RealTime ready RT-qPCR Assay Development and Qualification

On the RealTime ready Configurator [1], we offer function tested human, mouse, and rat qPCR assays for gene expression quantification based on the unique Universal ProbeLibrary technology (UPL) [2]. The Universal ProbeLibrary is a set of short hydrolysis probes (8–9-mers) containing modified bases (locked nucleic acids, or LNAs) in order to stabilize the binding and thus ensure high flexibility together with good amplicon specificity [2, 4]. UPL probes are labeled at the 5’ end with fluorescein (FAM) and at the 3’ end with a dark quencher dye. UPL assays are compatible with all real-time PCR instruments capable of detecting fluorescein (FAM) or SYBR Green I. Each assay is function tested according to stringent criteria [5]. The RealTime ready assays are available as Focus Panels and Custom Panels (pre-plated and dried-down on LightCycler® 480 Multiwell Plates, 96 or 384), or as liquid single assays. Focus Panels contain assays for predefined targets of a specific biochemical pathway, such as apoptosis, cell cycle, etc. Custom Panels are custom configured with RealTime ready assays of choice. Each assay consists of a primer pair (forward and reverse primer) and a UPL probe. The main advantage of RealTime ready RT-qPCR assays is that each individual assay is designed and function tested for specificity, PCR efficiency, and signal intensity according to stringent criteria and in general consistency with the recently published MIQE guidelines [6, 7].

In this document, we describe the processes involved in on-demand RealTime ready RT-qPCR assay development for human, mouse, and rat target genes, and provide an overview on key parameters of more than 10,000 RealTime ready RT-qPCR assays.

For life science research only.
Not for use in diagnostic procedures.
Production
For each new target, multiple assays are designed (see below). The individual primers for three selected assays are synthesized in-house using a dedicated Dr. Oligo™ Instrument from Biolytic® Inc. Following the synthesis (and after combining forward and reverse primer), each single assay is transferred to a fully automated storage and production framework. In order to safeguard the reagents’ purity and to avoid any contamination, the production facility operates in a dedicated clean room environment. The whole process is controlled by proprietary software. Barcode tracking provides complete traceability of used starting materials, and all pipetting steps of each plate and vial. During the wet chemistry testing workflow, the assays and plates are subjected to identical procedures on the same machine setup as the final products shipped to customers. For all multiwell plates (Custom and Focus Panels), this includes a proprietary drying process that maximizes shelf life and ease of use. To our knowledge, this interlink of end product QC and initial testing QC workflows is unrivaled in the field of customized RT-qPCR assay development, and is thought to warrant excellent performance for an off the shelf product.

Function Testing
Every assay is tested in duplicates with a ten-fold dilution series ranging from 50 ng to 5 pg cDNA (below details) and a no-template control (NTC). This means 12 individual data points for every assay or at least 36 reactions per target gene. If no functional assay could be identified for a target gene, the target is subjected to another design and testing round. At the end, the best assay is selected according to the criteria shown below.

The selection of each RealTime ready assay follows stringent criteria to ensure optimal performance. All assays are tested in RT-qPCR using the LightCycler® 480 System and each RealTime ready assay meets the following criteria:

- PCR efficiency 2.0 +/- 0.2 (equals 100 +/- 10 %)
- Cq of highest cDNA concentration <= 34
- Linear dynamic range of at least 3 logs
- High amplification specificity, no side products in gel analysis
- Sigmoidal amplification curve
- Fluorescence intensity of amplification curves between 5 and 50 fluorescence units

The general performance of RealTime ready assays is demonstrated in Figure 2 for the human peroxiredoxin 3 gene (amplification curves shown). The replicates show almost identical Cq values (see Table 1). The distances between the different dilutions are consistent and the steepness of all individual curves is almost identical.

The standard curve shows that the replicates are all on a line and that efficiency meets the above-mentioned criteria (see Figure 3 and Table 2).

The efficiency of each RealTime ready assay is checked via agarose gel electrophoresis. Aliquots of the 50 pg and 50 ng dilutions are loaded onto a 4% agarose gel. The assay is selected only if no nonspecific bands (i.e., bands which cannot be explained by, for example, splice variants) are seen. Figure 4 shows that there are no side products detectable for the selected assay for peroxiredoxin 3.

Since all assays can be combined on a RealTime ready Custom Panel, and because we recommend using the dynamic mode for data acquisition, it is obvious that the
fluorescence intensities of all assays have to be similar in a certain range. Therefore, we discard assays with a fluorescence hub above 50 or below 5.

In some instances, for example, when working with certain interleukins or tissue-specific genes or variants, we get late or no Cq signals. In those special cases, we increase the amount of starting material or test with specific material.

The RealTime ready Configurator “Assay Details” page provides detailed information about amplicon length, location of amplicon, and point of start of the primers. The UPL probe and primer sequence is provided online when the products are shipped. Therefore, RealTime ready assays allow MIQE compliant RT-qPCR data generation.

Table 1: Mean Cq and standard deviation of the different dilutions of peroxiredoxin 3.

<table>
<thead>
<tr>
<th>Concentration</th>
<th>Mean Cq</th>
<th>SD Cq</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 ng</td>
<td>22.14</td>
<td>0.019</td>
</tr>
<tr>
<td>5 ng</td>
<td>25.42</td>
<td>0.054</td>
</tr>
<tr>
<td>0.5 ng</td>
<td>28.69</td>
<td>0.030</td>
</tr>
<tr>
<td>0.05 ng</td>
<td>31.86</td>
<td>0.050</td>
</tr>
<tr>
<td>0.005 ng</td>
<td>35.62</td>
<td>0.043</td>
</tr>
<tr>
<td>NTC</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 2: Amplification efficiency and related data.

| Slope | – 3.25 |
|       |       |
| r²   | – 1.00 |
| Efficiency | 2.03 |
| Mean Cq 50 ng | 22.14 |
| Mean Fluo Hub 50 ng | 35.25 |

Figure 3: Standard curve of peroxiredoxin 3.
10-fold dilution series from 50 ng to 5 pg.

Figure 4: Gel analysis of peroxiredoxin 3 assay: 4% agarose gel.
Right lane: an aliquot of the 50 pg dilution is applied to the gel.
Middle lane: aliquot of 50 ng dilution.
Left lane: 50 bp ladder.
Reproducibility

As an example for inter-run reproducibility, we show data from two biological replicates (RNA isolation from 1x10^7 HLR-CHOP cells in two independent reactions). After cDNA synthesis, the samples were applied to three different RealTime ready panels: Reference Gene Panel (19 different target genes, 95 data points), Cell Cycle Regulation Panel (91 different genes, 455 data points), and Apoptosis Panel (378 different genes, 1512 data points). In Figures 5a – 5c, nearly perfect correlations for the biological replicates are demonstrated: r^2 = 0.98 – 0.99.

![Biological replicate correlation: house-keeping gene panel](image)

**Figure 5: Correlation between two biological replicates.** This plot shows the correlation between the first and second replicate for each sample in a) Reference Gene Panel, b) Cell Cycle Regulation Panel, and c) Apoptosis Panel. All data using HLR-CHOP cells were done by J. Zhang et al. [8].

Target Preparation and cDNA Synthesis

As a standardized template for our in-house testing, a commercially available universal total RNA is used for each of the individual species: Human Universal Reference Total RNA (Clontech, Cat # 636538), Mouse Universal Reference Total RNA (Clontech, Cat # 636657), Rat Universal Reference Total RNA (Clontech, Cat # 636658).

To obtain accurate and reproducible results, a high quality RNA is required. The total RNA must be non-degraded and free of residual genomic DNA, RNases, and RT or PCR inhibitors. The quality of the total RNA is checked using an Agilent 2100 Bioanalyzer to verify that the RNA is non-degraded. Only high quality RNA is used for cDNA preparation with the Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science, Cat # 04 379 012 001).

Reverse transcription of RNA can be primed using random hexamers, oligo(dT), or gene-specific primers. However, since the positions of the genes in the RealTime ready assays may be scattered throughout the complete transcripts, we use a combination of random hexamers and oligo(dT) priming to avoid (for example) any 3’ bias in the cDNAs. For our in-house assay development, we reserved a larger amount of a dedicated RNA lot and every newly synthesized cDNA is quality checked against the previous lot. For this purpose, we use three defined control assays: A 5’ and a more 3’ control assay for the same target and another control gene. Each of these assays is tested in the above-mentioned 5-step 10-fold dilution series and the no-template control. The 5’ and 3’ controls check for the quality of the initial total RNA and the efficiency of the reverse transcription step. The differences in Cq values of these assays must be less than one cycle. The Cq value of the third control gene must be in the same range as in the previous lot. With the dilution series, we assess whether the intervals between the different Cq values (concentrations) are equal and close to the theoretically expected 3.32 for a 10-fold dilution series (2^3.32 ≈ 10 for an assay with 100% qPCR efficiency) to make sure that there are no inhibitors present.

The PCR run protocol for the quality check of the cDNA is identical to the one used for all other assays (see Table 3). PCR reactions are run using a LightCycler® 480 Multiwell Plate 384 with 10 µl total reaction volume using the LightCycler® 480 Probes Master.

<table>
<thead>
<tr>
<th>Number of Cycles</th>
<th>Temperature</th>
<th>Hold Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Incubation</td>
<td>1x</td>
<td>95°C</td>
</tr>
<tr>
<td>Amplification</td>
<td>45x</td>
<td>95°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>72°C</td>
</tr>
<tr>
<td>Cooling</td>
<td>1x</td>
<td>40°C</td>
</tr>
</tbody>
</table>

Table 3: qPCR run protocol.
Control Concept
RealTime ready Custom Panels can be ordered with a control concept error detection feature similar equivalent to that used for internal qualification (see above).

RT-Positive Control:
The RT-positive control checks for the quality of the initial total RNA and the efficiency of the reverse transcription step. The control consists of three assays, each targeting different portions of the same transcript. When the Cq values of these three assays differ by more than one cycle, this may be an indicator of either incomplete or inefficient RNA transcription, degraded starting material or other problems with the experimental setup and should be clarified.

RT-Minus Control:
The RT-minus control detects residual genomic DNA. The control consists of two identical assays in two wells; the sample in the first well is transcribed cDNA, the sample in the second well is untranscribed RNA. When the Cq values of these two assays differ by less than ten cycles, or the absolute Cq value for the untranscribed RNA sample is less than 32, this result indicates significant DNA contamination.

Reference Genes:
In addition, each Custom Panel contains up to 3 species-specific reference genes, selectable by the user on the RealTime ready Configurator. These reference gene assays enable the user to perform relative quantification. To ease the selection of suitable reference genes, we also provide a human reference gene panel with 19 reference genes.

More than 10,000 RealTime ready assays have been qualified according to the guidelines described above for human, mouse, and rat target genes. The following section presents a first statistical overview of the RealTime ready qPCR assay set. Unless explicitly stated otherwise, the data, graphs, and visualizations are based on a snapshot of a corpus of more than 10,000 human, mouse, and rat RealTime ready assays or a representative subset of the ~6,000 most recently developed assays. Visualizations were created using the R-project statistical software package [9, 10].

A Census on 10,000+ Developed RealTime ready Assays

Selecting the Design Sequence
Our constantly growing qPCR assay stock currently consists of more than 10,000 pre-tested RealTime ready assays. If no RealTime ready assay is available in the Configurator for a particular gene, a customer can trigger a new design development and testing workflow for the new target. The first step in the RealTime ready qPCR assay design process is the identification of a suitable design sequence for the transcript.

Most of the human protein coding genes are structurally organized in exons and introns. Alternative splicing describes the tissue-specific and developmental stage-specific removal and retention of individual exons by the cellular splicing machinery during pre-mRNA processing; it is one of the key foundations for the functional plasticity of the human transcriptome [11, 12]. In general, the RealTime ready qPCR assay concept follows a “one gene, one assay” paradigm. In rare cases where multiple isoforms are present, the discrimination of individual splice variants might be mandatory for the biological function.

As a general rule, and in order to cover the targets for a broad range of applications and a widespread variety of sample materials, we identify a “canonical” transcript that is used for the assay design. The design sequence selection algorithm weighs and utilizes the different “importance” inherited by different annotations from varying public databases. For example, a matching entry in UniProt [13] for one particular transcript identifier will naturally outweigh transcript IDs without any additional cross-references. For more than 80% of the genes covered by RealTime ready assays, all RefSeq [14] annotated isoforms are covered (data not shown).
Assay Design

As mentioned before, all RealTime ready qPCR assays are based on UPL technology [3, 4 and 2]. A normal ProbeFinder design [15] usually takes advantage of using an organism-optimized subset of 90 out of 165 UPL probes. In the case of human assay designs, for example, this leads to more than 600,000 possible assays. Each probe covers ~7,000 transcripts on average and approximately 16 probes bind on one particular transcript. For the human probe set, this provides 99% coverage of all transcripts by at least one UPL probe [3, 2]. In order to get the best possible coverage and flexibility during RealTime ready assay design, we use all 165 probes for all organisms, resulting in even higher numbers and design success rates.

The multistep ProbeFinder design process starts by placing the UPL probes onto the design sequence. This step is followed by identifying suitable primers in the flanking regions of each probe by the Primer3 algorithm [16, 17]. Further processing and fine tuning (for example, high sequence specificity by in silico PCR, checking probe and primer fit/complementarity, and avoiding SNPs) leads to a ranked primer pair list for each probe.

For each new target, at least 3 assays are manually selected for testing. If possible, the different assays are selected to span different introns; no further restrictions are applied. These guidelines, as well as the QC based on experimental assay performance, form the enabling prerequisite for all RealTime ready assays to be combined and used with the same LightCycler® 480 Instrument run protocol.

Assays are defined as intron spanning if at least one exon/exon border (annotated in Ensembl [18, 19] or identified by BLAST-based sequence comparison) is either directly covered by one primer or contained between the primer binding sites. Compared to the median spanned intron size (2.1kb), the median RealTime ready assay amplicon size (75bp) is approximately 30 times shorter (see Figure 6 and 7). In conjunction with a short amplification time, this size difference can be exploited to gain specificity for mRNA-derived cDNA template versus template derived from residual genomic DNA.

In addition, short amplicons are also favorable in experimental setups where the RNA is fragmented or of lower quality. Formalin-fixed, paraffin-embedded tissues are a common example of inevitably compromised RNA integrity [20, 21]. Amplicons with more than 300 bp are known to be problematic in these setups, and the general recommendation is to aim for amplicons that are as short as possible (see e.g. Bustin and Nolan in 22). The average design amplicon size of RealTime ready assays is ~81 bp, with 99% of all assays having amplicons equal to or shorter than 132 bp (Figure 7). Having a well-balanced amplicon base composition is also known to be beneficial for good assay performance. The mean and the median GC content for the RealTime ready assay design amplicons is 52%.

### Table 4: ProbeFinder assay design default parameters for Primer3.

<table>
<thead>
<tr>
<th>Default parameters for the Primer3 design are set to:</th>
<th>Length</th>
<th>Tm</th>
<th>Amplicon</th>
<th>Intron Spanning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>18–27 bp</td>
<td>59–61°C</td>
<td>60–150 bp</td>
<td>Yes</td>
</tr>
<tr>
<td>(20 bp optimum)</td>
<td>(60°C optimum)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: Summarized log10 scaled size distribution of introns spanned by human RealTime ready assays. The median spanned intron size is longer than 2.1x10³ basepairs and 90% of all amplicons are spanning introns with more than 600 bp.

Figure 7: Amplicon lengths of designed assays and selected assays. The design-dependent minimum and maximum are 60 and 150 bp, respectively. The upper whisker-boxplot represents the 6,000+ set of selected assays. 50% of all amplicons are equal to or shorter than 75 bp and 99% are equal to or shorter than 132 bps. The lower whisker-boxplot represents a set of 43,000+ random ProbeFinder designs. The reduced amplicon size within the selected subset is largely attributed to the ProbeFinder intrinsic ranking algorithm which prefers short amplicons in concordance with general best practices for qPCR assay design.
**RealTime ready Assay Relative Position within the mRNA**

Assessing the relative positional distribution of our assays along the transcripts, the RealTime ready RT-qPCR assays reveal a slight accumulation towards the 5’ region of the transcripts (see Figure 7). However, this bias is not caused by explicit additional design scoring or selection preferences. It is merely a consequence of preferring intron spanning assays per and reflects a biological phenomenon known from many eukaryotic organisms. Abundance of introns increases toward the 5’ end of transcripts in general [23]. Consequently, the subset of non-intron spanning RealTime ready assays does not show this 5’ positional bias (see Figure 8).

**Amplification Efficiencies Results**

As stated in numerous publications and also recently published as part of the MIQE guidelines, stable amplification efficiency is key to reproducible and trustworthy RT-qPCR results [6, 7]. In the ideal case with 100% efficiency, the amount of generated amplicon doubles with every PCR cycle. The newly generated fluorescence in every cycle due to spatial separation of the quenching dye from the fluorophor by UPL probe cleavage is an indirect and proportional measurement of this amplicon generation. It is generally accepted that efficiencies for qPCR assays should be within the range of 90–110% (equivalent to 1.8 to 2.2-fold increase with each cycle). Several methods have been published to assess the amplification specificity from individual fluorescence curves. See, for example, M. Pfaffl [24, 25] for recent reviews. For RealTime ready assays, we calculate the efficiency from slopes of a calibration curve for each individual assay, as in the equation below.

\[ E = 10^{-1/slope} \]

Only assays with a linear range of >= 3 log10 ratios and no “Non Template Control” signal qualify as RealTime ready assays. During our QC, we tolerate 10% divergence from the optimum efficiency; however most of the assays perform significantly better (see Figure 9). The mean efficiency of the subset of 6,000+ RealTime ready assays is 1.99 with a standard deviation 0.034. 90% of the efficiencies are located within a window from 1.95 to 2.05. However, it is important to note that the PCR efficiency is not only a function of the primer, probe design, or quality, but also depends on the overall experimental setup and sample material.

**Figure 8:** Kernel density distribution of the relative position of the RealTime ready RT-qPCR assays along the transcripts.

The x axis represents the relative position within the transcript from 5’ to 3’ (0–100%). The red curve represents a subset of intron spanning RealTime ready assays whereas the blue line represents a subset of non-intron spanning RealTime ready assays.

**Figure 9:** Distribution of the PCR efficiencies of 6,000+ RealTime ready assays. The mean efficiency of a sample of 6k+ RealTime ready assays is 1.99 with a standard deviation 0.034. 90% of the efficiencies are located within a window from 1.95 to 2.05.
10. R-Project home page: http://cran.r-project.org/

For life science research only.
Not for use in diagnostic procedures.

License Disclaimer information is subject to change or amendment. For current information on license disclaimer, please refer to the Online Technical Support page for that product. (http://technical-support.roche.com).

HIGH PURE, LIGHTCYCLER, MAGNA PURE and REALTIME READY are trademarks of Roche. Exiqon, LNA, ProbeFinder and ProbeLibrary are registered trademarks of Exiqon A/S, Vedbaek, Denmark.
SYBR is a registered trademark of Molecular Probes, Inc. Other brands or product names are trademarks of their respective holders.

Published by Roche Diagnostics GmbH Sandhofer Straße 116 68305 Mannheim Germany

© 2011 Roche Diagnostics. All rights reserved.