

⁴R-squared > 0.985 denotes 98.5% of the plotted data is explained by the regression used for the determination of quantity of starting material in each corresponding reaction.

Second, high-throughput qPCR involves simultaneous quantification of multiple transcripts and samples at the same time. Therefore, a robust universal protocol is required to produce specific, sensitive and comparable assays for each transcript on the same 384-well plate. Our results show ~70% of qPCR assays were successfully carried out using the same amplification condition without any optimization. The Universal ProbeLibrary provides multiple choices of hydrolysis probes, which have a high melting temperature due to the underlying locked nucleic acids chemistry [4]. The high temperature ramping rate of the LightCycler[®] 480 System, the intron-spanning primers, and Universal ProbeLibrary probe chemistry enable qPCR to be performed on cDNA template at a high annealing temperature, efficiently and with low background signal, thus improving sensitivity and specificity (Figure 1e). In addition, the 8- to 9-bases Universal ProbeLibrary probes occur frequently throughout the mouse genome. This provides flexibility in assay or probe selection during experimental design (a replacement assay or probe can be easily selected).

Third, high-throughput quantification of transcripts from multiple tissues at different developmental stages causes difficulty in choosing a housekeeping gene for normalization. To date, no single housekeeping gene has been reported to have consistent expression across different tissues through different developmental stages, at least not in an *in vivo* system. The best solution is to include multiple housekeeping genes for normalization purposes. However, the inclusion of 2–3 housekeeping genes will at least double or triple the total reactions per sample per run. Furthermore, this does not take into consideration the additional reactions necessary to construct standard curves for every qPCR assay. An increase in the number of reactions leads to problems in sample and data handling. Inclusion of reactions for controls, standard curves and all samples is considered crucial to reduce batch-to-batch variation if these reactions are performed on separate plates. In order to include all reactions on the same plate without complicating the post-analysis data handling, systematic placement of samples was used in all the expression profiling studies. This method is only feasible only with the LightCycler[®] 480 System because its

thermal block has been proven to generate homogeneous temperature across the platform [5], and therefore, positional randomization of samples is not required.

Finally, analysis of large-scale qPCR data remains the most challenging part of such a high-throughput project. Data analysis involves data conversion, normalization between biological replicates and with multiple reference genes, and statistical comparison. To date, there is no one-step application that supports our customized workflow and stringent analysis criteria. Therefore, we used a customized pipeline designed in-house (not discussed here) for multiple comparative analyses.

Conclusions

In the development and revolution of qPCR analysis, both the Universal ProbeLibrary and LightCycler[®] 480 System provide an avenue to perform large-scale confirmatory and sensitive high-throughput qPCR analysis. This system generates vast amounts of expression data in a rapid format without compromising the quality or reliability of the final quantification results. ■

References

1. Velculescu VE *et al.* (1995) *Science* 270:484–487
2. Schroeder A *et al.* (2006) *BMC Mol Biol* 7:3
3. Smyth GK (2004) *Stat Appl Genet Mol Biol* 3:Article3
4. Peterson M, Wengel J (2003) *Trends Biotechnol* 21:74–81
5. Hoebeek J *et al.* (2007) *Biochemica* 2:7–9
6. Vencio RZ *et al.* (2004) *BMC Bioinformatics* 5:119
7. Benjamini Y, Hochberg Y (1995) *J R Statist Soc* 57:289–300

Product	Pack Size	Cat. No.
LightCycler[®] 480 Instrument II	1 instrument (96 well)	05 015 278 001
LightCycler[®] 480 Instrument II	1 instrument (384 well)	05 015 243 001
LightCycler[®] 480 Probes Master	5 x 1ml (2x conc.)	04 707 494 001
Universal ProbeLibrary Set, Mouse	1 set with 90 probes	04 683 641 001

