Quality control for quantitative PCR based on amplification compatibility test

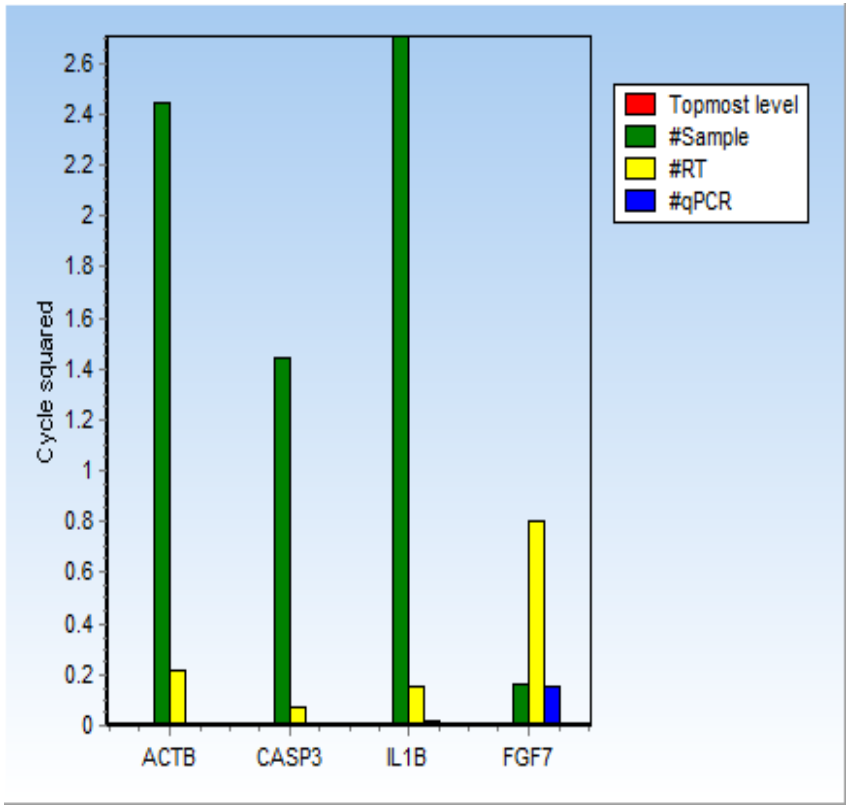
Ales Tichopad

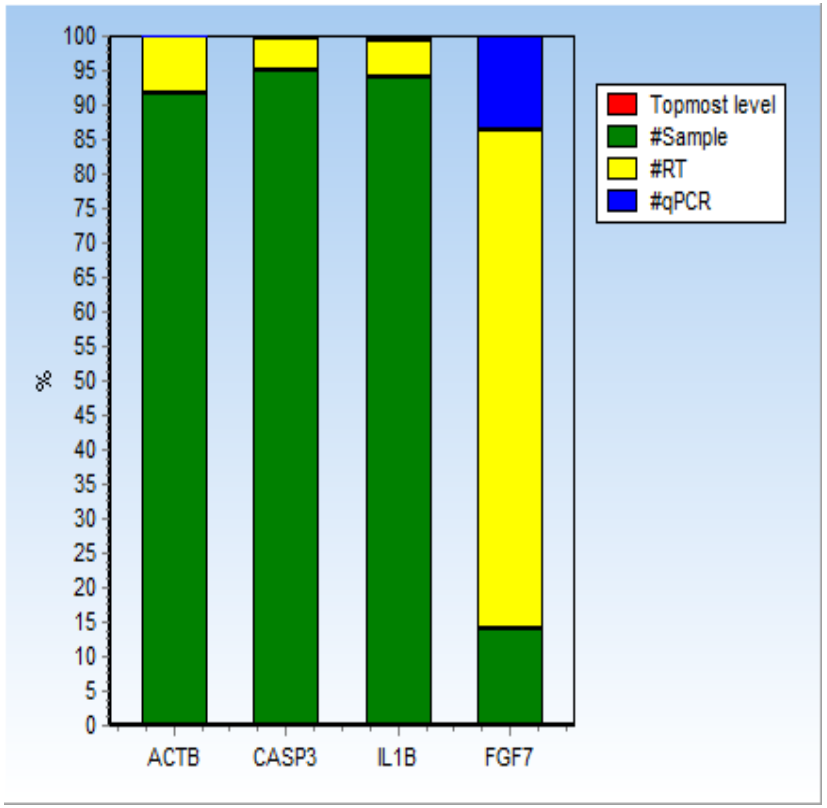
Methods 50 (2010) 308–312

Error stratification throughout preanalytics

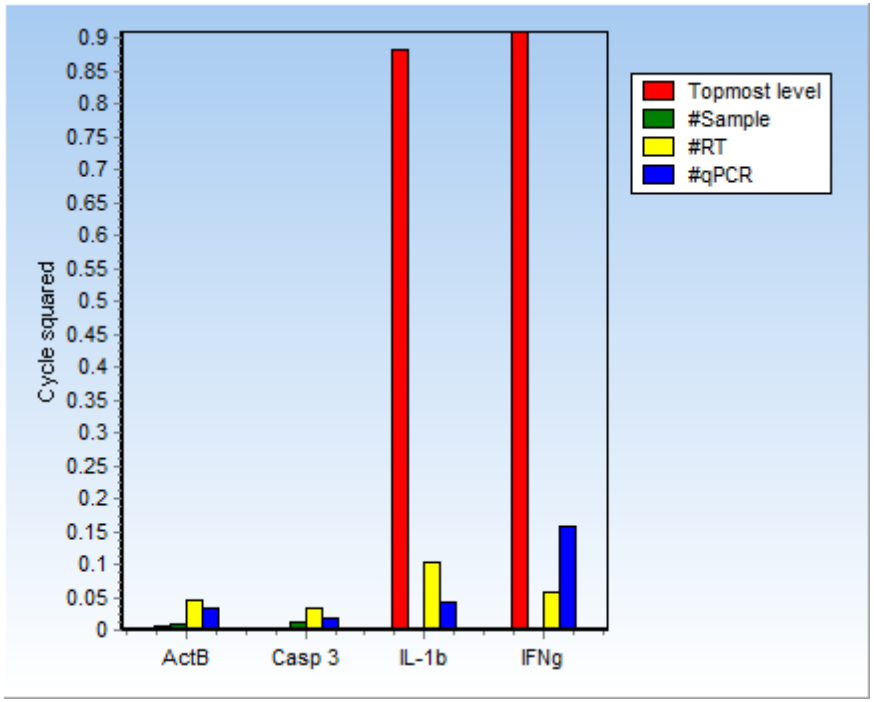
tissue			liver				blood				cell culture				single cell
gene			ACTB	IL1B	CASP3	FGF7	ACTB	IL1B	CASP3	IFNG	ACTB	H3	IL8	BCL2	18S
Mean Cq			20.41	26.76	27.25	31.52	16.05	17.6	24.71	32.2	15.87	20.1	23.4	28.5	29.95
SD (cycles)	I.S. var.	Subject	0.00	0.00	0.00	0.00	0.07	0.94	0.00	0.95	0.00	0.00	0.00	0.00	0.00
	Processing noise	Sampling	1.56	1.64	1.20	0.40	0.10	0.00	0.11	0.00	0.37	0.20	0.29	0.20	1.90
		RT	0.46	0.39	0.27	0.90	0.21	0.32	0.18	0.24	0.35	0.35	0.31	0.21	0.30
		qPCR	0.07	0.12	0.08	0.39	0.18	0.20	0.13	0.40	0.21	0.10	0.09	0.16	0.51
	Total noise		1.63	1.70	1.23	1.06	0.31	1.01	0.25	1.06	0.55	0.42	0.44	0.33	1.99
				~1.5				~0.66				~0.44			~2

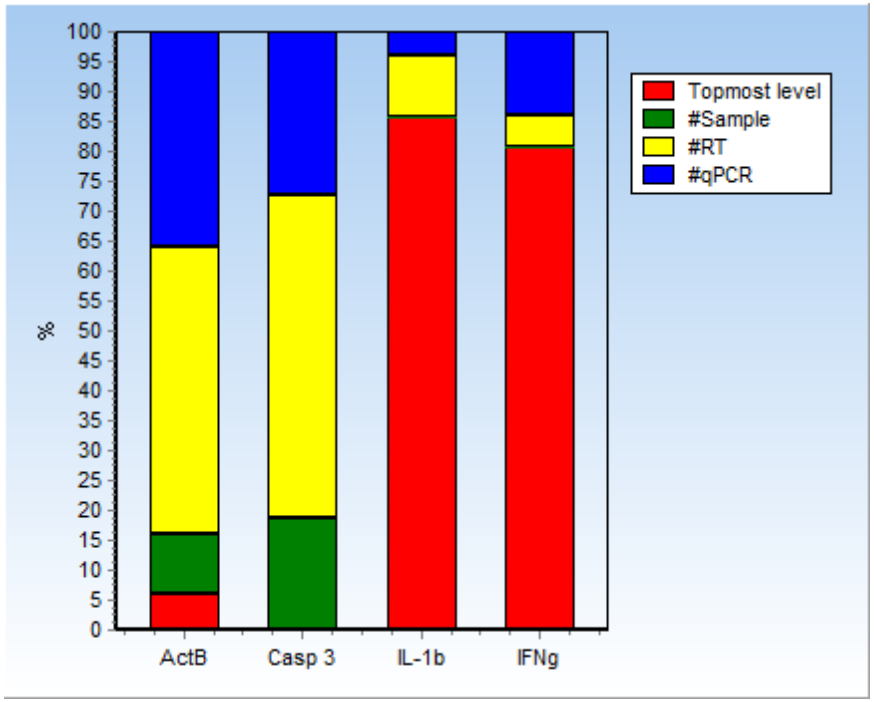
Liver tissue



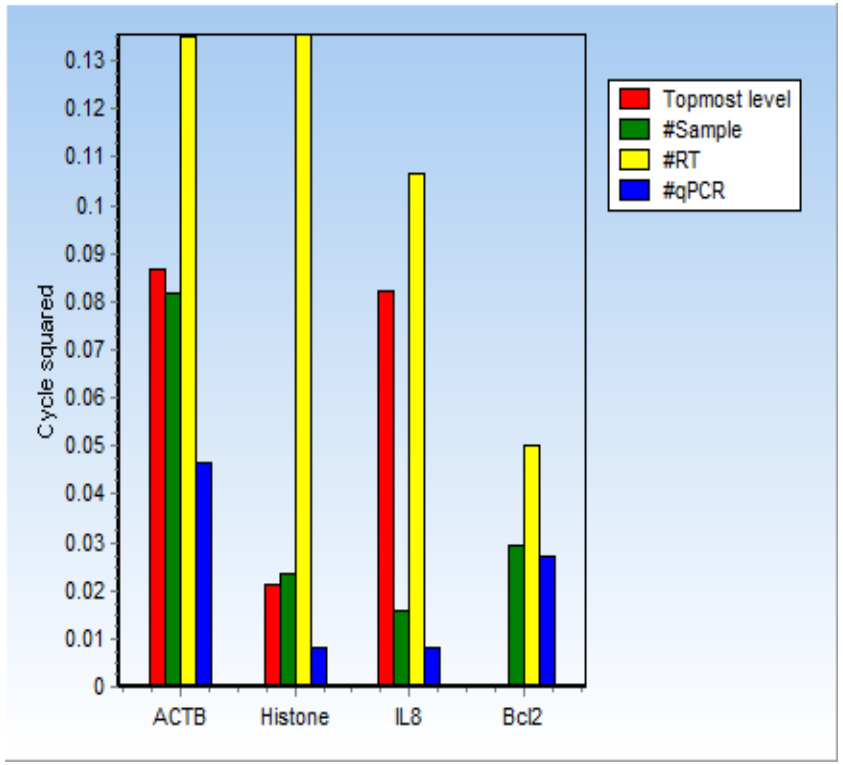


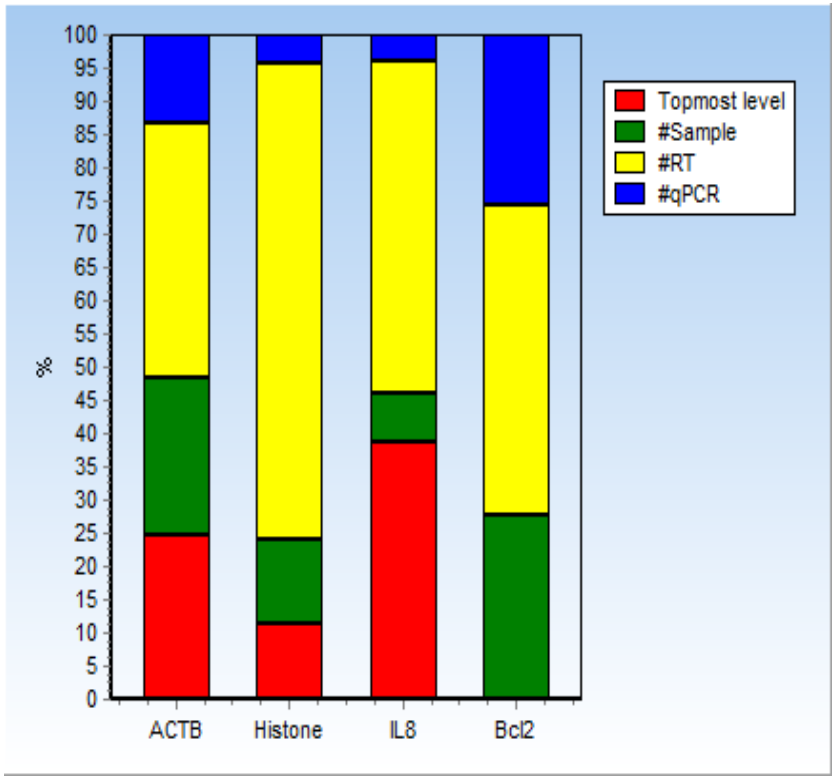
Blood samples



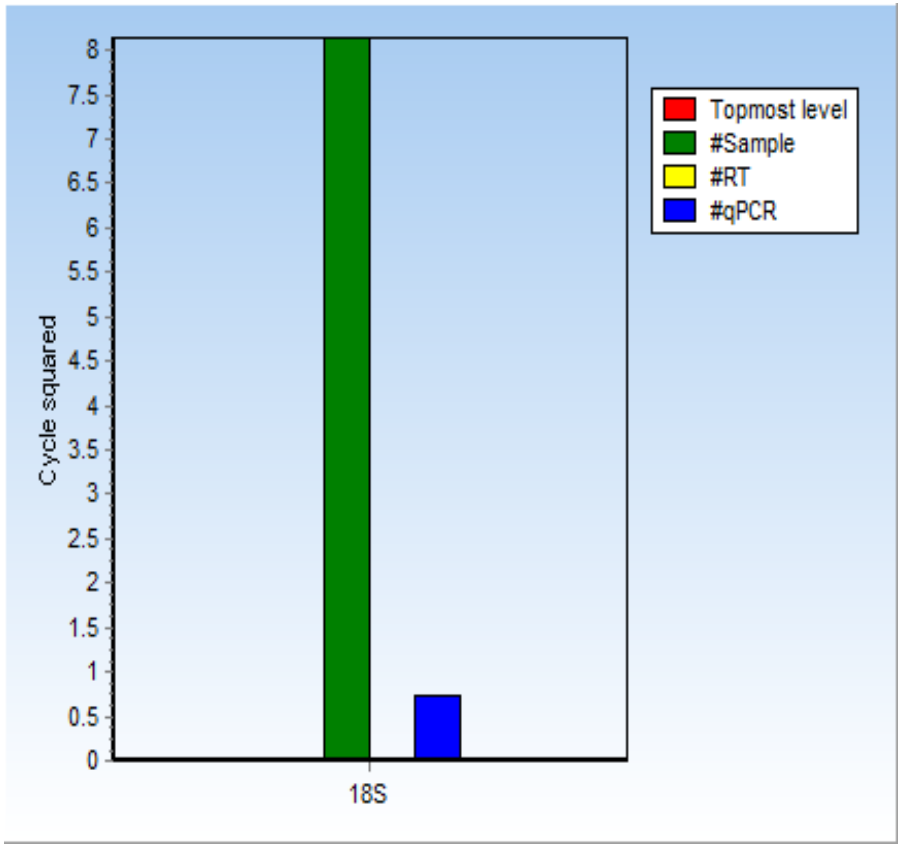


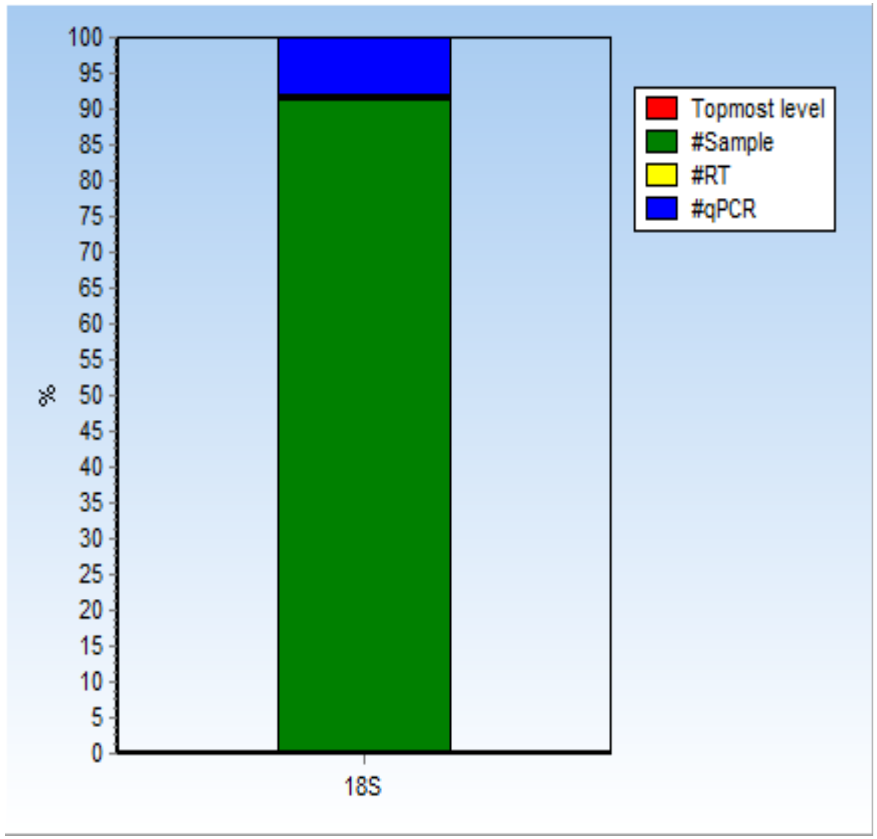
Cell culture



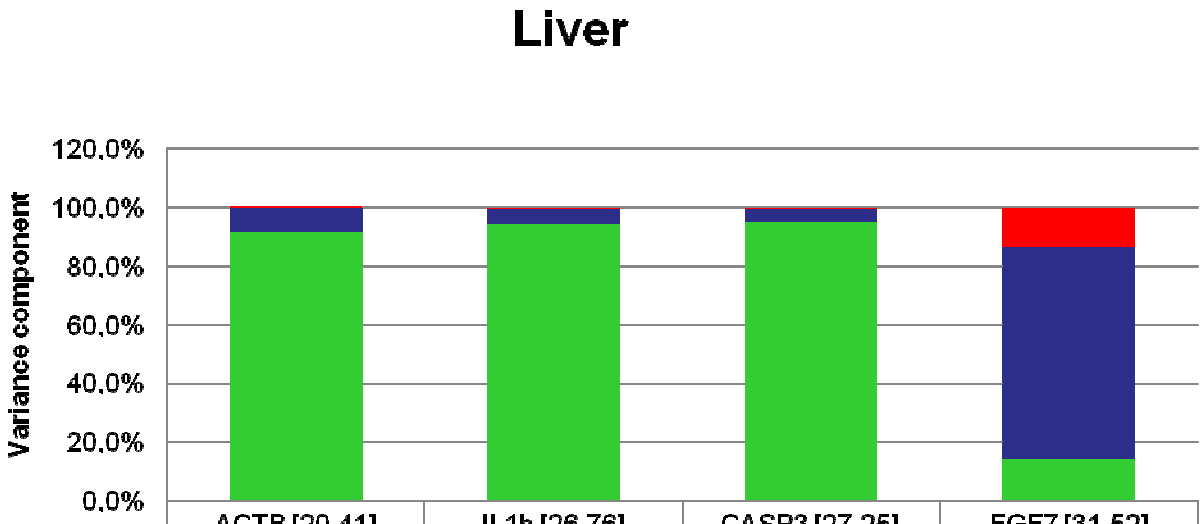


Single cell


Cycle squared
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
18S
Topmost level
#Sample
#RT
#qPCR

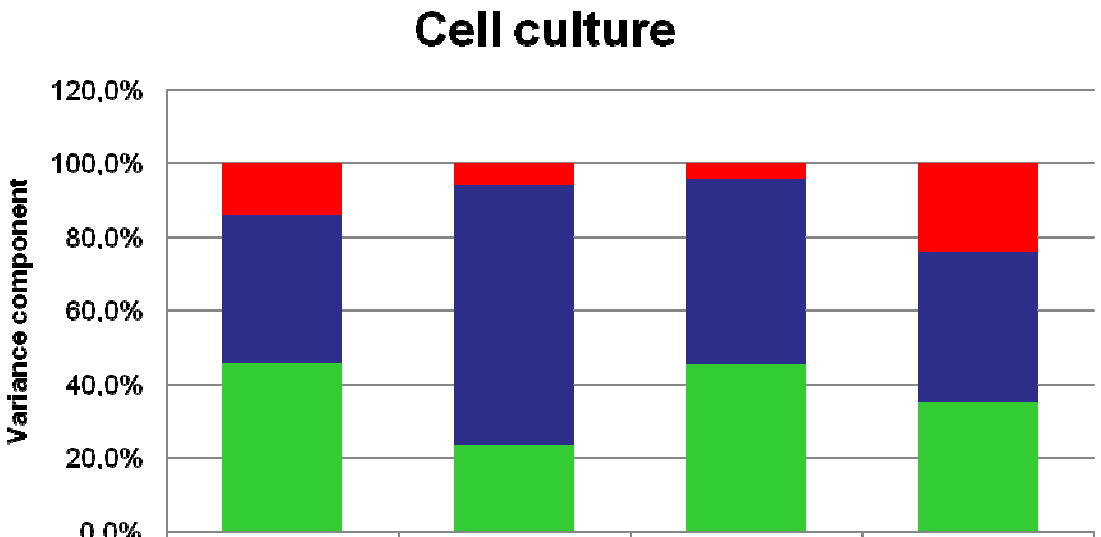

%
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
18S
Topmost level
#Sample
#RT
#qPCR

Noise contribution by various sample processing steps



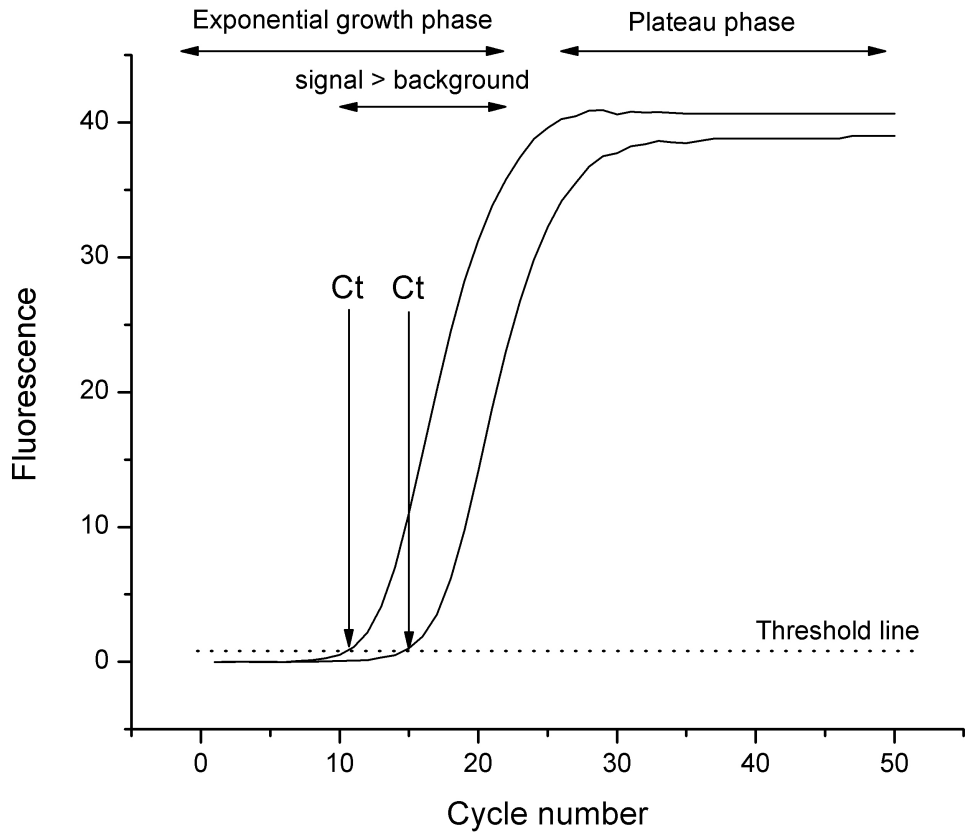
	ACTB [20.41]	IL1b [26.76]	CASP3 [27.25]	FGF7 [31.52]
qPCR	0.2%	0.5%	0.4%	13.8%
RT	8.0%	5.3%	4.7%	72.0%
Sampling	91.8%	94.1%	94.9%	14.2%

Noise contribution by various sample processing steps



	ACTB [15.87]	H3 [20.11]	IL8 [23.41]	BCL2 [28.48]
qPCR	13.9%	6.0%	4.3%	24.2%
RT	40.6%	70.7%	50.3%	40.5%
Sampling	45.5%	23.3%	45.4%	35.2%

Real-time PCR response curve - Cq values



If a biological sample is inhibited, technical replicates WILL NOT protect us from the error

Any difference in the Cq between different samples may be due to a biological effect or due to INHIBITION!!!

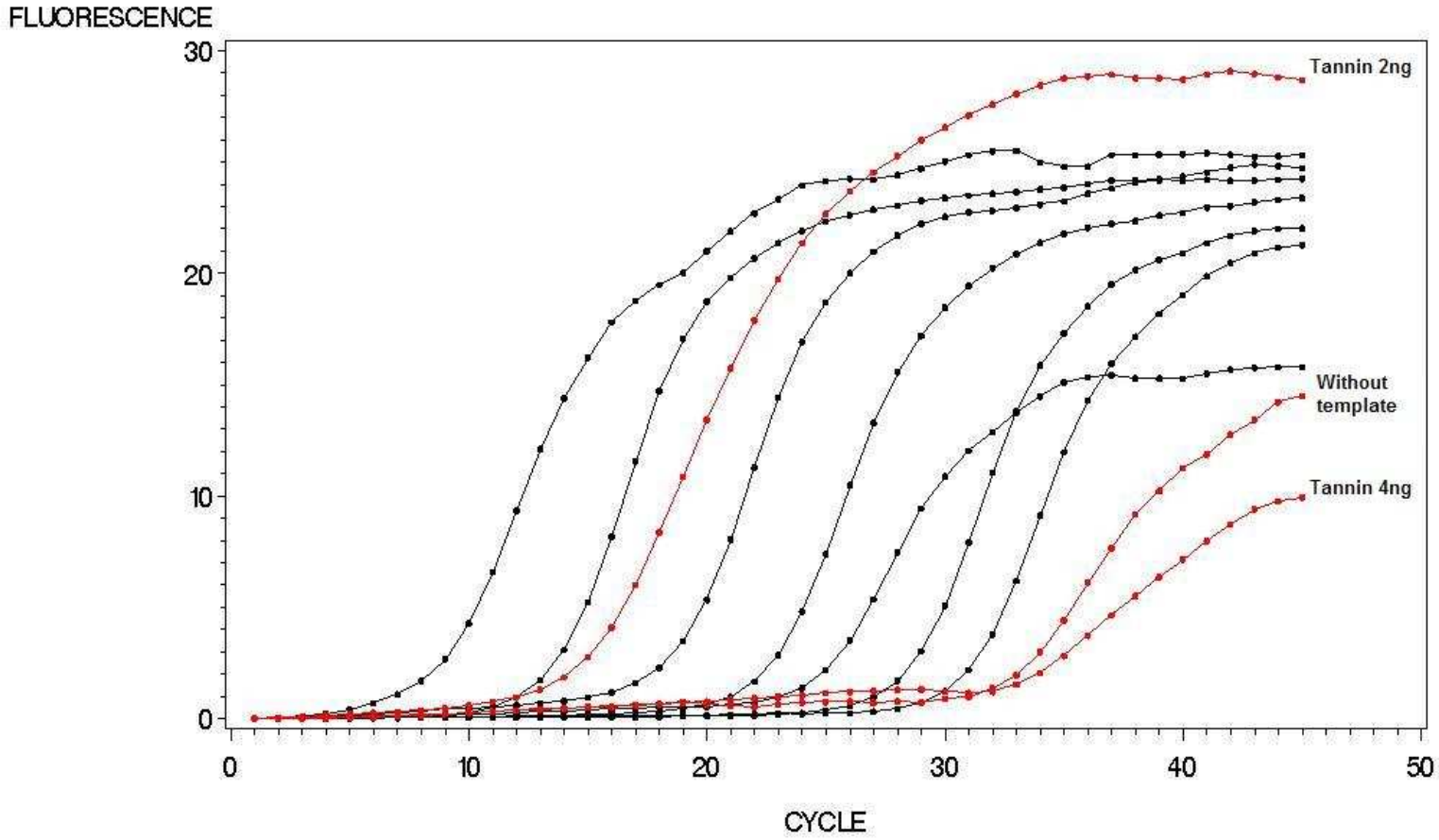
Therefore, Cq is not suitable as a quality control measure

Kinetics of the reactions is much more reliable

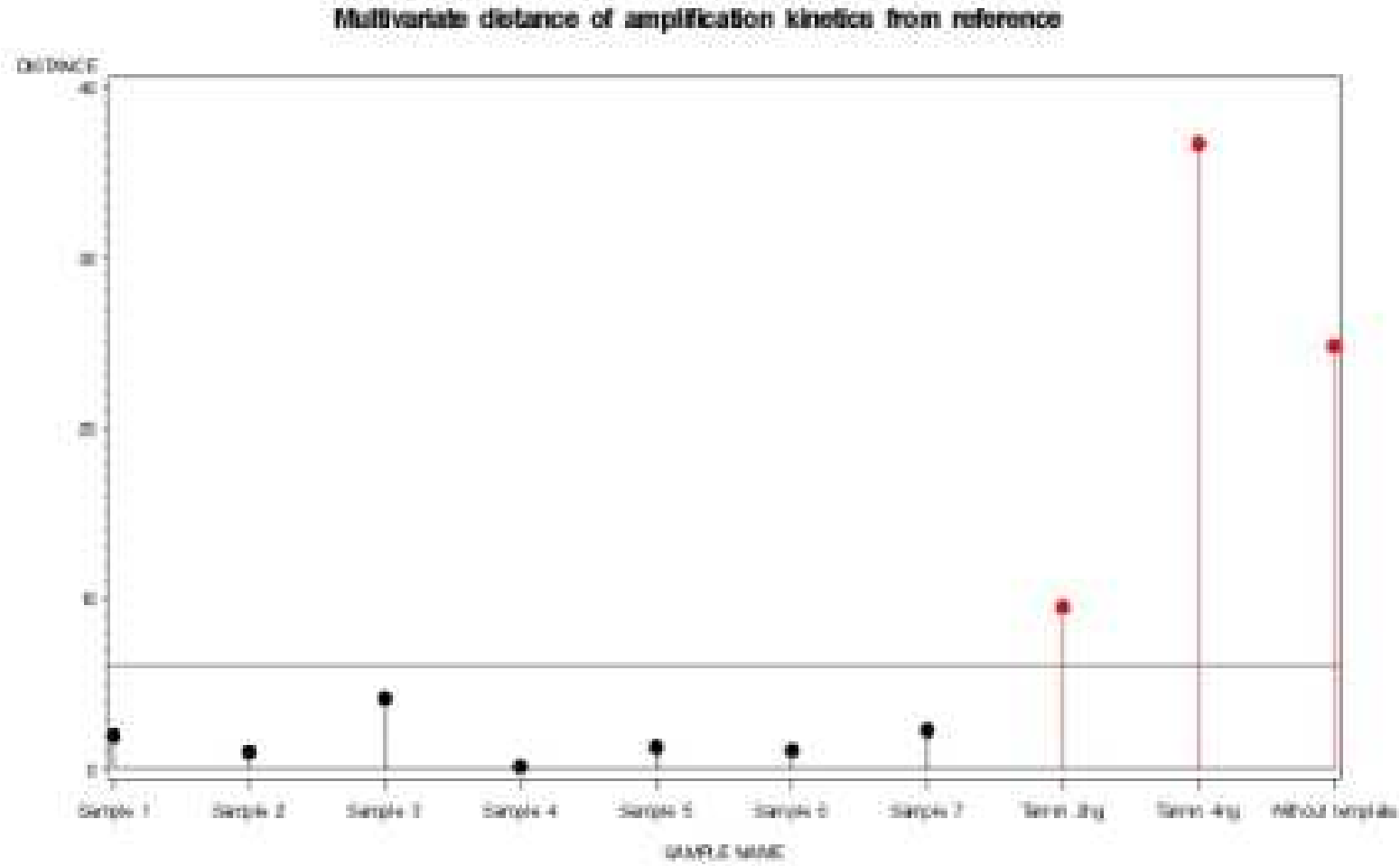
Because kinetics must be compatible among samples, regardless of the initial DNA concentration

Amplification kinetics

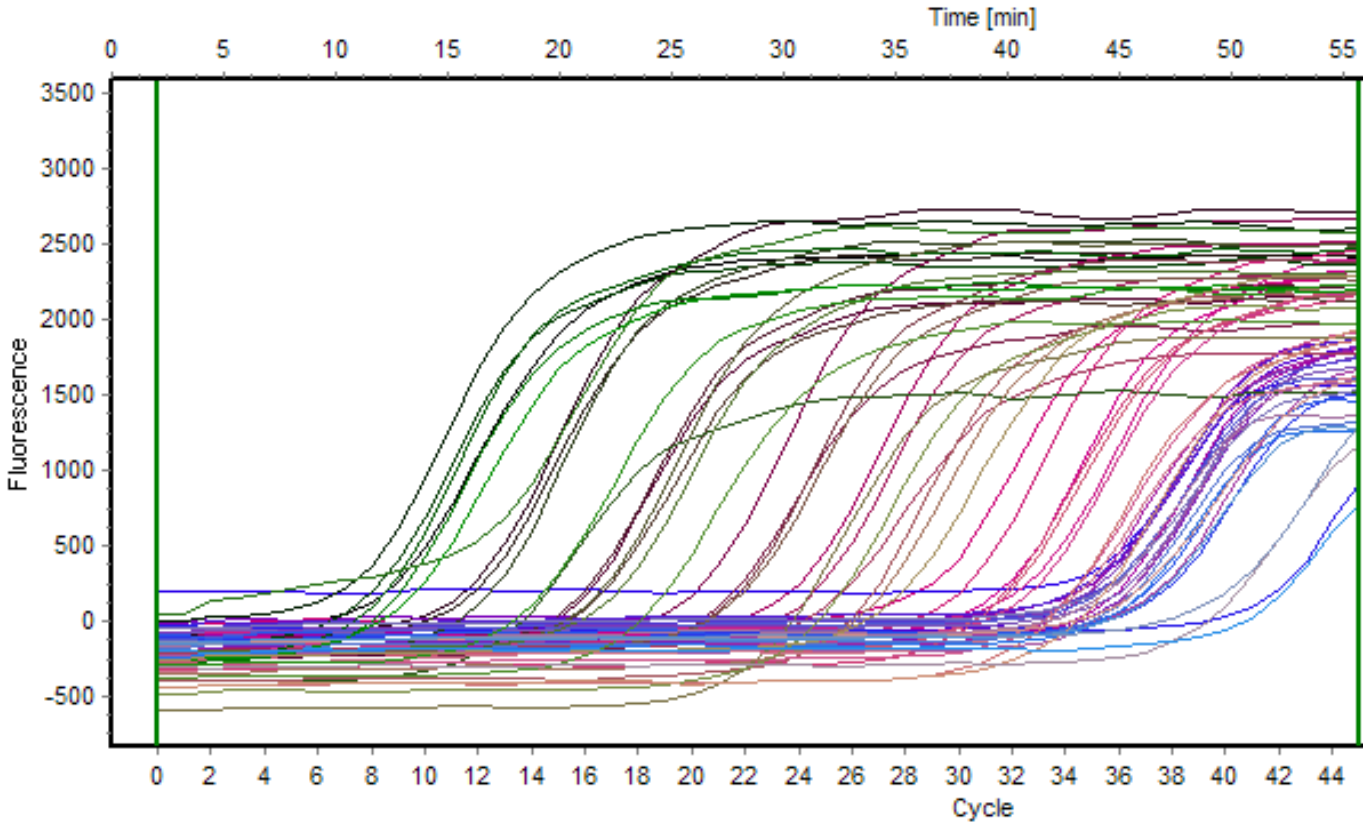
example of incompatible samples



The multivariate distance from the centre of the reference set



Visual check may sometimes be impossible



The kinetics must be digitalised and the obtained parameters compared statistically.

Discrepancy between methods for amplification efficiency estimation from single sample

Comparison with E obtained from standard curve

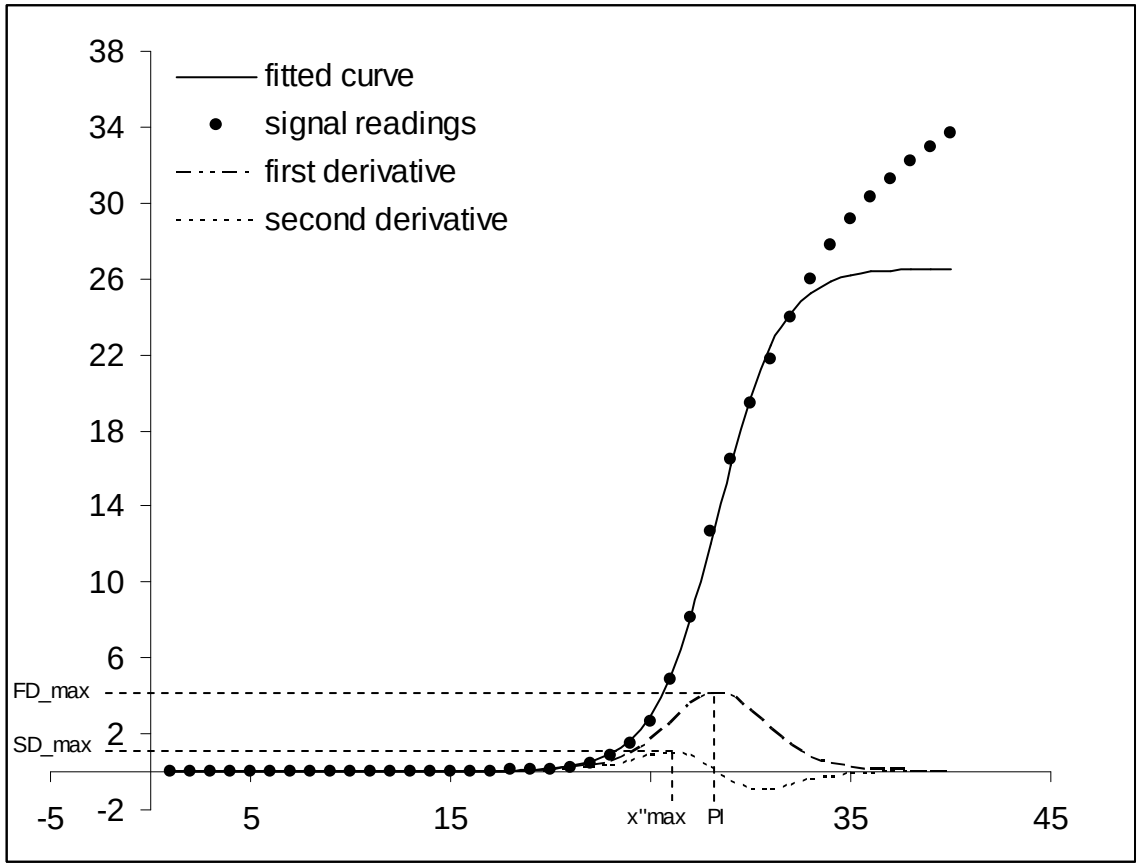
Tichopad et al.		**Ramakers et al.**		**Peirson et al.**		**Wilhelm et al.**		**Liu & Saint**	
ΔE	SD	ΔE	SD	ΔE	SD	ΔE	SD	ΔE	SD
0.44	0.076	0.26	0.102	0.24	0.118	0.31	0.076	0.33	0.071

$E_{std} = 10^{-1/slope} - 1$

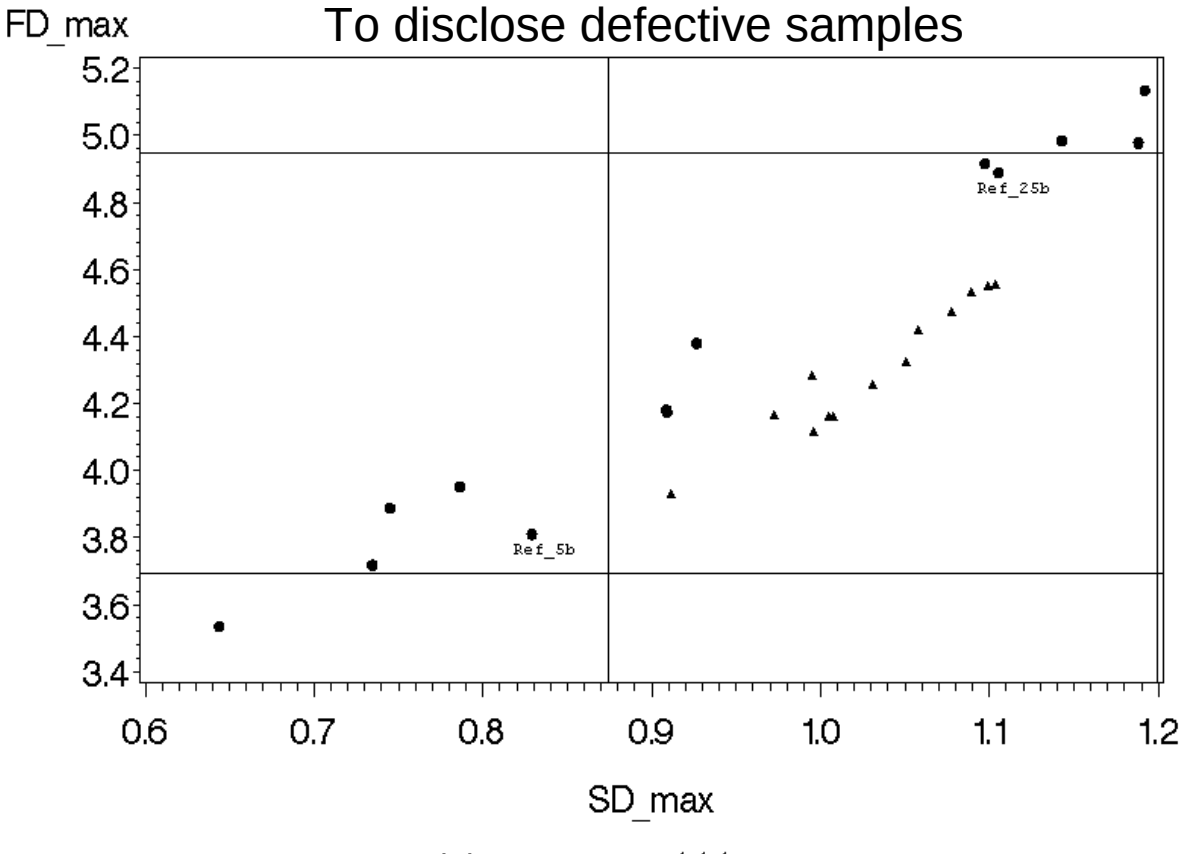
$\Delta E = E_{std} - \bar{E}_{individual}$

Kinetics parameters for amplification outlier detection

Maximum of the first [FD_max] and second derivative [SD_max] are used to identify amplification kinetics in 2D space



Multivariate outlier detection



Test samples ● vs. reference set ▲. Flagged points were excluded from the reference set. The inner lines define traditional univariate boundaries for outliers obtained as upper quartile plus 1.5 times interquartile range and lower quartile minus 1.5 times interquartile range.

Validation experiments

EXPERIMENT 1: One assay varying inhibition strength

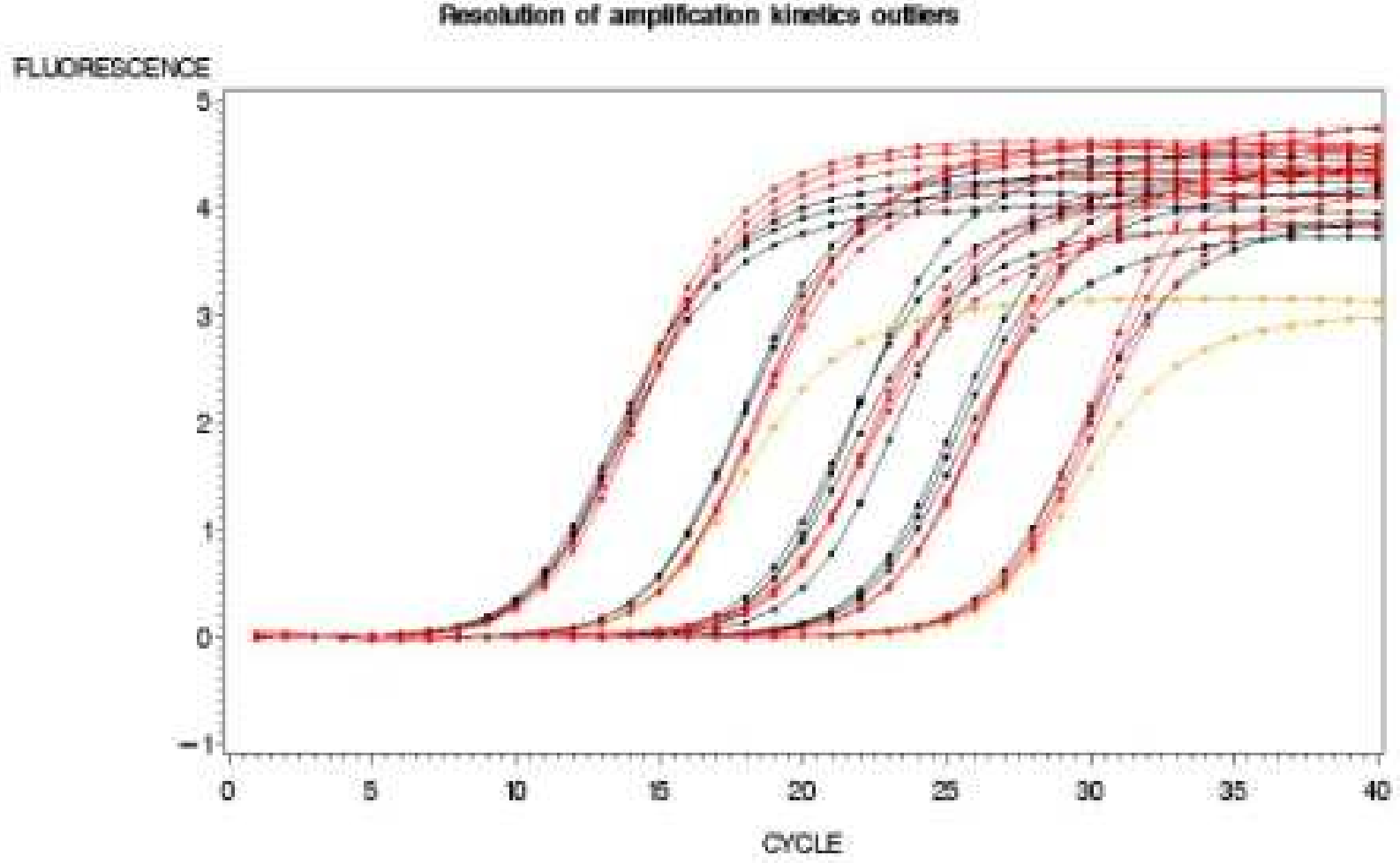
3 x 5 serial dilutions were produced as non-inhibited reference set (n=15) and inhibited sets (each n=15) with 1%, 2%, 4%, 8%, and 16% of primer competamers added to regular primer concentration. **Primer competimers** were used to introduce the inhibition.

EXPERIMENT 2: Three assays constant inhibition strength

Three assays as standard curves were performed. Each standard curve consisted of 5-fold dilutions (1-, 5-, 25-, 125-, and 625-fold) in triplicates (total 15 reactions). Two standard curves were produced from the same cDNA stock solution, one without inhibitor and one with 2.0 ng **tannic acid** added per 15 µl reaction mix.

EXPERIMENT 1

Effect of the inhibition by **primer competimers** on the Cq value



EXPERIMENT 1

Effect of the inhibition by **primer competimers** on the Cq value

	Differences from reference [ΔCq] for various inhibition strengths				
DNA conc.	**1%**	**2%**	**4%**	**8%**	**16%**
x*10000 (n=3)	0.05	-0.29	-0.223	-0.257	-0.363
x*1000 (n=3)	0.233	0.327	0.143	-0.077	-0.307
x*100 (n=3)	0.213	0.017	-0.073	-0.303	-0.47
x*10 (n=3)	-0.23	-0.173	-0.457	-0.753	-0.737
x (n=3)	NA	NA	NA	NA	NA
p of t-test (H0: Dif<0)	**0.58**	**0.84**	**0.31**	**0.09**	**0.02**

NA – too large scatter of the reference to reliably calculate the ΔCq.

EXPERIMENT 1

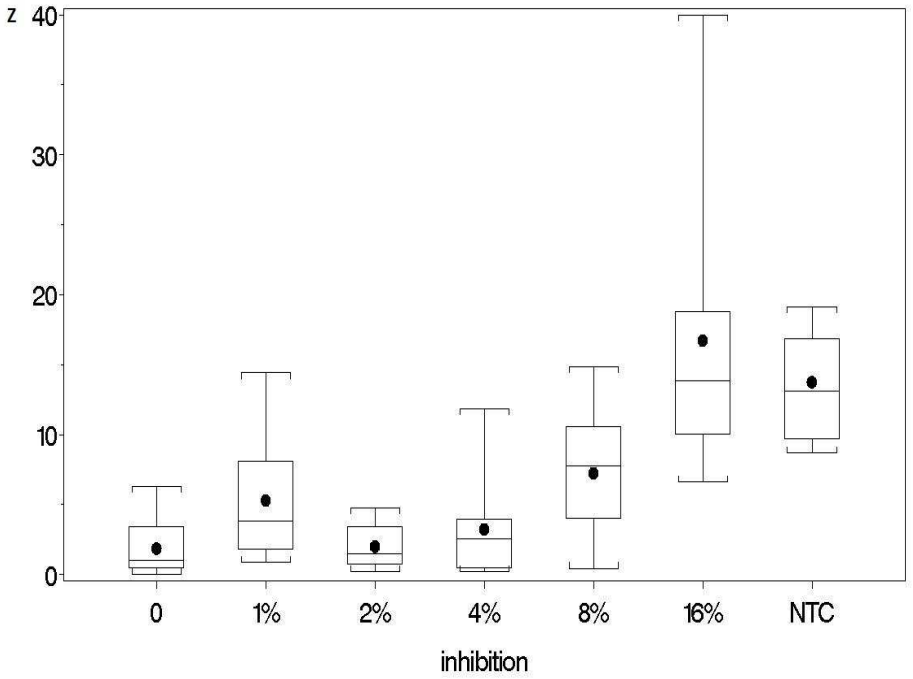
Retrieval of samples inhibited by **primer competimers** by the multivariate vs. univariate test

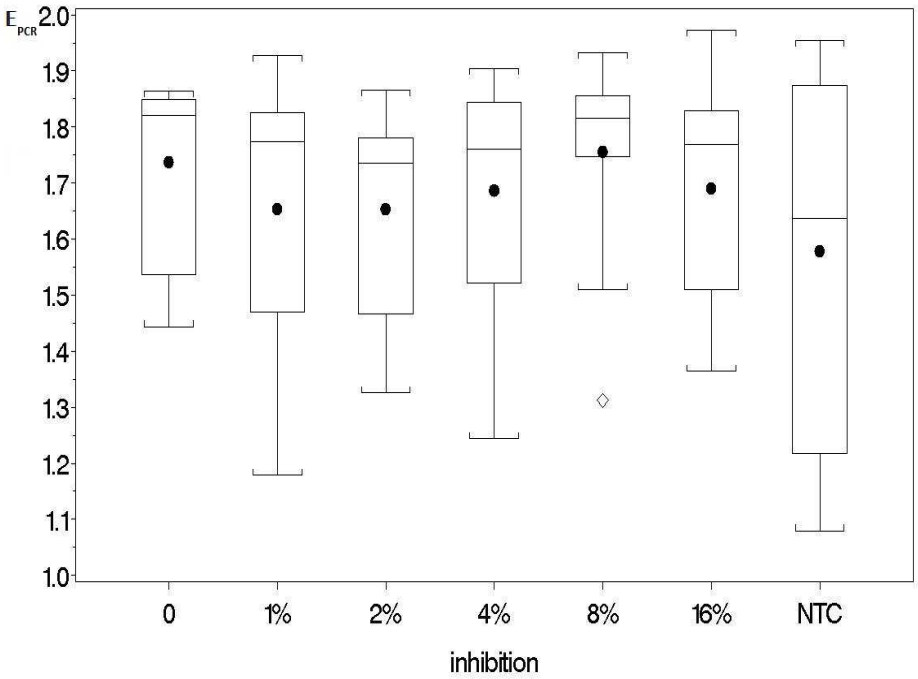
Multivariate (Z)	**1%**	**2%**	**4%**	**8%**	**16%**	**NTC**
N/Total	6/15	2/15	2/15	11/15	15/15	6/6
Retrieval [%]	40%	13%	13%	73%	**100%**	**100%**
Univariate (E)						
N/Total	4/15	5/15	2/15	1/15	2/15	2/6
Retrieval [%] *Bar et al (2003)*	27%	33%	13%	7%	13%	13%

Bar T, Stahlberg A, Muszta A, Kubista M. (2003). Nucleic Acids Res. 31, e105

EXPERIMENT 1

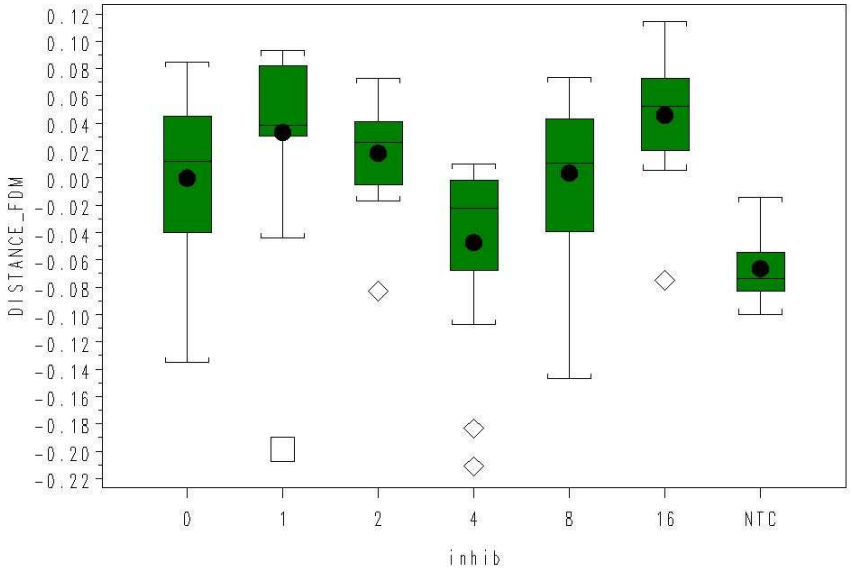
Retrieval of samples inhibited by **primer competimers** by the multivariate vs. univariate test



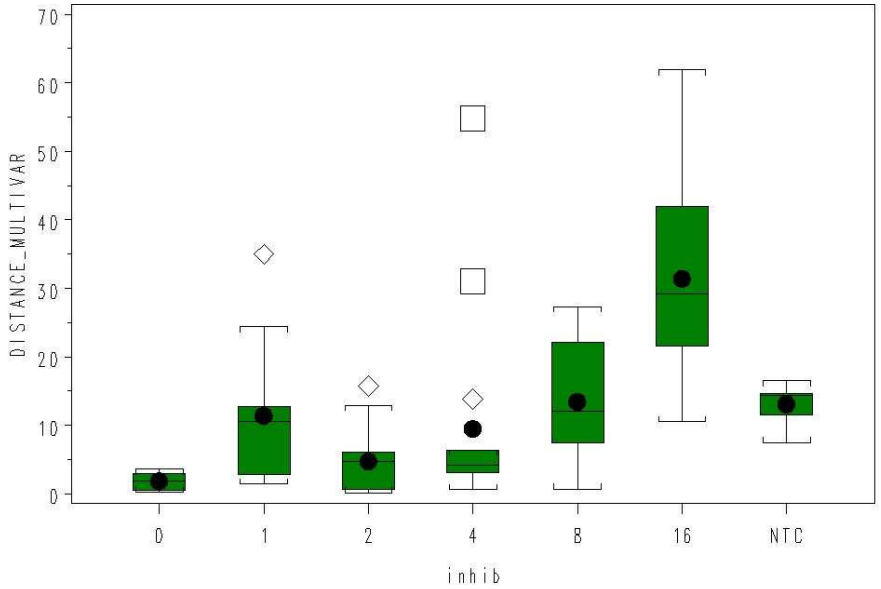


Bar T, Stahlberg A, Muszta A, Kubista M. (2003). Nucleic Acids Res. 31, e105

One parameter vs. two parameters in detecting kinetics outliers



Maximum of the first derivative (FDM)



Multivariate Mahalanobis DISTANCE calculated from the maximum of the first derivative (FDM) and the maximum of the second derivative (SDM)

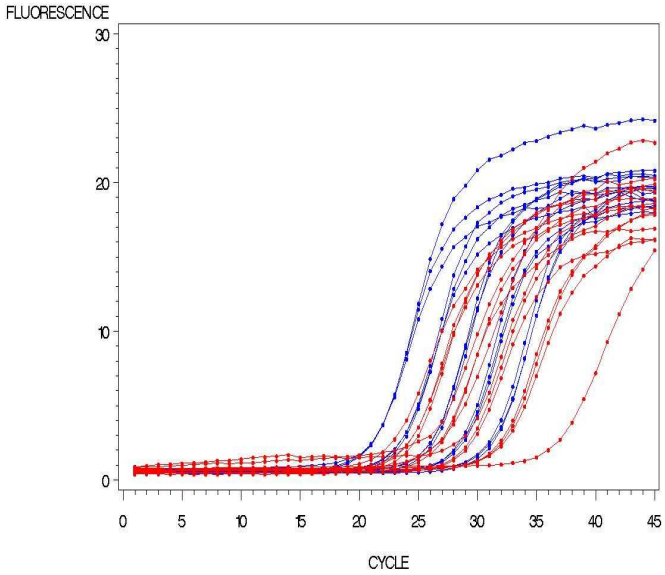
EXPERIMENT 2

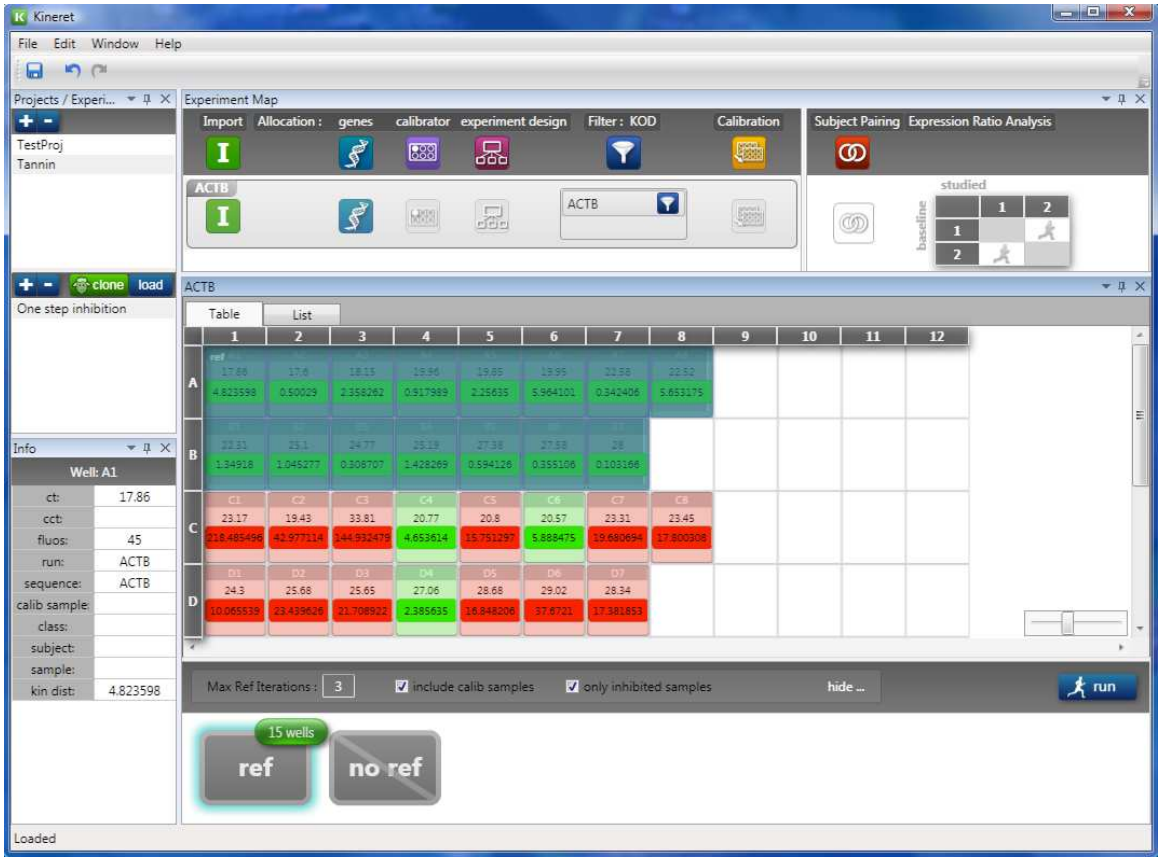
Retrieval of samples inhibited by **tannic acid** by the multivariate vs. univariate test

Multivariate (Z)	**ACTB**	**H3**	**IGF**
N/Total	12/15	15/15	10/15
Retrieval [%]	80%	100%	67%
Univariate (E)			
N/Total	1/15	5/15	2/15
Retrieval [%]	7%	33%	13%

Bar T, Stahlberg A, Muszta A, Kubista M. (2003). Nucleic Acids Res. 31, e105

Multivariate KOD using Kineret software





www.labonnet.com

Calculation

Assuming that n variables (X_1 ,...,X_n) are independent, approximately normally distributed Z-score can be calculated:

$$Z = X_1^2 + ... + X_n^2$$

where Z is approximately χ2 (chi-square) distributed with n degrees of freedom

***Problem**: SD_max and FD_max are correlated*

*FD_max=SD_max * b + a + τ*

Where a and b are the linear coefficients and τ is the residual error from the predicted regression line.

$\tau = FD\_max - \widehat{FD\_max}$

The final Z-score: $Z = SD\_\max^2 + ... + \tau^2$

Critical values for the 95% and 99% percentily of the χ2 distribution for 2 df are 5.991 and 9.210, respectively.

Use of Kineret within SPIDIA

Objective: amplification compatibility as an additional RNA quality

Kineret Version 1.0.5 was used

The reference set: samples REFpool RNA + REF RNA from all the participating laboratories.

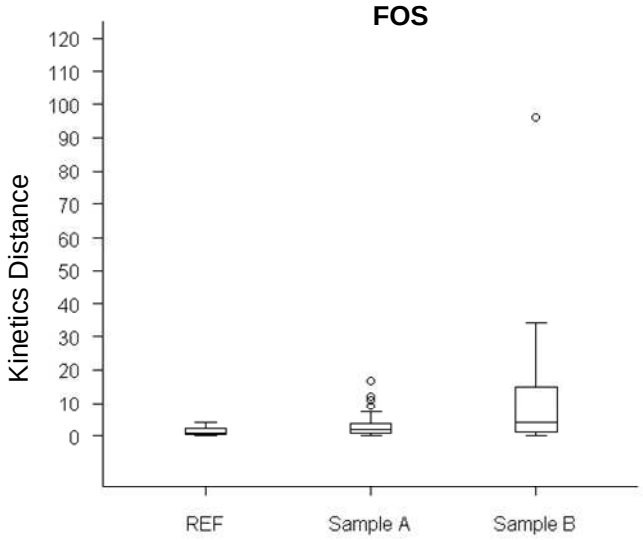
The Z score (called Kinetics Distance – KD) of the three qPCR technical replicates of each sample were averaged.

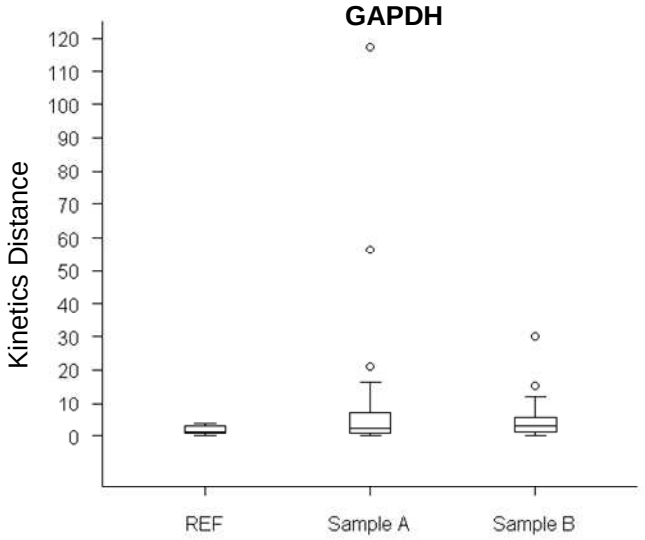
The test set: RNA from samples A and B were compared with the reference set

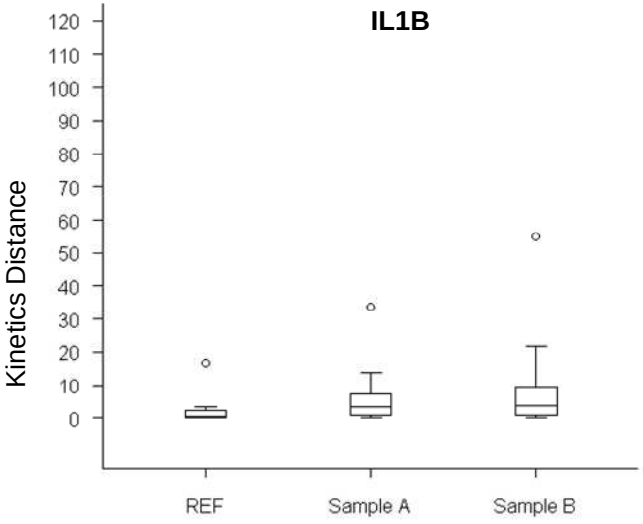
KDs of each group are presented in box-whisker plots

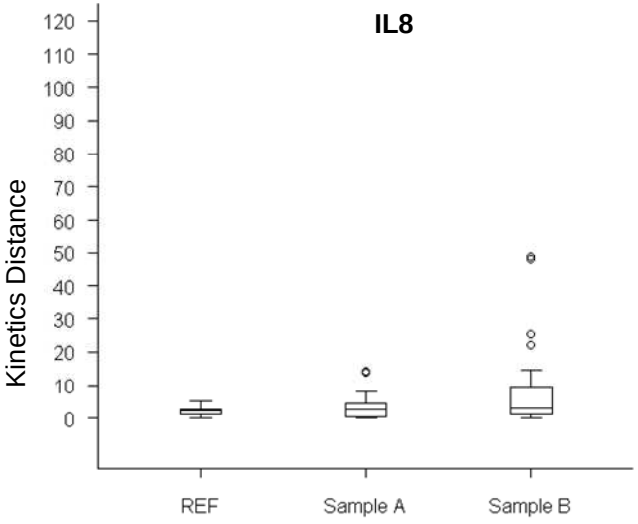
KD distribution as calculated by the Kineret software of the four gene transcripts in each sample

Gene	Sample	N	min	median	max	IQR
FOS	Sample A	10	0.0	2.0	16.9	3.2
	Sample B	9	0.1	4.3	96.1	14.5
	REF	8	0.0	0.8	4.1	1.8
GAPDH	Sample A	9	0.0	2.4	117.5	6.4
	Sample B	9	0.3	2.9	30.0	4.4
	REF	8	0.1	1.2	3.9	2.2
IL1B	Sample A	10	0.2	3.2	33.8	6.7
	Sample B	9	0.1	3.7	55.1	8.6
	REF	8	0.2	0.7	16.9	2.1
IL8	Sample A	10	0.1	2.6	14.3	3.8
	Sample B	9	0.1	3.2	48.8	8.2
	REF	8	0.2	2.1	5.1	1.7


FOS
Kinetics Distance
120
110
100
90
80
70
60
50
40
30
20
10
0
REF
Sample A
Sample B


GAPDH
Kinetics Distance
120
110
100
90
80
70
60
50
40
30
20
10
0
REF
Sample A
Sample B

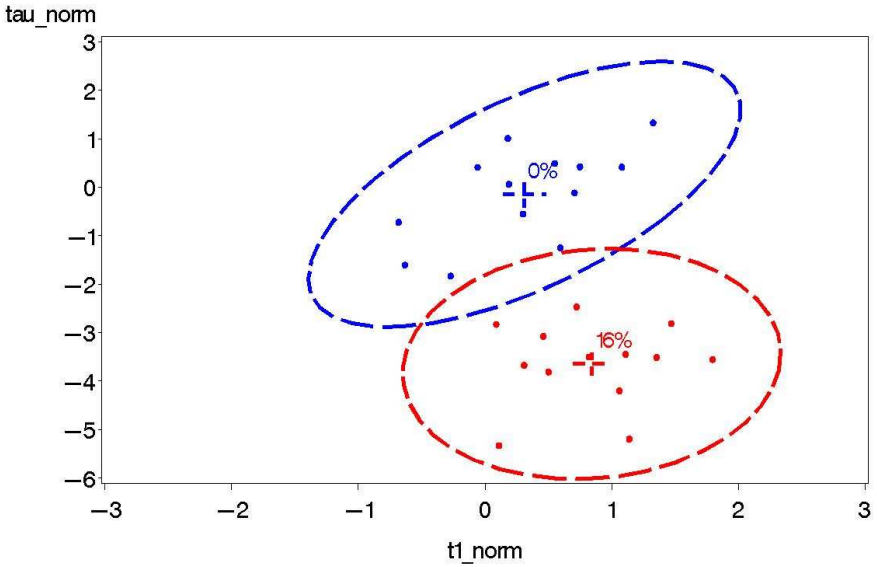

IL1B
Kinetics Distance
120
110
100
90
80
70
60
50
40
30
20
10
0
REF
Sample A
Sample B


IL8
Kinetics Distance
120
110
100
90
80
70
60
50
40
30
20
10
0
REF
Sample A
Sample B

Conclusion

Generally, multivariate methods perform better in separating defective reactions than univariate methods.

Several methods can be used; e.g. the Mahalanobis distance is the uncorrected sum of squares of the principal component scores calculated from the center of the reference data set.
Also other multivariate approaches may be employed such as the Kohonen self-organising networks, K-means or support vector machines.



Acknowledgements

Financial grants were obtained from:

•European Community seventh framework project SPIDIA (www.spidia.eu) Grant Agreement **N°: 222916**
•European Community sixth framework project SmartHEALTH (www.smarthealthip.com) Grant Agreement **N°: 016817**
•National R&D incubator program of Ministry of Industry and Trade of the State of Israel

Original paper available in Methods 50 (2010) 308–312