Prognostic Multigene Expression Classification of Cancer Patients: a Route for Success

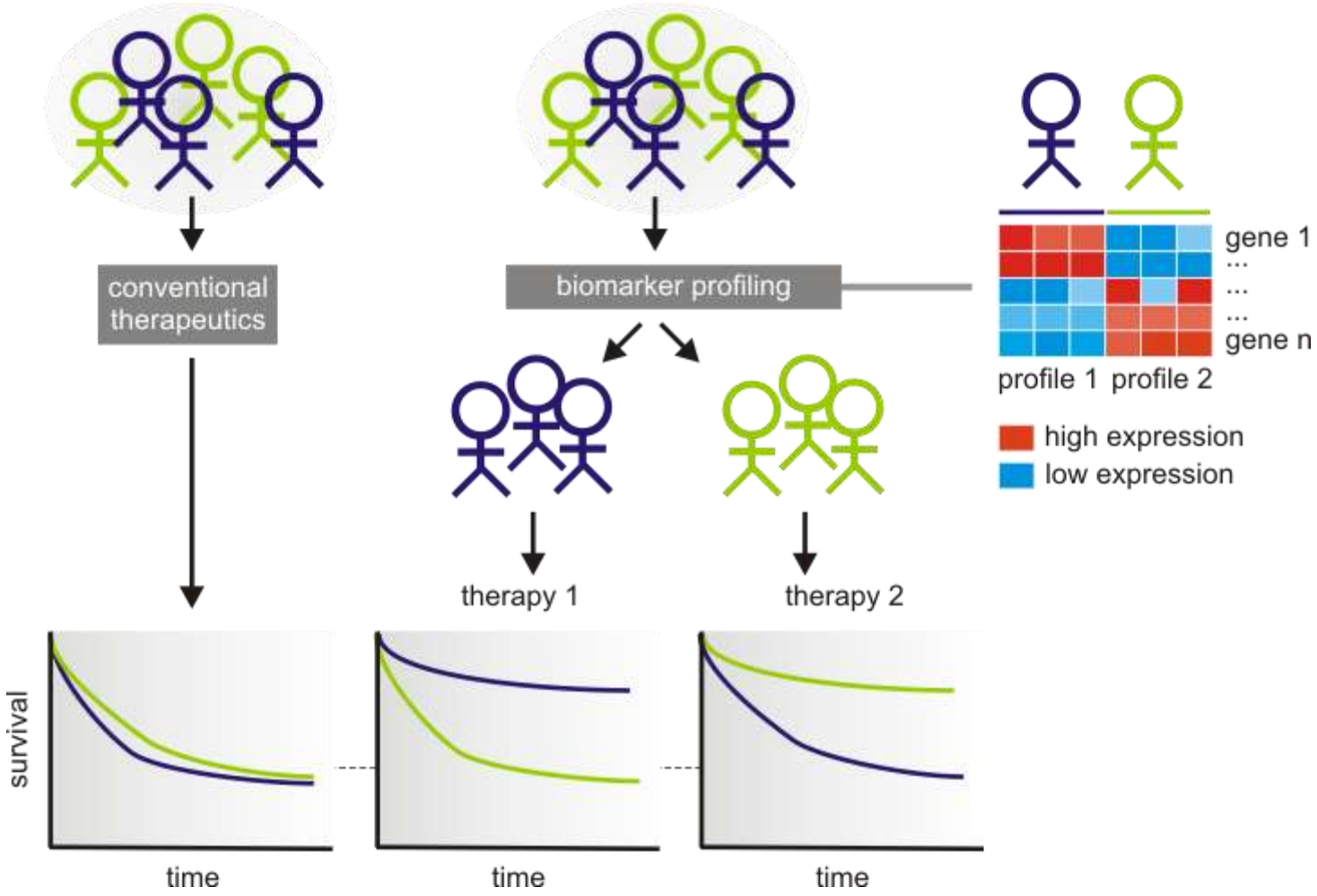
Jo Vandesompele
professor, Ghent University
co-founder and CEO, Biogazelle

4th International qPCR Symposium
Freising, March 9, 2009

outline

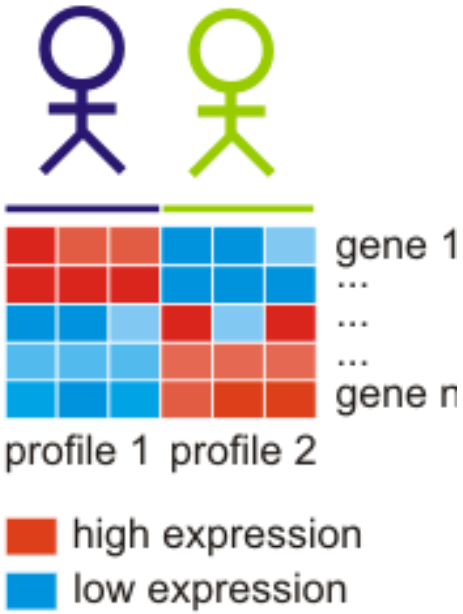
- background research & goals
- neuroblastoma
- prognostic marker selection
- study design and workflow
  - RNA quality control
  - sample pre-amplification
  - normalization
- data-analysis and results

biomarker signature based stratification



aim

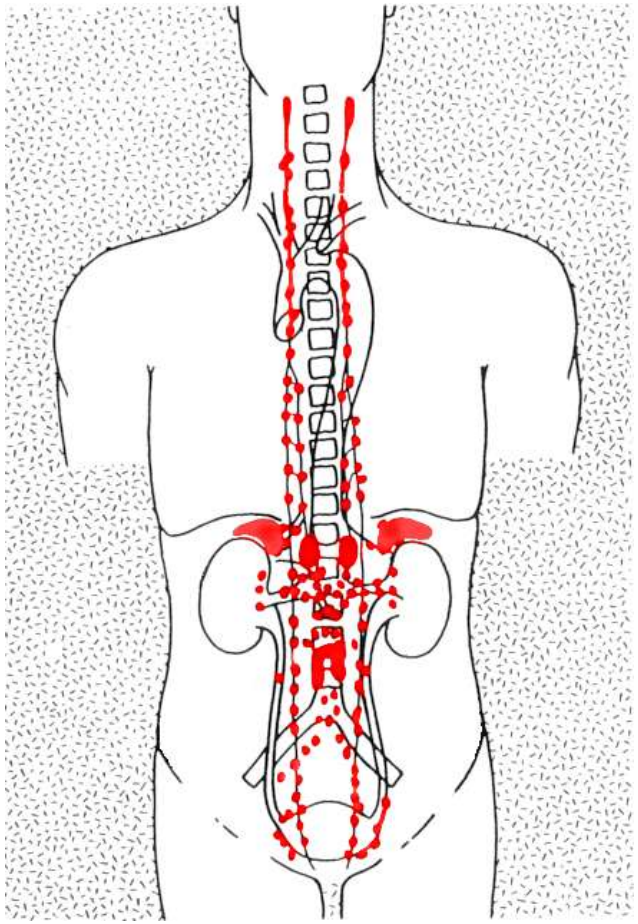
- development and validation of a robust prognostic gene signature for neuroblastoma using **real-time qPCR**
- identifying patients with
  - increased risk in the current low risk and high risk group
  - good molecular signature in the current high risk group
- better choice of risk-related therapy



neuroblastoma

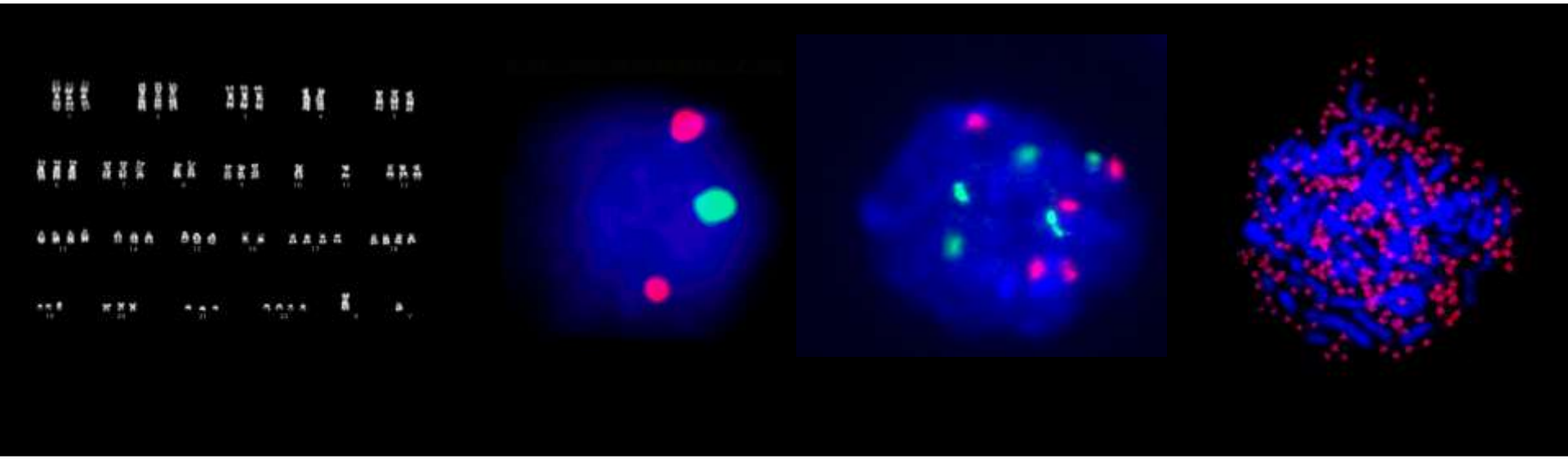
- most frequent extra-cranial solid tumor in children
- 95% <10 years at diagnosis
- 1:100,000 children (< 15 years)
  - 20 cases/year Belgium
  - 700 cases/year USA
- 15% of childhood cancer deaths
- originates from primitive (immature) sympathetic nervous system cells

- remarkably variable clinical course
  - spontaneous regression or maturation
  - widespread metastasis and rapid death



neuroblastoma

- prognosis is dependent on
  - tumor stage (localized vs. metastatic disease)
  - age at diagnosis (< or > 1 year)
  - genetic defects: amplification MYCN, ploidy, loss of 1p, gain of 17q



prognostic classification

- misclassifications resulting in overtreatment or undertreatment
- need for additional tumor-specific prognostic markers
- current microarray gene expression studies
  - data overfitting
  - unstable gene lists
  - lack of overlap

  - biological & technical noise
  - much more genes than samples
  - probe annotation / platform
  - different risk definition
  - different data processing and analysis

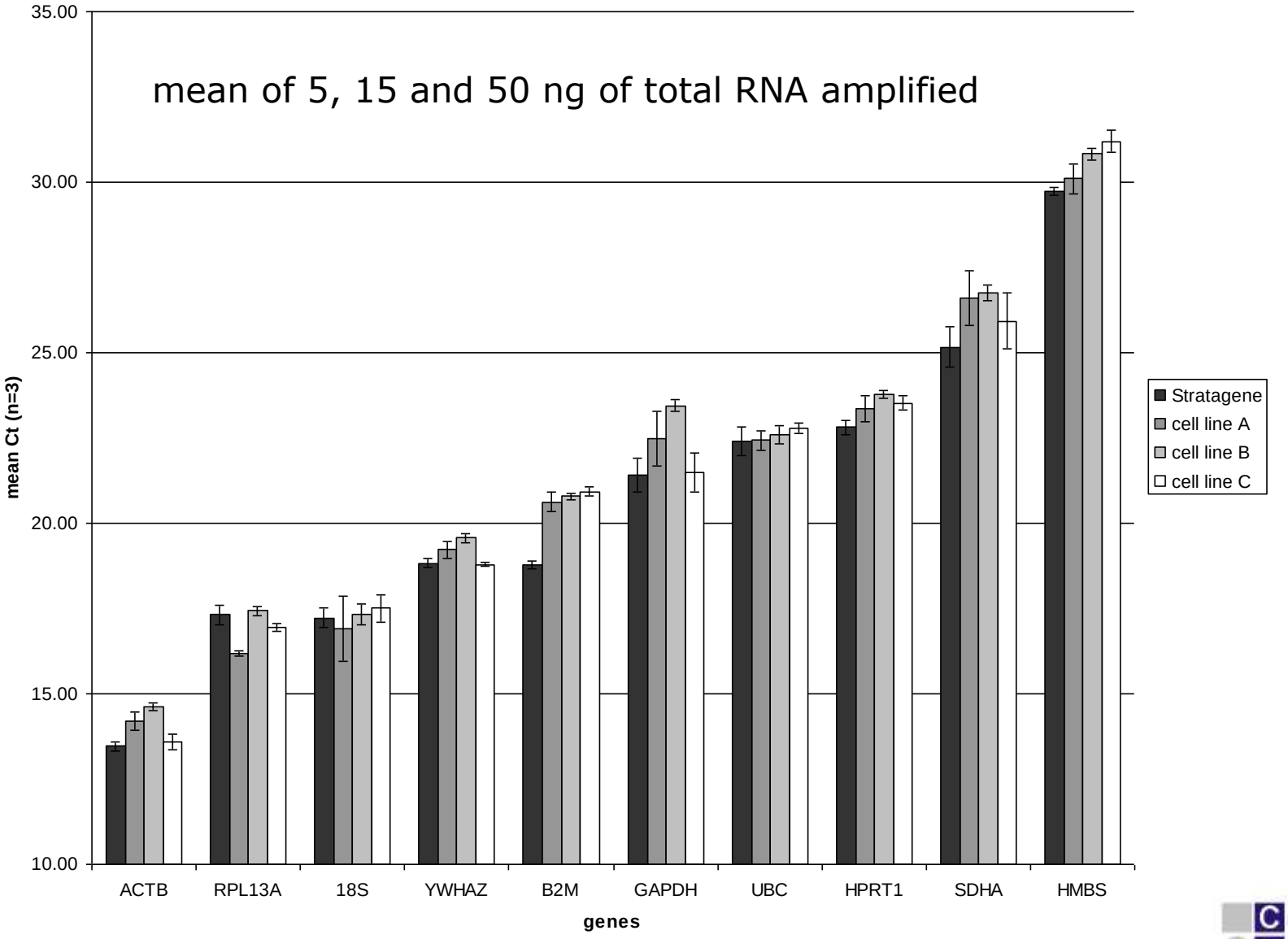
study workflow

selection of a top ranking list of 59 prognostic markers	• meta-analysis of 7 published microarray gene expression studies • literature screening of almost 800 abstracts from single-gene studies
RNA quality control 423 samples	• two PCR-based assays • capillary gel electrophoresis (Experion)
sample pre-amplification (WT-Ovation)	
analysis of 366 primary untreated neuroblastoma tumours using real-time qPCR	
data-analysis	• Prediction Analysis of Microarrays • Kaplan-Meier • Cox proportional hazards

towards real-time PCR signature profiling

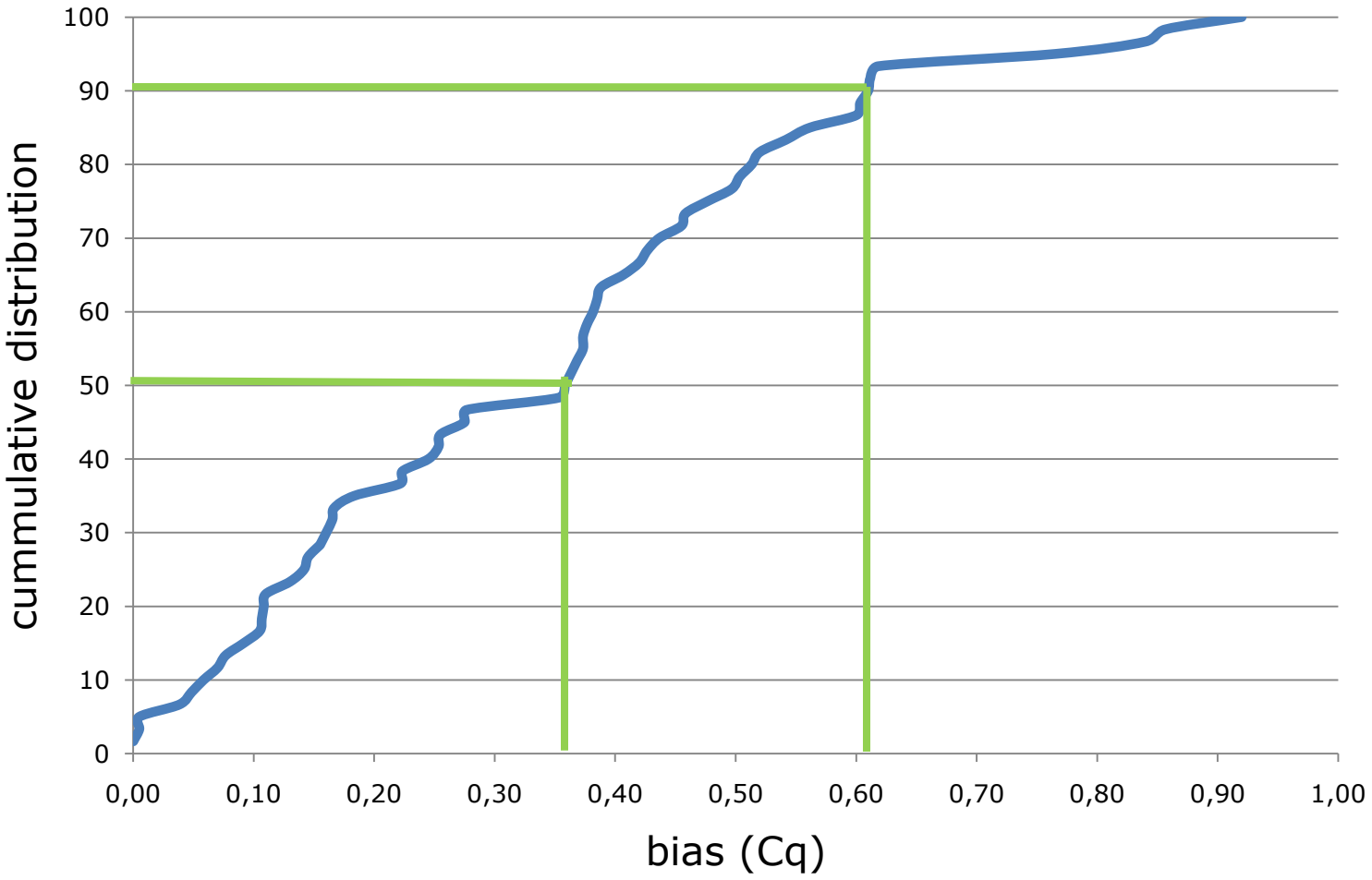
- 100 ng total RNA
  - 30 ng quality control
  - 10 ng unbiased amplification
    WT-Ovation (NuGEN)
- PCR assay design and validation
  - sensitivity, specificity and efficiency
    RTPrimerDB
    (Pattyn et al., 2006, NAR; Lefever et al, 2009, NAR)
  - absolute standards
- real-time PCR using 384-well format
  - sample maximization strategy
    (Hellemans et al., Genome Biology, 2007)
  - 360 tumors and 1 gene/plate

WT-Ovation reproducibility



Center for Medical Genetics

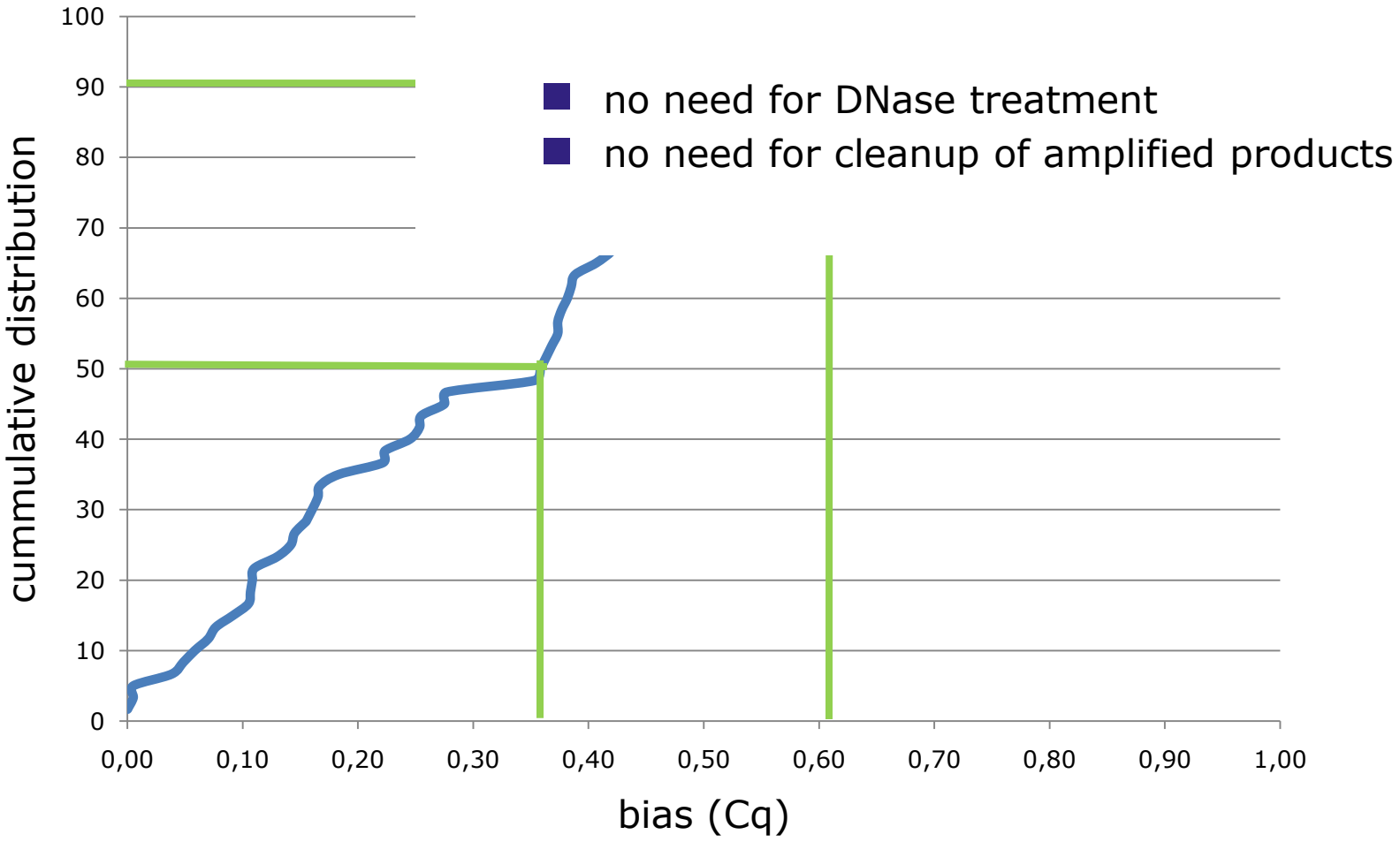
WT-Ovation amplification bias



median bias = 0.36, 90%tile bias = 0.61

Center for Medical Genetics

WT-Ovation amplification bias

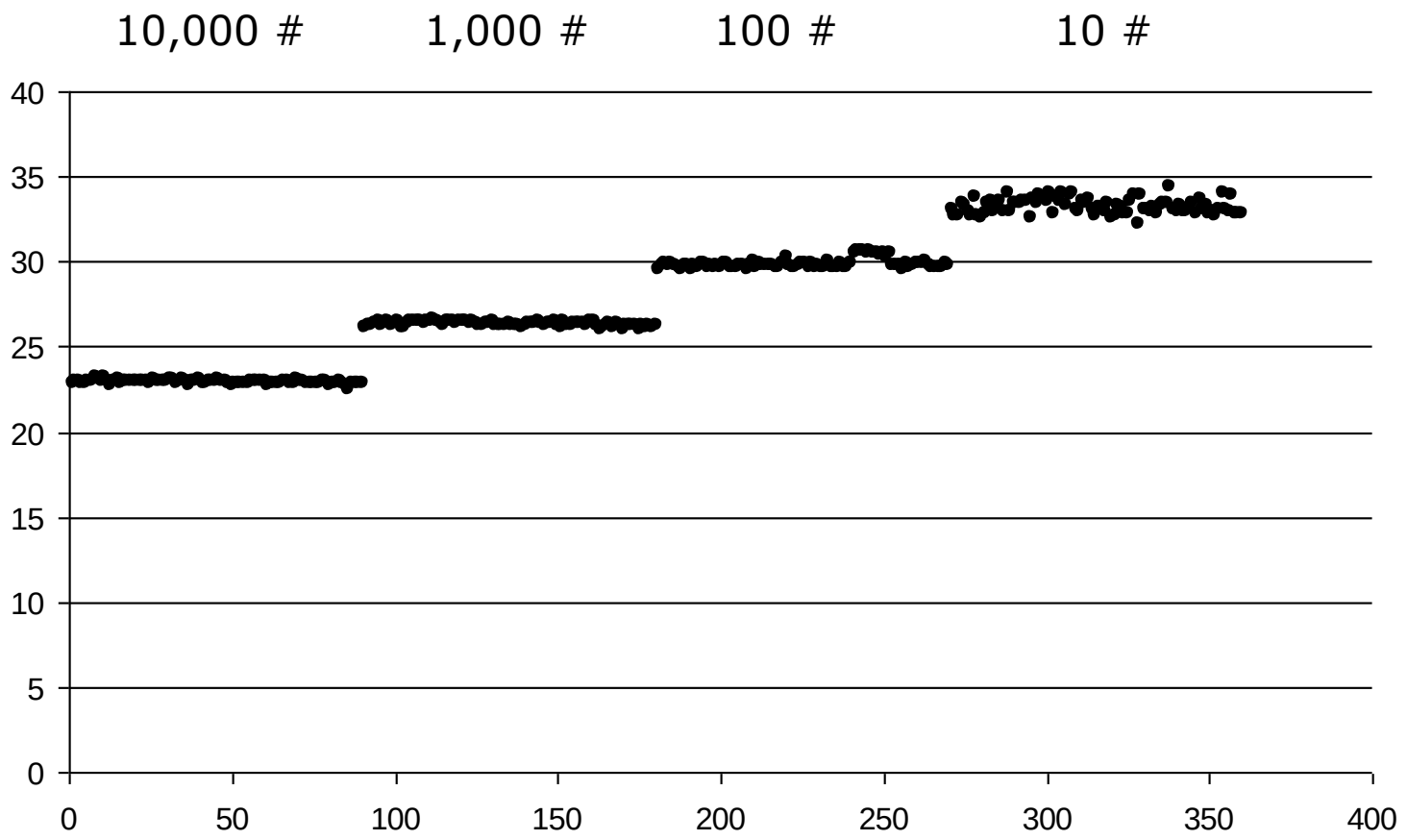


median bias = 0.36, 90%tile bias = 0.61

Center for Medical Genetics

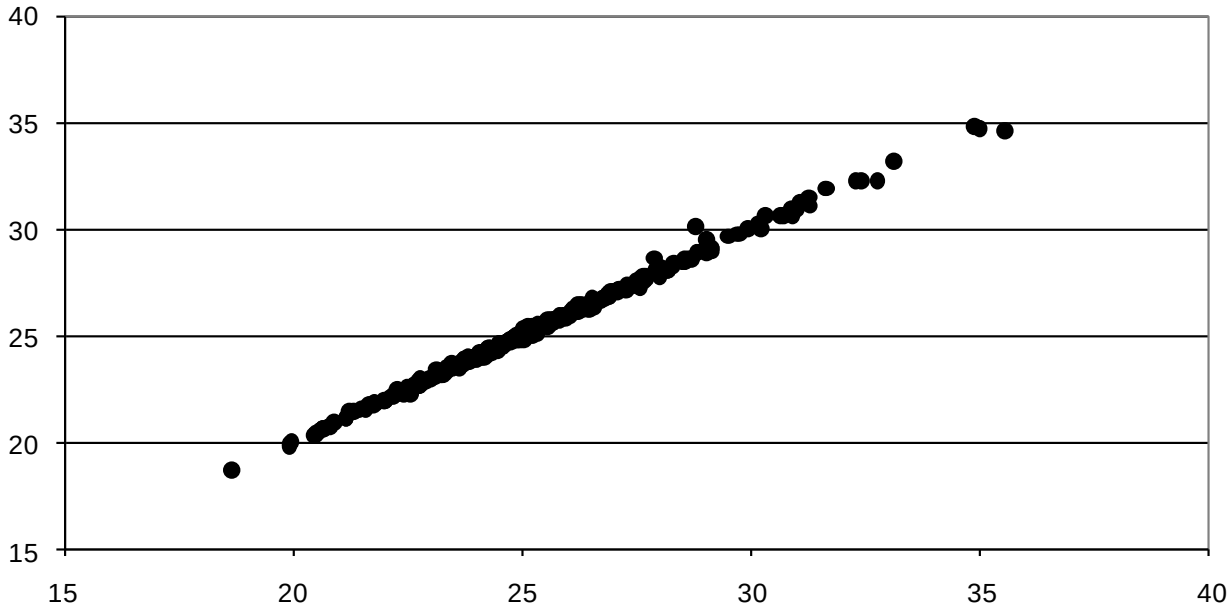
qPCR reproducibility

- within a 384-well plate: 4 x 96 replicates



qPCR reproducibility

- between two identical 384-well plates
- maximum ΔCq: 0.45

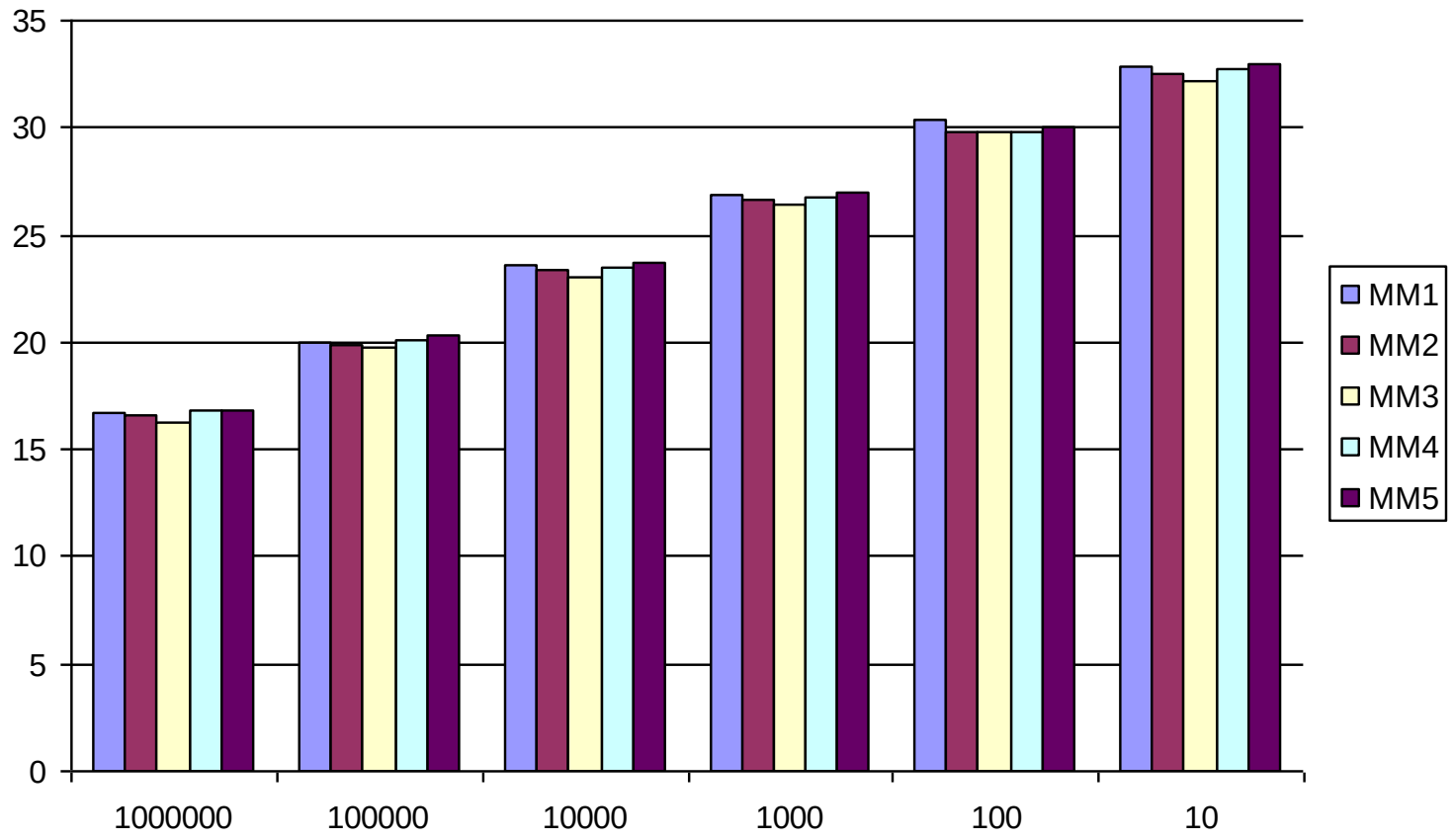


absolute standards

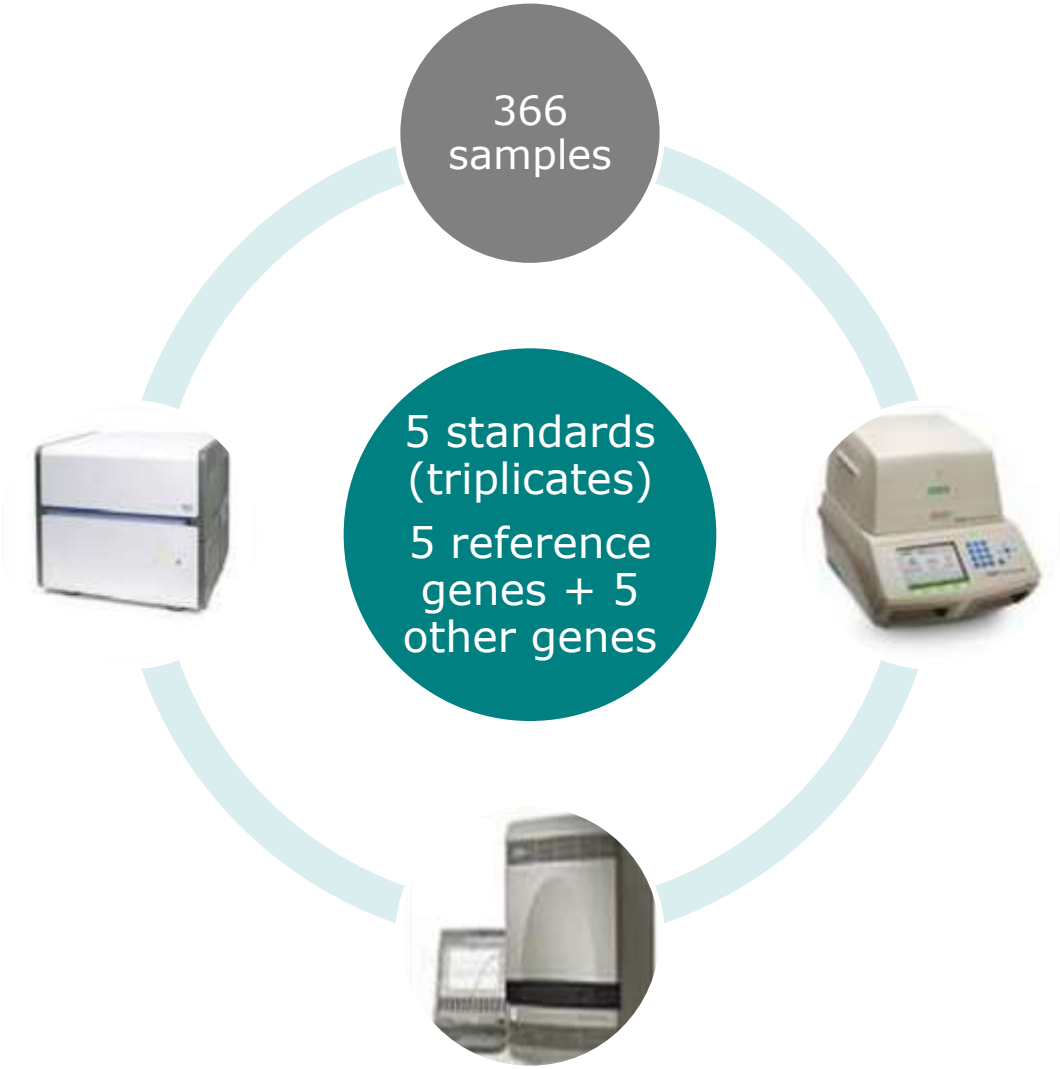
- synthetic control
  - 55 nucleotides
  - PAGE purification
  - blocking group
  - 5 points dilution series: 15 molecules > 150.000 molecules

absolute standards

- reproducibility across master mixes (5) and instruments (2)

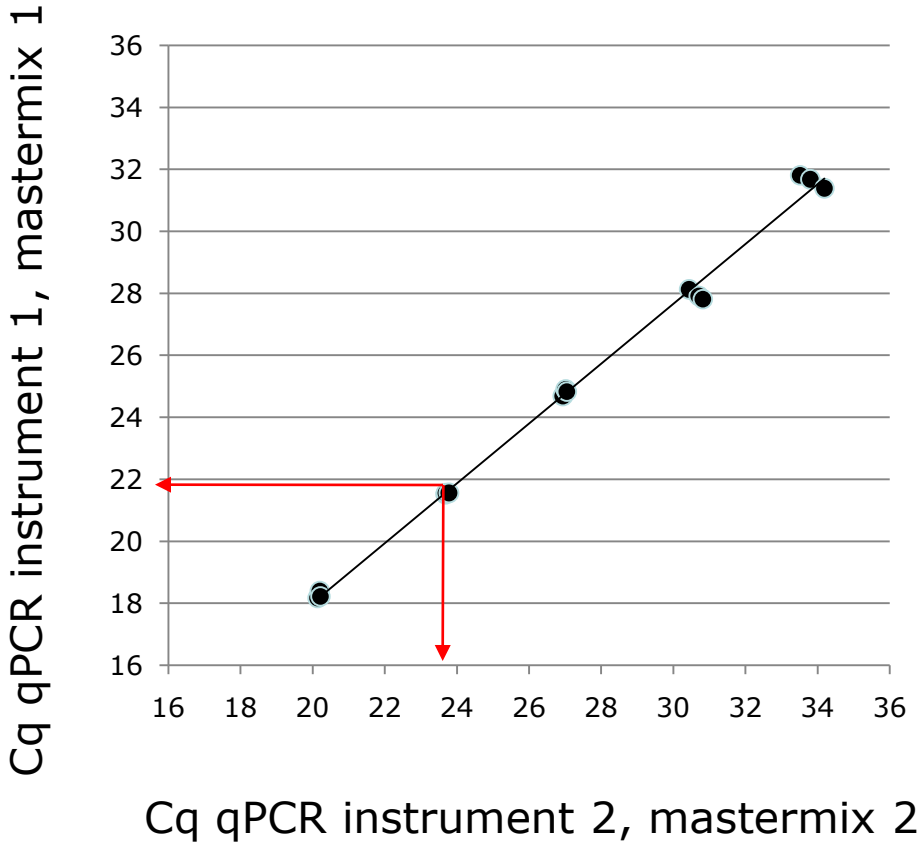


absolute standards cross lab comparison



absolute standards cross lab comparison

- 5 standards (triplicates)



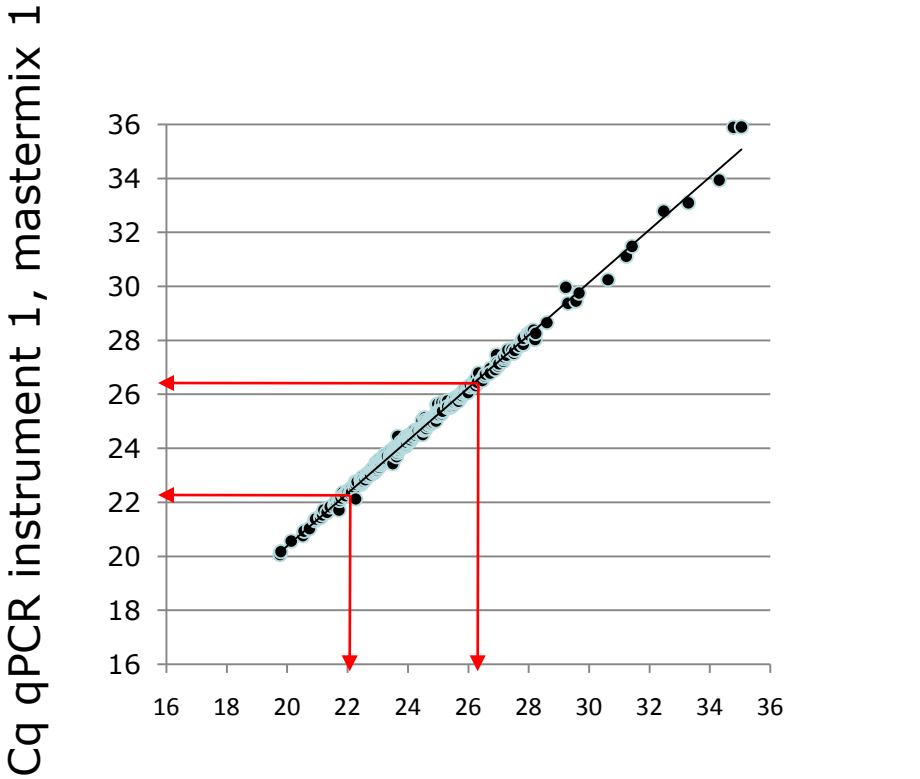
average ΔCq standards

⇓

correction Cq samples

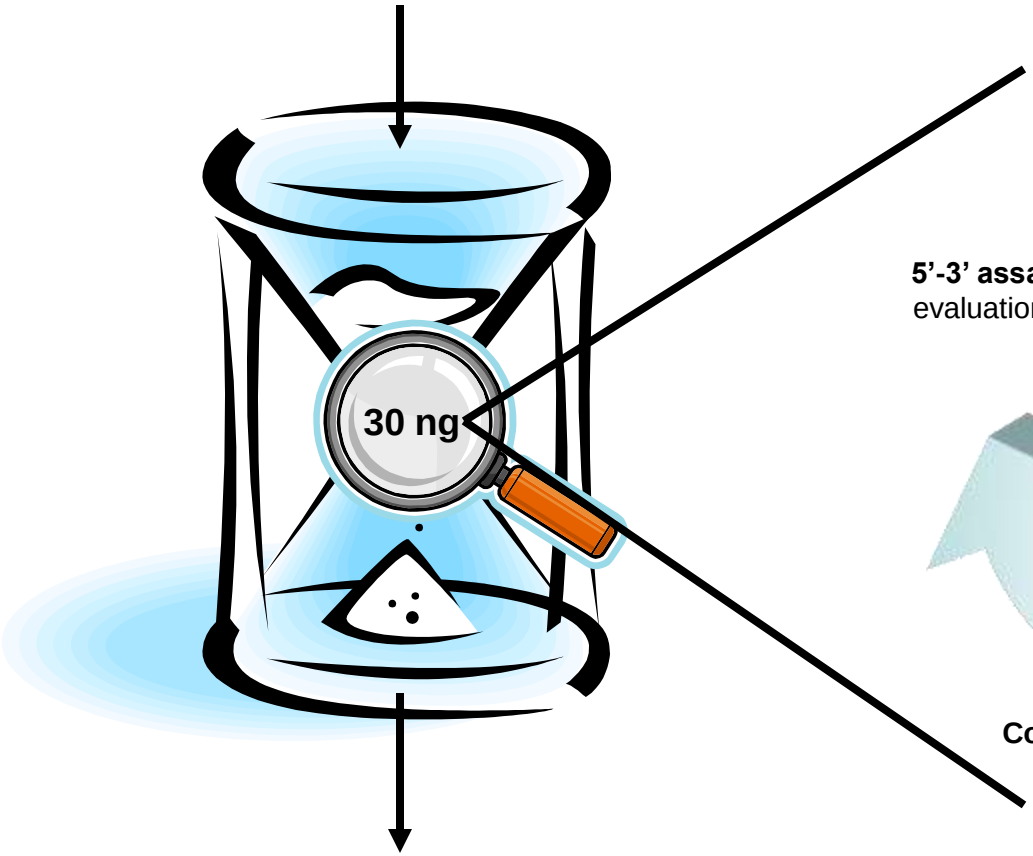
absolute standards cross lab comparison

- ARHGEF7 gene
  - 366 samples
  - use of 5 standards (triplicates) for correction



rigorous control of RNA quality

423 primary untreated NB (100 ng total RNA)



353 RNA samples

5'-3' assay (HPRT1):
evaluation of mRNA integrity

SPUD assay (Nolan et al, 2006):
detection of inhibitors

Computed gel analysis (Experion, Biorad):
evaluation of total RNA quality

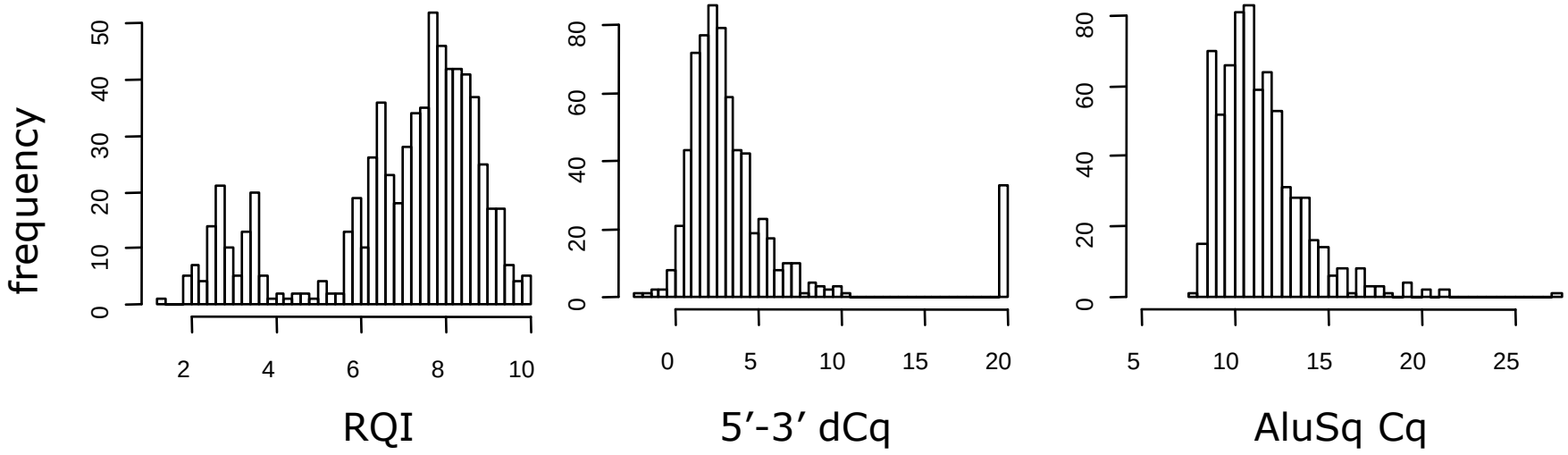
Center for Medical Genetics

impact of RNA quality on expression stability

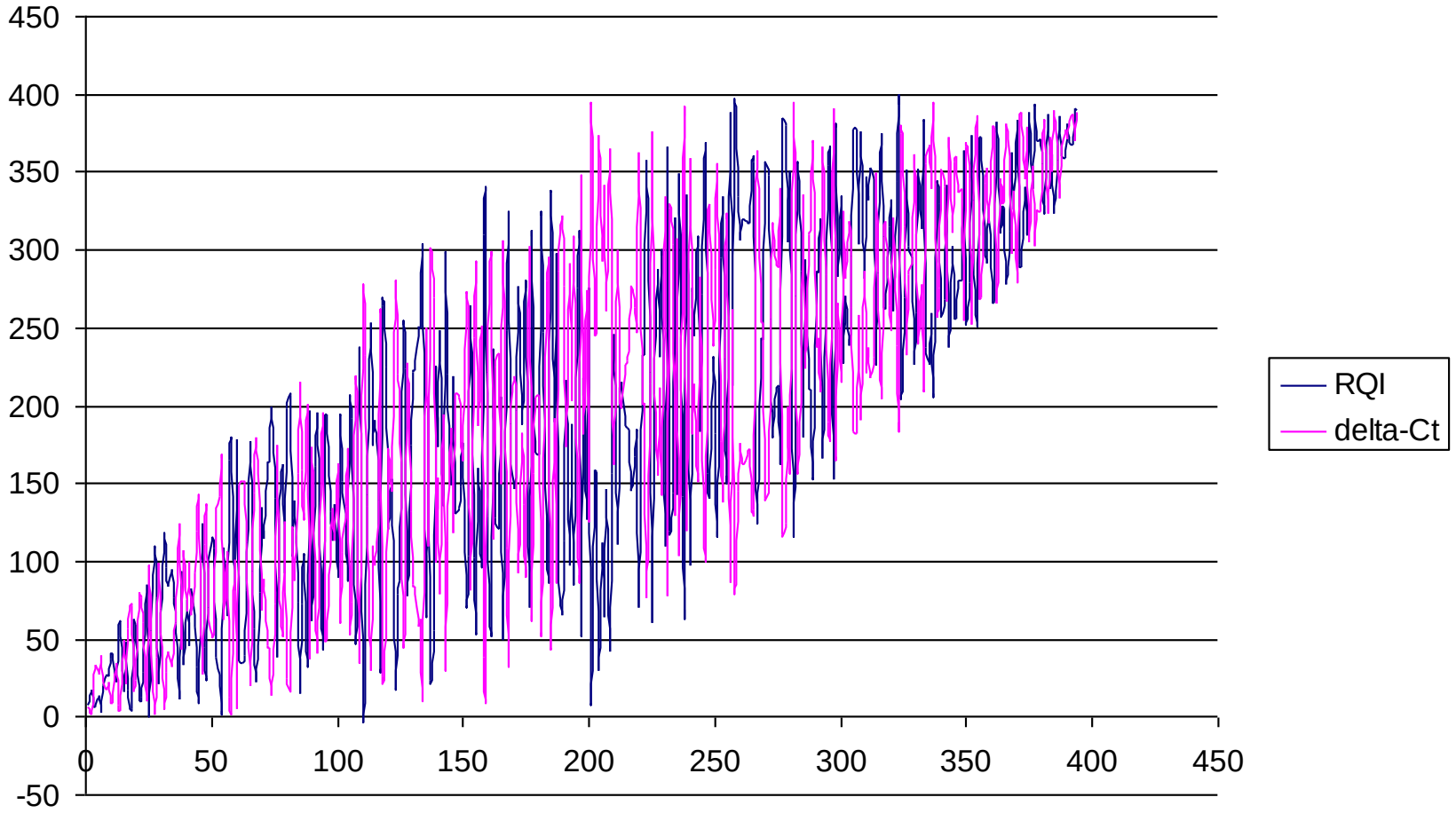
- differences in reference gene ranking between intact and degraded RNA (Perez-Novo et al., Biotechniques, 2005)

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	RPL3IA	RPL3IA	RPL3IA
4	TBP	B2M	GAPD	UBC
5	RPL3IA	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

RNA quality parameters



delta-Cq 5’-3’ vs. RQI



RNA samples ordered by average rank (good -> worse)

RNA quality control & sample selection

- 423 samples:
  - 4 samples < DOT/DOO without event
  - 5 samples < presence of enzymatic inhibitors (SPUD)
  - 20 samples < lack of mRNA integrity (no ΔCq 5'-3')
    - o *12/14 failed WT-Ovation*
    - o *all low RQI values*
  - 28 samples < poor RNA quality (RQI + ΔCq 5'-3')

- 366 best samples (86.5 %)
  - RQI:
    - o *average = 7.4*
    - o *median = 7.6*
    - o *90%-tile > 6.1*
  - ΔCq 5'-3':
    - o *average = 2.36*
    - o *median = 2.06*
    - o *90%-tile < 4.75*

normalisation using geNorm technology

- 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

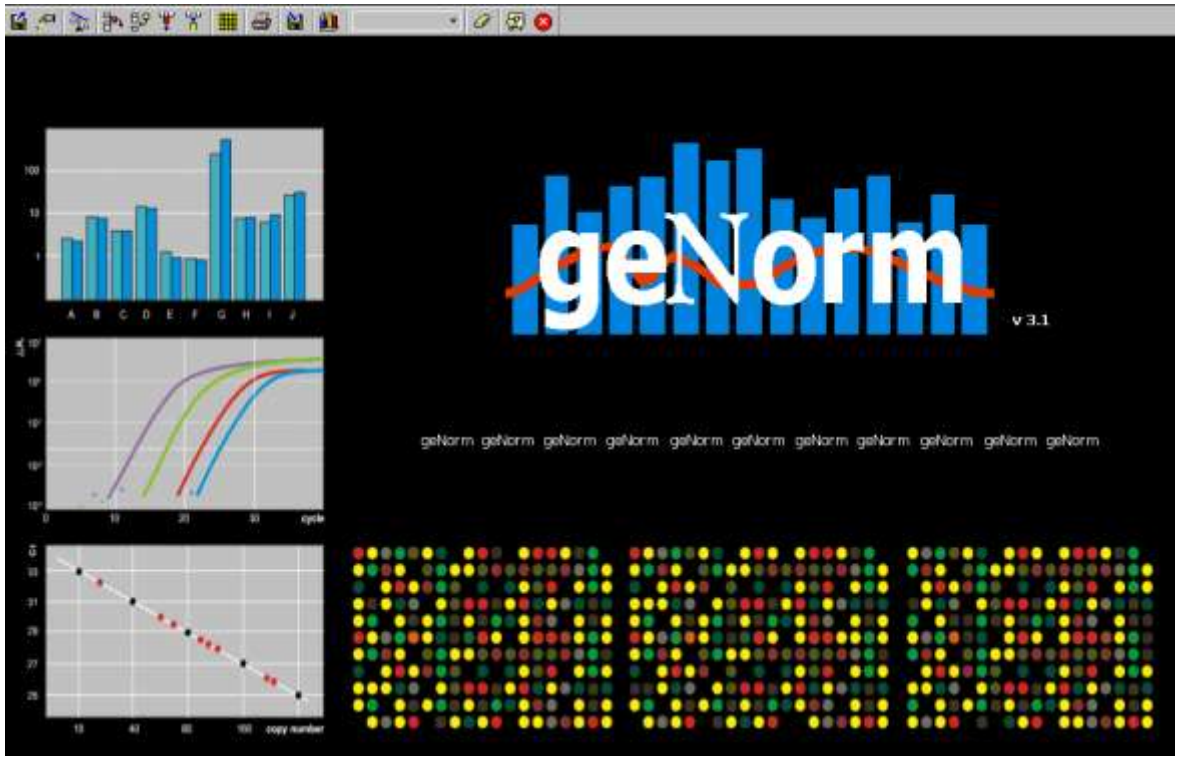
Research

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 software

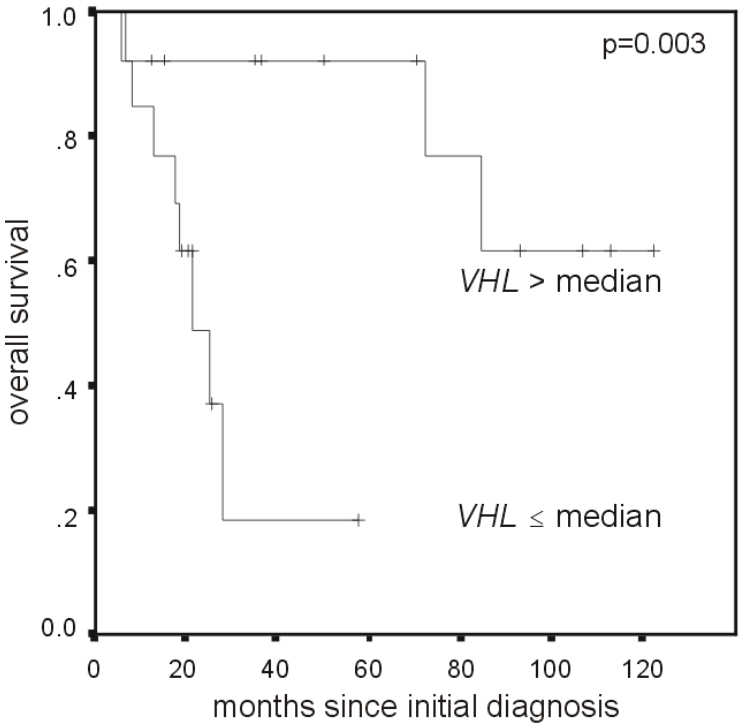
- automated analysis
  - ranking of candidate reference genes according to their stability
  - determination of how many genes are required for reliable normalization



http://medgen.ugent.be/genorm

geNorm validation (I)

- cancer patients survival curve
  statistically more significant results



log rank statistics

NF4

0.003

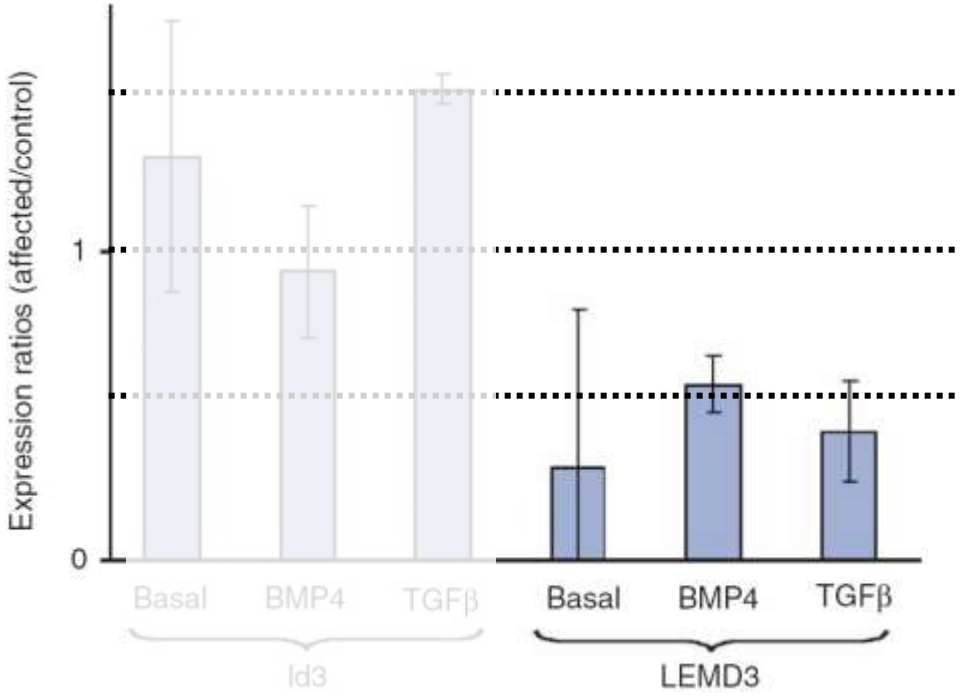
NF1

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

normalisation using multiple stable reference genes

- use of multiple references is now well established
  - > 1250 citations of our geNorm technology in PubMed
  - > 8000 geNorm downloads in 100 countries

data analysis

- data analysis using 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

qbase PLUS

http://www.qbaseplus.com – booth in room S1

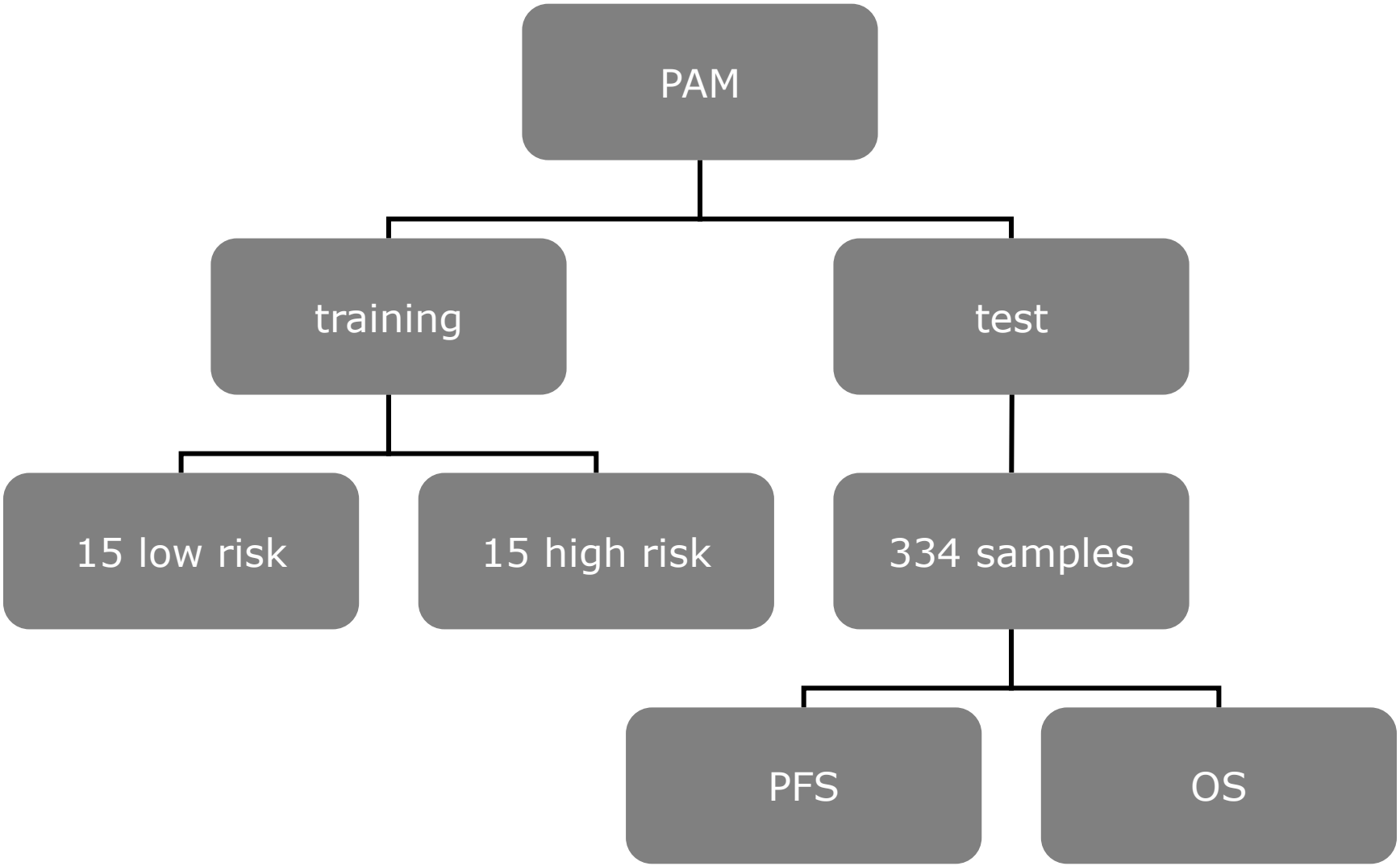
data analysis

- data analysis using 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

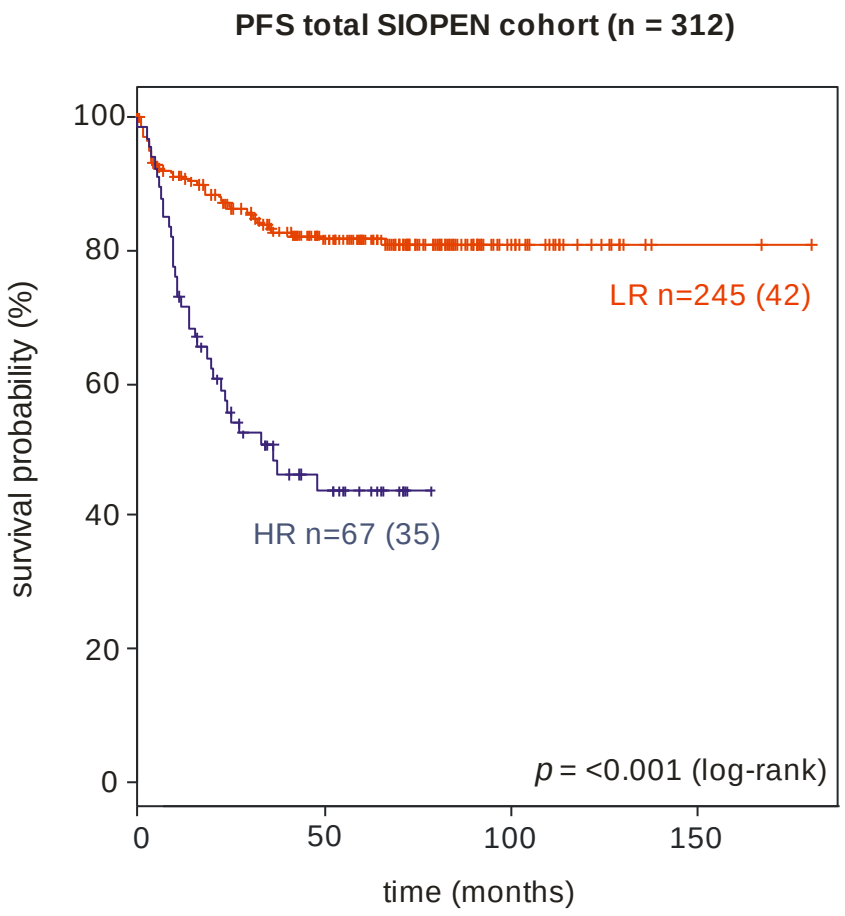
- 59 prognostic markers + 5 reference genes
- 364 samples (2 failed for 1 reference gene)

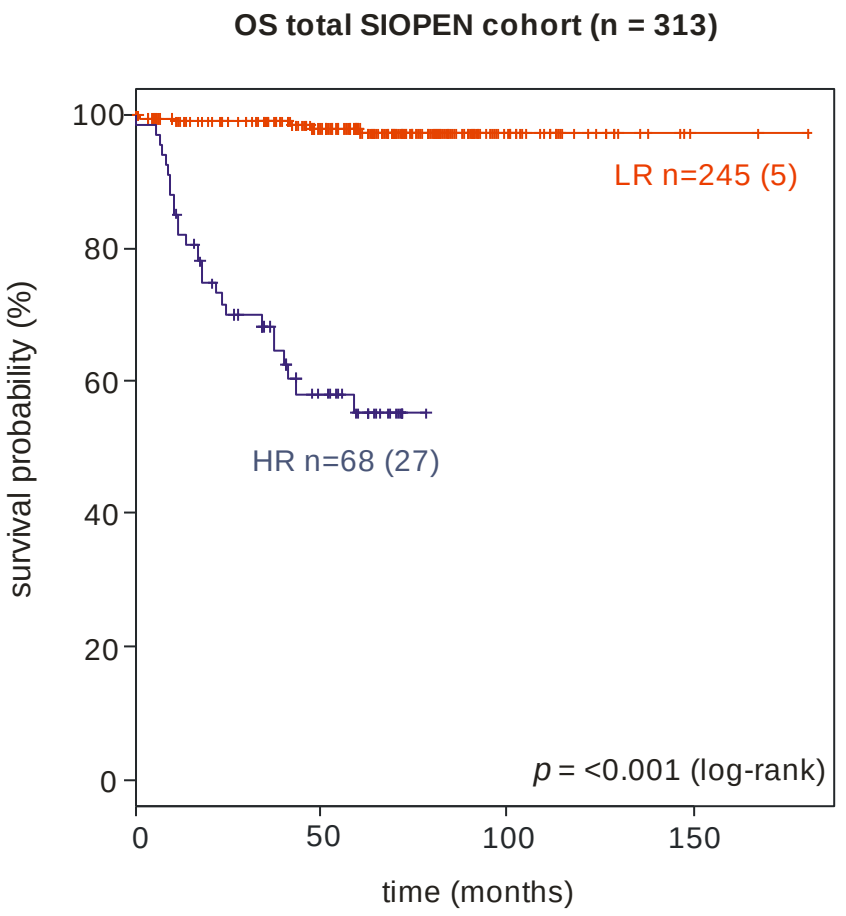
- hierarchical clustering
- survival analysis
  - Prediction Analysis of Microarrays
  - Cox proportional hazards modeling
  - Kaplan-Meier

Prediction Analysis of Microarrays



classification of patients with respect to PFS and OS

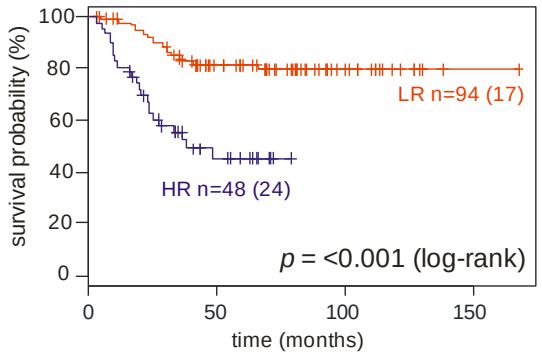




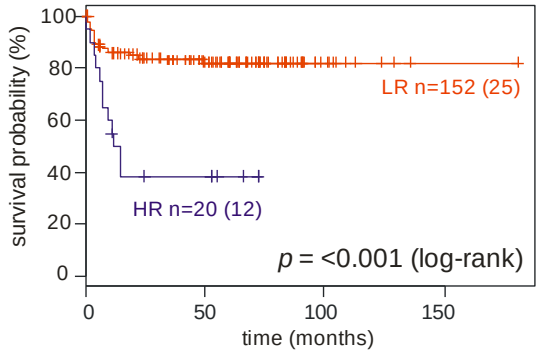
Center for Medical Genetics

value of the classifier in relation to currently used risk factors: PFS

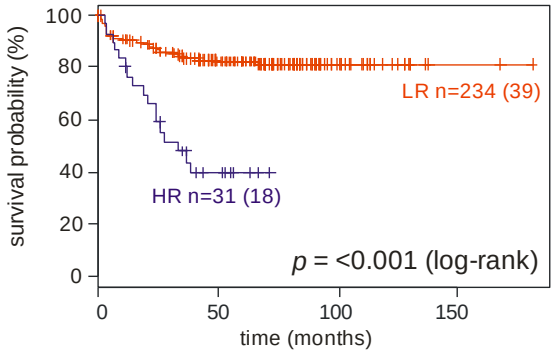
PFS age > 12 months (n = 142)



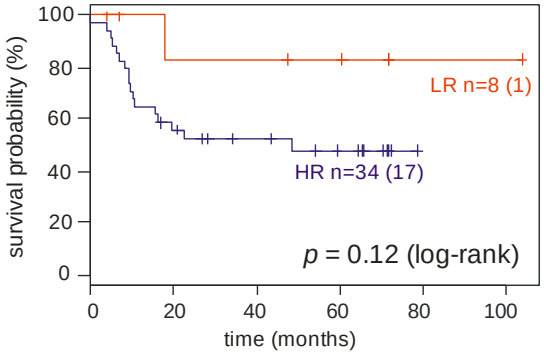
PFS age <=12 months (n = 172)



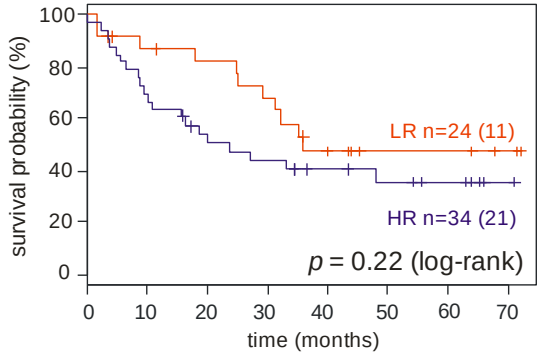
PFS *MYCN* single copy (n = 265)



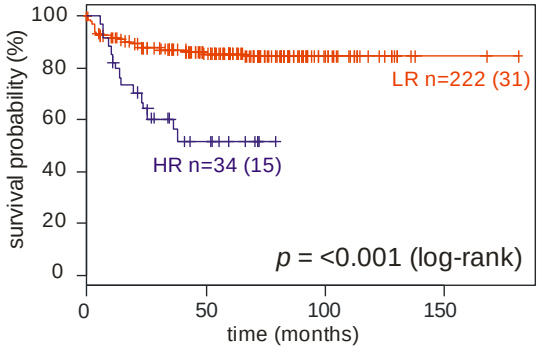
PFS *MYCN* amplification (n = 42)



PFS stage 4 (n = 58)



PFS not stage 4 (n = 256)



Center for Medical Genetics

Cox multivariate analysis

PAM classifier

multivariate cox analysis

independent predictor (age, stage, MYCN)

strong independent predictor:

patients with high molecular risk have

a **19-fold higher risk** to die from disease
a **4-fold higher risk** for relapse/progression

compared to patients with low molecular risk

RNA quality control cut-off

- depends on the application
  - microarray vs. qPCR
  - expression difference of the target(s)
  - abundance & stability of the target(s)
  - fresh frozen vs. FFPE

- based on the performance of our classifier

  ROC AUC accuracy analysis

	bad	good
RQI	<3: 0.27	≥3: 0.82
5’-3’ dCq	>7: 0.43	≤7: 0.79
AluSq Cq	>15: 0.13	≤15: 0.81

conclusions (I)

- validation matters – quality control along the entire workflow
  - assay performance
  - template quality
  - normalization
  - data-analysis

http://www.sabustin.org

conclusions (II)

- largest qPCR gene-expression study (rigourous RNA quality control)
  - optimized workflow
  - using minimal amounts of RNA (100 ng)
  - use of absolute standards (cross-lab comparison)
  - selected gene list (59) on a large panel of tumours (589)
- robust multigene expression prognostic classifier
  - **validated on an independent set of tumours**
  - **independent after controling for other known risk factors**
  - suitable for routine lab tests
- this study might form the basis for future research, i.e. prospective studies
- cDNA library source for future qPCR gene expression studies

acknowledgements

Center for Medical Genetics, Ghent, Belgium

- Frank Speleman
- Jo Vandesompele
- Nadine Van Roy
- Katleen De Preter
- Jasmien Hoebeeck
- Filip Pattyn
- Tom Van Maerken
- Joëlle Vermeulen
- Nurten Yigit
- Els De Smet
- Liesbeth Vercruysse
- Anne De Paepe

Collaborators

- Geneviève Laureys
- Gianpaolo Tonini
- Olivier Delattre
- Jean Bénard
- Valérie Combaret
- Raymond Stallings
- Angelika Eggert
- Akira Nakagawara
- Matthias Fischer

Grants: Childhood Cancer Fund, Emmanuel van der Schueren foundation, UGent-GOA, FWO, IUAP, IWT

Center for Medical Genetics

acknowledgements

Center for Medical Genetics, Ghent, Belgium

- Frank Speleman
- Jo Vandesompele
- Nadine Van Roy
- Katleen De Preter
- Jasmien Hoebeeck
- Filip Pattyn
- Tom Van Maerken
- Joëlle Vermeulen
- Nurten Yigit
- Els De Smet
- Liesbeth Vercruysse
- Anne De Paepe

BIO-RAD

SIOPEN R NET

Collaborators

- Geneviève Laureys
- Gianpaolo Tonini
- Olivier Delattre
- Jean Bénard
- Valérie Combaret
- Raymond Stallings
- Angelika Eggert
- Akira Nakagawara
- Matthias Fischer

Grants: Childhood Cancer Fund, Emmanuel van der Schueren foundation, UGent-GOA, FWO, IUAP, IWT

Center for Medical Genetics