




Accurate normalization of gene expression using multiple housekeeping genes

Jo Vandesompele
Gent University Hospital - Belgium



- 
- SYBR Green I assay
 - normalization + demo geNorm
 - RTPrimerDB

SYBR Green I assay

- optimized protocol for gene expression analysis
 - two-step RT-PCR (eliminate primer dimers)
 - DNase treatment of RNA (on column vs. in solution)
 - standard reaction conditions (temp. profile, primer conc. 250 μ M)
 - 20 ng input cDNA (random primed) in 15 μ l reactions
 - very small intra-assay and inter-assay variation
- Analytical Biochemistry, 2002

NOTES & TIPS

Elimination of Primer-Dimer Artifacts and Genomic Coamplification Using a Two-Step SYBR Green I Real-Time RT-PCR¹

Jo Vandesompele, Anne De Paepe, and Frank Speleman²

normalization

Northern blot, microarray and quantitative RT-PCR:
necessity to adjust for experimental differences

RNA quantity & quality
overall transcriptional activity
cDNA synthesis & PCR efficiency

strategies

- RNA mass quantity / 18S-28S (Northern blot)
- sum of all signals / median ratio / Lowess fit (array methods)
- internal endogeneous control (housekeeping gene)

normalization

search for the holy gene

- housekeeping gene
maintenance gene
internal endogenous control
- a suitable reference to which expression can be normalized
constant expression in the cells or tissues under investigation, no response to experimental treatment
- beta actin (ACTB)*
glyceraldehyde-3-phosphate dehydrogenase (GAPD)
18S rRNA
>90% studies - presumed stable expression

normalization

real life internal controls

There is no such gene as a housekeeping gene, only genes with a small variation in expression level.

Thellin et al., 1999 (J Biotech)

“Commonly used internal controls can quantitatively vary in response to various factors.”

Warrington et al., 2000 (Physiol Genomics)

“A set of genes frequently used in standard expression analysis were found to vary in expression level by 7- to 23-fold (including *ACTB* and *GAPD*).” (comparing 7 adult and 4 fetal normal human tissues)

...

normalization

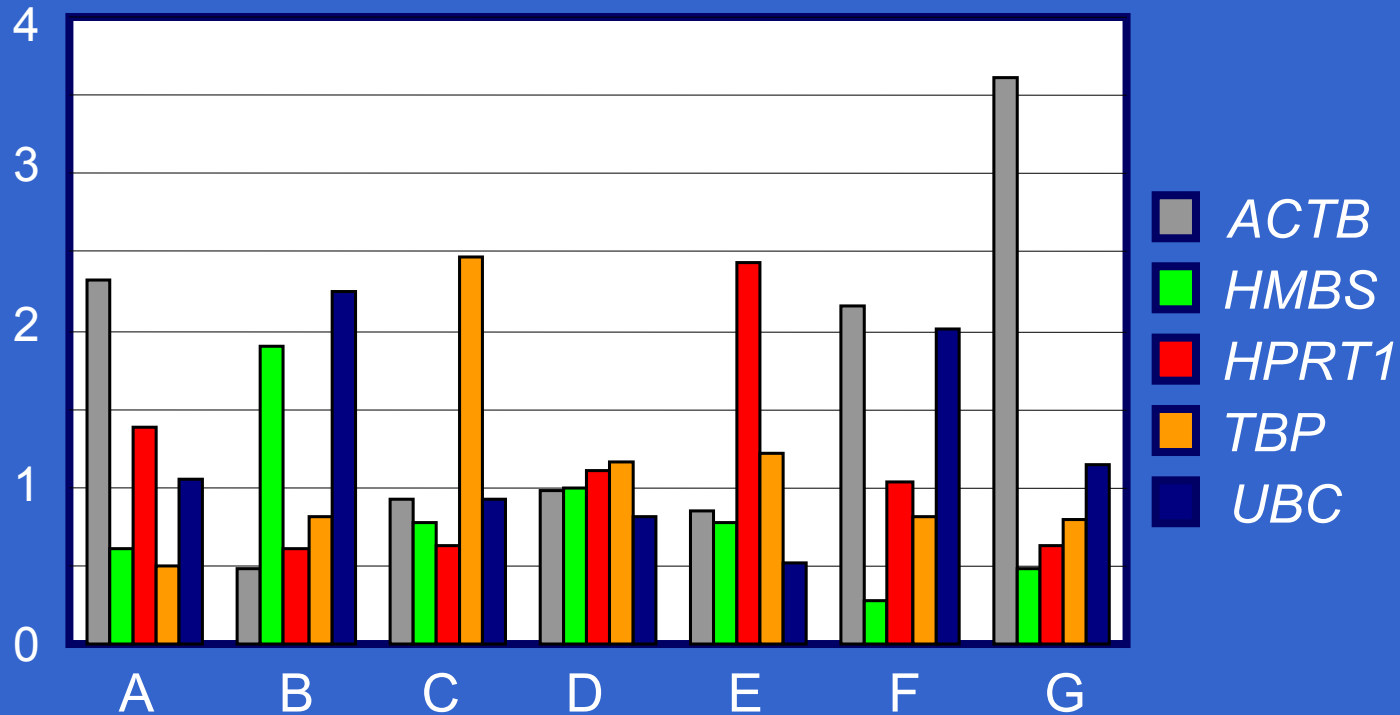
real life internal controls

quantitative RT-PCR analysis of 10 internal controls
(belonging to different functional and abundance classes)
on 85 samples < 13 different human tissues

- | | |
|----------------|-----------------|
| ■ <i>TBP</i> | ■ <i>RPL13A</i> |
| ■ <i>HMBS</i> | ■ <i>YWHAZ</i> |
| ■ <i>UBC</i> | ■ <i>SDHA</i> |
| ■ <i>HPRT1</i> | ■ <i>B2M</i> |
| ■ <i>ACTB</i> | ■ <i>GAPD</i> |

normalization

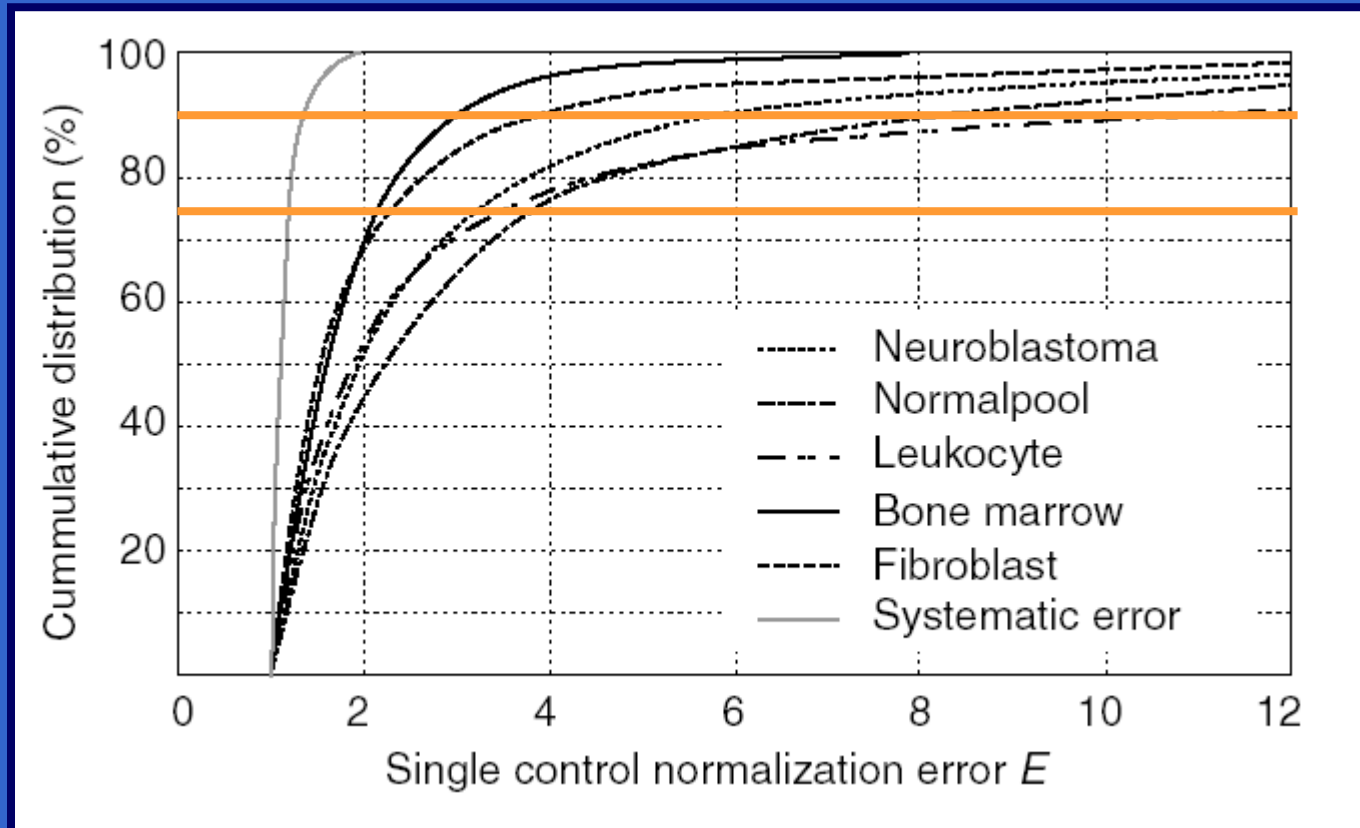
real life internal controls



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

normalization

single control normalization errors



normalization

multiple internal controls

- given the extreme sensitivity, reproducibility and large dynamic range of quantitative RT-PCR
- the observed expression differences between so-called housekeeping genes
- absence of sufficient data to determine the biological significance of 2- to 3-fold expression differences

we propose the use of **multiple** internal controls
for accurate normalization




which ? how many? how?



normalization



robust algorithm

-  identify the most stable control genes in a given tissue/cell
-  determine how many control genes are required for accurate/reliable normalization
-  assumption-free

normalization

how



assess the (standard) variation of the control gene
assume equal input of equal quality RNA



compare 2 (or more) housekeeping genes

normalization

gene stability measure M

pairwise variation V (between 2 genes)

	gene A	gene B	
sample 1	a1	b1	$\log_2(a_1/b_1)$
sample 2	a2	b2	$\log_2(a_2/b_2)$
sample 3	a3	b3	$\log_2(a_3/b_3)$
...
sample n	a _n	b _n	$\log_2(a_n/b_n)$

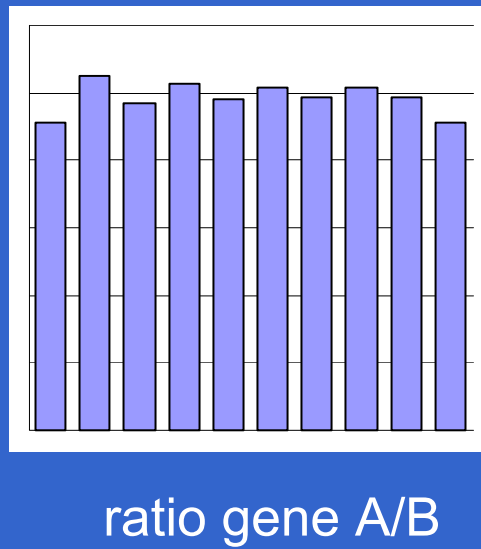
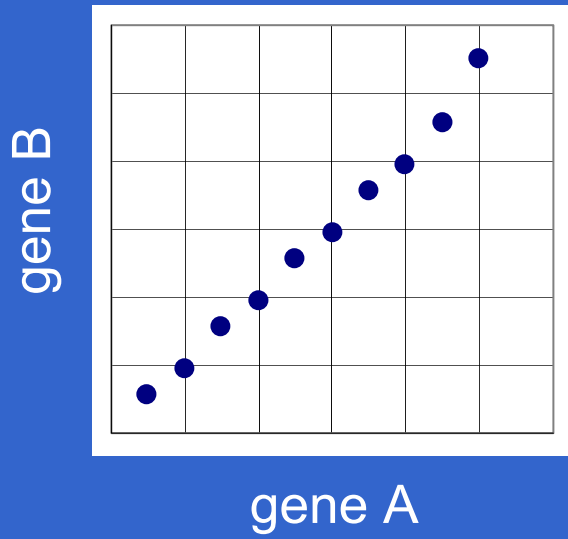


st.dev = V

gene stability measure M

average pairwise variation V of a gene with all other genes

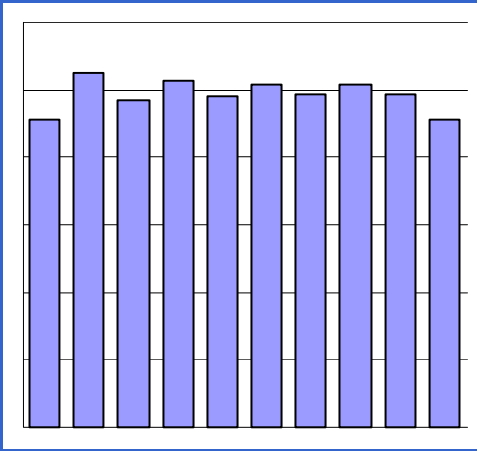
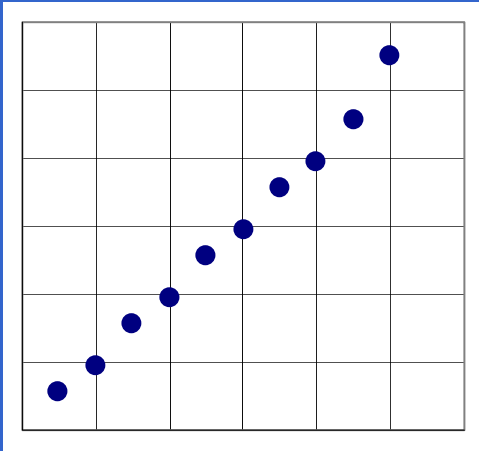
normalization



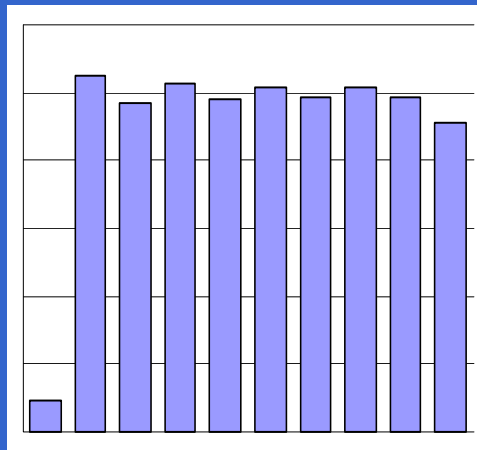
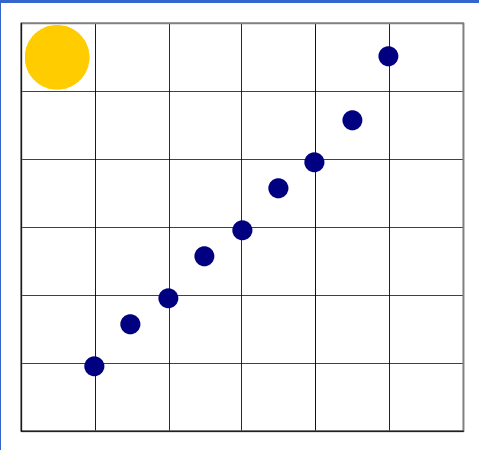
pearson r	0.99 (p<0.01)
spearman r	1.00 (p<0.01)
geNorm V	0.07

normalization

gene B

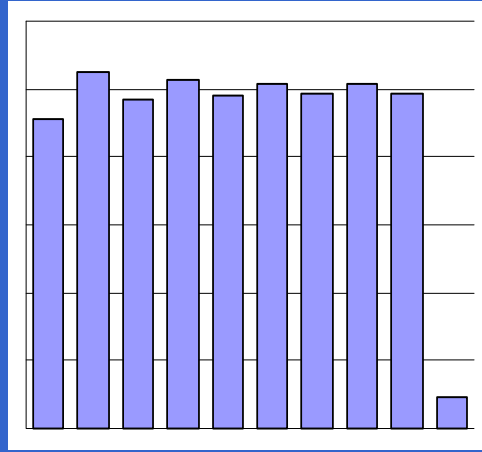
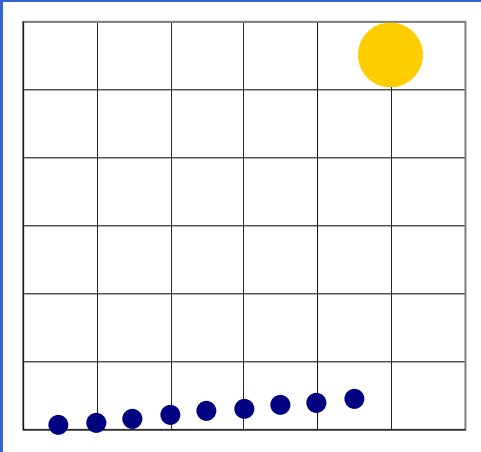



pearson r 0.99 (p<0.01)
spearman r 1.00 (p<0.01)
geNorm V 0.07

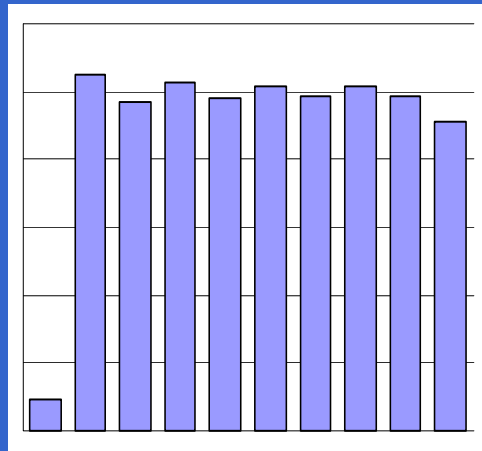
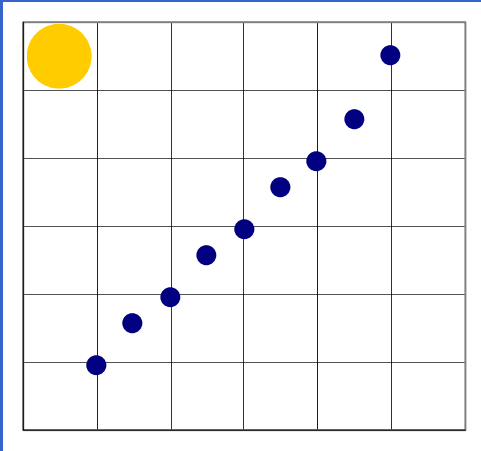



pearson r 0.49 (p=0.15)
spearman r 0.46 (p=0.13)
geNorm V 1.09

normalization



pearson r 0.59 (p=0.08)
spearman r 1.00 (p<0.01)
geNorm V 



pearson r 0.49 (p=0.15)
spearman r 0.46 (p=0.13)
geNorm V 

normalization

gene stability measure M

pairwise variation V (between 2 genes)

	gene A	gene B	
sample 1	a1	b1	$\log_2(a_1/b_1)$
sample 2	a2	b2	$\log_2(a_2/b_2)$
sample 3	a3	b3	$\log_2(a_3/b_3)$
...
sample n	a _n	b _n	$\log_2(a_n/b_n)$



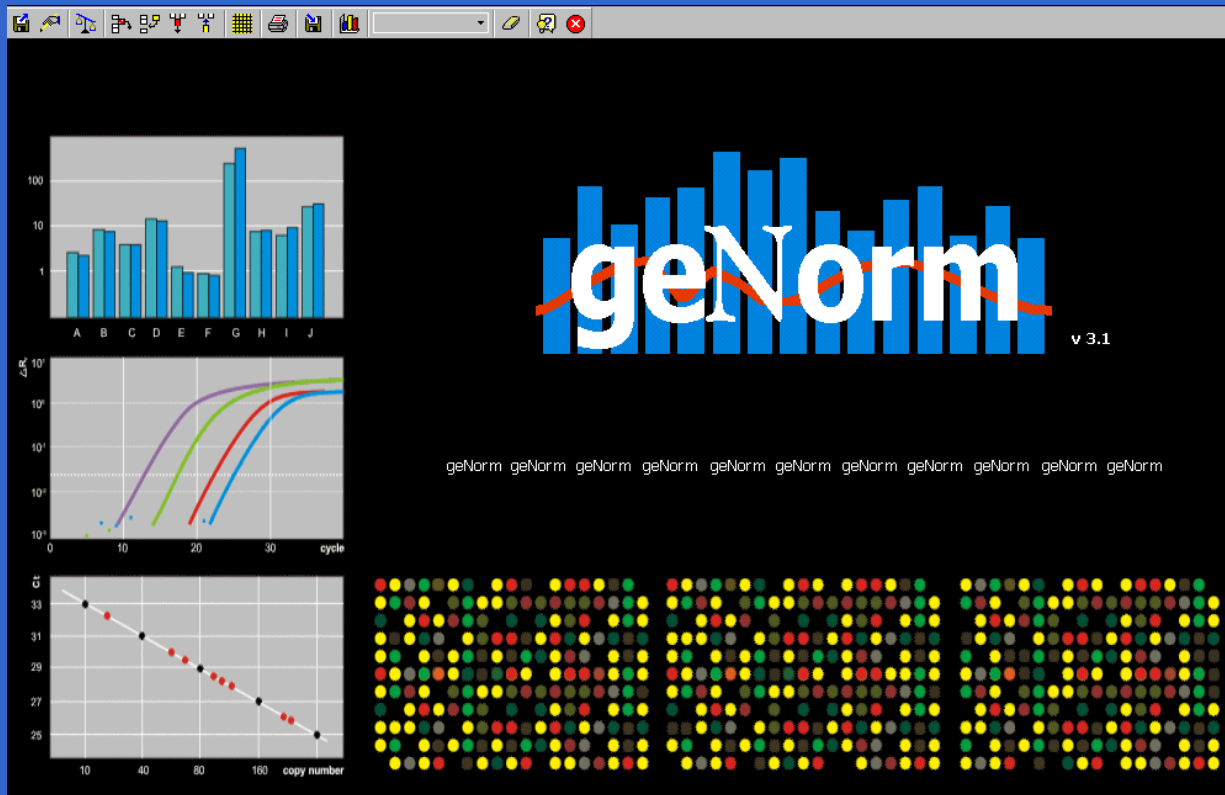
st.dev = V

gene stability measure M

average pairwise variation V of a gene with all other genes

normalization

geNorm VBA applet

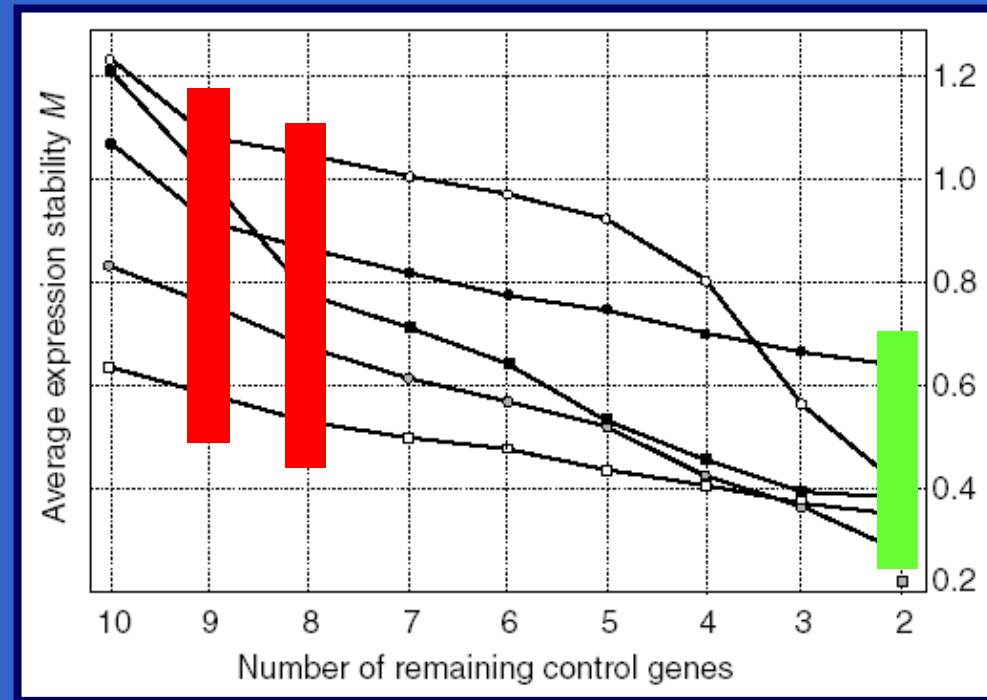


allserv.ugent.be/~jvdesomp/genorm/

normalization

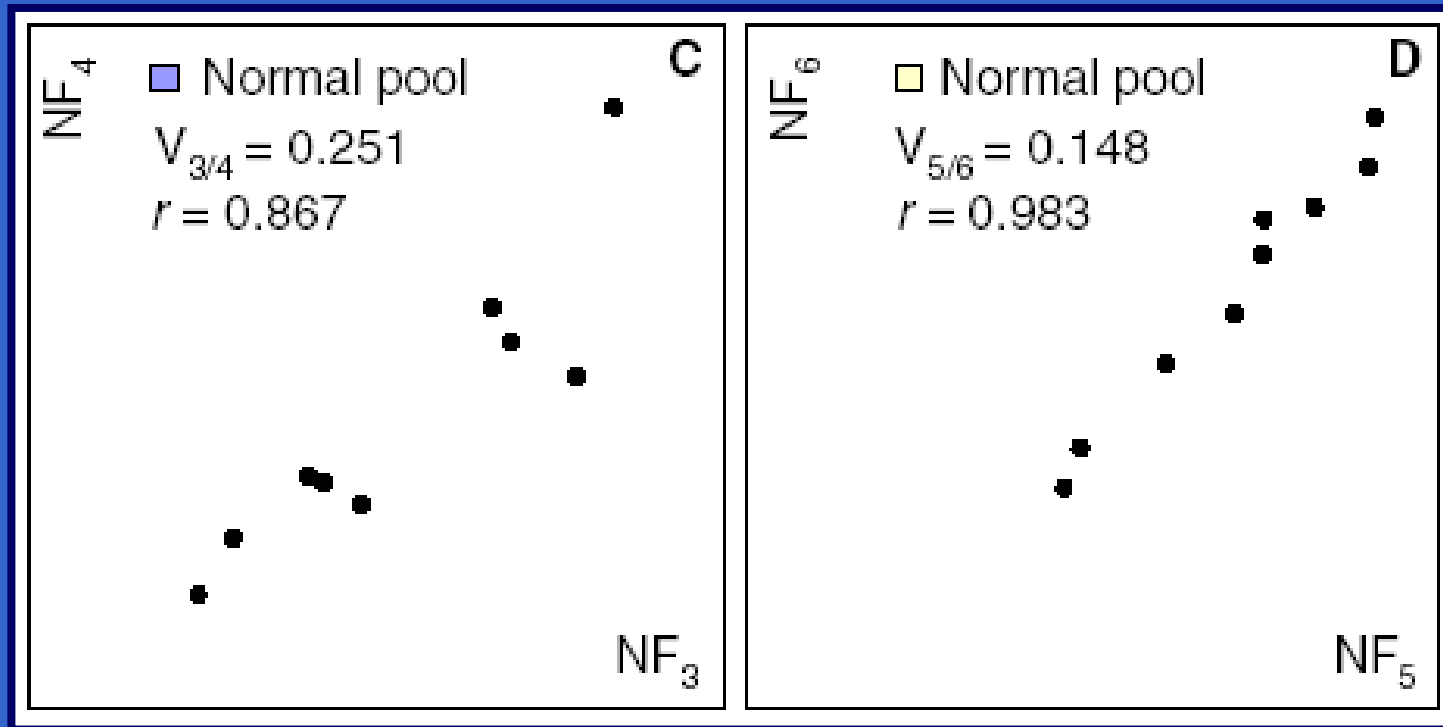
- stepwise elimination of worst scoring control gene (which)

Neuro-blastoma	Fibroblast	Leukocyte	Bone marrow	Normal pool
<i>ACTB</i>	<i>RPL13A</i>	<i>HPRT1</i>	<i>HMBS</i>	<i>YWHAZ</i>
<i>TBP</i>	<i>SDHA</i>	<i>SDHA</i>	<i>TBP</i>	<i>RPL13A</i>
<i>YWHAZ</i>	<i>TBP</i>	<i>TBP</i>	<i>SDHA</i>	<i>UBC</i>
<i>HMBS</i>	<i>ACTB</i>	<i>RPL13A</i>	<i>GAPD</i>	<i>TBP</i>
<i>UBC</i>	<i>UBC</i>	<i>GAPD</i>	<i>HPRT1</i>	<i>HPRT1</i>
<i>SDHA</i>	<i>YWHAZ</i>	<i>B2M</i>	<i>YWHAZ</i>	<i>HMBS</i>



normalization

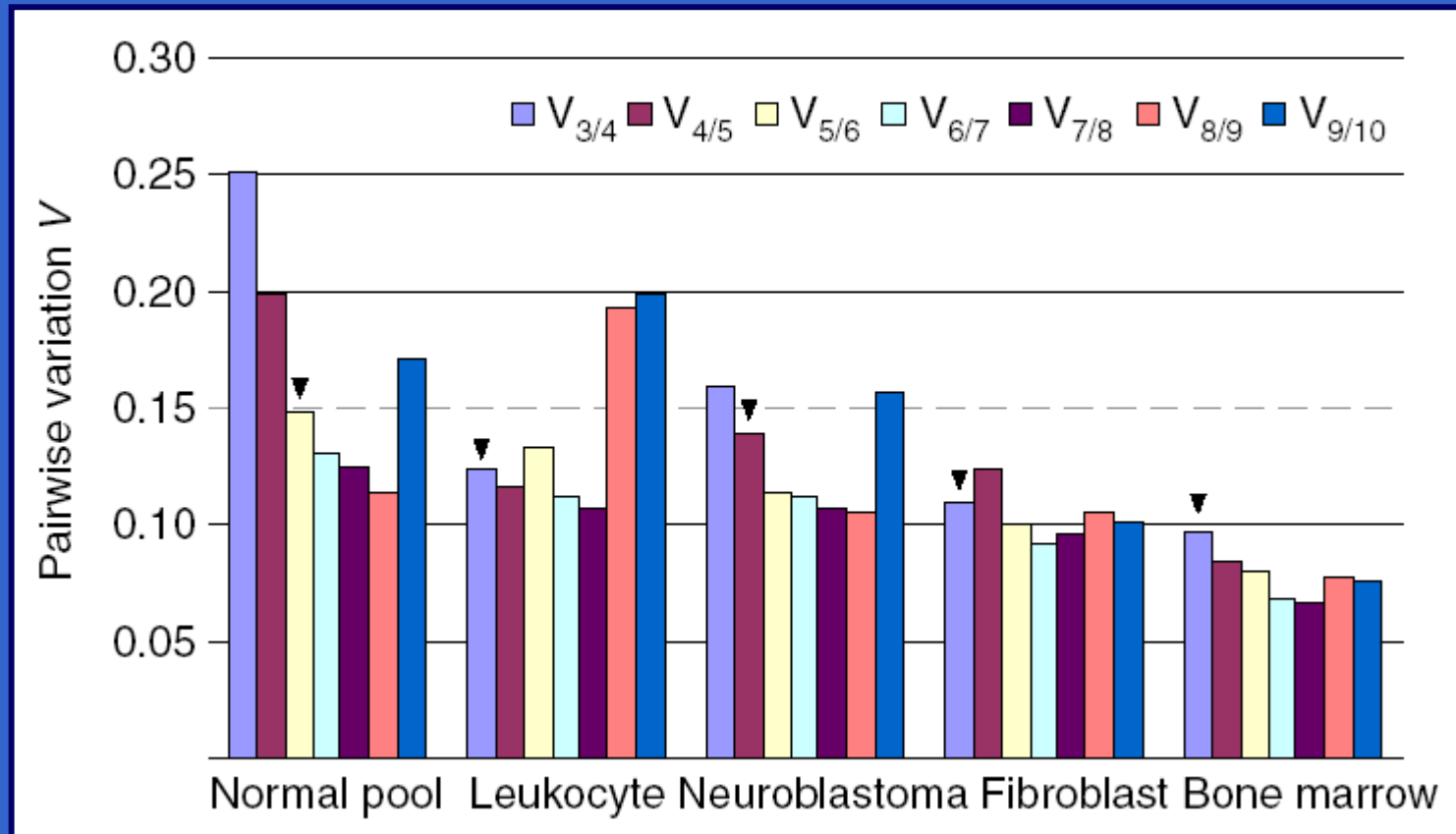
- determination of the optimal number of control genes (**how many**)



3 – 5 stable housekeepers are needed

normalization

- determination of the optimal number of control genes (**how many**)



normalization

- calculation of normalization factor (**how?**)

geometric mean of 3 housekeeping genes levels

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

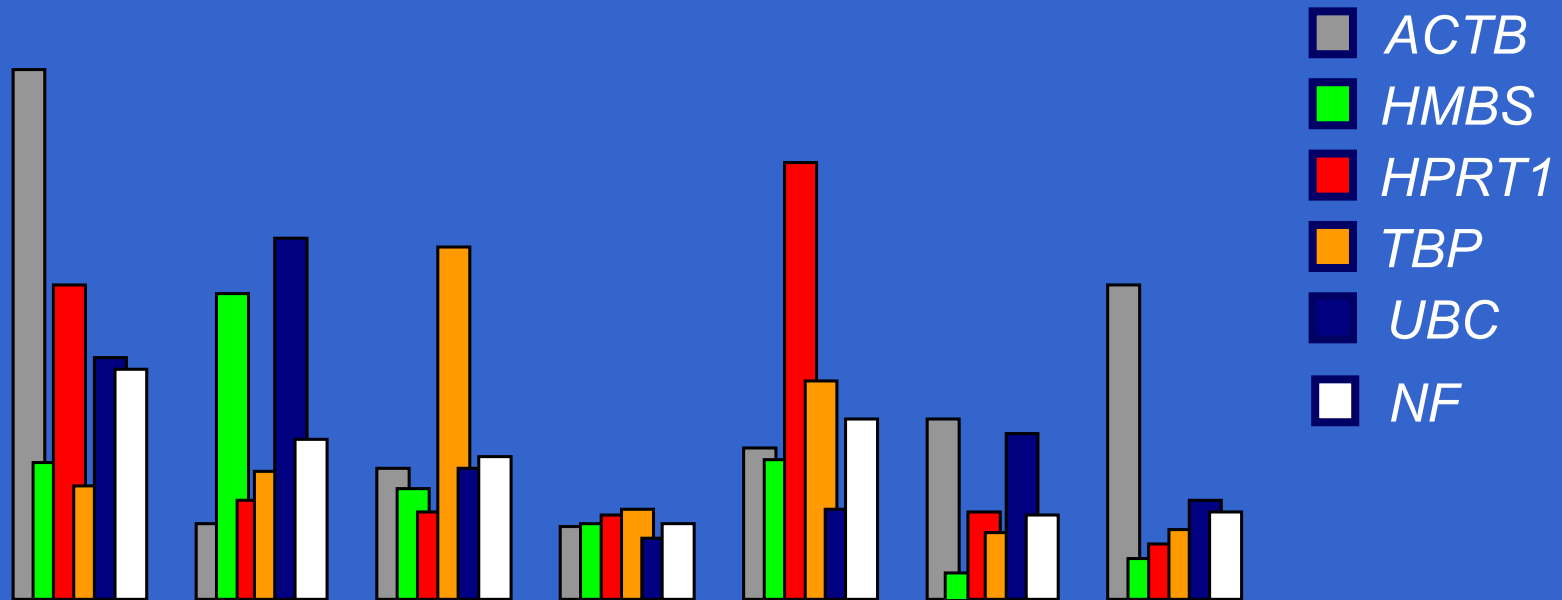
geomean < arithmetic mean

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

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

normalization

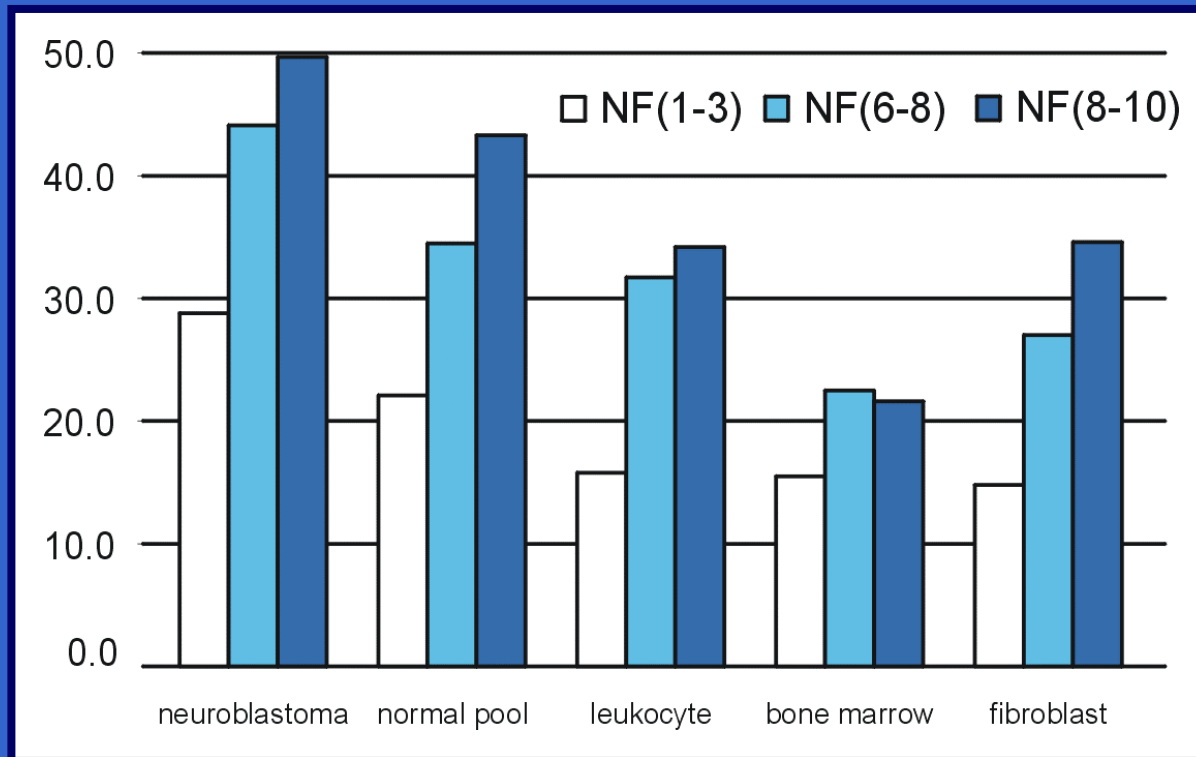
calculation of normalization factor (NF)



normalization

validation of geNorm

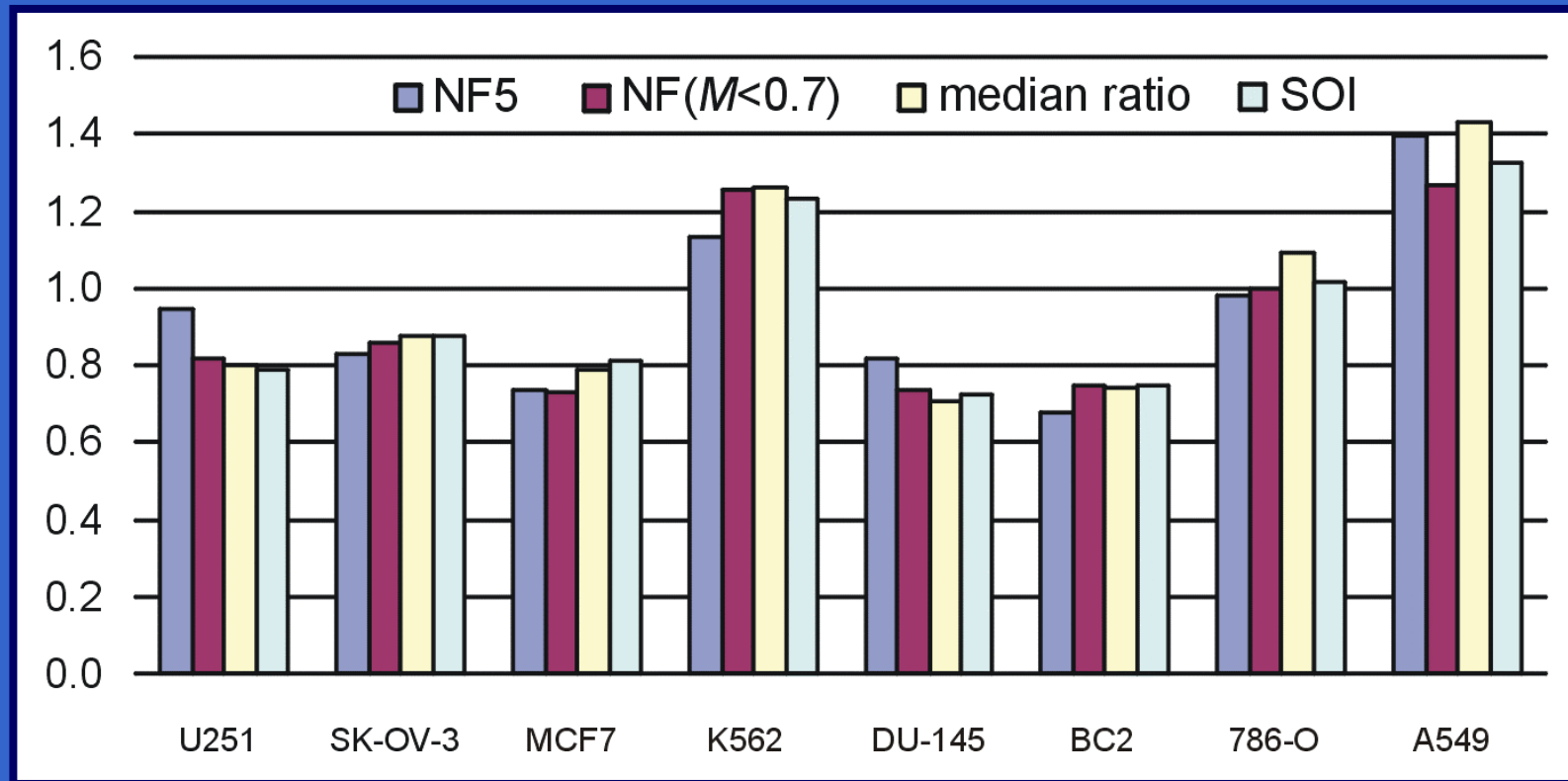
gene specific variation



normalization

validation of geNorm

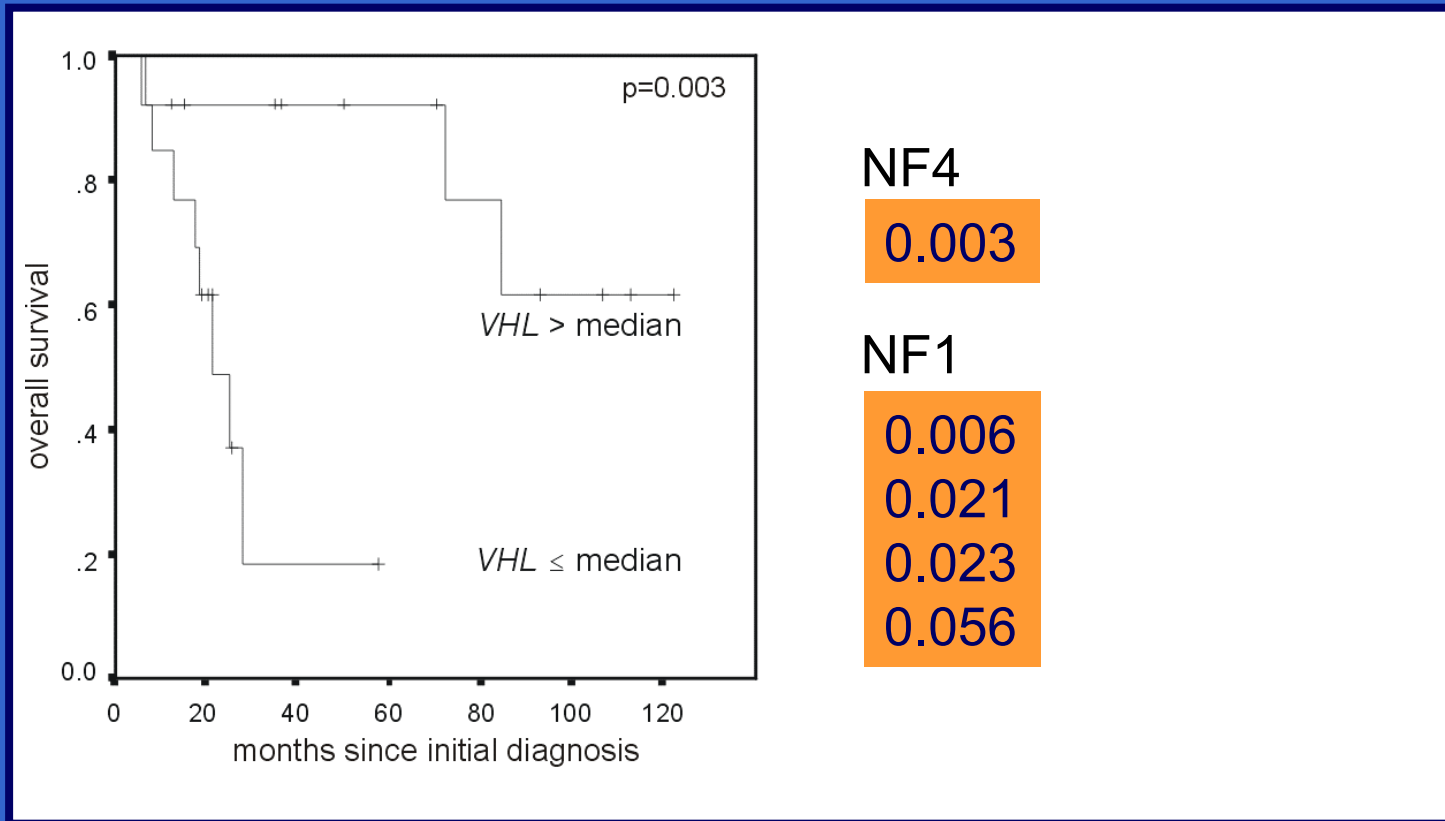
microarray normalization factors



normalization

validation of geNorm

cancer patient survival analysis



normalization

conclusion

- large expression differences between so-called housekeeping genes
- use of only one gene leads to erroneous normalization
 - up to 3 fold in 25% tested samples
 - up to 6.4 fold in 10% of tested samples
- calculation of normalization factor based on the geometric mean of at least 3 internal control genes
- Vandesompele et al., Genome Biology, 2002

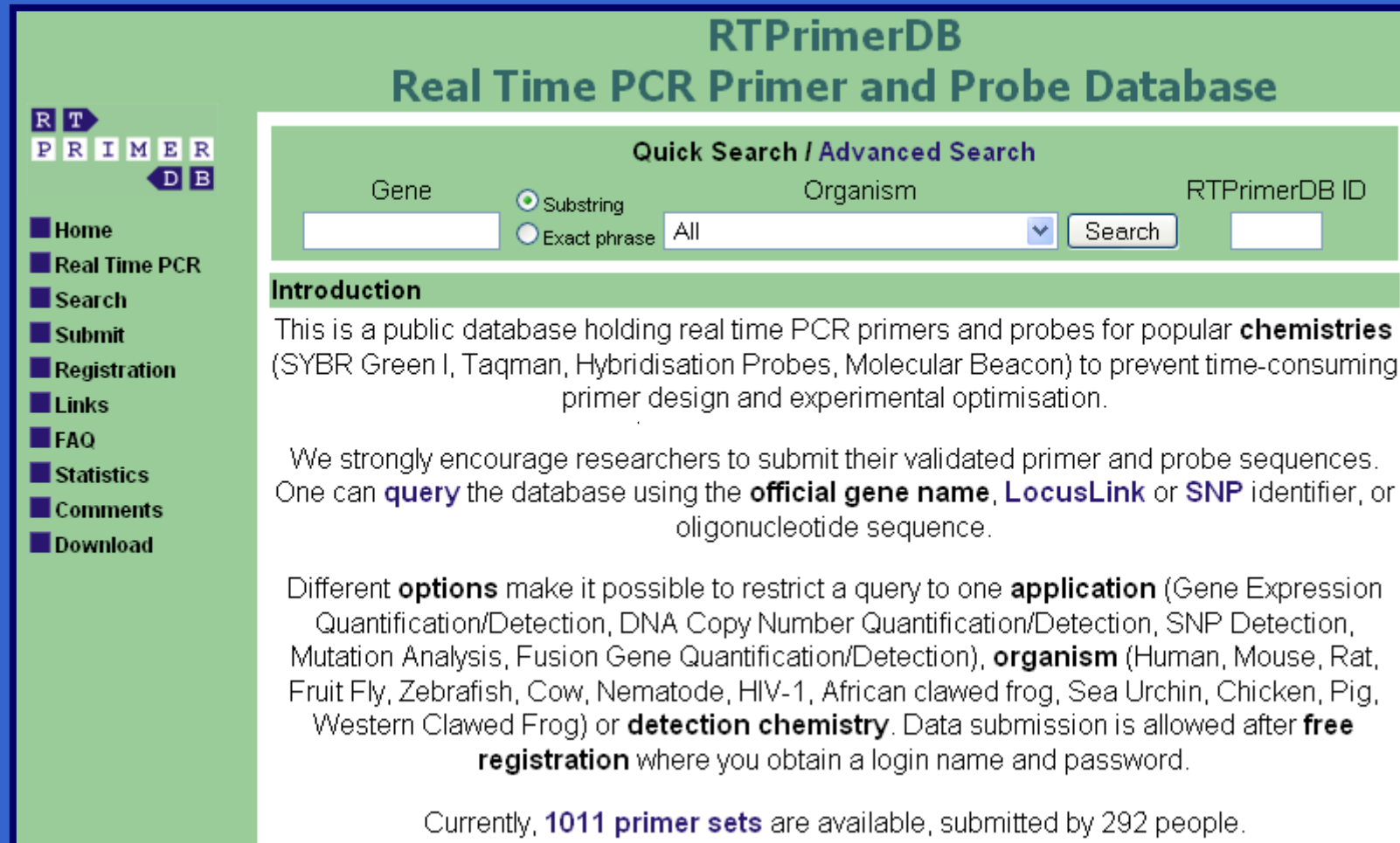
Research

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

normalization

- future perspectives: from holy gene to repeat
 - using a repetitive sequence in the human transcriptome as a measure for the mRNA fraction
 - rationale: repeat sequences are present in the 3' UTR of many genes, and the differential expression of a small number of genes won't influence the overall repeat abundance in the transcriptome
 - the most abundant repeats found in 3' UTR of genes are Alu repeats
 - validating Alu-J, Alu-Sx and Alu-Sq repeats for mRNA normalization



RTPrimerDB

Real Time PCR Primer and Probe Database

Quick Search / Advanced Search

Gene Substring Exact phrase All Organism RTPrimerDB ID

Introduction

This is a public database holding real time PCR primers and probes for popular **chemistries** (SYBR Green I, Taqman, Hybridisation Probes, Molecular Beacon) to prevent time-consuming primer design and experimental optimisation.

We strongly encourage researchers to submit their validated primer and probe sequences. One can **query** the database using the **official gene name**, **LocusLink** or **SNP** identifier, or oligonucleotide sequence.

Different **options** make it possible to restrict a query to one **application** (Gene Expression Quantification/Detection, DNA Copy Number Quantification/Detection, SNP Detection, Mutation Analysis, Fusion Gene Quantification/Detection), **organism** (Human, Mouse, Rat, Fruit Fly, Zebrafish, Cow, Nematode, HIV-1, African clawed frog, Sea Urchin, Chicken, Pig, Western Clawed Frog) or **detection chemistry**. Data submission is allowed after **free registration** where you obtain a login name and password.

Currently, **1011 primer sets** are available, submitted by 292 people.



- Home
- Real Time PCR
- Search
- Submit
- Links
- FAQ
- Statistics
- Comments
- Download

RTPrimerDB Real Time PCR Primer and Probe Database

Quick Search/Advanced Search

Gene Substring Organism RTPrimerDB ID

Exact phrase

8 records found

This is page 1 of 1

Go to Page

ID	Org	Symbol	Name	App	Detec	Remarks
3	Hs	GAPD	glyceraldehyde-3-phosphate dehydrogenase	GXP	SYB	
55	Hs	GAPD	glyceraldehyde-3-phosphate dehydrogenase	GXP	TQ	
167	Mm	Gapd	glyceraldehyde-3-phosphate dehydrogenase	GXP	TQ	
473	Mm	Gapd	glyceraldehyde-3-phosphate dehydrogenase	GXP	SYB	remarks
256	Rn	Gapd	Glyceraldehyde-3-phosphate dehydrogenase	GXP	SYB	
192	Rn	Gapd	Glyceraldehyde-3-phosphate dehydrogenase	GXP	SYB	remarks
440	Rn	Gapd	Glyceraldehyde-3-phosphate dehydrogenase	GXP	TQ	remarks
379	Bt	GAPD	glyceraldehyde-3-phosphate dehydrogenase	GXP	TQ	

RTPrimerDB ID: 3

Official Gene Symbol and Name

Homo sapiens (Hs, Human)

GAPD: glyceraldehyde-3-phosphate dehydrogenase

Alias Symbol(s): G3PD, GAPDH

Locuslink ID: [2597](#)

Assay

Application: Gene Expression Quantification/Detection GXP

Detection: SYBR Green I SYB

Template: cDNA

Primer/Probe Sequences (5' → 3')

Forward Primer: TGCACCACCAACTGCTTAGC

Reverse Primer: GGCATGGACTGTGGTCATGAG

Annealing Temperature: 60 °C

Amplicon Length: 86

BLAST sequences

PubMed ID

PubMed ID: [12184808](#)

Submitter

Jo Vandesompele

Ghent University Hospital, Center for Medical Genetics

De Pintelaan 185

9000 Ghent

acknowledgements

Center for Medical Genetics
Gent University Hospital - Belgium



Jo Vandesompele
Katleen De Preter
Filip Pattyn
Els De Smet
Bruce Poppe
Anne De Paepe
Frank Speleman

<http://medgen.ugent.be>
Joke.Vandesompele@UGent.be

