Kinetic Outlier Detection (KOD) in real-time PCR

Tzachi Bar1, Anders Muszta2 , Jose Manuel Andrade-Garda3 and Mikael Kubista1,4
1Department of Chemistry and Bioscience Chalmers University of Technology Medicinargatan 7B 405 30, Gothenburg, SWEDEN
2Department of Mathematical statistics, Chalmers University of Technology Eklandagatan 86, 412 96, Gothenburg, SWEDEN
3Department of Analytical Chemistry, University of A Coruna, A Zapateira s/n E-15071 A Coruna, Spain
4TATAA Biocenter, Medicinargatan 7B 405 30, Gothenburg, SWEDEN
tzachi.bar@molbiotech.chalmers.se

1 Introduction

The exponential nature of PCR makes it sensitive to differences in the efficiency of the compared reactions. Kinetic Outlier Detection (KOD) is a statistical method to identify test samples with high probability for dissimilar efficiency. Table 1 summarizes the requirements for similarity in absolute and relative quantification.

	Absolute	Relative
Characteristics	Each test sample stands alone (e.g., quantification of viral load) and quantified relatively to the standard curve samples.	Test samples are quantified one relatively to the other (e.g., comparison of gene expression analysis after treatment).
The efficiency of a test sample should be similar to the mean efficiency of the	Standard curve samples Also referred to as **Training set**	Other test samples
Criterion for outlier detection	Variance of efficiency of high quality samples (standard curve) samples or a nominal value from a previous study.	

Materials, methods and experimental design as in the poster "Using the variance of efficiency for QA in real-time PCR" (P50)

2

Mathematical model

E_{test} → Efficiency of a test sample

$\overline{E}_{train}$ → Estimated mean efficiency of a training set
S^2 →Estimated variance of efficiency of high quality samples
σ^2 →Nominal value for variance from previous study
$c_p = \left|\overline{E}_{train} - E_{test}\right|$ → Critical value for decision on outlier with probability p
Φ→ Cumulative normal distribution function
t→ Cumulative t-distribution function
$p = P\left(\left|\overline{E}_{train} - E_{test}\right| > c_p\right)$ for some small p, E_{test} is considered an outlier.

$$c_p = \sqrt{\sigma^2\left(1+\frac{1}{n}\right)} * \Phi^{-1}\left(1-\frac{p}{2}\right) \quad [1]$$ σ^2 is known (Nominal KOD)

$$c_p = \sqrt{S^2\left(1+\frac{1}{n}\right)} * t_{n-1}^{-1}\left(1-\frac{p}{2}\right) \quad [2]$$ If σ^2 is unknown (Comparative KOD)

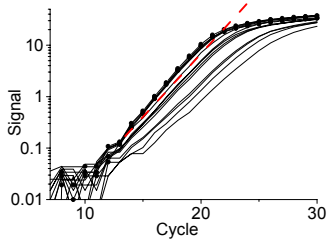
Equation [3] is the confidence interval for the error in quantification associated with dissimilar efficiencies. Here N_0^1 and N_0^2 are the initial copy number if calculated by the test sample efficiency and the mean efficiency of the training set, respectively.

$$1-p = P\left(\left(1-\frac{c_p}{1+\overline{E}_{train}+c_p}\right)^{CT} \le \frac{N_0^1}{N_0^2} \le \left(1+\frac{c_p}{1+\overline{E}_{train}-c_p}\right)^{CT}\right) \quad [3]$$

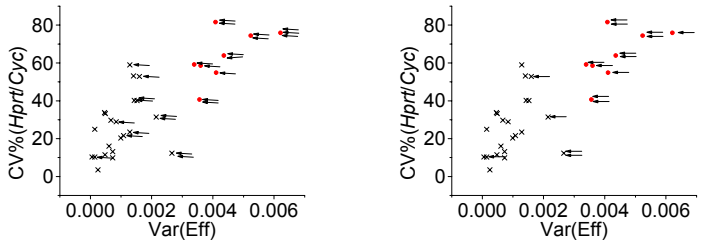
Results

3

KOD in absolute quantification. Equal initial numbers of DNA molecules were amplified with different efficiencies. The red line represents a theoretical sample with the same number of molecules and efficiency equal to the critical value. Samples to the right of the red line are outliers.

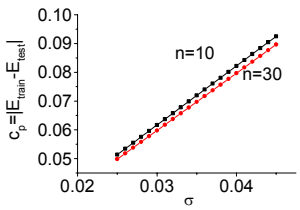


4

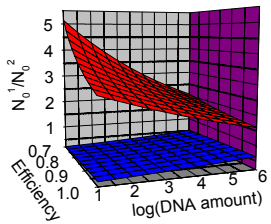


Outlier samples (arrows) detected by Comparative KOD (left) and Nominal KOD (right) in high (x) or low (•) quality replicate sets on CV-Var(Eff) plot (CV = 100*SD/Average, see poster P50 for details). The proportion of outliers in the low quality sets was significantly higher ($p<0.01$) comparing to high quality sets 36% vs. 6% for Comparative KOD, and 36% vs. 2% for Nominal KOD.

5



Minimal difference in efficiencies, $\left|\overline{E}_{train} - E_{test}\right|$ KOD detects. Calculated by Equation [1], with p=0.05.



The distance between the blue and red sheets is the confidence interval for the error in quantification associated with dissimilar efficiencies (Equation [3] with σ=0.035, p=0.05, n=30).

6

Conclusion

KOD can be used to draw attention of real-time PCR user to suspected samples.