Better appreciation of true biological miRNA expression differences using an improved version of the global mean normalization strategy

Jo Vandesompele
professor, Ghent University
co-founder and CEO, Biogazelle

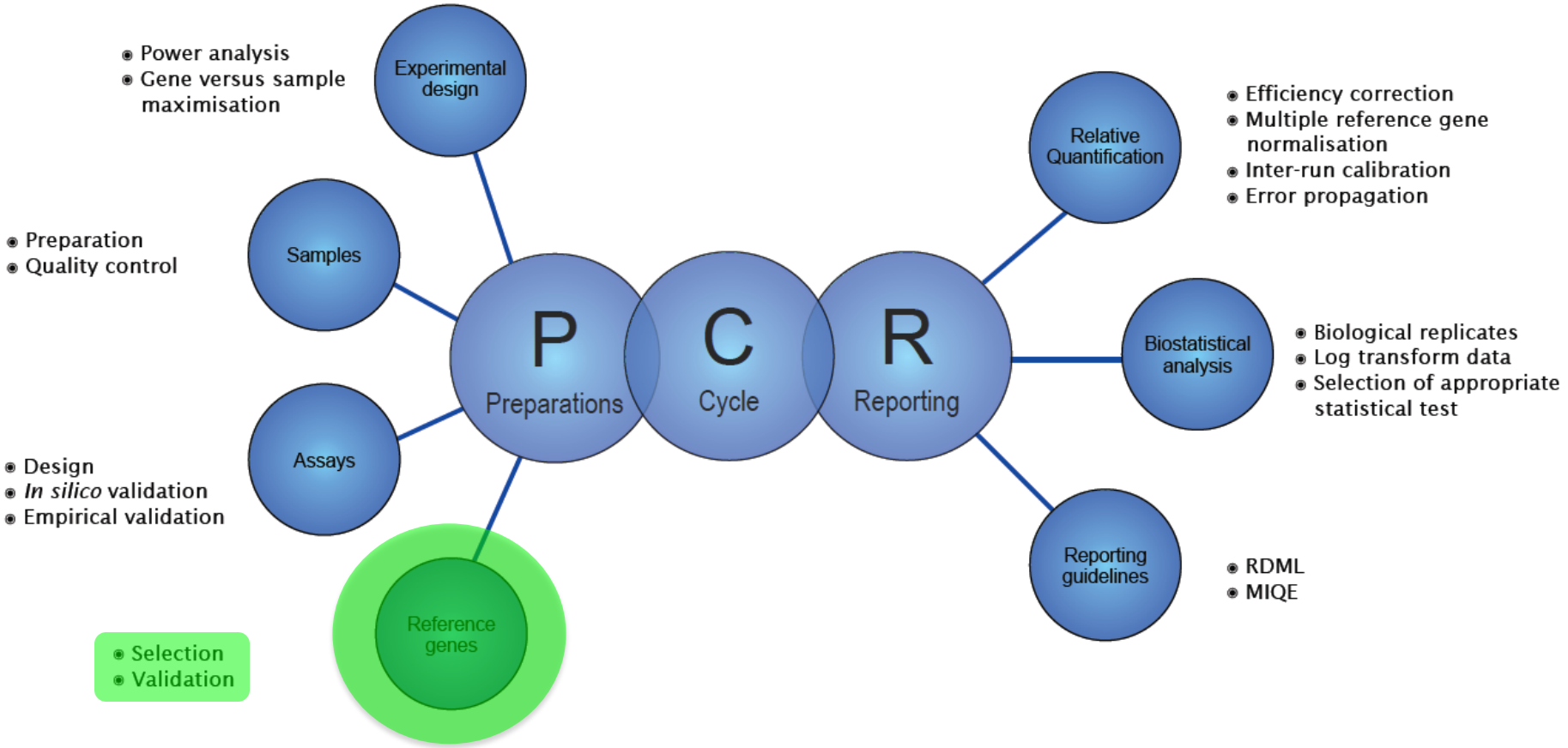
RNAi and miRNA world congres
Boston, April 27, 2011

Biogazelle – a real-time PCR company

qbase^{PLUS} software, courses, miR profiling, data mining service

Center for Medical Genetics

How to do successful gene expression analysis?



Derveaux et al., Methods, 2010

biogazelle > resources > articles

biogazelle
accelerating your analysis

mybiogazelle

products | courses | resources | support | company

you are here: start > resources > articles

articles
presentations
databases
tools

Selected key papers and book chapters on qPCR developments and applications.

2010

book chapter qPCR data analysis – unlocking the secret to successful results

PCR Troubleshooting and Optimization: The Essential Guide
Horizon Press 2010
keywords: data-analysis, relative quantification, statistics, experiment design

How to do successful gene expression analysis using real-time PCR [PDF]

Methods (in press)
keywords: relative quantification, gene expression, workflow

Accurate and objective copy number profiling using real-time quantitative PCR [PDF]

Methods (in press)
keywords: copy number analysis

2009

book chapter Reference gene validation software for improved normalization [PDF]

Real-time PCR: Current Technology and Applications
Academic Press 2009, ISBN 978-1-904455-39-4
keywords: validation of reference genes (formerly housekeeping genes)

RTPrimerDB: the portal for real-time PCR primers and probes [PDF]

Nucleic Acids Res. 2009 Jan;37(Database issue):D942-5
keywords: qPCR assay database, in silico assay quality control

http://www.biogazelle.com

Center for Medical Genetics

outline - normalization

- what is normalization
- reference genes: gold standard for normalization
- global mean normalization and selection of stable references

introduction to normalization

- 2 sources of variation in gene expression results
  - biological variation (true fold changes)
  - experimentally induced variation (noise and bias)

- purpose of normalization is reduction of the experimental variation
  - input quantity: RNA quantity, cDNA synthesis efficiency, …
  - input quality: RNA integrity, RNA purity, …

- gold standard is the use of multiple stably expressed reference genes
  - which genes?
  - how many?
  - how to do the calculations?

normalization: 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

geNorm expression stability parameter

- pairwise variation V (between 2 genes)

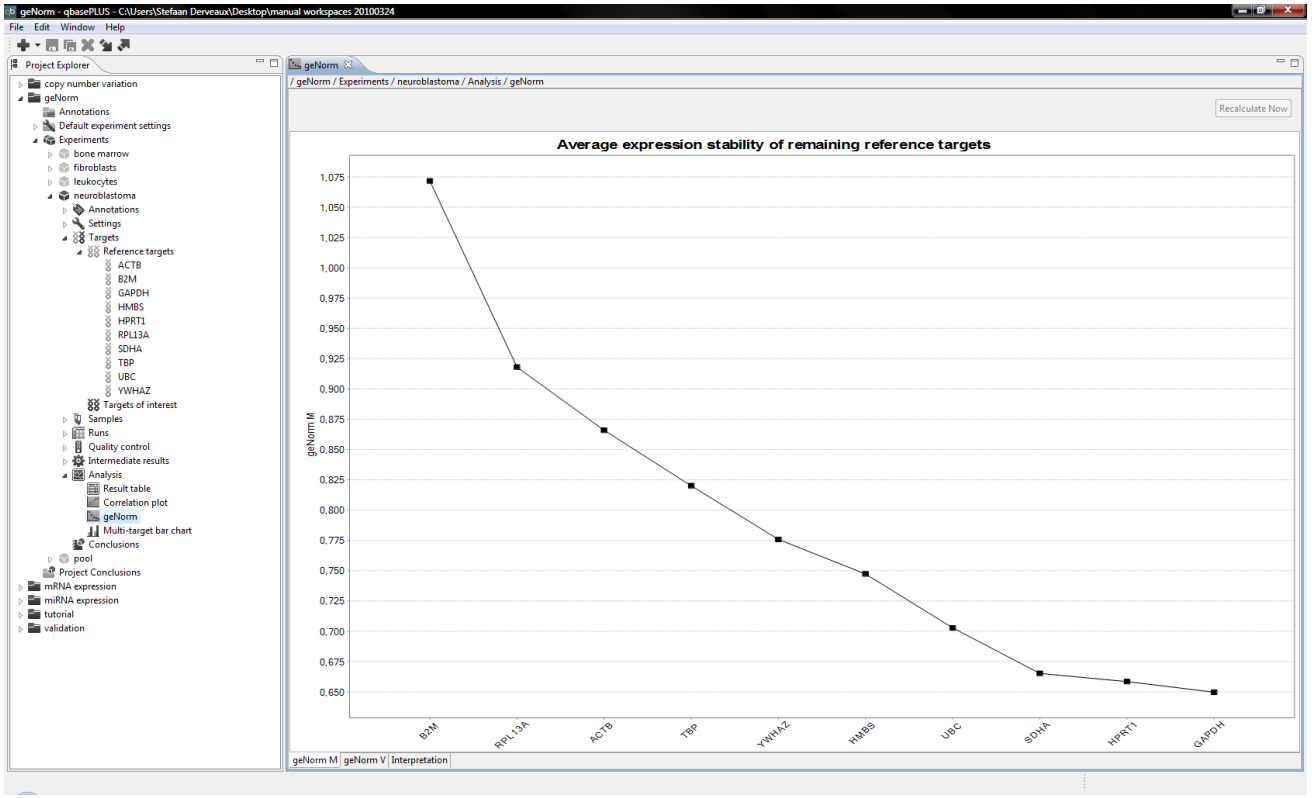
	gene A	gene B	
sample 1	a1	b1	log2(a1/b1)
sample 2	a2	b2	log2(a2/b2)
sample 3	a3	b3	log2(a3/b3)
…	…	…	…
sample n	an	bn	log2(an/bn)

standard deviation = V

- gene stability measure M
  average pairwise variation V of a gene with all other genes

geNorm software

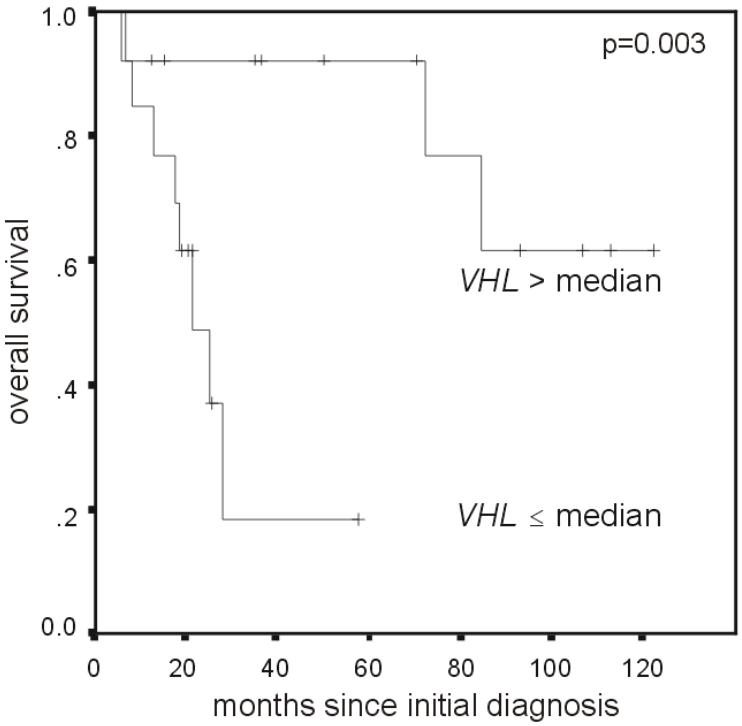
- ranking of candidate reference genes according to their stability
- determination of how many genes are required for reliable normalization
- http://www.genorm.info



Center for Medical Genetics

geNorm validation (I)

- cancer patients survival curve
  statistically more significant results



log rank statistics

NF4

0.003

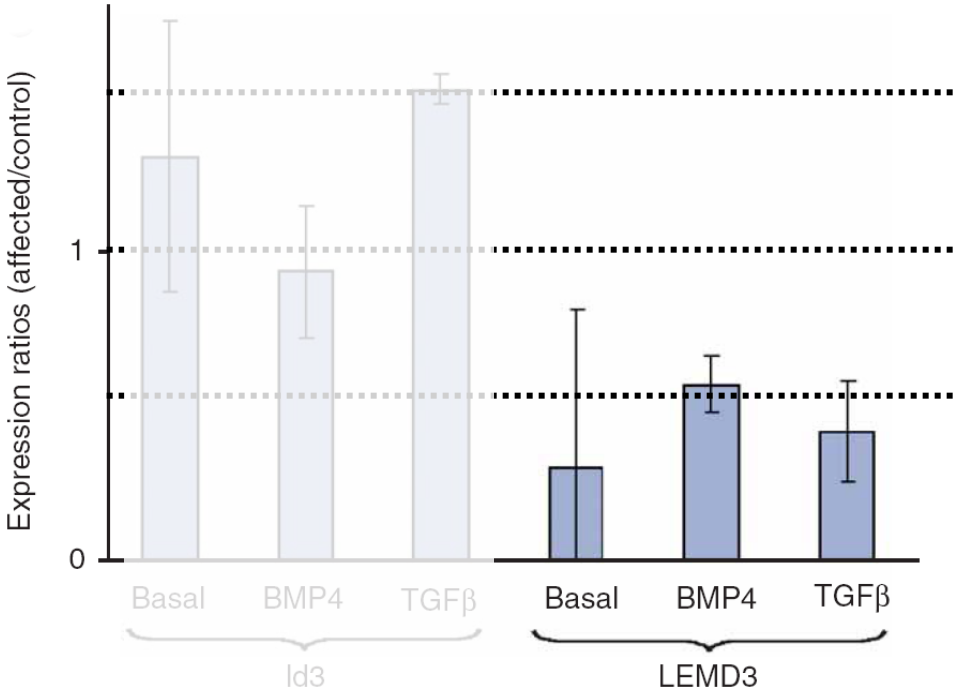
NF1

0.006
0.021
0.023
0.056

Hoebeeck et al., Int J Cancer, 2006

geNorm validation (II)

- mRNA haploinsufficiency measurements
  accurate assessment of small expression differences



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

Hellemans et al., Nature Genetics, 2004

normalization using multiple stable reference genes

- geNorm is the *de facto* standard for reference gene validation and normalization
  - > 3,000 citations of our geNorm technology
  - > 15,000 geNorm software downloads in 100 countries

Center for Medical Genetics

improved geNorm > genorm^{PLUS}

	classic geNorm	genorm^{PLUS}
platform	Excel Windows	qbase^{PLUS} Win, Mac, Linux
speed	1x	20x
interpretation	-	+
ranking best 2 genes	-	+
handling missing data	-	+
raw data (Cq) as input	-	+

a new normalization method: global mean normalization

- hypothesis: when a large set of genes are measured, the **average expression level** reflects the input amount and could be used for normalization
  - microarray normalization (lowess, mean ratio, …)
  - RNA-seq read counts

- the set of genes must be sufficiently **large** and **unbiased**

- we test this hypothesis using genome-wide **microRNA** data from experiments in which Biogazelle quantified a large number of miRNAs (450-750) in a given sample series

  - cancer biopsies & serum
    - *neuroblastoma, T-ALL, EVI1 leukemia, retinoblastoma*
  - pool of normal tissues, normal bone marrow set
  - induced sputum of smokers vs. non-smokers

How to validate a normalization method?

- geNorm ranking global mean vs. candidate reference genes (small RNA controls, such as snRNA and snoRNA)
- reduction of experimental noise
- balancing of expression differences (up vs. down)
- identification of truly differentially expressed genes

- original global mean (Mestdagh et al., Genome Biology, 2009)
- improved global mean (D'haene et al., in press)
  - mean center the data > equal weight to each gene
  - allow PCR efficiency correction

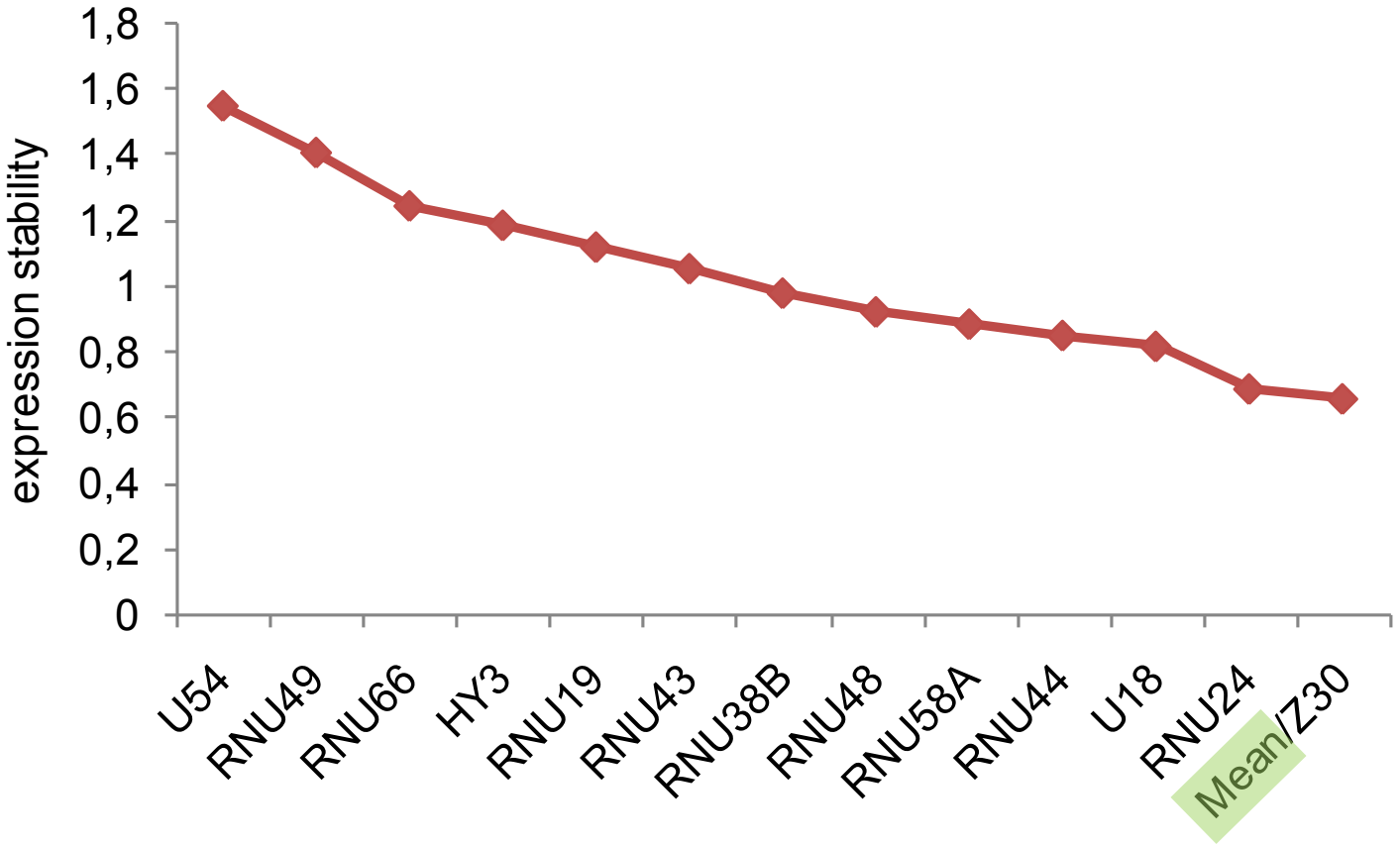
small RNA controls

- How 'stable' is the global mean compared to (small RNA) controls?
  - geNorm analysis using controls and global mean as input variables
  - exclusion of potentially co-regulated controls

HY3	7q36
RNU19	5q31.2
RNU24	9q34
RNU38B	1p34.1-p32
RNU43	22q13
RNU44	1q25.1
RNU48	6p21.32
RNU49	17p11.2
RNU58A	18q21
RNU58B	18q21
RNU66	1p22.1
RNU6B	10p13
U18	15q22
U47	1q25.1
U54	8q12
U75	1q25.1
Z30	17q12
RPL21	13q12.2

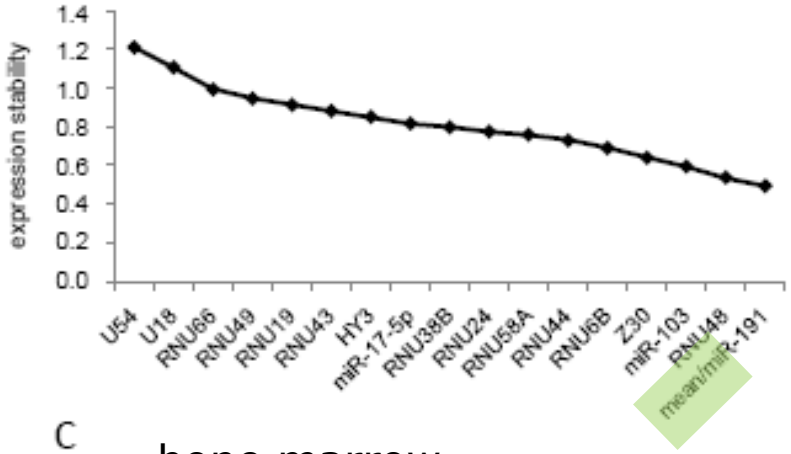
geNorm ranking (T-ALL) (I)

- lower M-value means better stability

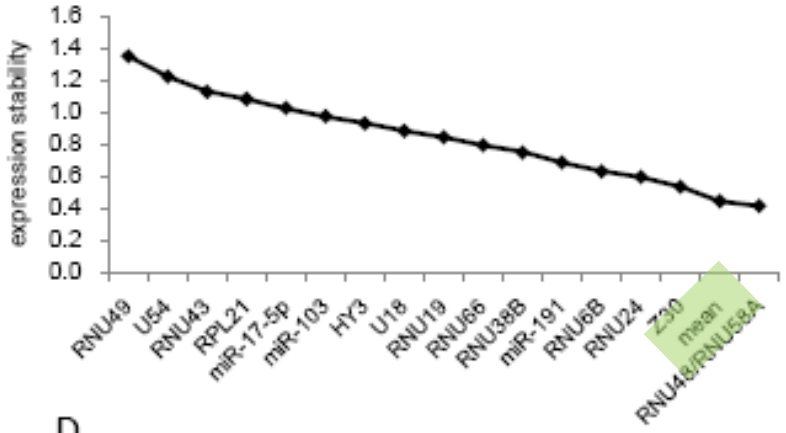


geNorm ranking (I)

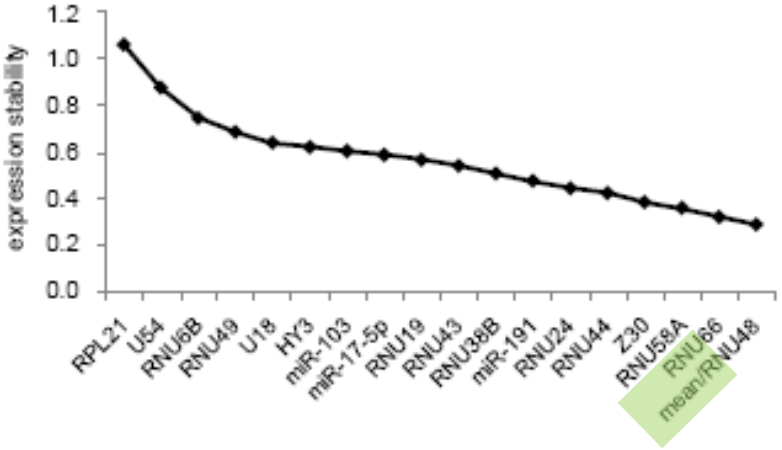
A neuroblastoma



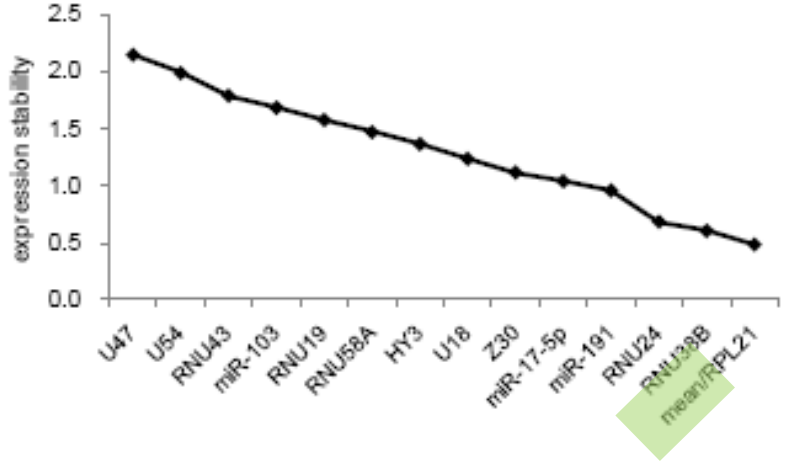
B leukemia EVI1 overexpression



C bone marrow



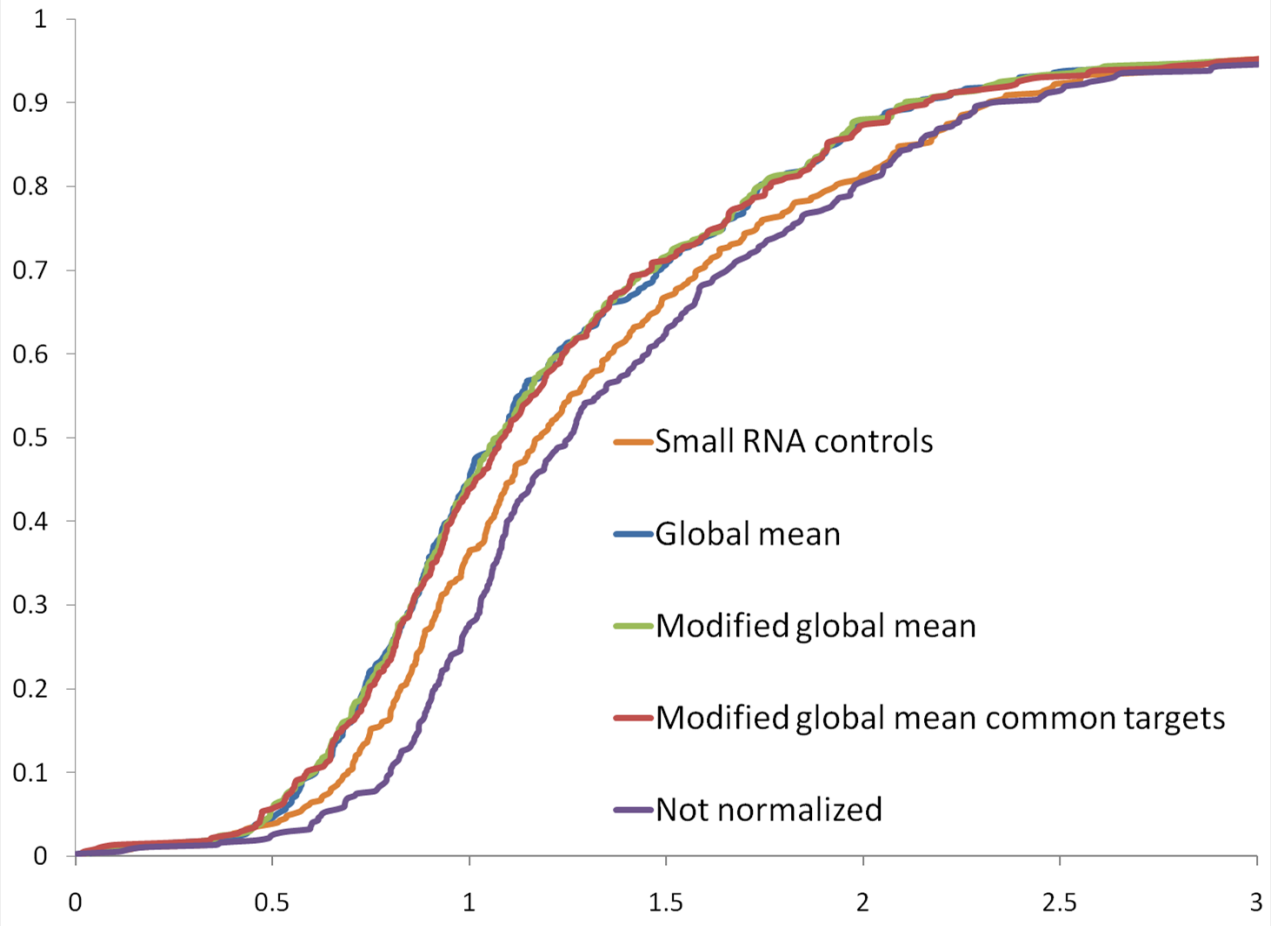
D pool normal tissues



Center for Medical Genetics

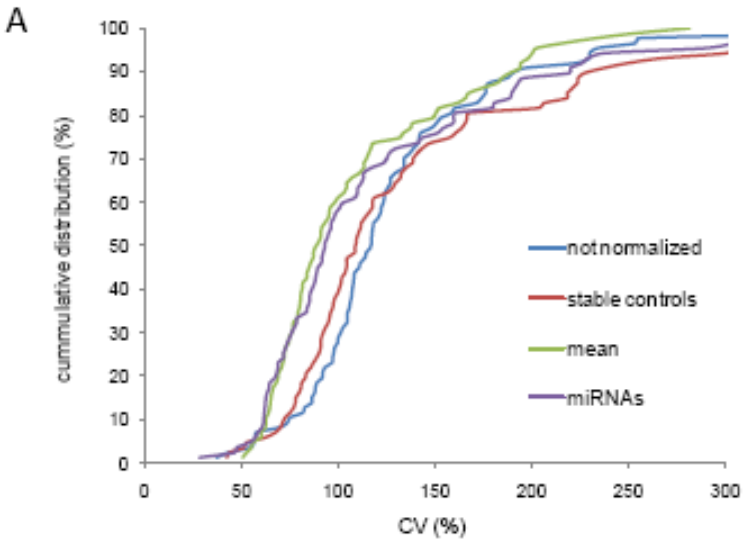
reduction of experimental variation (neuroblastoma) (II)

- cumulative noise distribution plot (more to left is better, less noise)
  - global mean methods remove more experimental noise

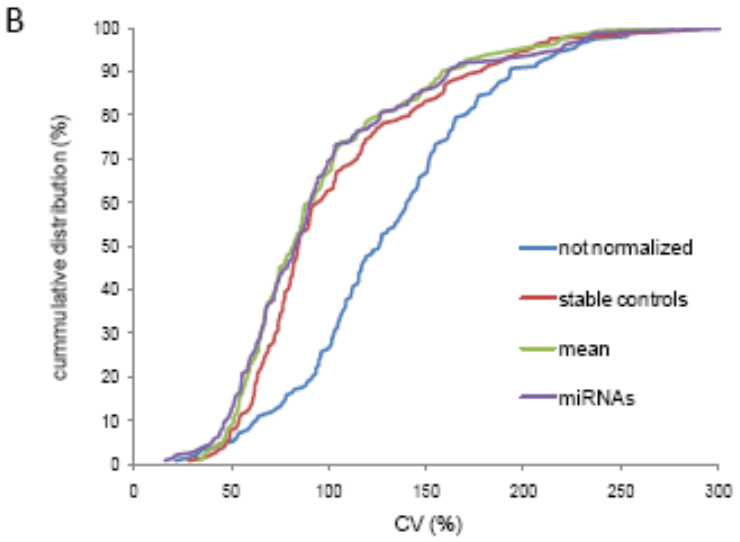


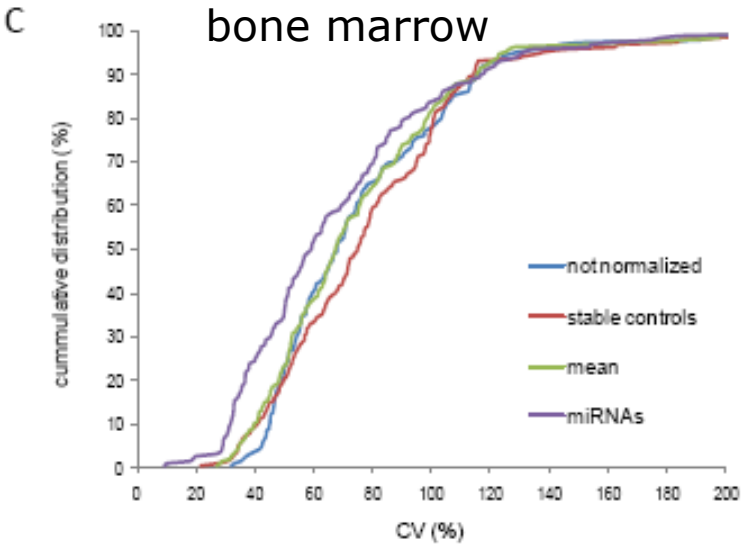
reduction of experimental variation (II)

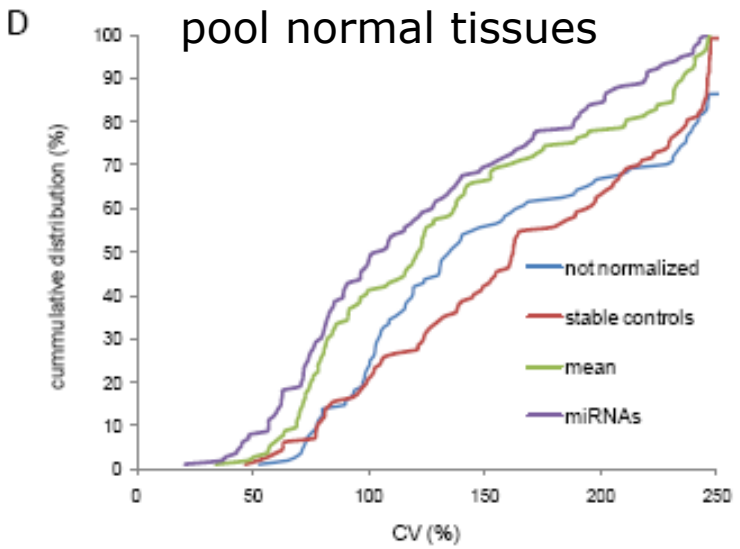
T-ALL



leukemia EVI1 overexpression



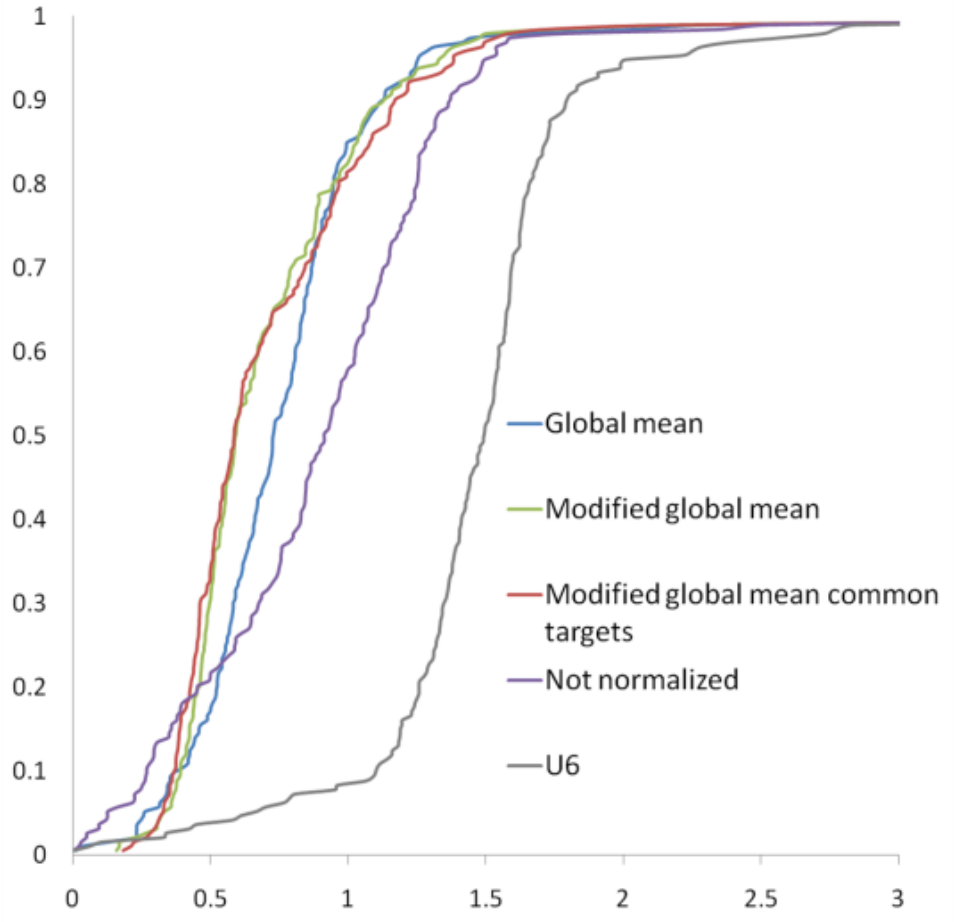




Center for Medical Genetics

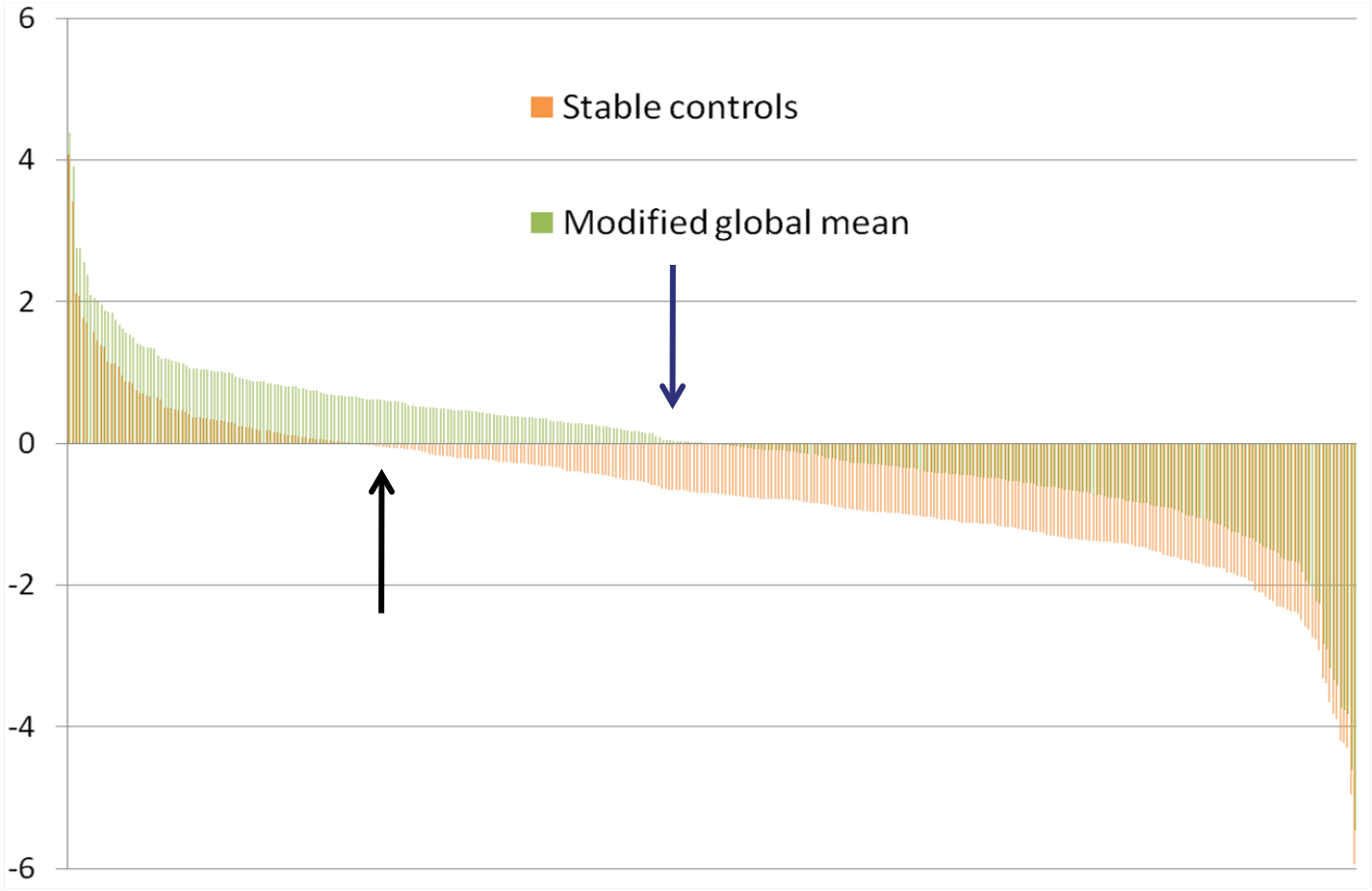
reduction of experimental variation (induced sputum) (II)

- U6 normalization (only expressed small RNA) induces more noise than not normalizing
- modified global mean is better than original global mean method



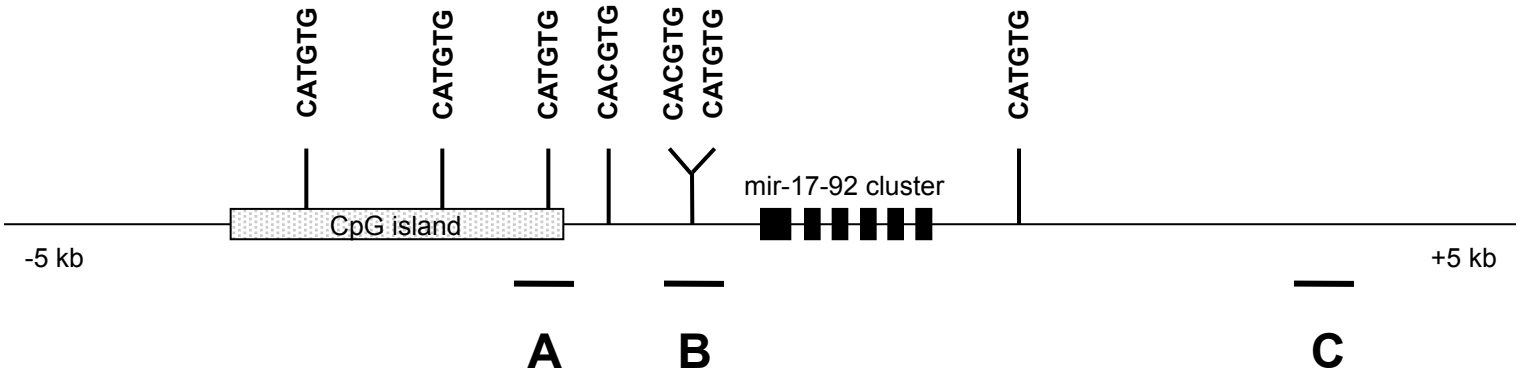
balancing differential expression (III)

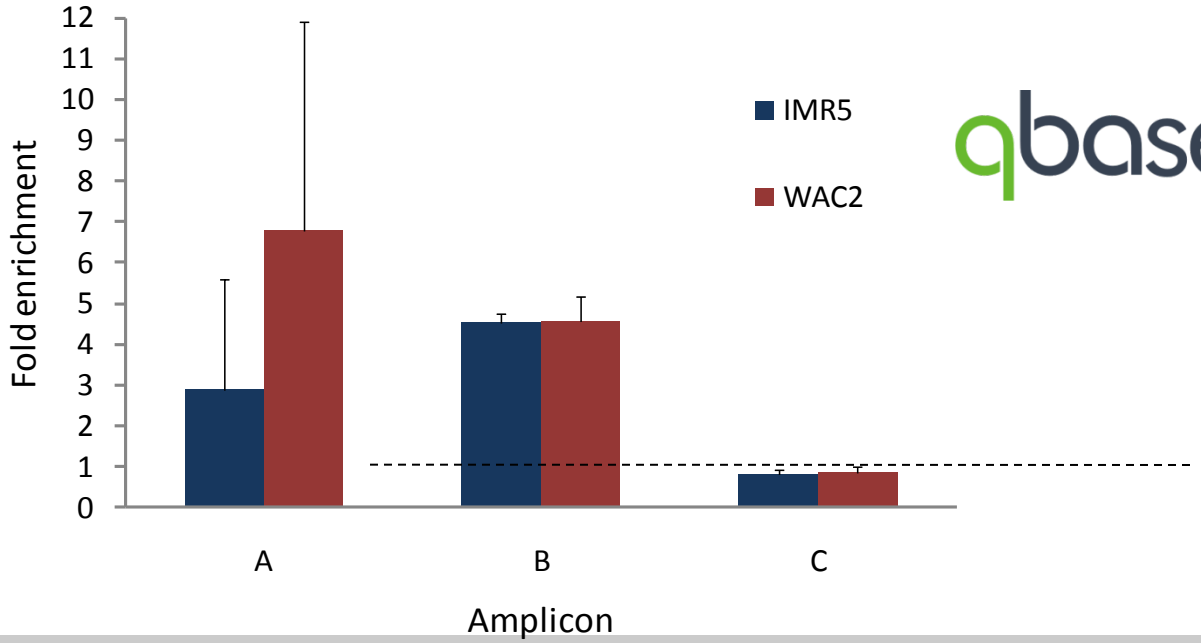
- fold changes in 2 cancer patient subgroups
- global mean normalization results in equal number of downregulated and upregulated miRs



better identification of differentially expressed miRs (IV)

- MYCN binds to the mir-17-92 promoter (poster 407)



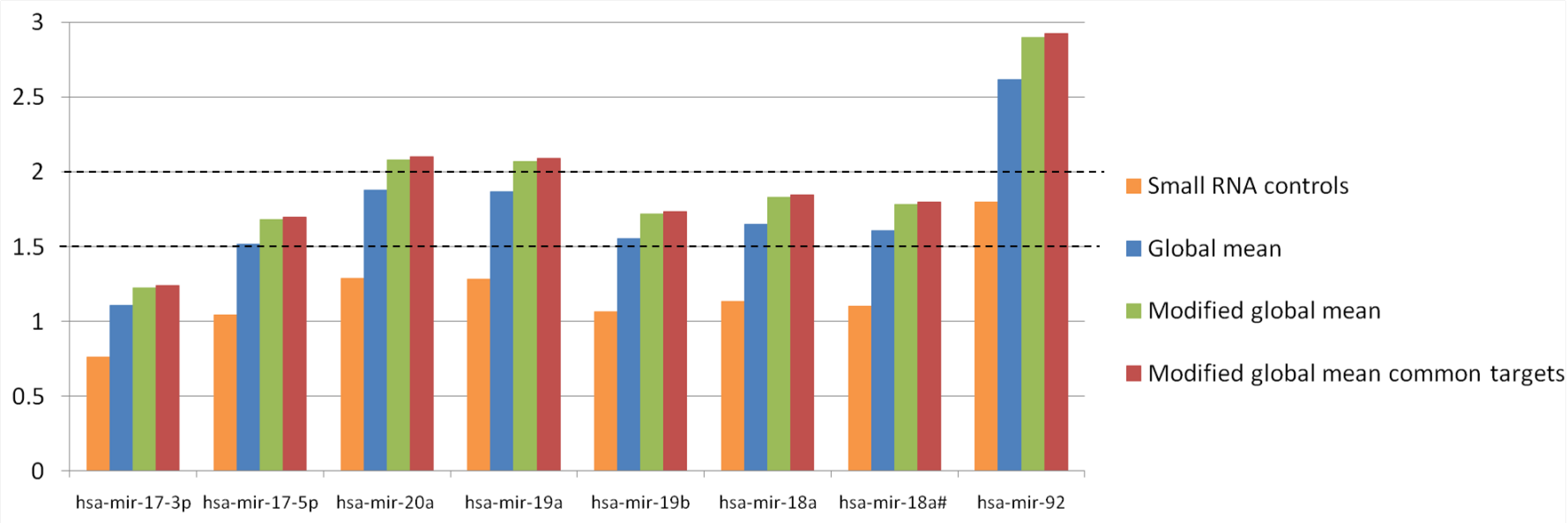


qbase PLUS

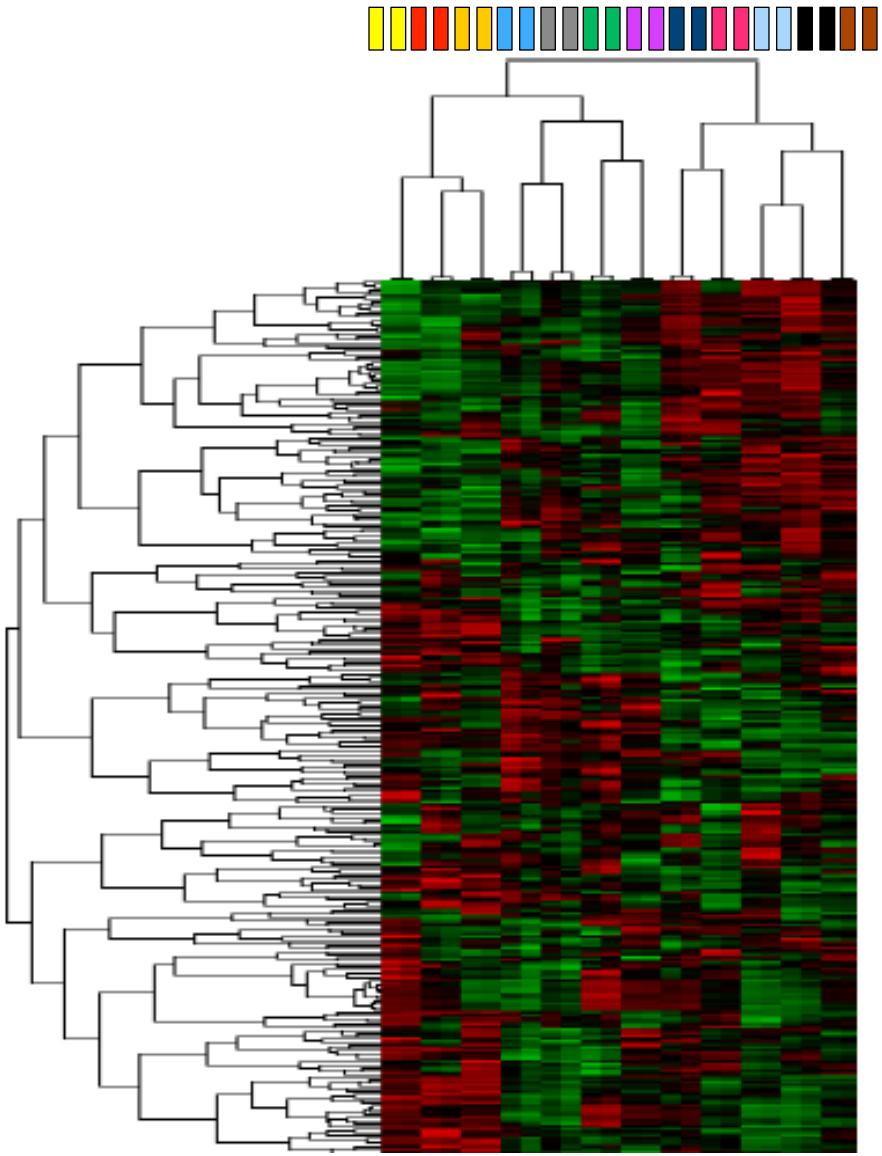
Center for Medical Genetics

better identification of differentially expressed miRs (IV)

- miR-17-92 expression in 2 subgroups of neuroblastoma (MYCN amplified vs. MYCN normal)
- global mean enables better appreciation of upregulation



strategy also works for microarray data

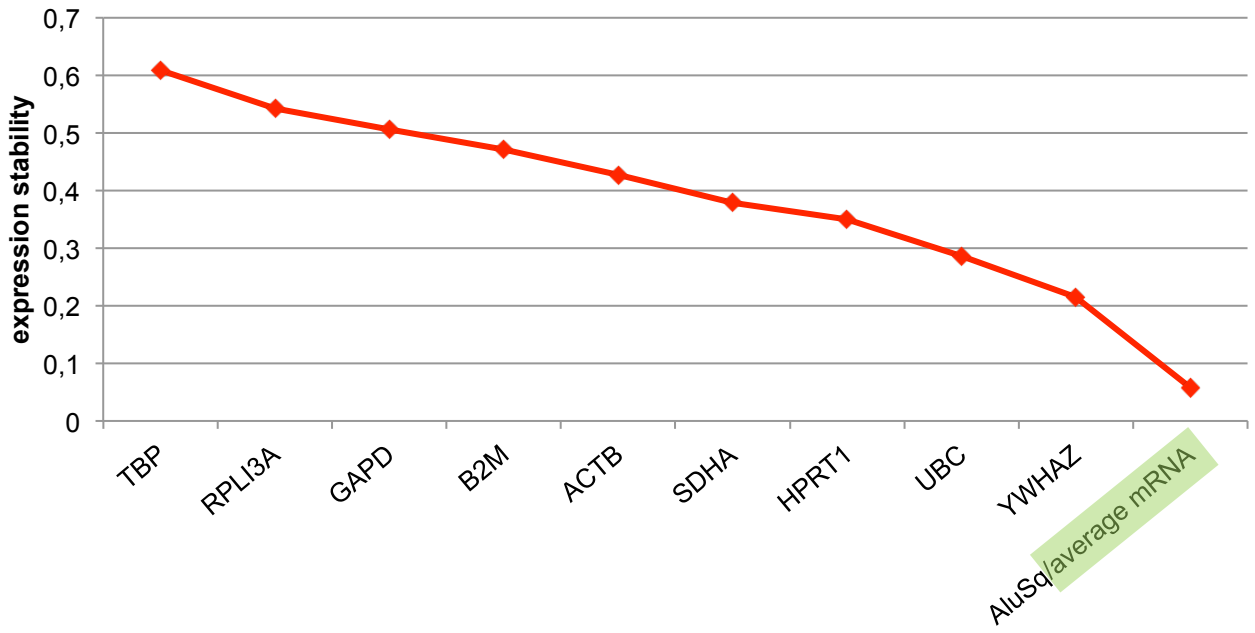


- each sample is measured by **RT-qPCR** and **microarray**
- global mean normalization
- standardization per method
- hierarchical clustering

- samples cluster by sample (and NOT by method)

strategy also works for mRNA data

- 4 MAQC samples (Canales et al., Nature Biotechnology, 2006)
- 201 MAQC consensus genes are measured
- geNorm analysis
  - 10 classic reference genes
  - global mean of 201 mRNAs



conclusions

- novel and powerful (miRNA) normalization strategy
  - best ranking according to geNorm
  - maximal reduction of experimental noise
  - balancing of differential expression
  - improved identification of differentially expressed genes

  - Mestdagh et al., Genome Biology, 2009
  - D'haene et al., in press (improved global mean)

Open Access

Method

A novel and universal method for microRNA RT-qPCR data normalization

Pieter Mestdagh*, Pieter Van Vlierberghe*, An De Weer*, Daniel Muth†, Frank Westermann†, Frank Speleman* and Jo Vandesompele*

Addresses: *Center for Medical Genetics, Ghent University Hospital, De Pintelaan 185, Ghent, Belgium. †Department of Tumour Genetics, German Cancer Center, Im Neuenheimer Feld 280, Heidelberg, Germany.

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

normalization in practice

- most powerful, flexible and user-friendly real-time PCR data-analysis software
  - based on Ghent University's geNorm and qBase technology
  - **state of the art normalization procedures**
    - o *one or more classic reference genes*
    - o *global mean normalization*
  - detection and correction of inter-run variation
  - dedicated error propagation
  - fully automated analysis; no manual interaction required

qbase PLUS

http://www.qbaseplus.com

acknowledgments

- UGent
  - Pieter Mestdagh
  - Filip Pattyn
  - Katleen De Preter
  - Frank Speleman

- Biogazelle
  - Barbara D'haene
  - Gaëlle Van Severen
  - Jan Hellemans