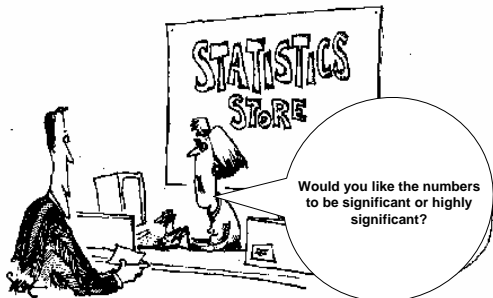


# Analysis of Quantitative Real Time PCR

Gregory D. Ayers,  
Christian Kis,  
Ryan Splittgerber

Division of Cancer Biostatistics  
Department of Surgical Oncology  
Department of Cancer Biology  
Vanderbilt University Medical Center

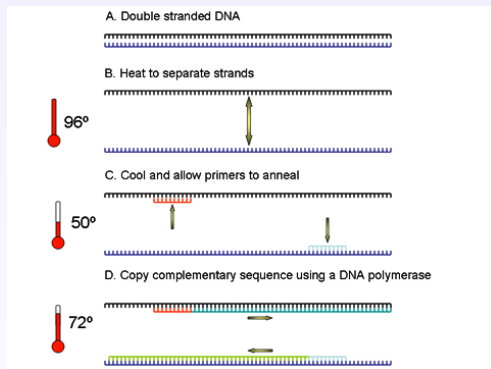
Cancer Biostatistics Workshop  
August 21st, 2009



By Signe Wilkinson, Philadelphia Daily News, Cartoonists & Writers Syndicate

# Introduction

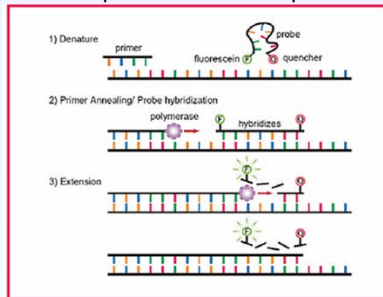
- **Quantitative Real-time PCR (qRT-PCR)** is a highly sensitive technique enabling amplification and quantification of a specific nucleotide (e.g., DNA, cDNA, RNA) sequence in real time.
- **Basic PCR**



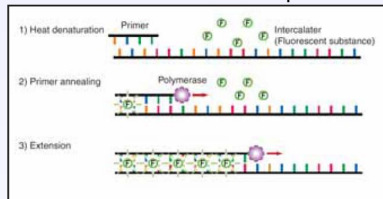
# Introduction

## ■ Fluorophore (Signal) Detection

### TaqMan Probe Principle

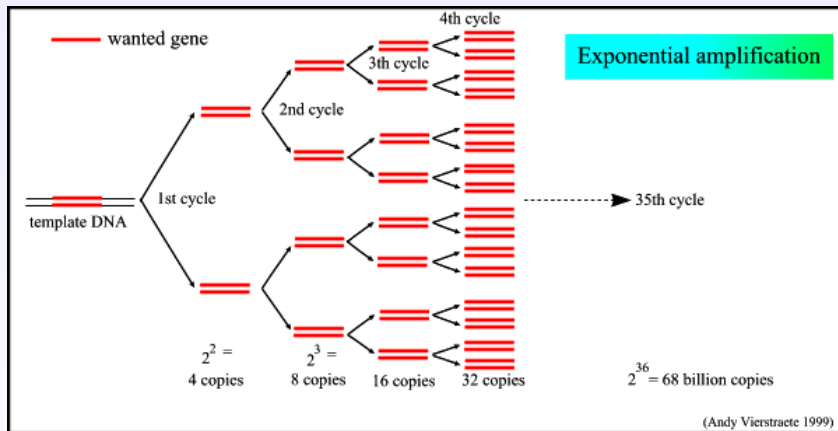


### SYBR Green Principle



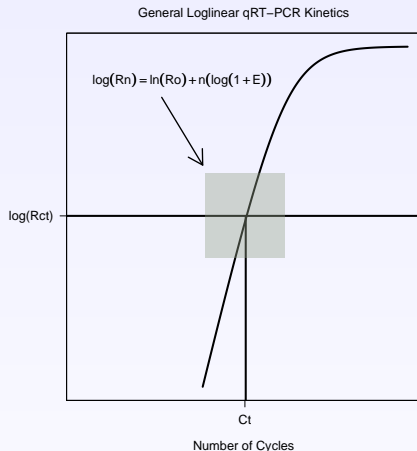
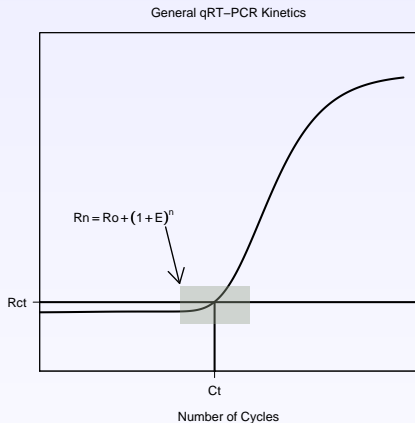
# Introduction

## ■ Amplification Rate



# Introduction

## ■ qRT-PCR Kinetics



## ■ The First Equation

$$X_{T,Ct} = X_{T,0}(1 + E_T)^{Ct,T}$$

where,

- $X_{T,Ct}$  is the fluorescence value at the threshold cycle (Ct) for the treatment group (T),
- $X_{T,0}$  is the fluorescence value at cycle zero proportional to the true copy number of the treatment group, and
- E is the efficiency of the amplification process.

## ■ A First Equation for Each Group

- Treated group reference

$$R_{T,Ct} = R_{T,0}(1 + E_{R_T})^{Ct,R_T}$$

- Control group

$$X_{C,Ct} = X_{C,0}(1 + E_C)^{Ct,C}$$

- Control group reference

$$R_{C,Ct} = R_{C,0}(1 + E_{R_C})^{Ct,R_C}$$

# Analysis

## ■ Fold Change

$$\frac{\frac{X_{T,0}}{R_{T,0}}}{\frac{X_{C,0}}{R_{C,0}}} = \frac{\frac{X_{T,Ct}}{R_{T,Ct}}}{\frac{X_{C,Ct}}{R_{C,Ct}}} * \frac{(1+E_{RT})^{Ct,RT}}{(1+E_T)^{Ct,T}} * \frac{(1+E_{RC})^{Ct,RC}}{(1+E_C)^{Ct,C}} \quad (1)$$

dividing fractions and rearranging terms gives,

$$\frac{X_{T,0}}{X_{C,0}} * \frac{R_{C,0}}{R_{T,0}} = \frac{(1+E_T)^{Ct,C}}{(1+E_C)^{Ct,T}} * \frac{(1+E_{RT})^{Ct,RT}}{(1+E_{RC})^{Ct,RC}} \quad (2)$$

## Analysis

- From equation 2,

$$\frac{X_{T,0}}{X_{C,0}} * \frac{R_{C,0}}{R_{T,0}} = \frac{(1+E_T)^{Ct,C}}{(1+E_C)^{Ct,T}} * \frac{(1+E_{R_T})^{Ct,R_T}}{(1+E_{R_C})^{Ct,R_C}}$$

- If  $E_T = E_C = E_{R_T} = E_{C_T}$  equation (2) becomes,

$$(1+E)^{-\Delta\Delta Ct} \tag{3}$$

- If  $\frac{R_{T,0}}{R_{C,0}} = 1$  equation (2) becomes,

$$\frac{(1+E_T)^{Ct,T}}{(1+E_C)^{Ct,C}} \tag{4}$$

## Impact of the IF's (Efficiency Not Equal To 1)

- Is there an impact of assuming efficiency's don't =1 (but you assume they do)?

- Error Percent=  $(100 \times \frac{2^n}{(1+E)^n}) - 100$

- Assume E is actually 0.9 in 25 cycles

- Error Percent=  $(100 \times \frac{2^{25}}{(1+0.9)^{25}}) - 100 = 261\%$

- The calculated expression level will be  $\frac{2^{25}}{(1+0.9)^{25}} = 3.6$  fold higher than the actual value

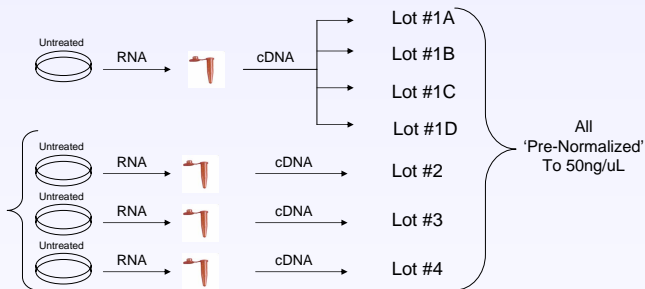
## ■ What is the impact of normalization?

- The use of an endogenous reference gene is supposed to correct for variation in,
  - RNA content,
  - reverse-transcriptase efficiency,
  - degradation or presence of inhibitors,
  - nucleic acid recovery, and
  - sample handling
- The use of an endogenous reference gene also increases variation in the experiment

# Impact of the IF's (Endogenous reference genes increase experimental variation)

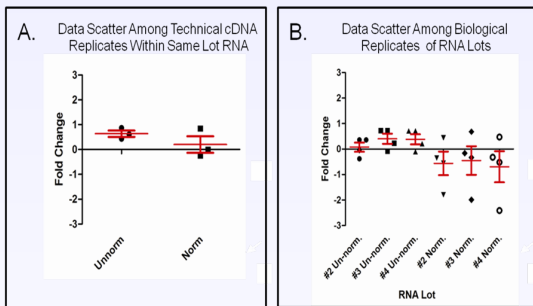
- In vitro experiments (Christian Kis)

## Scheme for RNA Samples



# Impact of the IF's (Endogenous reference genes increase experimental variation)

- In vitro cell line experiments (Christian Kis)



## Impact of the IF's (Endogenous reference genes increase experimental variation)

- In vitro cell line experiments (Christian Kis)
  - Ratios of normalized to unnormalized standard deviations ranged from 2.4 to 3.1
- Every 2-fold increase in standard deviation requires 4 times the number of experimental units to have equivalent statistical power to detect a statistically significant, biologically meaningful difference between 2 groups.
  - standard deviation of 2 compared to a standard deviation of 4
  - interested in detecting a  $\Delta=2$ -fold change in RNA expression
  - sample size of 4 ( $\Delta/\frac{2}{\sqrt{4}}$ ) required for a standard deviation of 2
  - sample size of 16 ( $\Delta/\frac{4}{\sqrt{16}}$ ) required for a standard deviation of 4
- If we can develop methods so that  $\frac{R_{T,0}}{R_{C,0}} = 1$  then,
  - Things that endogenous controls do are the same in both groups
  - Potentially huge savings in experimental costs
  - Analysis becomes much easier, fold change is  $\frac{(1+E_T)^{Ct_T}}{(1+E_C)^{Ct_C}}$

# Aurora A Knockdown Study

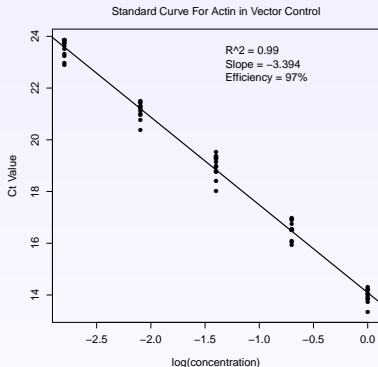
- In vitro cell line experiments (Ryan Splittgerber)
- Replicated over 3 plates in separate preparations
- 5-fold dilution series (1, 0.2, 0.04, 0.008, and 0.0016  $\mu\text{g}$ ) replicated four times for each condition on each plate

Condition: CONTROL Primer: AURKA		
		Samples
		Condition: shRNA Primer: ACTIN
	Samples	
	Condition: shRNA Primer: AURKA	
Samples		
Condition: CONTROL Primer: ACTIN		
		Samples

# Aurora A Knockdown Study (Efficiency Estimation with Standard Curves)

## ■ Efficiency Estimation with Standard Curves

- Conduct dilution series and regress Ct vs  $\log_{10}$ (amount of standard).
- Regression line should be strongly linear, efficiency should not depend on concentration.
- Efficiency =  $10^{\frac{-1}{\text{slope}}} - 1$



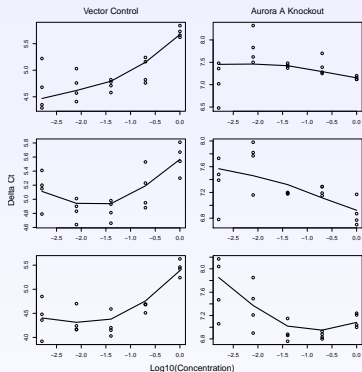
Slope	< -3.322	-3.322	> -3.22
Efficiency	< 1	1	> 1

Slope	Efficiency
-3.3	1.09
-3.8	0.83

# Aurora A Knockdown Study (Comparing Efficiency)

## ■ Is Efficiency The Same Between A Gene and Its Endogenous Control?

- Plot Differences in Ct values vs  $\log_{10}$ (amount of standard).
- Slope Should Be Approximately 0



## Aurora A Knockdown Study (Is Normalization Necessary?)

### ■ Analyzing Copy Number

The following Analysis Method is Easy and Interpretable

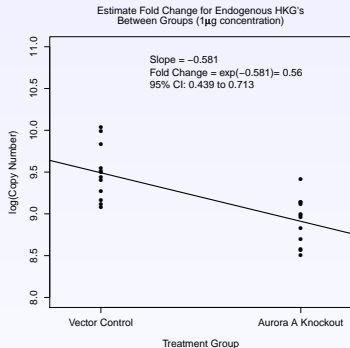
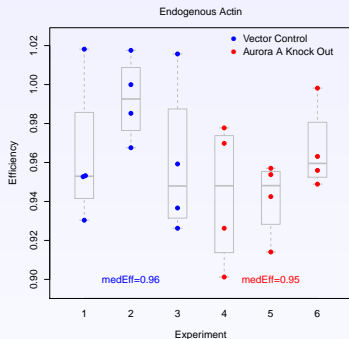
As An Exponential process, PCR Kinetics are often symmetric on the log scale

$$\begin{aligned}E[Y] &= \beta_0 + \beta_1(X) \\E[\log(Y)] &= \gamma_0 + \gamma_1(X) \\ \gamma_1 &= E[\log(Y)|x = 1] - E[\log(Y)|x = 0] \\ &= \text{med}(\log(Y)|x = 1) - \text{med}(\log(Y)|x = 0) \\ &= \log(\text{med}(Y|x = 1)) - \log(\text{med}(Y|x = 0)) \\ \gamma_1 &= \log\left(\frac{\text{med}(Y|x = 1)}{\text{med}(Y|x = 0)}\right) \\ \exp(\gamma_1) &= \frac{\text{med}(Y|x = 1)}{\text{med}(Y|x = 0)} = \text{FoldChange}\end{aligned}$$

# Aurora A Knockdown Study (Is Normalization Necessary?)

## ■ Is Normalization Necessary? (i.e., $\frac{R_{T,0}}{R_{C,0}} \approx 1$ ?)

- Efficiency for each replicate and well Ct were used to estimate per replicate copy number
- $\frac{R_{T,0}}{R_{C,0}} = 0.56$  (95% CI= 0.439 to 0.713)



# Analysis of Aurora A Knockdown Study

## ■ Analyze Assuming A Single Efficiency Overall ( $\Delta\Delta Ct$ Method)

- Calculate  $\Delta Ct$  between gene of interest and its endogenous control for each replicate
- There is no natural pairing between treatments, so the  $\Delta\Delta Ct$  part comes from comparing differences in the distributions of the  $\Delta Ct$  values
- Compare  $\Delta Ct$  values between treatments using non-parametric Wilcoxon rank sum test, or
- If we can assume that the  $\Delta Ct$  values are normally distributed, then,

- $\overline{\Delta Ct_T} - \overline{\Delta Ct_C} \sim N(0, \frac{\sigma_{\Delta Ct_T}^2}{n} + \frac{\sigma_{\Delta Ct_C}^2}{n})$

- Calculate difference in mean  $\Delta Ct$  between groups and its confidence interval

$$\overline{\Delta\Delta Ct} \pm 1.96 \times \sqrt{\left(\frac{\hat{\sigma}_{\Delta Ct_T}^2}{n} + \frac{\hat{\sigma}_{\Delta Ct_C}^2}{n}\right)}$$

- Exponentiate mean  $\Delta\Delta Ct$  and confidence limits according to  $2^{\Delta\Delta Ct}$  or  $(1+E)^{\Delta\Delta Ct}$

## Analysis of Aurora A Knockdown Study

### ■ $\Delta C_t$ Values

Vector Control	Aurora A Knockdown
5.66	7.20
5.62	7.16
5.73	7.12
5.84	7.12
5.30	7.17
5.54	6.77
5.67	6.77
5.81	6.87
5.42	7.04
5.24	7.21
5.63	7.24
5.46	7.00

## Analysis of Aurora A Knockdown Study

### Analyze Assuming Single Efficiency

Statistic	Vector Control	Aurora A Knockdown
$\overline{\Delta Ct}$	5.58	7.05
$\hat{\sigma}_{Ct}^2$	0.0361	0.0324

Statistic	LCL	Estimate	UCL
$\Delta\Delta Ct$	1.32	1.47	1.62
$E = 1$	.400	0.361	0.326
$E = 0.96$	0.411	0.372	0.336

- Conclusion: Aurora A Knockout Elicits Significantly Less RNA Than Control Group

## Analysis of Aurora A Knockdown Study

### ■ Analyze Assuming a Specific Efficiency for Each Replicate

- Calculate a copy number for each replicate well
- Calculate fold change for each replicate GOI and HKG pair
- Compare log of fold change values between treatments as previously discussed assuming fold change is log normal

Statistic	Vector Control	Aurora A Knockdown
$\overline{\ln FC}$	-5.03	-4.13
$\hat{\sigma}_{FC}^2$	0.49	0.314

Statistic	LCL	Estimate	UCL
$\overline{\ln FC}$	0.393	0.9	1.407
$e^{\overline{\ln FC}}$	1.48	2.46	4.08

- **Conclusion: Aurora A Knockout Elicits Significantly More RNA than Control Group**

# Analysis of Aurora A Knockdown Study

## ■ Completely Different Conclusions!

- Single Efficiency: Aurora A Knockdown reduces gene expression
- Replicate Efficiency: Aurora A increased gene expression
- Why?

Average Value	Control Group	Aurora A Knockdown
HKG Efficiency	0.97	0.95
Treatment Efficiency	1.1	0.90
Normalized Fold Change	0.008	0.019

## Summary

- Throughout, we have recommended analysis methods that depend on Normality assumptions because they tend to be better understood by our collaborators and have nice features like confidence intervals. However, in the usual setting, sample sizes are small, and normality cannot be verified. We recommend larger studies (more than 3 replications) so see your statistician for better, reproducible results.
- qRT PCR follows an exponential growth curve so efficiency matters
- Variable efficiencies are the rule, not the exception
- Lab methods for eliminating endogenous controls will improve every level of qRT PCR - Major Area for Improvement!
- Don't Use Standard Curves
- Utilize Well Level Estimates of Efficiency