Different Approaches to Normalization of Gene Expression RT-qPCR Data

Jan Hellemans, Jo Vandesompele
Ghent University
CEO and founder Biogazelle

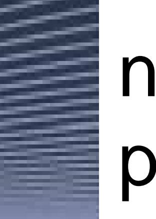
RethinkPCR Scientific Conferences, Europe

Rethink the Way You do Real-time PCR

- normalization
- data analysis
- assay validation

outline part normalization

- problem
- multiple reference gene normalization
  - geNorm
  - other methods
- expressed Alu repeat normalization



normalization: what's the problem?

- gene-specific (biological) variation
- non-specific (technical) variation
  - RNA extraction yield
  - RNA quantity & quality
  - RT efficiency
  - PCR efficiency

normalization: what's the solution (part I)?

Genes and Immunity (2005) 6, 279–284
© 2005 Nature Publishing Group All rights reserved 1466-4879/05 $30.00
www.nature.com/gene

REVIEW

Real-time RT-PCR normalisation; strategies and considerations

J Huggett[1,2], K Dheda[1,2,3], S Bustin[4] and A Zumla[1,2]

[1]Centre for Infectious Diseases and International Health, University College London, London, UK; [2]Royal Free Medical School, London, UK; [3]Department of Thoracic and HIV Medicine, Royal Free Hospital, London, UK; [4]Centre for Academic Surgery, Barts and the London, Queen Mary's School of Medicine and Dentistry, London, UK

Real-time RT-PCR has become a common technique, no longer limited to specialist core facilities. It is in many cases the only method for measuring mRNA levels of vivo low copy number targets of interest for which alternative assays either do not exist or lack the required sensitivity. Benefits of this procedure over conventional methods for measuring RNA include its sensitivity, large dynamic range, the potential for high throughout as well as accurate quantification. To achieve this, however, appropriate normalisation strategies are required to control for experimental error introduced during the multistage process required to extract and process the RNA. There are many strategies that can be chosen; these include normalisation to sample size, total RNA and the popular practice of measuring an internal reference or housekeeping gene. However, these methods are frequently applied without appropriate validation. In this review we discuss the relative merits of different normalisation strategies and suggest a method of validation that will enable the measurement of biologically meaningful results.

Genes and Immunity (2005) **6,** 279–284. doi:10.1038/sj.gene.6364190
Published online 7 April 2005

normalization: what's the solution (part I)?

- sampling size (number of cells, volume or mass of the sample)
  - reproducible extraction yields
  - not always possible (e.g. microdissected tissue)
- total RNA amount
  - not always possible (e.g. embryo)
  - quality (inhibitors)
  - cDNA synthesis efficiency is not taken into account
  - total RNA (rRNA) is not always representative of the mRNA fraction
- spiking (alien RNA)
  - corrects for enzymatic efficiency differences
  - not assumption-free (equal input template)

normalization: what's the solution (part I)?

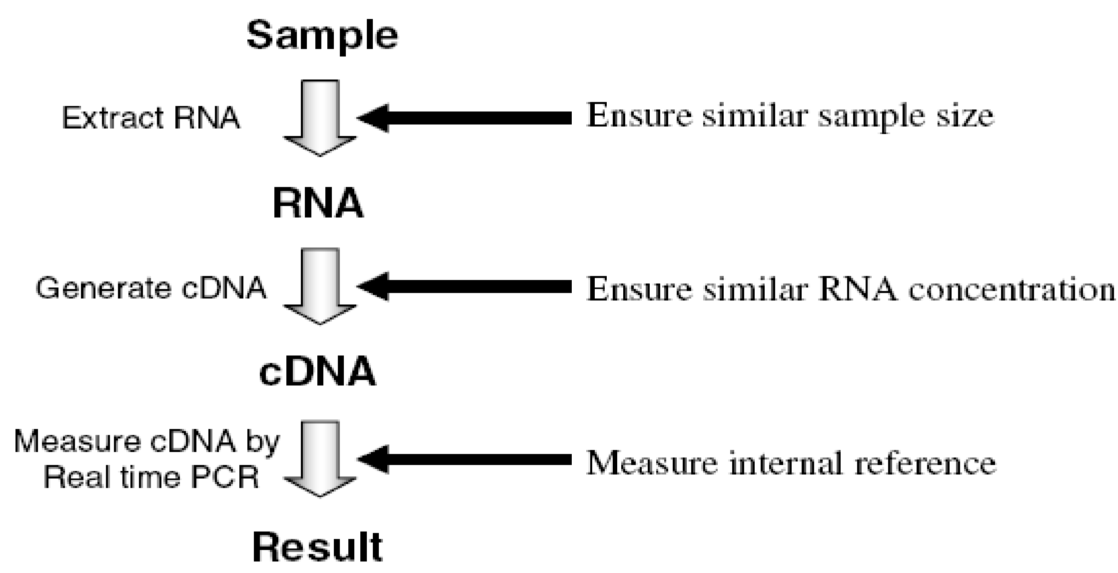
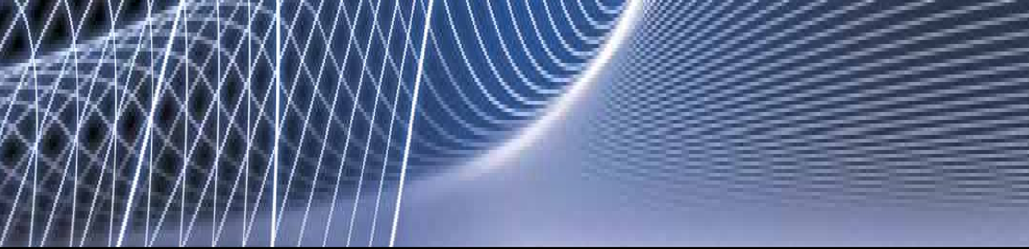


Figure 1 Processes required to generate a real time RT-PCR result. Black arrows indicate points, which should be considered for a good normalisation strategy.



normalization: what's the solution (part II)?

- reference genes
  - most popular
  - capture most variation
- attention!
  - reference genes (might) vary in expression
  - until recently, non-validated reference genes were used (assuming stable expression)
- normalization against 3 or more validated reference genes is considered as the most appropriate and universally applicable method
  - 3rd London qPCR Symposium (April 2005)
  - which genes?
  - how to do the calculations?

normalisation: our geNorm solution

- framework for qPCR gene expression normalisation using the reference gene concept:
  - quantified errors related to the use of a single reference gene (> 3 fold in 25% of the cases; > 6 fold in 10% of the cases)
  - developed a robust algorithm for assessment of expression stability of candidate reference genes
  - proposed the geometric mean of at least 3 reference genes for accurate and reliable normalisation
  - Vandesompele et al., Genome Biology, 2002

Research

Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes

Jo Vandesompele, Katleen De Preter, Filip Pattyn, Bruce Poppe, Nadine Van Roy, Anne De Paepe and Frank Speleman

rethink your reference genes

- quantitative RT-PCR analysis of candidate reference genes

4
3
2
1
0

A B C D E F G

ACTB
HMBS
HPRT1
TBP
UBC

15 fold difference between A and B if normalized by only one gene (*ACTB* or *HMBS*)

geNorm expression stability parameter

- pairwise variation V (between 2 genes)

	gene A	gene B	
sample 1	a1	b1	log2(a1/b1)
sample 2	a2	b2	log2(a2/b2)
sample 3	a3	b3	log2(a3/b3)
...	...	...	...
sample n	an	bn	log2(an/bn)

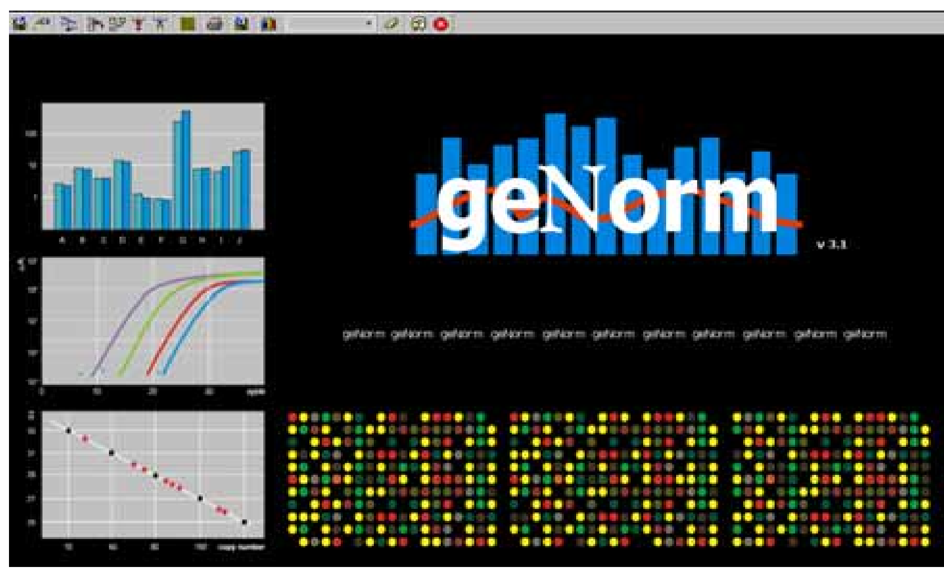
standard deviation = V

- gene stability measure M

  average pairwise variation V of a gene with all other genes

geNorm

- automated analysis
  - ranking of candidate reference genes according to their stability
  - determination of how many genes are required for reliable normalization



http://medgen.ugent.be/genorm

geNorm

- ranking of candidate reference genes according to their stability

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2

HMBS B2M RPL13A SDHA TBP ACTB UBC YHWAZ GAPD HPRT1

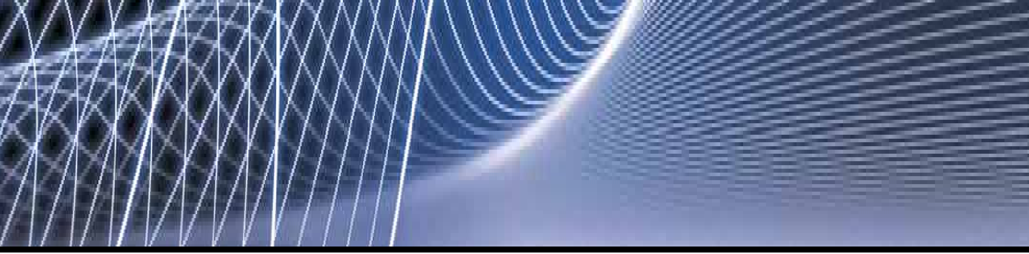
calculation of the normalization factor

- geometric mean of 3 reference gene expression levels

$$\text{geometric mean} = (a \times b \times c)^{1/3}$$

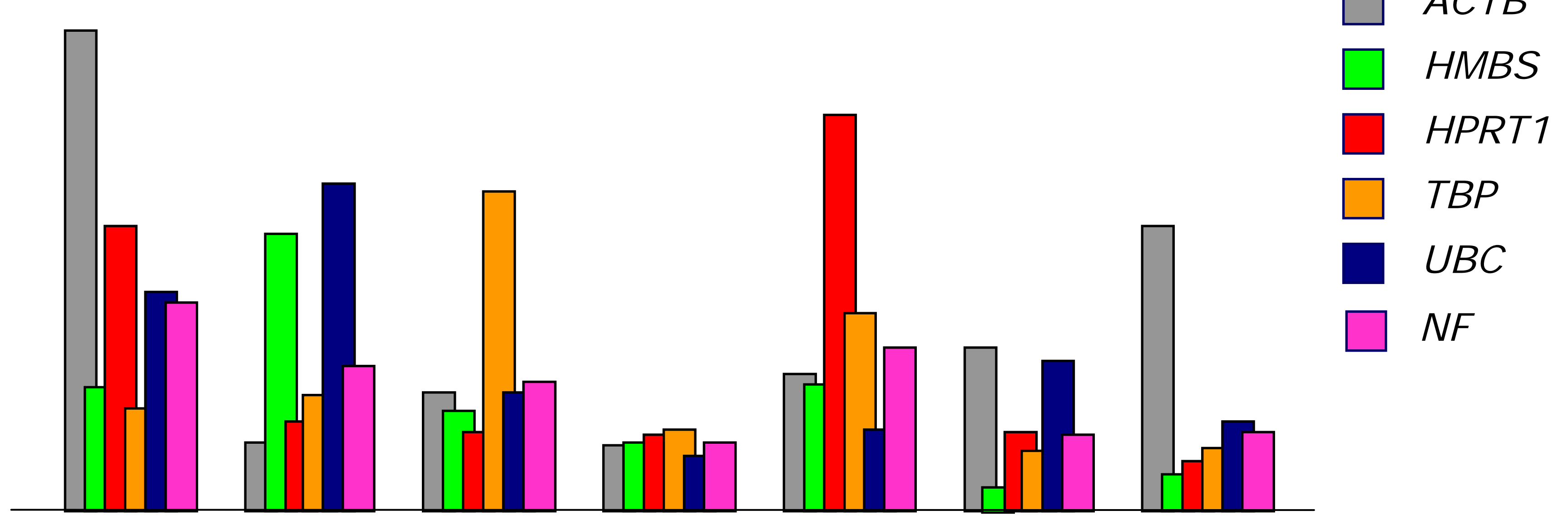
$$\text{arithmetic mean} = \frac{a + b + c}{3}$$

- controls for outliers
- compensates for differences in expression level between the reference genes



geNorm validation (I)

- robust – insensitive to outliers



geNorm validation (II)

- purpose of normalization: reduction of non-specific variation

  only geNorm best reference genes are able to reduce most of the variation

50.0
40.0
30.0
20.0
10.0
0.0

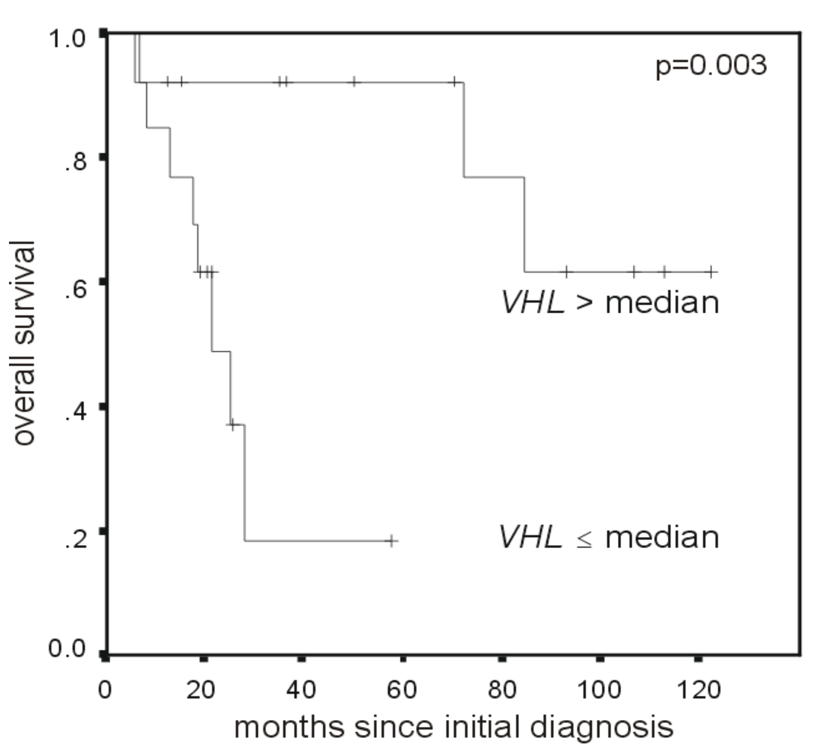
neuroblastoma normal pool leukocyte bone marrow fibroblast

3 good references
3 average references
3 bad references

geNorm validation (III)

- cancer patients survival curve

  statistically more significant results



log rank statistics

NF4

0.003

NF1

0.006
0.021
0.023
0.056

Hoebeeck et al., Int J Cancer, 2006

geNorm validation (IV)

- mRNA haploinsufficiency measurements
  accurate assessment of small expression differences

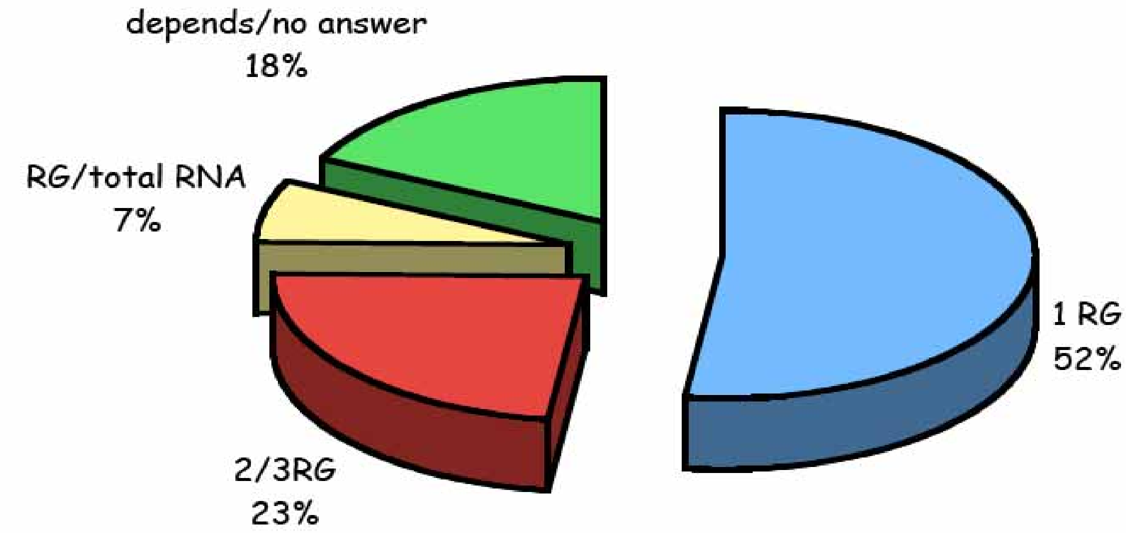
Expression ratios (affected/control)
0
1
Basal BMP4 TGFβ
Id3
Basal BMP4 TGFβ
LEMD3

- patient / control
- 3 independent experiments
- 95% confidence intervals

Hellemans et al., Nature Genetics, 2004

normalisation using multiple stable reference genes

- multiple reference gene normalisation has become common practice:
  - > 1000 citations of our geNorm technology in PubMed
  - > 6000 geNorm downloads worldwide
  - 3rd London qPCR Symposium survey / EMBO 2005 qPCR course



selection of stable reference genes

- other approaches
  - Global Pattern Recognition (Akilesh et al., Genome Research, 2003)
  - BestKeeper (Pfaffl et al., Biotechnology Letters, 2004)
  - Equivalence test (Haller et al., Analytical Biochemistry, 2004)
  - ANOVA test (Brunner et al., BMC Plant Biology, 2004)
  - Normfinder (Andersen et al., Cancer Research, 2004)
  - Szabo et al., Genome Biology, 2004
  - Abruzzo et al., Biotechniques, 2005

  present mathematical (linear mixed-effects) models to analyze candidate reference genes
  $\log y_{ij} = \mu + T_i + G_j + \varepsilon_{ij}$

  - Vandesompele, Kubista & Pfaffl
    Reference gene validation software for improved normalization
    in “Real-time PCR: an essential guide” (Horizon Bioscience, 2nd edition, 2009)

impact of RNA quality on expression stability

- differences in reference gene ranking between intact and degraded RNA

Step*	Degraded RNA (CRS samples)	Intact RNA (CRS samples)	Degraded RNA (NP samples)	Intact RNA (NP samples)
1	HPRT1	GAPD	HPRT1	YWHAZ
2	YWHAZ	YWHAZ	ACTB	B2M
3	B2M	RPL31A	RPL31A	RPL31A
4	TBP	B2M	GAPD	UBC
5	RPL31A	UBC	TBP	GAPD
6	UBC	HPRT1	YWHAZ	HMBS
7	ACTB	TBP	HMBS	HPRT1
8	GAPD	ACTB	SDHA	SDHA
9	HMBS- SDHA	HMBS- SDHA	B2M- UBc	ACTB- TBP

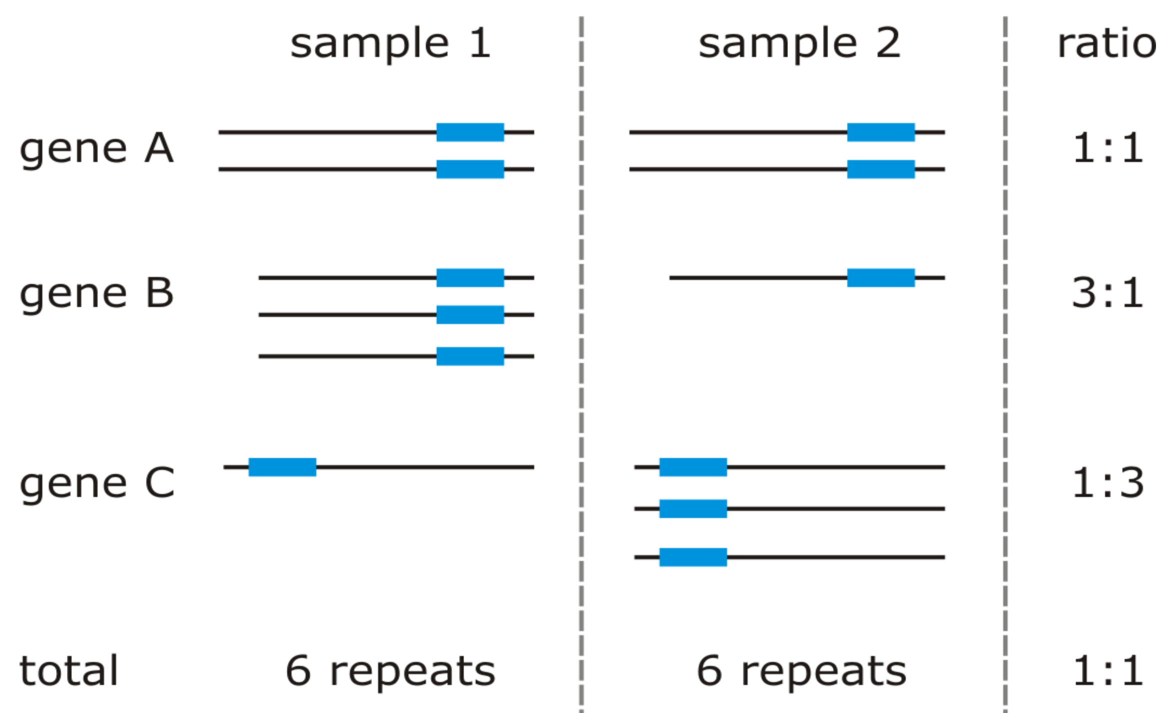
new strategies for normalization

- need for something new – rethink
  - reference gene validation requires (extensive) experimental work
  - sometimes not possible (lack of sample material, funding, time or devotion)

- there must be something better
  - EAR normalization (Expressed Alu Repeat)

    “using a repetitive sequence in the human transcriptome as a measure for the mRNA fraction”

EAR normalization - principle



rationale: repeat sequences are present in the UTR of many genes, and the differential expression of a small number of genes won't influence the overall repeat abundance in the transcriptome

Alu repeat elements

- by far the most abundant repeats in the human genome
- 1 million copies (10% of the genome), 31 subfamilies (well conserved)
- short interspersed elements (SINE) replicating via retrotransposition
- ~280 bp long, followed by a variable poly-A tail
- no known biological function
- implicated in human disease (unequal recombination)

Alu repeat element sequence conservation

```
         1                                                                                             95
Alu      GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCC
AluJo    ........................................................A.....G.T....CC......................G.
AluSx    ........................................................T......................................
AluSq    ........................................................T......................................
AluSp    ............................................................................................A..
AluY     ................................................................--...........A.........T......T
AluYa5   ................................................................--...........A.........T..C...T
AluSx_3  ...............................T.............................TG......C.......................G.
AluSx_5  ........................................................T.....G......C.................C.....G.
AluSq_3  ...............................................................................................
AluSg_4  ........................................................T.......--..............A..............
AluSc_8  ................................................................--...........A.........T......T
AluY_8   ...............................T...................A............--...........A.........T.T....T

         96                                                                                            189
Alu      AACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAA-TTAGCCGGGCGTGGTGGCGCGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCA
AluJo    .....A....G.....A.......A.......A.....-......A...A.........AT.......G........................TG
AluSx    ......................................-....................AT..................................
AluSq    ......................................-...................G....................................
AluSp    .......A..............................-....................AT..................................
AluY     ....C.................................A...................G.........G..........................
AluYa5   ..A.C.................................A............A......G.........G...........T..............
AluSx_3  ......................................-...................T.........G..........................
AluSx_5  ......................................-.............................G..........................
AluSq_3  ......................................-..................AG....................................
AluSg_4  ..G...................................-...................G....................................
AluSc_8  ....C.................................A..................A..........G..........................
AluY_8   ....C.............T...................A.............T.....G.........G...........T..............

         190                                                                                          282
Alu      GGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACA-GAGCGAGACTCCGTCTC
AluJo    ....G.........G...A.....TC....C..........TAT.....T.....................T....-....A....C.T.....
AluSx    ............................................................................-.................
AluSq    ......................................................T.................A...A......A..........
AluSp    ......................................................T.................A...A......A..........
AluY     .......G..G..................C..............................................-.................
AluYa5   .......G..G..................C..................C...........................-.................
AluSx_3  .......T....................................................................-.................
AluSx_5  C.............G...................................................T.....T...-...T.......T.....
AluSq_3  ......................................................T................G....A..........T......
AluSg_4  -......T....................................................................-.................
AluSc_8  ............................................................................-.................
AluY_8   .......G..A......T...........C..................................A......A....-C................
```

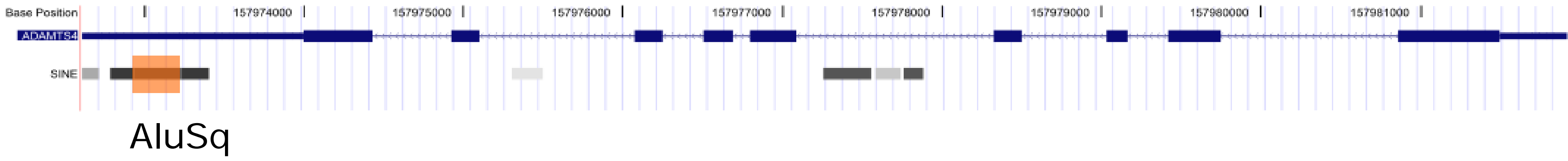
in silico transcriptome analysis

- extraction of all Alu repeat elements in the human genome
  - UCSC genome browser table function
- database with repeat element info and gene structure information for all human genes -> ‘expressed Alu repeats’
  - MySQL
- Alu subfamily sequence alignment
  - PHP script ‘Alu FASTA generator’
  - wEMBOSS clustalW alignment
- primer design
- roughly 1500 human genes contain one or more Alu repeats

AluSx	532
AluJo	250
AluJb	236
AluSq	178
AluY	169
AluSg	161
FLAM_C	102

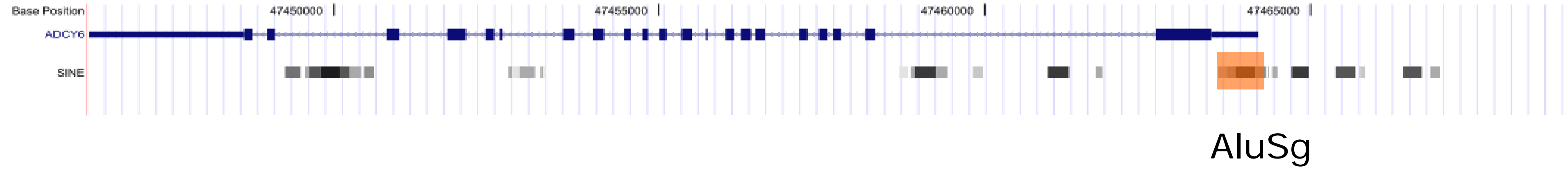
examples Alu containing genes

ADAMTS4 (1q23.3)



AluSq

ADCY6 (12q13.12)



AluSg

Alu repeat assay evaluation

- AluSx assay (AluSq | AluJ)

Correlation Coefficient: 0.997 Slope: -3.514 Intercept: 18.709 Y = -3.514 X + 18.709
PCR Efficiency: 92.6 %

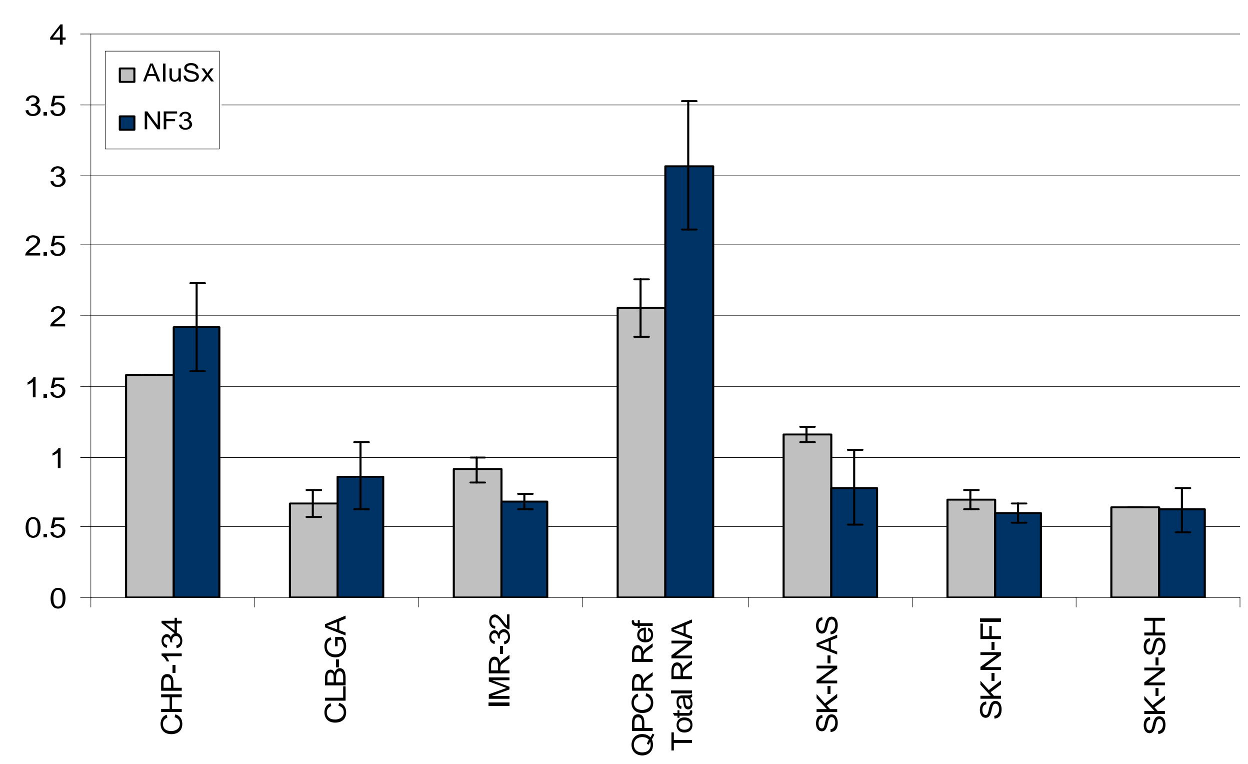
Unknowns
Standards

Threshold Cycle

64, 16, 4 and 1 ng
QPCR Reference Total RNA
(Stratagene)

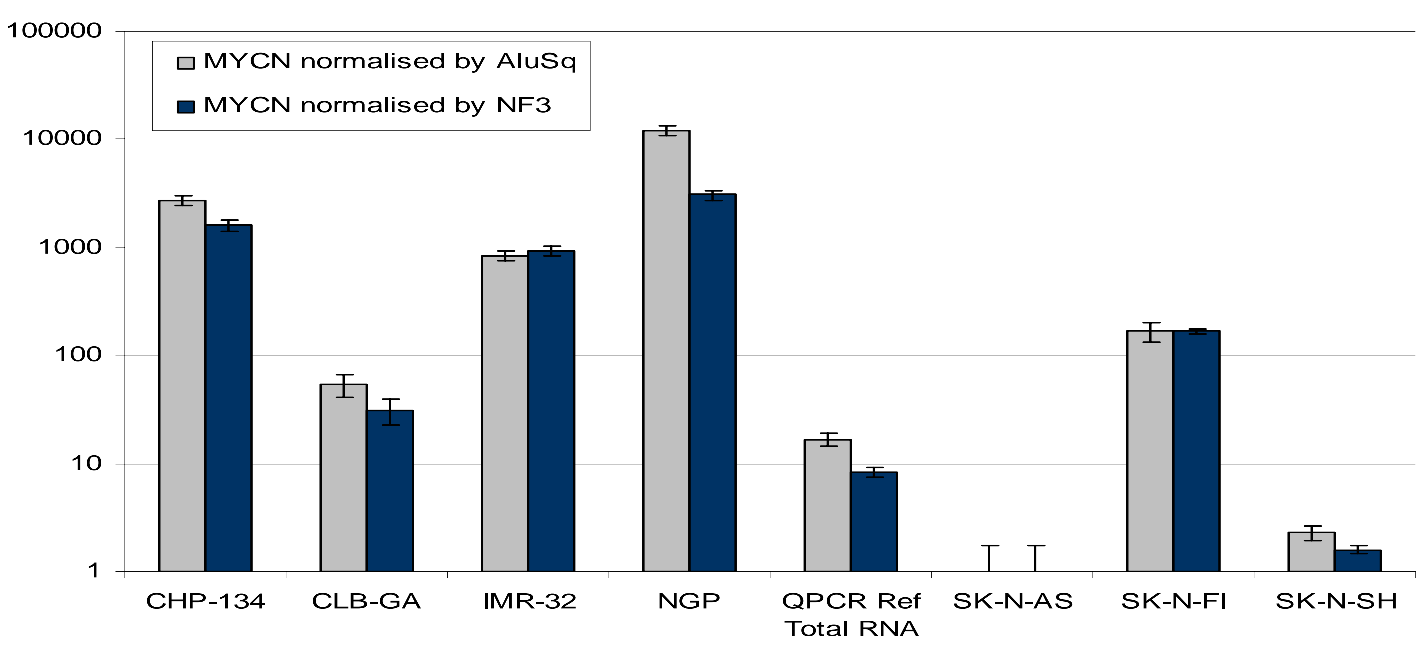
EAR normalization

- comparison of Alu repeat levels and NF based on 3 best reference genes
  Pearsons correlation 0.943 (p=0.0014)



EAR normalization

- MYCN expression levels normalized by Alu repeat or NF3



geNorm ranking of candidate reference genes

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

TATAA-TUBB
TATAA-HPRT1
TATAA-TBP
PD-SF3A1
PD-SDHA
TATAA-GUSB
PD-YWHAZ
TATAA-YWHAZ
TATAA-PPIA
PD-ACTB
PD-TOP1-MyiQ
PD-GAPDH
TATAA-GAPDH
TATAA-ACTB
PD-EIFA2
TATAA-B2M
PD-B2M
TATAA-UBC
TATAA-RPLP
PD-Alu-Sq-Epp
PD-TOP1-Rotorgene
TATAA-RRN18S
PD-18S-ABI7300
PD-UBC
PD-18S-ABI7500
Alu-Sx-Rotorgene
Alu-J-ABI7300
Alu-Sx-MyiQ
Alu-J-ABI7500 Alu-Sq-ABI7500

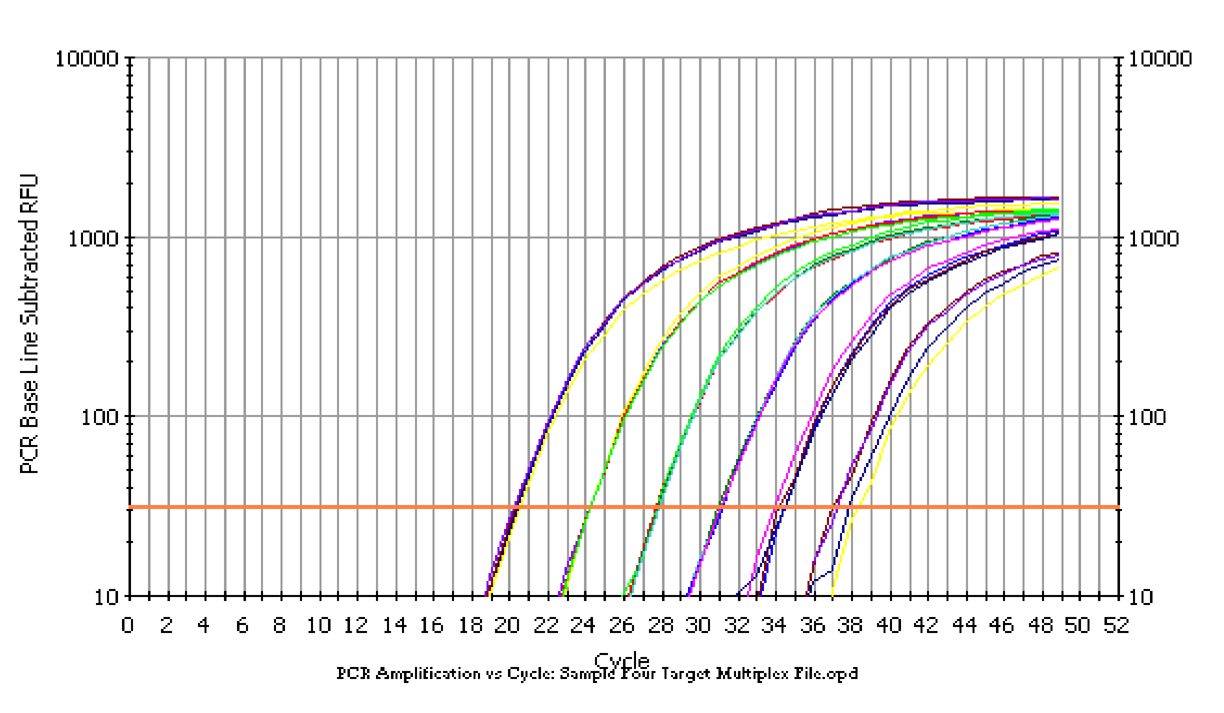
3rd EMBO qPCR course, Heidelberg, 2006

conclusions Alu repeat normalization

- simple and convenient normalization strategy for
  - gene expression analysis (cDNA) (EAR normalization)
  - gene copy number quantification (DNA)
- no (extensive) experimental validation required
- only limited sample amount required
- strategy could be expanded to other expressed repeats in other organisms

problem of data-analysis

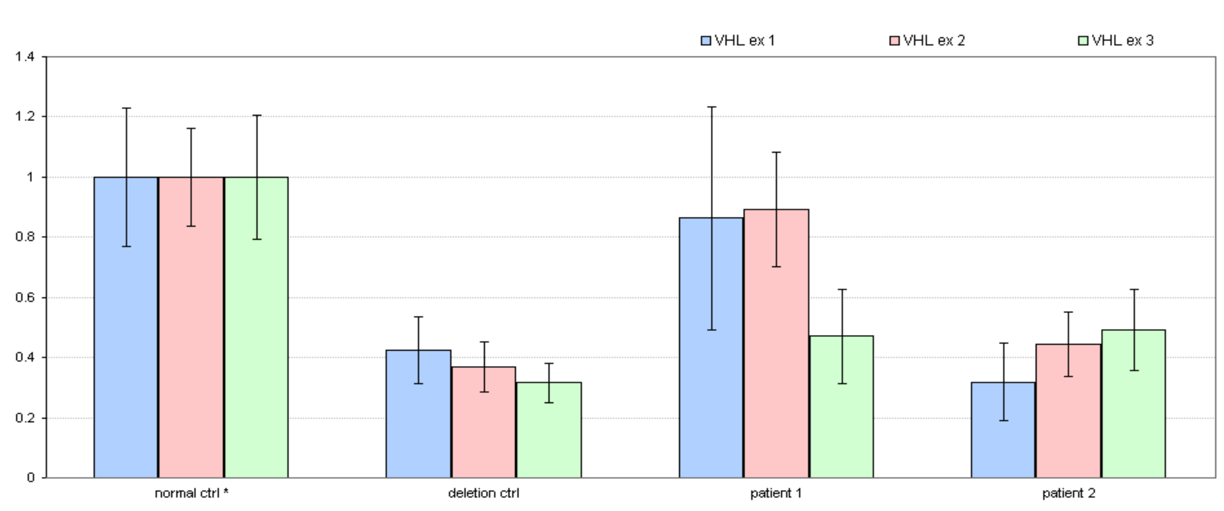
- extraction of meaningful biological information from qPCR data



	A	B		C	D
1	Well	Threshold Cycle	Ct	Set Point	Identifier
2	A5		24		normal ctrl
3	A6		23.8		normal ctrl
4	B5		24.1		normal ctrl
5	B6		23.5		normal ctrl
6	C5		25.6		patient 1
7	C6		25.5		patient 1
8	C11		N/A		
9	C12		N/A		
10	D5		24.7		patient 1
11	D6		25.3		patient 1
12	E5		24.9		patient 2
13	E6		25.3		patient 2
14	F5		25.4		patient 2
15	F6		25.4		patient 2
16	G5		28.9		deletion ctrl
17	G6		29.1		deletion ctrl
18	H5		29.4		deletion ctrl
19	H6		29.3		deletion ctrl
20	A1		24.6		normal ctrl
21	A2		24.6		normal ctrl
22	A11		N/A		
23	A12		N/A		
24	B1		24.2		normal ctrl
25	B2		24		normal ctrl
26	C1		24.5		patient 1
27	C2		24.5		patient 1
28	D1		25.7		patient 1
29	D2		24.9		patient 1
30	E1		26.3		patient 2
31	E2		26.4		patient 2
32	F1		25.8		patient 2
33	F2		27		patient 2
34	G1		28.8		deletion ctrl
35	G2		29.6		deletion ctrl

problem of data-analysis

- extraction of meaningful biological information from qPCR data



	A	B	C	D
1	Well	Threshold Cycle Ct	Set Point	Identifier
2	A5	24		normal ctrl
3	A6	23.8		normal ctrl
4	B5	24.1		normal ctrl
5	B6	23.5		normal ctrl
6	C5	25.6		patient 1
7	C6	25.5		patient 1
8	C11	N/A		
9	C12	N/A		
10	D5	24.7		patient 1
11	D6	25.3		patient 1
12	E5	24.9		patient 2
13	E6	25.3		patient 2
14	F5	25.4		patient 2
15	F6	25.4		patient 2
16	G5	28.9		deletion ctrl
17	G6	29.1		deletion ctrl
18	H5	29.4		deletion ctrl
19	H6	29.3		deletion ctrl
20	A1	24.6		normal ctrl
21	A2	24.6		normal ctrl
22	A11	N/A		
23	A12	N/A		
24	B1	24.2		normal ctrl
25	B2	24		normal ctrl
26	C1	24.5		patient 1
27	C2	24.5		patient 1
28	D1	25.7		patient 1
29	D2	24.9		patient 1
30	E1	26.3		patient 2
31	E2	26.4		patient 2
32	F1	25.8		patient 2
33	F2	27		patient 2
34	G1	28.8		deletion ctrl
35	G2	29.6		deletion ctrl

data-analysis: 3 generations of quantification models

- Livak and Schmittgen (2001)
  100% PCR efficiency, 1 reference gene

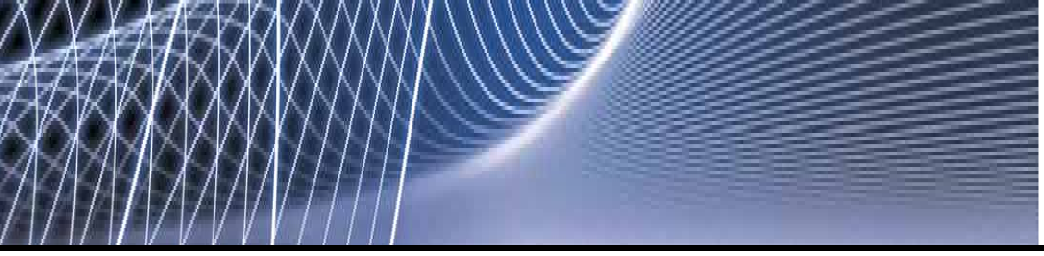
$$NRQ = 2^{\Delta\Delta Ct}$$

- Pfaffl (2001)
  adjusted PCR efficiency, 1 ref. gene

$$NRQ = \frac{E_{goi}^{\Delta Ct,goi}}{E_{ref}^{\Delta Ct,ref}}$$

- qBase model (2007)
  adjusted PCR efficiency & multiple reference genes

$$NRQ = \frac{E_{goi}^{\Delta Ct,goi}}{\sqrt[n]{\prod_{i}^{n} E_{ref_i}^{\Delta Ct,ref_i}}}$$



universal quantification model with proper error propagation

$$s_{e,jl} = \sqrt{\frac{\sum_{q=1}^{h}\left(Cq_{qjl,measured} - Cq_{qjl,predicted}\right)^2}{h-2}} \quad \text{(formula 2)}$$

$$s_{x,jl} = \sqrt{\frac{1}{h-1}\sum_{q=1}^{h}\left(Q_{qjl} - \overline{Q_{jl}}\right)^2} \quad \text{(formula 3)}$$

$$SE\left(slope_{jl}\right) = \frac{s_{e,jl}}{s_{x,jl}(h-1)} \quad \text{(formula 4)}$$

The base for exponential amplification E, and its standard error $SE(E)$ are calculated from these values:

$$E_{jl} = 10^{\left(\frac{-1}{slope_{jl}}\right)} \quad \text{(formula 5)}$$

$$SE\left(E_{jl}\right) = \frac{E_{jl} \cdot \ln(10) \cdot SE\left(slope_{jl}\right)}{slope_{jl}^2} \quad \text{(formula 6)}$$

Conversion of Cq values into relative quantities

Step 1

Calculation of the average Cq value for all replicates of the same gene/sample combination jk within a given run l:

$$\overline{Cq_{jkl}} = \frac{\sum_{i=1}^{n} Cq_{ijkl}}{n} \quad \text{(formula 7)}$$

$$SE\left(Cq_{jkl}\right) = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(Cq_{ijkl} - \overline{Cq_{jkl}}\right)^2} \quad \text{(formula 8)}$$

Step 2

Transformation of mean Cq value into RQ using the gene specific PCR efficiency E_{jl}, with minimization of the overall error:

$$Cq_{reference,jl} = \overline{Cq_{jl}} = \frac{\sum_{k=1}^{s} Cq_{jkl}}{s} \quad \text{(formula 9)}$$

$$\Delta Cq_{jkl} = Cq_{reference,jl} - Cq_{jkl} \quad \text{(formula 10)}$$

$$RQ_{jkl} = E_{jl}^{\Delta Cq_{jkl}} \quad \text{(formula 11)}$$

$$SE\left(RQ_{jkl}\right) = \sqrt{RQ_{jkl}^2\left[\left(\frac{\Delta Cq_{jkl} \cdot SD\left(E_{jl}\right)}{E_{jl}}\right)^2 + \left(\ln\left(E_{jl}\right) \cdot SD\left(\overline{Cq_{jkl}}\right)\right)^2\right]} \quad \text{(formula 12)}$$

Normalization: inter-run calibration

The procedures for normalization and inter-run calibration are highly analogous and are therefore described in parallel.

Step 1

Calculation of the normalization factor NF for sample k based on the RQs of the reference genes p.

Step 1'

Calculation of the calibration factor CF for gene j in run l based on the NRQs of the IRCs m:

$$NF_k = \sqrt[f]{\prod_{p=1}^{f} RQ_{pk}} \quad \text{(formula 13)}$$

$$CF_{jl} = \sqrt[c]{\prod_{m=1}^{c} NRQ_{jlm}}$$

(formula 13'; for definition of NRQ, see formula 15)

$$SE\left(NF_k\right) = NF_k\sqrt{\sum_{p=1}^{f}\left(\frac{SE\left(RQ_{pk}\right)}{f \cdot RQ_{pk}}\right)^2} \quad \text{(formula 14)}$$

$$SE\left(CF_{jl}\right) = CF_{jl}\sqrt{\sum_{m=1}^{c}\left(\frac{SE\left(NRQ_{jlm}\right)}{c \cdot NRQ_{jlm}}\right)^2} \quad \text{(formula 14')}$$

Step 2

Conversion of RQs into NRQs.

Step 2'

Conversion of NRQs into CNRQs:

$$NRQ_{jk} = \frac{RQ_{jk}}{NF_k} \quad \text{(formula 15)}$$

$$CNRQ_{jkl} = \frac{NRQ_{jkl}}{CF_{jl}} \quad \text{(formula 15')}$$

$$SE\left(NRQ_{jk}\right) = NRQ_{jk}\sqrt{\left(\frac{SE\left(NF_k\right)}{NF_k}\right)^2 + \left(\frac{SE\left(RQ_{jk}\right)}{RQ_{jk}}\right)^2} \quad \text{(formula 16)}$$

$$SE\left(CNRQ_{jkl}\right) = CNRQ_{jkl}\sqrt{\left(\frac{SE\left(CF_{jl}\right)}{CF_{jl}}\right)^2 + \left(\frac{SE\left(NRQ_{jkl}\right)}{NRQ_{jkl}}\right)^2} \quad \text{(formula 16')}$$

Coefficient of variation of NRQs of a reference gene

Step 1

Calculation of the mean NRQ for all samples k and a given reference gene p:

$$\overline{NRQ_p} = \frac{\sum_{k=1}^{s} NRQ_{pk}}{s} \quad \text{(formula 17)}$$

R19.12 Genome Biology 2007, Volume 8, Issue 2, Article R19 Hellemans et al. http://genomebiology.com/2007/8/2/R19

$$s_{e,jl} = \sqrt{\frac{\sum_{q=1}^{h}\left(Cq_{qjl,measured} - Cq_{qjl,predicted}\right)^2}{h-2}} \quad \text{(formula 2)}$$

$$s_{x,jl} = \sqrt{\frac{1}{h-1}\sum_{q=1}^{h}\left(Q_{qjl} - \overline{Q_{jl}}\right)^2} \quad \text{(formula 3)}$$

$$SE\left(slope_{jl}\right) = \frac{s_{e,jl}}{s_{x,jl}(h-1)} \quad \text{(formula 4)}$$

The base for exponential amplification E, and its standard error $SE(E)$ are calculated from these values:

$$E_{jl} = 10^{\left(\frac{-1}{slope_{jl}}\right)} \quad \text{(formula 5)}$$

$$SE\left(E_{jl}\right) = \frac{E_{jl} \cdot \ln(10) \cdot SE\left(slope_{jl}\right)}{slope_{jl}^2} \quad \text{(formula 6)}$$

Conversion of Cq values into relative quantities

Step 1

Calculation of the average Cq value for all replicates of the same gene/sample combination jk within a given run l:

$$\overline{Cq_{jkl}} = \frac{\sum_{i=1}^{n} Cq_{ijkl}}{n} \quad \text{(formula 7)}$$

$$SE\left(Cq_{jkl}\right) = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(Cq_{ijkl} - \overline{Cq_{jkl}}\right)^2} \quad \text{(formula 8)}$$

Step 2

Transformation of mean Cq value into RQ using the gene specific PCR efficiency E_{jl}, with minimization of the overall error:

$$Cq_{reference,jl} = \overline{Cq_{jl}} = \frac{\sum_{k=1}^{s} Cq_{jkl}}{s} \quad \text{(formula 9)}$$

$$\Delta Cq_{jkl} = Cq_{reference,jl} - Cq_{jkl} \quad \text{(formula 10)}$$

$$RQ_{jkl} = E_{jl}^{\Delta Cq_{jkl}} \quad \text{(formula 11)}$$

$$SE\left(RQ_{jkl}\right) = \sqrt{RQ_{jkl}^2\left[\left(\frac{\Delta Cq_{jkl} \cdot SD\left(E_{jl}\right)}{E_{jl}}\right)^2 + \left(\ln\left(E_{jl}\right) \cdot SD\left(\overline{Cq_{jkl}}\right)\right)^2\right]} \quad \text{(formula 12)}$$

Normalization: inter-run calibration

The procedures for normalization and inter-run calibration are highly analogous and are therefore described in parallel.

Step 1

Calculation of the normalization factor NF for sample k based on the RQs of the reference genes p.

Step 1'

Calculation of the calibration factor CF for gene j in run l based on the NRQs of the IRCs m:

$$NF_k = \sqrt[f]{\prod_{p=1}^{f} RQ_{pk}} \quad \text{(formula 13)}$$

$$CF_{jl} = \sqrt[c]{\prod_{m=1}^{c} NRQ_{jlm}}$$

(formula 13'; for definition of NRQ, see formula 15)

$$SE\left(NF_k\right) = NF_k\sqrt{\sum_{p=1}^{f}\left(\frac{SE\left(RQ_{pk}\right)}{f \cdot RQ_{pk}}\right)^2} \quad \text{(formula 14)}$$

$$SE\left(CF_{jl}\right) = CF_{jl}\sqrt{\sum_{m=1}^{c}\left(\frac{SE\left(NRQ_{jlm}\right)}{c \cdot NRQ_{jlm}}\right)^2} \quad \text{(formula 14')}$$

Step 2

Conversion of RQs into NRQs.

Step 2'

Conversion of NRQs into CNRQs:

$$NRQ_{jk} = \frac{RQ_{jk}}{NF_k} \quad \text{(formula 15)}$$

$$CNRQ_{jkl} = \frac{NRQ_{jkl}}{CF_{jl}} \quad \text{(formula 15')}$$

$$SE\left(NRQ_{jk}\right) = NRQ_{jk}\sqrt{\left(\frac{SE\left(NF_k\right)}{NF_k}\right)^2 + \left(\frac{SE\left(RQ_{jk}\right)}{RQ_{jk}}\right)^2} \quad \text{(formula 16)}$$

$$SE\left(CNRQ_{jkl}\right) = CNRQ_{jkl}\sqrt{\left(\frac{SE\left(CF_{jl}\right)}{CF_{jl}}\right)^2 + \left(\frac{SE\left(NRQ_{jkl}\right)}{NRQ_{jkl}}\right)^2} \quad \text{(formula 16')}$$

Coefficient of variation of NRQs of a reference gene

Step 1

Calculation of the mean NRQ for all samples k and a given reference gene p:

$$\overline{NRQ_p} = \frac{\sum_{k=1}^{s} NRQ_{pk}}{s} \quad \text{(formula 17)}$$

Genome Biology 2007, 8:R19

qBase paper

Open Access

Method

qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data

Jan Hellemans, Geert Mortier, Anne De Paepe, Frank Speleman and Jo Vandesompele

Address: Center for Medical Genetics, Ghent University Hospital, De Pintelaan, B-9000 Ghent, Belgium.

Correspondence: Jo Vandesompele. Email: Joke.Vandesompele@UGent.be

Published: 9 February 2007

*Genome **Biology*** 2007, **8:**R19 (doi:10.1186/gb-2007-8-2-r19)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2007/8/2/R19

Received: 31 August 2006
Revised: 7 December 2006
Accepted: 9 February 2007

© 2007 Hellemans *et al.*; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hellemans et al., Genome Biology, 2007

qBase*Plus*

- based on Ghent University's geNorm and qBase technology
- up to fifty 384-well plates
- multiple reference genes for accurate normalization
- detection and correction of inter-run variation
- dedicated error propagation
- automated analysis; no manual interaction required

- basic version is free, available from Biogazelle (http://www.biogazelle.com)

run editor (Windows XP)

Biogazelle qBasePlus Projects - Run1 - Biogazelle qBasePlus - C:\Documents and Settings\Jo\workspace

File Edit Window Help

Project Explorer

- Biogazelle qBasePlus Demo Project
  - Annotations
  - Default experiment settings
    - Default experiment calculation pa
    - Default experiment quality contro
  - Experiments
    - Data experiment 1_1
      - Annotations
      - Settings
        - Calculation parameters
        - Quality control settings
      - Targets
      - Samples
      - Runs
        - Run1
        - Run2
        - Run3
        - Run4
        - Run5
        - Run6
      - Quality control
        - Replicates
        - +/- controls
        - Reference genes
      - Intermediate results
        - Amplification efficiency
        - Normalization
        - Interrun calibration
          - Interrun calibrators
          - Calibration factors
      - Analysis
        - Quantity matrix
      - Conclusions
  - Project Conclusions

Run1 | Interrun calibrators | Calculation parameters | Duvel

/ Biogazelle qBasePlus Demo Project / Experiments / Data experiment 1_1 / Runs / Run1

Sample	Target	Quantity	Cq	Sample Type
	Stable			unknown

Clear Wells | Include Wells | Exclude Wells

	1	2	3	4	5	6	7	8	9	10	11	12
A	Sample1 Nonregulat 27,990 unknown	Sample1 Nonregula 28,120 unknown	Sample2 Stable 29,100 unknown	Sample2 Stable 29,030 unknown	Sample3 Stable 28,340 unknown	Sample3 Stable 28,660 unknown	Sample4 Stable 29,920 unknown	Sample4 Stable 29,910 unknown	Sample5 Stable 29,590 unknown	Sample5 Stable 29,700 unknown	Sample6 Stable 28,760 unknown	Samp Stabl 28,4: unkn
B	Sample7 Stable 29,700 unknown	Sample7 Stable 29,440 unknown	Sample8 Stable 28,390 unknown	Sample8 Stable 28,000 unknown	Sample9 Stable 28,350 unknown	Sample9 Stable 28,620 unknown	Sample10 Stable 28,710 unknown	Sample10 Stable 28,550 unknown	Sample11 Stable 28,000 unknown	Sample11 Stable 27,890 unknown	Sample12 Stable 28,530 unknown	Samp Stabl 28,3 unkn
C	Sample13 Stable 33,070 unknown	Sample13 Stable 33,160 unknown	Sample14 Stable 28,340 unknown	Sample14 Stable 27,980 unknown	Sample15 Stable 28,600 unknown	Sample15 Stable 28,980 unknown	Sample16 Stable 29,620 unknown	Sample16 Stable 29,470 unknown				
D	Standard1 Stable 256,000 27,080 standard	Standard1 Stable 256,000 27,000 standard	Standard2 Stable 64,000 29,210 standard	Standard2 Stable 64,000 29,470 standard	Standard3 Stable 16,000 31,340 standard	Standard3 Stable 16,000 31,330 standard	Standard4 Stable 4,000 33,470 standard	Standard4 Stable 4,000 33,290 standard	Standard5 Stable 1,000 38,600 standard	Standard5 Stable 1,000 standard	Water Stable negative con	Wate Stabl nega
E	Sample1 Flexible 23,780 unknown	Sample1 Flexible 24,150 unknown	Sample2 Flexible 24,560 unknown	Sample2 Flexible 24,810 unknown	Sample3 Flexible 24,020 unknown	Sample3 Flexible 24,280 unknown	Sample4 Flexible 25,800 unknown	Sample4 Flexible 25,520 unknown	Sample5 Flexible 25,340 unknown	Sample5 Flexible 25,020 unknown	Sample6 Flexible 24,800 unknown	Samp Flexi 24,5: unkn
F	Sample7 Flexible 25,790 unknown	Sample7 Flexible 26,090 unknown	Sample8 Flexible 23,880 unknown	Sample8 Flexible 23,920 unknown	Sample9 Flexible 24,580 unknown	Sample9 Flexible 24,490 unknown	Sample10 Flexible 24,420 unknown	Sample10 Flexible 24,630 unknown	Sample11 Flexible 23,690 unknown	Sample11 Flexible 23,760 unknown	Sample12 Flexible 24,300 unknown	Samp Flexi 24,0: unkn
G	Sample13 Flexible	Sample13 Flexible	Sample14 Flexible	Sample14 Flexible	Sample15 Flexible	Sample15 Flexible	Sample16 Flexible	Sample16 Flexible				

Properties | Matrix

PCR efficiency correction and results (Windows Vista)

Biogazelle qBasePlus Projects - Amplification efficiency - Biogazelle qBasePlus - C:\Users\Jo\workspace

File Edit Window Help

Amplification efficiency

/ Demo project 1 / Experiments / Demo experiment 1 / Intermediate results / Amplification efficiency

	E computed
Duvel	1,948
Flexible	1,963
GAPDH	1,875
Headache	1,966
Leffe	1,970
Nonregulated	2,054
Palm	1,962
Stable	1,726

Standard Curve

Cq

Quantity

Computed Amplification Efficiency

E : 1,948 SE (E) : 0,005 r^2 : 0,999 slope :

Included	Sample	Quantity	Cq
☑	Standard1	256,000	19,000
☑	Standard1	256,000	18,940
☑	Standard2	64,000	20,840
☑	Standard2	64,000	21,000
☑	Standard3	16,000	22,840
☑	Standard3	16,000	23,060
☑	Standard4	4,000	25,230
☑	Standard4	4,000	25,120

Duvel

/ Demo project 1 / Experiments / Demo experiment 1 / Targets / Targets of interest / Duvel

A.U.

Y axis: Linear, Logarithmic

Ordering: Alphabetic, Small to big, Big to small

Results Properties

RethinkPCR Scientific Conferences, Europe

Rethink the Way You do Real-time PCR

inter-run calibration (Suse Linux)

File Edit Window Help

Normalization factors

/ Demo project 1 / Experiments / Experiment 1 / Intermediate results / Normali

	Nonregulated	Flexible	Stable	Normalization Factor
Sample 1	5.973	4.882	4.810	5.196
Sample 2	2.420	1.951	1.879	2.070
Sample 3	3.963	2.599	3.303	3.240
Sample1	2.412	1.289	1.506	1.673
Sample10	1.093	0.883	1.031	0.998
Sample11	1.881	1.515	1.620	1.665
Sample12	1.285	1.111	1.165	1.185
Sample13	0.002	0.068	0.053	0.019
Sample14	1.457	1.070	1.406	1.299
Sample15	1.226	0.917	0.928	1.014
Sample16	0.644	0.485	0.564	0.560
Sample17	3.036	2.004	2.398	2.444
Sample18	1.411	1.063	0.898	1.104
Sample19	1.888	1.757	1.647	1.761
Sample2	0.733	0.793	0.774	0.766
Sample20	3.407	1.337	1.719	1.986
Sample21	4.228	2.652	3.425	3.374
Sample22	1.848	0.951	1.236	1.295
Sample23	1.489	1.632	1.557	1.558
Sample24	2.044	1.637	2.657	2.071
Sample3	1.543	1.137	1.123	1.254
Sample4	0.570	0.411	0.442	0.469
Sample5	0.651	0.568	0.528	0.580
Sample6	1.104	0.806	1.055	0.979
Sample7	0.702	0.340	0.555	0.510
Sample8	1.888	1.346	1.374	1.517

Amplification effici | Run1 | Run2 | Run3 | Run4 | Interrun calibration

/ Demo project 1 / Experiments / Experiment 1 / Intermediate results / Interrun calibration

Targets
- Asperin
- Duvel
- Flexible
- Headache
- Leffe
- Nonregulated
- Palm
- Stable

Run2

CF : 1.123 ± 0.052

Run2
- 1.383 ± 0.063 (Sample1 - IRC 1)
- 1.081 ± 0.045 (Sample2 - IRC 2)
- 0.948 ± 0.117 (Sample3 - IRC 3)

Run5

CF : 0.939 ± 0.076

Run5
- 1.773 ± 0.296 (Sample 1 - IRC 1)
- 0.320 ± 0.045 (Sample 2 - IRC 2)
- 1.457 ± 0.159 (Sample 3 - IRC 3)

Interrun Calibrators | Calibration Factors

Error Log | Alerts

Message	Source	Reference
A normalization factor could not be calculated for this sample because it lacks a Cq value for one of the reference targets. No results Affected Targets: Stable;	Sample	Demo project 1/Experiment 1/Standard5

results in 24 seconds

biogazelle

Batch import 4 run files (13 se

Batch import 4 run files (13 sec

Select multiple refe for normalization (9

View results for one of the targets (2 seconds)

http://www.biogazelle.com/filmpjes/24seconds.swf

RTPrimerDB 2003

- Nucleic Acids Research, Vol. 31, No. 1 122-123, 2003
- http://medgen.ugent.be/rtprimerdb

- database of experimentally validated real-time PCR assays
  - SYBR Green I (61%), TaqMan (38%), others (1%)
  - human (72%), mouse (16%), rat (10%), others (2%)
  - gene expression (97%), others (3%)

- **standardization**
- **time saving**

gene expression assay viewer

Gene Information

MYCN: v-myc myelocytoma
Alias Symbol(s): MODED, N-
Organism: ***Homo sapiens*** (
Entrez Gene ID: 4613
Ensembl Gene ID:

RTPrimerDB gene express
1 transcript is known for MYC

ENST00000281043 (Exons

assays
peptide domain
transcript 5' 3'
SNP positions
1 500 1000 1500 2000 2500 bp

ubiquitin
Home Sear

Figure 1: 1r4
Cell cycle
Appbp1-uba3-
e1-ubiquitin-lik

Key:

Domain	Chain	Start Residue	End Residue
ThiF	A	29	165
UBACT	B	248	318

e! Ensembl Hum
Ensembl v32 - Jul 2005

ENST00000339647

- Gene info.
- Gene splice site image
- Gene variation info.
- Genomic sequence
- Exon info.
- Transcript info.
- Export data
- Peptide info.

Chromosome 12
123,921,074 - 123,924,077

- View of Chromosome 12
- Graphical view
- Graphical overview
- Export information about region
- Export sequence as FASTA
- Export EMBL file

NCBI Entrez Gene

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books OMIM

My NCBI
[Sign In] [Register]

Search Gene for Go Clear current records only

Limits Preview/Index History Clipboard Details

Display Graphics Show 5 Send to

All: 1 Genes Genomes: 1 SNP GeneView: 1

1: UBC ubiquitin C [*Homo sapiens*]
GeneID: 7316 Locus tag: HGNC:12468; MIM: 191340 updated 07-Aug-2005

Summary

Official Symbol: UBC **and Name:** ubiquitin C **provided by** HUGO Gene Nomenclature Committee
Gene type: protein coding
Gene name: UBC
Gene description: ubiquitin C
RefSeq status: Validated
Organism: *Homo sapiens*
Lineage: *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo*
Gene aliases: HMG20

Transcripts and products

to other proteins, where these modifications alter the function, location or trafficking of the protein, or targets it for destruction by the 26S proteasome PUBMED:15454246. The terminal glycine in the C-terminal 4-residue tail of ubiquitin can form an isopeptide bond with a lysine residue in the target protein, or with a lysine in another ubiquitin molecule to form a ubiquitin chain that attaches itself to a target protein. Ubiquitin has seven lysine residues, any one

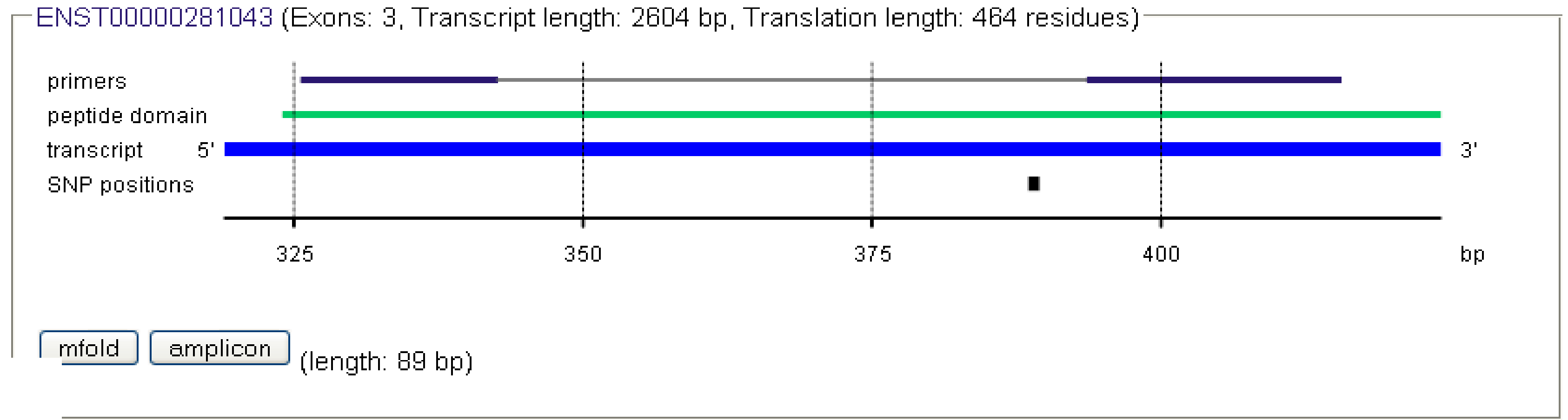
Assay Details

Application: Gene Expression Quantification/Detection GXP

Detection: SYBR Green I SYB

Template: cDNA

This primer pair amplifies part of the following transcript:



Forward Primer: CCGGGCATGATCTGCAA (17 bp)

Reverse Primer: CCGCCGAAGTAGAAGTCATCTT (22 bp)

Annealing Temperature: 60 °C

BLAST primers/probes

assay details (2)

Publication

PubMed ID: 12545167

Submitter's Remarks

primers in coding sequence of MYCN

Submitter

Jo Vandesompele (joke.vandesompele@ugent.be)
Ghent University Hospital, Center for Medical Genetics
De Pintelaan 185, 9000 Ghent, Oost-Vlaanderen, Belgium
http://medgen.ugent.be

Users' Feedback [login to add feedback]

13-SEP-05 - Filip Pattyn , Belgium - (filip.pattyn@ugent.be)

Melt curve: single peak | **Agarose gel**: correct band |

Amplification efficiency: 95% - 100% | **Template**: single sample cDNA | **Number of dilution points**: 4 | **Dilution factor**: 4-fold | **Ct range**: 25-31 | **Correlation coefficient**: 0.999

RTPrimerDB update 2006

- Nucleic Acids Research, 2006, Vol. 34, D684-D688
- *in silico* assay evaluation pipeline
  - SNP or sequence errors
  - secondary structure analysis

D684–D688 Nucleic Acids Research, 2006, Vol. 34, Database issue
doi:10.1093/nar/gkj155

RTPrimerDB: the real-time PCR primer and probe database, major update 2006

Filip Pattyn, Piet Robbrecht, Anne De Paepe, Frank Speleman and Jo Vandesompele*

Center for Medical Genetics Ghent (CMGG), Ghent University Hospital, De Pintelaan 185, 9000 Ghent, Belgium

Received September 15, 2005; Revised and Accepted October 31, 2005

in silico assay evaluation pipeline (1)

***in silico* assay evaluation** (step 1: sequence input)

Fill in the input fields

Organism:	Homo sapiens	
Gene Symbol/Name:	PSAP	◉ Substring ○ Exact phrase
Gene ID:		
forward primer sequence	GGCTTTCCCGTGTCTTTCC	
reverse primer sequence	CTTCGGAGAGCTAGCAGGTTACA	
probe 1 sequence		(optional)
probe 2 sequence		(optional)
Template	cDNA	
Annealing temperature (°C)	60	(between 0 en 100)

Sequence Evaluation

in silico assay evaluation pipeline (2)

In silico assay evaluation (step 2: sequence verification)

***In silico* assay ID: 237**

Gene Information

PSAP: prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)

Organism: ***Homo sapiens*** (Hs, Human)

Entrez Gene ID: 5660

Assay Details

Template: cDNA

Annealing Temperature: 60 °C

Sequence Evaluation

Forward Primer:

your forward primer: GGCTTTCCCGTGTCTTTCC (19 bp) **perfect match with reference sequence**

SNPs: none detected

Reverse Primer:

your reverse primer: CTTCGGAGAGCTAGCAGGTTACA (23 bp) **perfect match with reference sequence**

SNPs: none detected

BLAST primers/probes

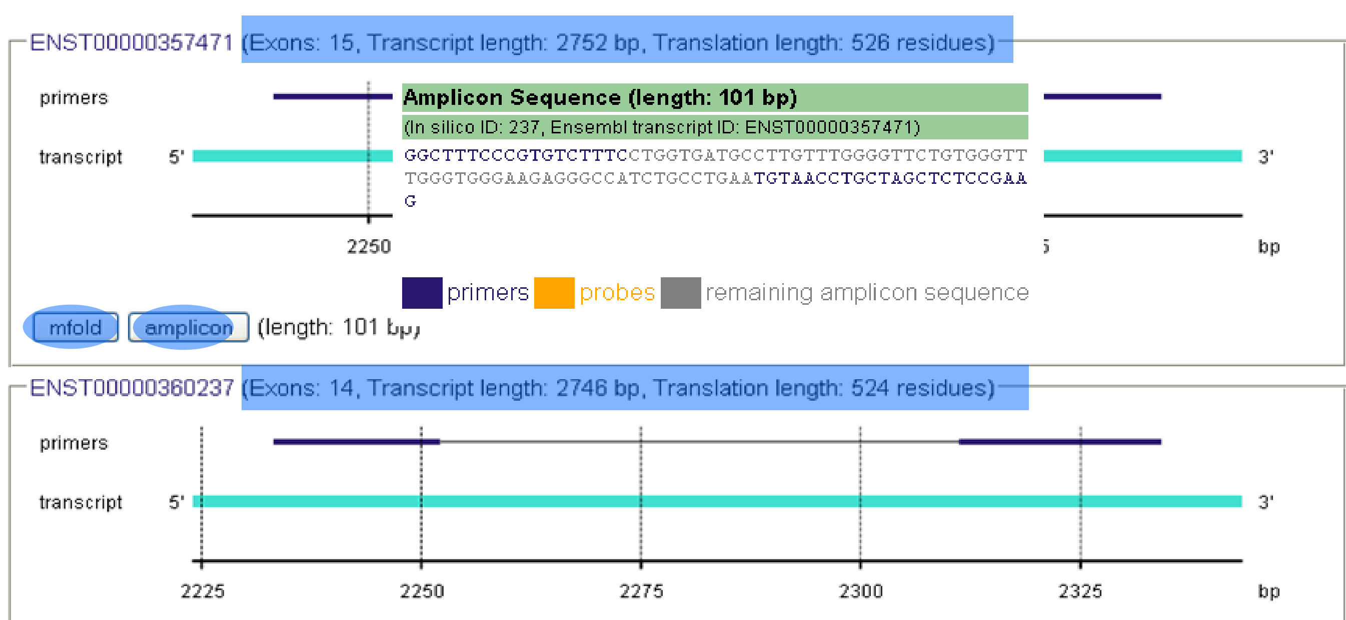
Alignment attempts on 2 PSAP transcript variants:

ENST00000357471 : alignment succesfull

ENST00000360237 : alignment succesfull

Change input | proceed

in silico assay evaluation pipeline (3)



in silico assay evaluation pipeline (unafold)

- unafold
- Zuker et al., Nucleic Acids Research, 2003

mfold results

RTPrimerDB ID: 8, Ensembl transcript ID: ENST00000339647

Temperature (°C)	60	(between 0 en 100)
Mg concentration (mM)	3	(≤ 100)
Na Concentration (mM)	50	(≥ 10) and (≤ 300 if Mg conc is > 0)

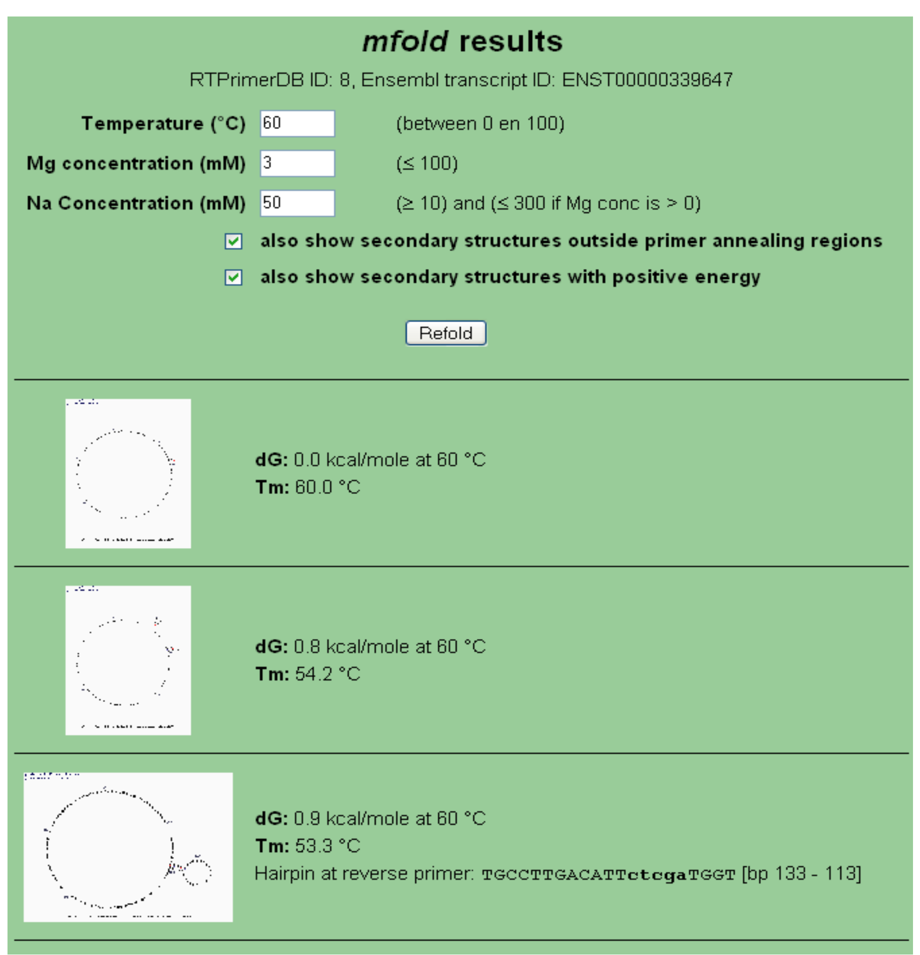
☐ **also show secondary structures outside primer annealing regions**

☐ **also show secondary structures with positive energy**

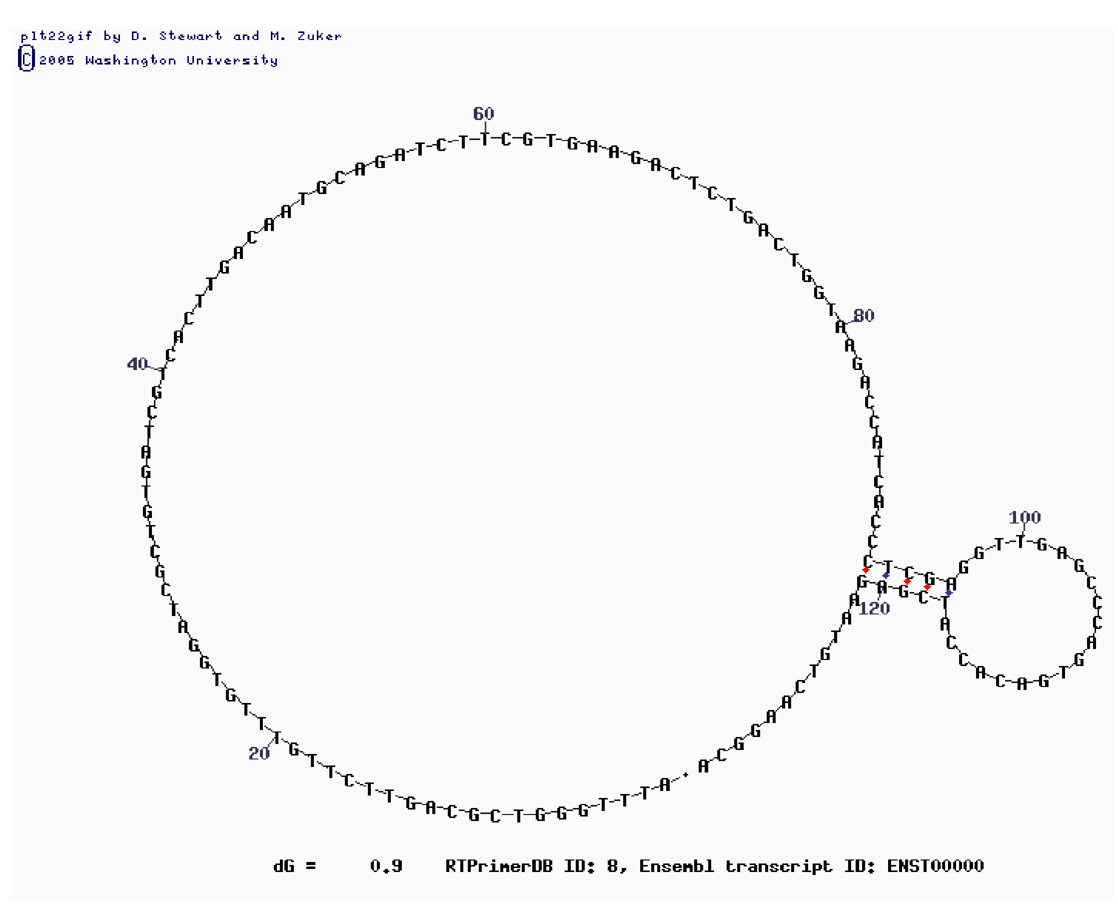
Refold

No significant secondary structure where primers anneal

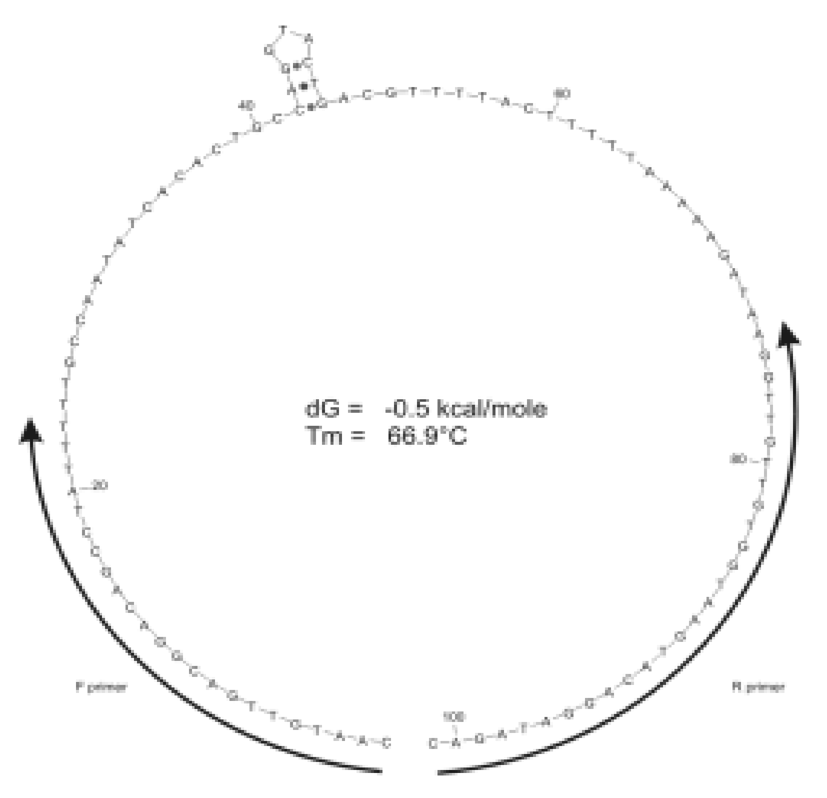
in silico assay evaluation pipeline (unafold)

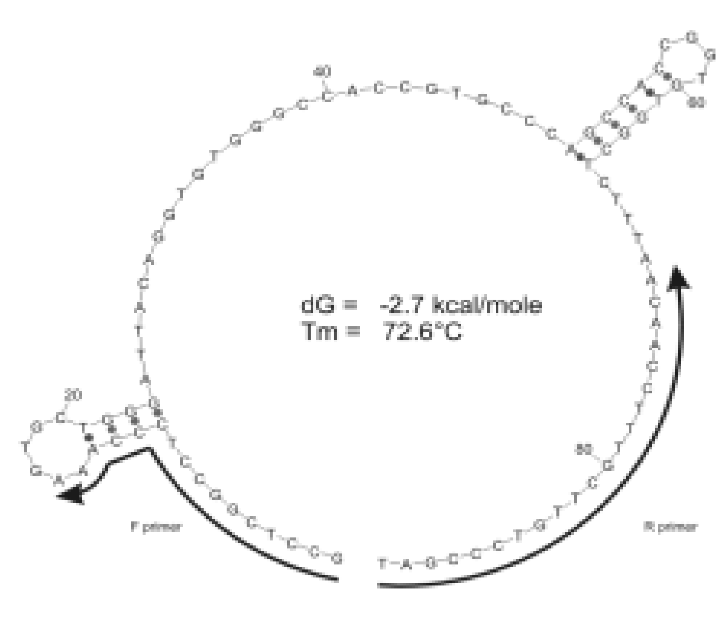


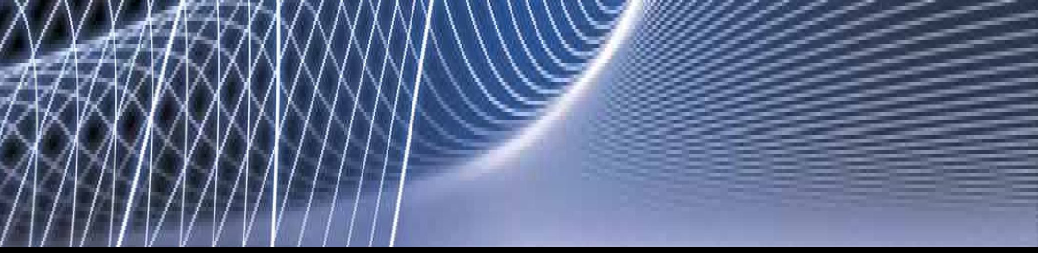
in silico assay evaluation pipeline (unafold)



in silico assay evaluation pipeline (unafold)







in silico assay evaluation pipeline (unafold)

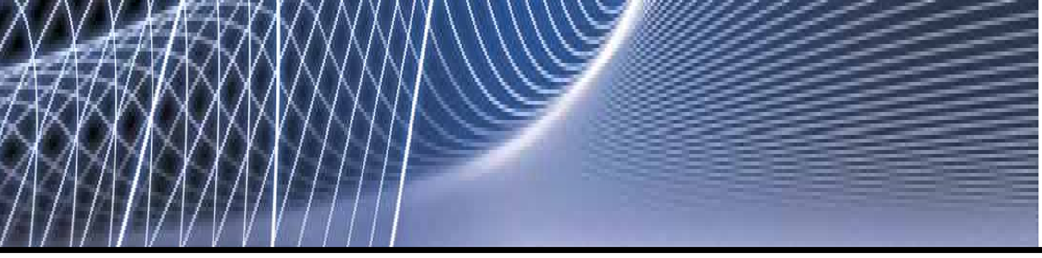
PCR Base Line Subtracted RFU
Cycle

Correlation Coefficient: 0.988 Slope: -3.296 Intercept: 26.042 Y = -3.296 X + 26.042
PCR Efficiency: 101.1 %

Unknowns
Standards

Right mouse click for

Threshold Cycle
Log Starting Quantity, nanograms



in silico assay evaluation pipeline (unafold)

PCR Base Line Subtracted RFU

Cycle

PCR Base Line Subtracted RFU

Right mouse click fo

Cycle

Correlation Coefficient: 0.909 Slope: -0.991 Intercept: 22.110 Y = -0.991 X + 22.110
PCR Efficiency: 920.5 %

Unknowns
Standards

Threshold Cycle

Log Starting Quantity, nanograms

new features 2008

- all NCBI organisms
- RDML compliant
- primer design pipeline
  - primer3
  - high-throughput

RDML (www.rdml.org)

RDML

Real-time PCR Data Markup Language

Home | News | Data Standard | Development | Guidelines | Tools | Consortium

Tools

- Create RDML file
- Validate RDML file
- Browse libraries

Join or support

The Real-time PCR Data Markup Language (RDML) is a structured and universal data standard for exchanging quantitative PCR (qPCR) data. Together with the accompanying guidelines for Minimal Information (MIqPCR), the data standard should contain sufficient information to understand the experimental setup, re-analyse the data and interpret the results. The data standard is a flat text file in Extensible Markup Language (XML) and enables transparent exchange of annotated qPCR data between instrument software and third-party data analysis packages, between colleagues and collaborators, and between authors, peer reviewers, journals and readers. To support the public acceptance of this standard, both an on-line RDML file generator is available for end users, as well as RDML software libraries to be used by software developers, enabling import and export of RDML data files.

instrument software

public repositories
journals

colleagues
collaborators

third party software
databases

Contact : info@rdml.org

© 2008 RDML Consortium

RDML

- European Biotechnology News, Issue 03-04/2008

TECHNOLOGY

News

Barcoding update

Constance – GATC Biotech has developed a platform-independent barcoding system that allows an additional level of parallel processing with a virtually unlimited increase in the number of samples processed. GATC's system is suitable for use with the Roche GS FLX and Illumina Genome

PCR

A new standard for qPCR data: RDML

Andreas Untergasser, Wageningen University, and Jo Vandesompele, Univ. Ghent

RDML

- Nature Biotechnology (submitted)
- http://www.mibbi.org

Promoting coherent minimum reporting requirements for biological and biomedical investigations: The MIBBI project

Chris F Taylor[1,2,], Dawn Field[2,3,*], Susanna-Assunta Sansone[1,2,*], Jan Aerts[4], Rolf Apweiler[1], Michael Ashburner[5], Catherine A Ball[6], Pierre-Alain Binz[7,8], Molly Bogue[9], Tim Booth[2], Alvis Brazma[1], Ryan R Brinkman[10], Adam Michael Clark[11], Eric W Deutsch[12], Oliver Fiehn[13], Jennifer Fostel[14], Peter Ghazal[15], Frank Gibson[16], Tanya Gray[2,3], Graeme Grimes[15], John M Hancock[17], Nigel W Hardy[18], Henning Hermjakob[1], Randall K Julian, Jr.[19], Matthew Kane[20], Carsten Kettner[21], Christopher Kinsinger[22], Eugene Kolker[23,24], Martin Kuiper[25a,b,c], Nicolas Le Novère[1], Jim Leebens-Mack[26], Suzanna E Lewis[27], Phillip Lord[16], Ann-Marie Mallon[17], Nishanth Marthandan[28], Hiroshi Masuya[29], Ruth McNally[30], Alexander Mehrle[31], Norman Morrison[2,32], Sandra Orchard[1], John Quackenbush[33], James M Reecy[34], Donald G Robertson[35], Philippe Rocca-Serra[1,36], Henry Rodriguez[22], Heiko Rosenfelder[31], Javier Santoyo-Lopez[15], Richard H Scheuermann[28], Daniel Schober[1], Barry Smith[37], Jason Snape[38], Chris J Stoeckert[39], Keith Tipton[40], Peter Sterk[1], Andreas Untergasser[41], Jo Vandesompele[42], Stefan Wiemann[31]*

general conclusions

- rethinking is useful
- validation matters

- reference gene validation is mandatory
- advanced quantification models and proper software enable accurate and precise results
- RTPrimerDB enables straightforward and automated qPCR assay quality control

acknowledgements

- Jo Vandesompele (geNorm)
- Jan Hellemans (qBase)
- Filip Pattyn (RTPrimerDB)
- Jasmien Hoebeeck
- Katleen De Preter
- Nurten Yigit
- Frank Speleman
- Rob Powel
- Stephen Bustin, Michael Pfaffl, Vladimir Benes