

The background is a dark chalkboard with various white chalk sketches. On the left, there is a large sketch of a microscope. Above it is a globe showing the continents. Below the globe is a large plus sign. At the bottom, there are sketches of an open book, a percentage sign, a division sign, and a less-than sign.

# Quantifying RNA levels from RTPCR curves: A New Method for Single Cell Analysis

Speakers:

Savion Smith, Esther Fevrier, and Havell Markus

Faculty Mentor:

Rafael Irizarry, Ph.D.

Professor, HSPH, Biostatistics and Computational Biology, and Dana-Farber

Graduate HSPH Student Mentor:

Stephanie Chan

Summer Program In Quantitative Sciences at HSPH

July 24<sup>th</sup>, 2014



# Overview

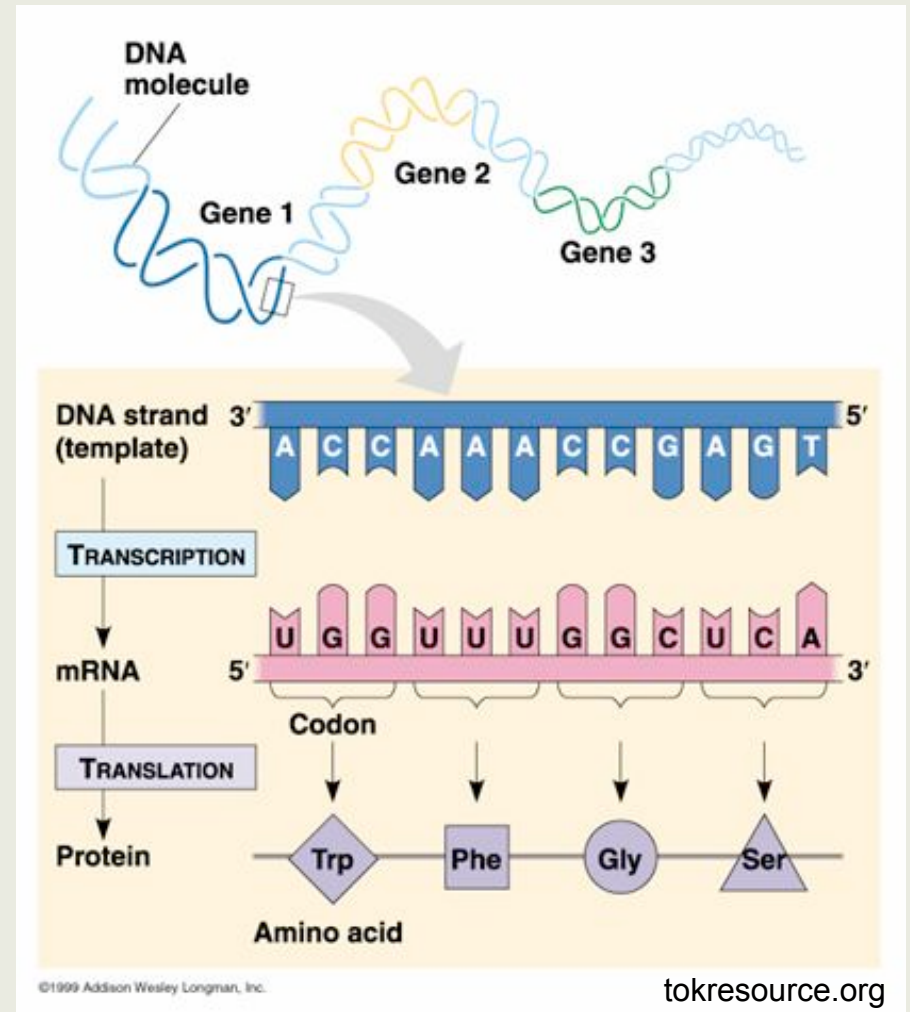
- Background
  - What is “gene expression”?
  - Basic biological background
  - Current applications of RT-PCR technology
- Our Study
  - Single analysis
  - RT-PCR for Single Cell Analysis
- Results and Conclusions
- Future Works
  - Optimizing the potential of Real-Time PCR technology for quantitative analysis



# Background

## Gene Expression

- Decoding information from a gene to create functional gene product
- Products include
  - Amino Acids
  - Ribosomal RNA (rRNA)
  - Transfer RNA (tRNA)

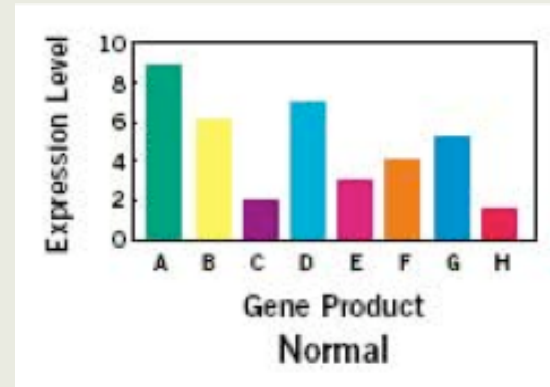




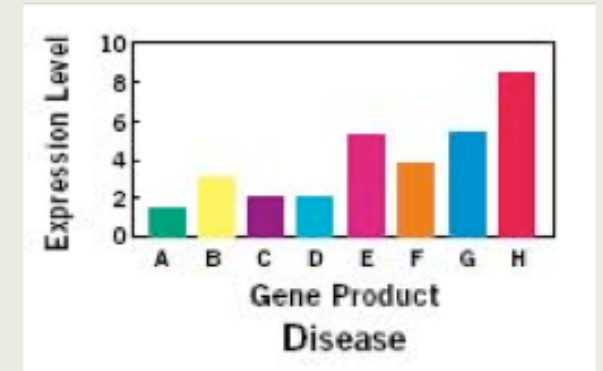
# Background

## Benefits of Quantifying Gene Expression

- Quantifying expression levels of a particular gene within an organism, tissue, or cell provides a lot of information to aid with:
  - Diagnostic Study
  - Prevention Interventions
  - Specific Treatment Options



**Vs.**





# Background

## Real Time - PCR (RT-PCR)

- Detects and quantifies gene expression in a given sample through PCR
- Polymerase Chain Reaction (PCR)
  - Process for amplifying a given sequence of DNA

## Applications of RT-PCR

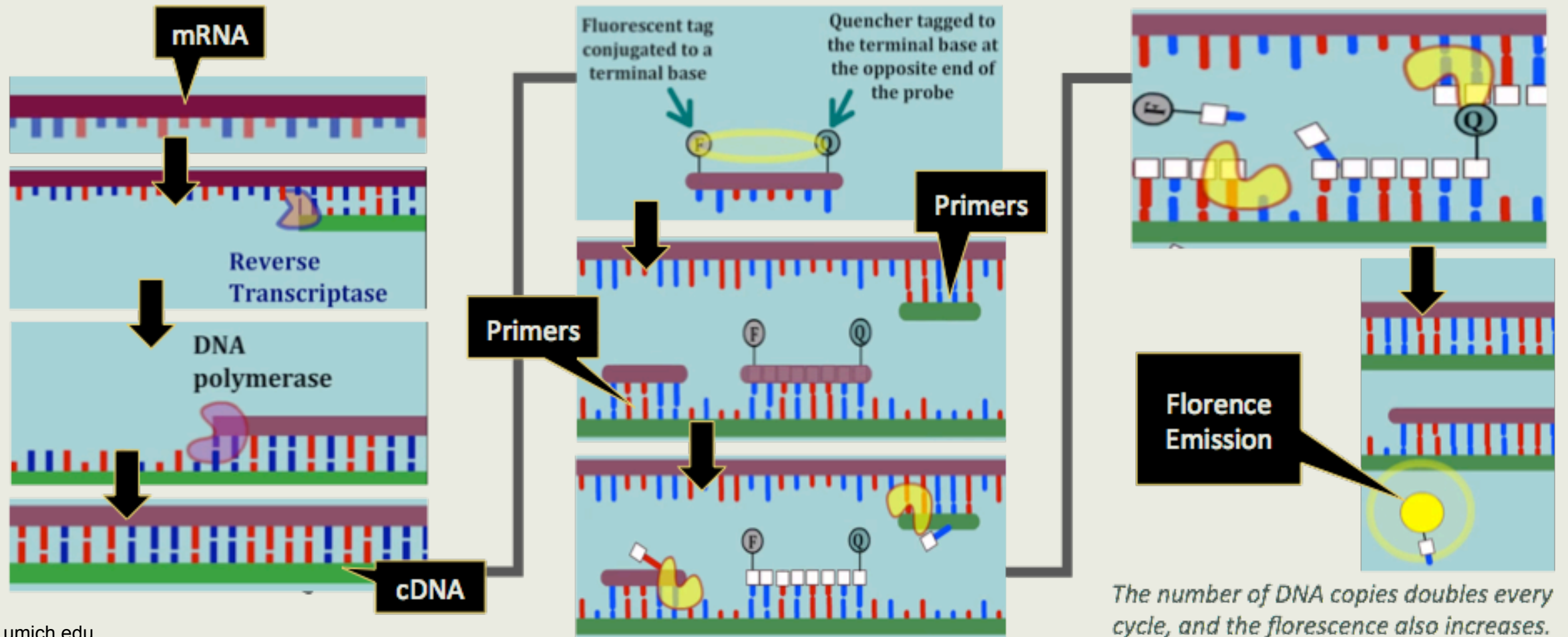
- Quantification of gene expression
- Pathogen detection
- Viral quantification
- Drug therapy efficacy
- DNA damage measurement
- Genotyping



gene-quantification.de



# Background: Real Time - PCR

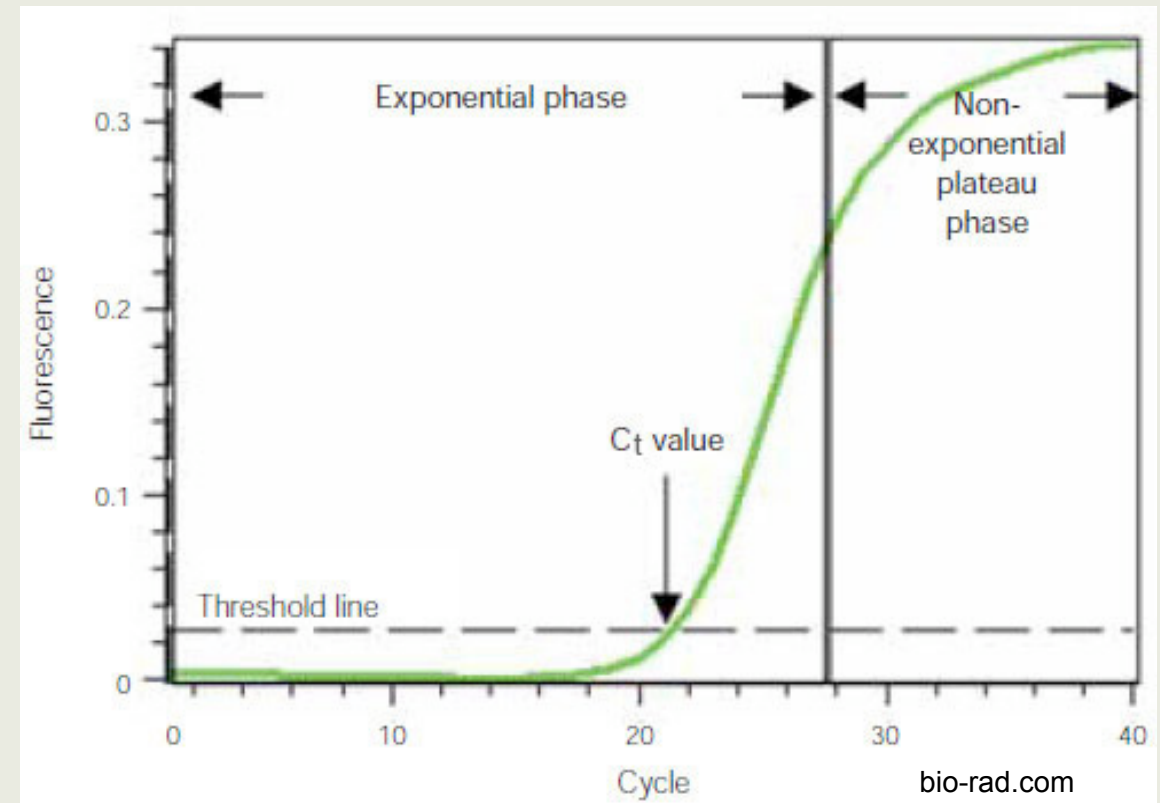




# Background

## Collecting Data w/ RT-PCR

- Stores fluorescence emitted over different PCR cycles
- Measures how many cycles it takes to reach a predetermined threshold of detection or, the  $C_t$  Value
- Building a Standard Curve
  - $C_t$  Values of a sample with known quantities of mRNA
  - Using it to predict the initial copies of mRNA
    - Copies per nanolitre of a homogenized tissue or copies per cell





A chalkboard background with various scientific and mathematical sketches in white chalk. On the left, there is a large 'V' shape, a telescope, a stack of books, and a microscope. In the upper center, there is a globe showing the continents. Below the globe, there is a plus sign. In the lower center, there is an equals sign and an open book with some handwritten text. On the right, there is a percentage sign, a division sign, and a less-than sign.

# Our Study



# Our Study

## Quantitative RT-PCR & Single Cell Analysis

### Benefits

- Allows analysis of cellular variability
- Detects genetic abnormalities, viral pathogens, and certain kinds of cancer on a cellular level
- Allows PCR amplification to be monitored and quantified in real time

### Limitations

- Minimal amounts of RNA available
- Efficiency of RT-PCR depends on the quality of its primers

# Our Study

## Underlying Question

- Limited amount of research conducted on RT-PCR and single cell analysis
- Testing efficiency of RT-PCR's current method for single cell analysis
  - Accuracy and Precision
- If not, proposing a new efficient statistical method for quantification





# Our Study

## Data Origin

- Conducted by Dr. Livak from the Fluidigm Corporation
- Fluidigm Corporation
  - Works with Single-Cell Gene Expression



## Data Set

- 9216  $C_t$  Values
  - 96 Genes
  - 96 Samples

Experiment	Experiment	Experiment	Experiment	Experiment	Experiment	EvaGreen	EvaGreen	EvaGreen	EvaGreen	EvaGreen	EvaGreen	EvaGreen
Chamber	Sample	Sample	Sample	EvaGreen	EvaGreen	Ct	Ct	Ct	Ct	Tm	Tm	Tm
ID	Name	Type	rConc	Name	Type	Value	Quality	Call	Threshold	In Range	Out Range	Peak Ratio
S96-A01		40 Unknown	1	ABCC1	Test	13.09257	1	Pass	0.016724	84.77468	999	1
S96-A02		40 Unknown	1	B2M	Test	6.813834	1	Pass	0.016724	82.61949	999	1
S96-A03		40 Unknown	1	CCNB1	Test	16.393	1	Pass	0.016724	77.29589	999	1
S96-A04		40 Unknown	1	CDKN1A	Test	19.1015	1	Pass	0.016724	88.03547	999	1
S96-A05		40 Unknown	1	CRADD	Test	18.05104	1	Pass	0.016724	84.03517	999	1
S96-A06		40 Unknown	1	E2F1	Test	19.31003	1	Pass	0.016724	85.13116	999	1
S96-A07		40 Unknown	1	HNRNPH3	Test	11.18747	1	Pass	0.016724	80.5318	999	1
S96-A08		40 Unknown	1	NDUFA4	Test	11.36832	1	Pass	0.016724	85.37928	999	1

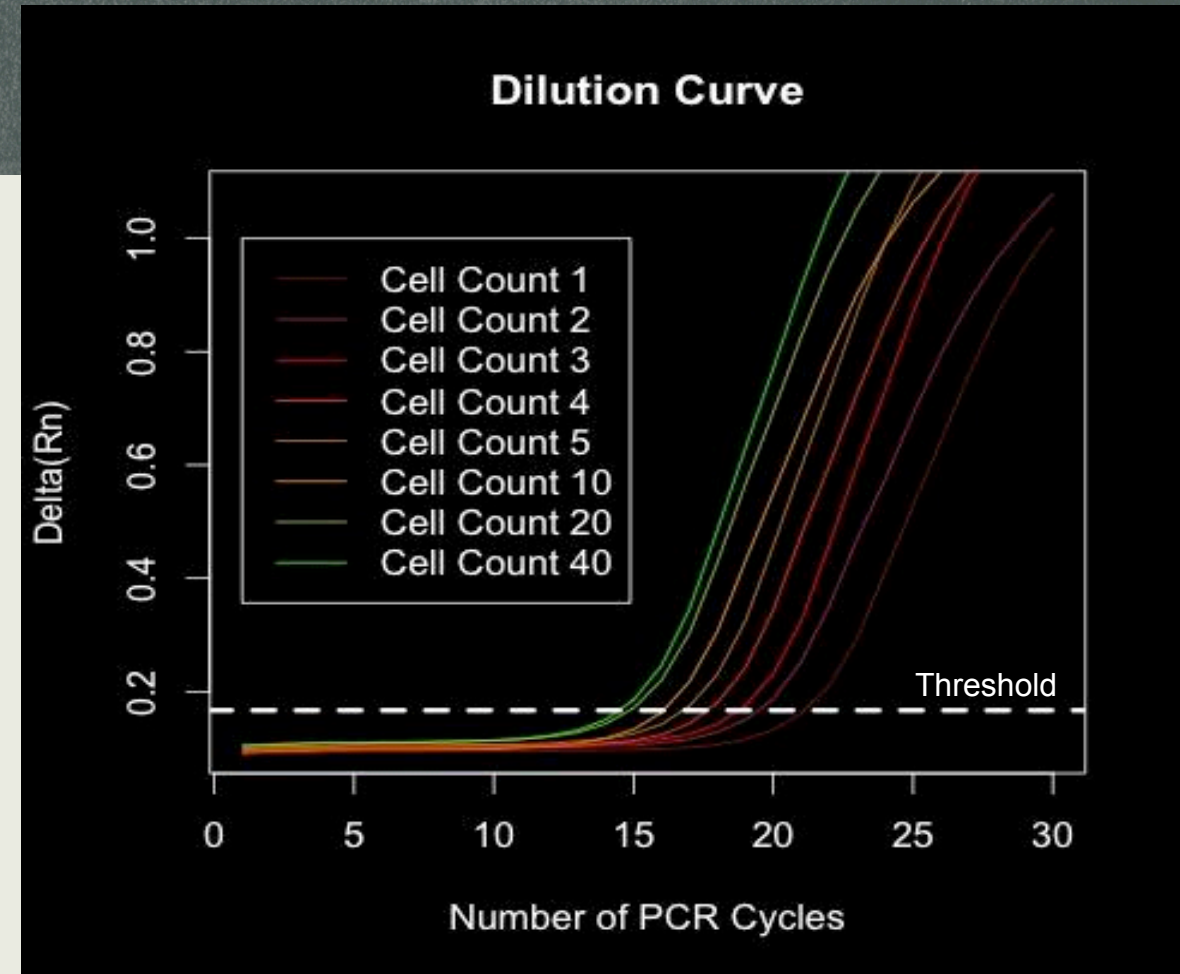




# Our Study

## Determining Relationship

- Samples:
  - Different sample cell count in sets of 12 for each gene
    - 1, 2, 3, 4, 5, 10, 20, 40
- $\uparrow$ Sample Cell Count  $\downarrow$   $C_t$  Value
- $X \cdot 2^{C_t} = Y$ 
  - X: Sample Cell Count
  - Y: Relative Fluorescence Threshold



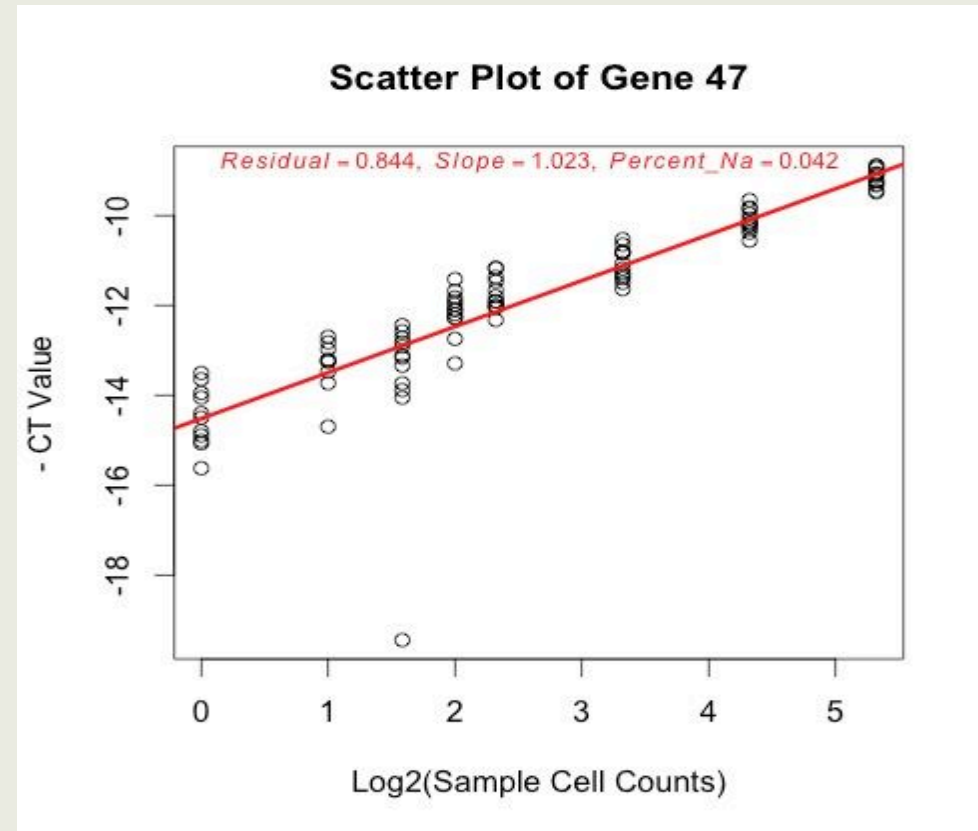
Dilution Curve for Gene 1  
Relative Fluorescence (Delta Rn) vs. PCR Cycle Number



# Our Study

## Examining Relationship

- $C_t$  are logged
  - RNA doubles each cycle
- Measured  $C_t$  vs.  $\log_2$ (Sample Cell Count)
- Used  $-C_t$  Value for positive slope
- **Model:**
  - $\log_2(\text{Sample Cell Count}) - a = -C_t$ 
    - “a” is a relative constant



**Standard Curve**  
 $-C_t$  Value vs.  $\log_2$ (Sample Cell Count)

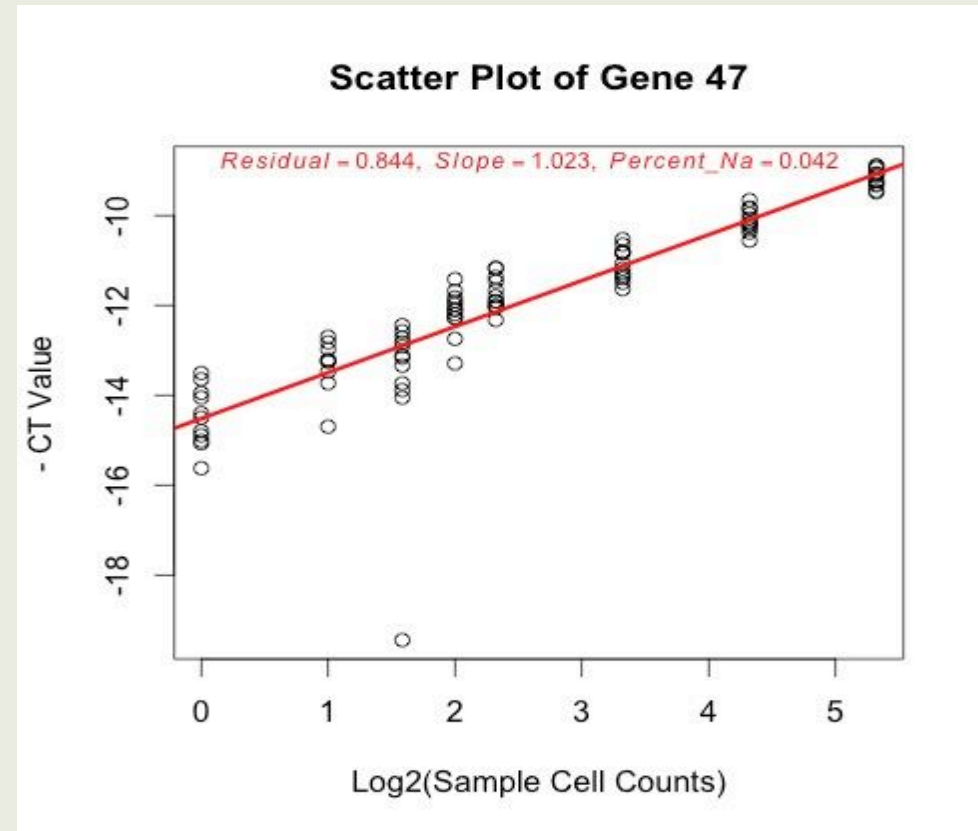


# Our Study

## Examining Data

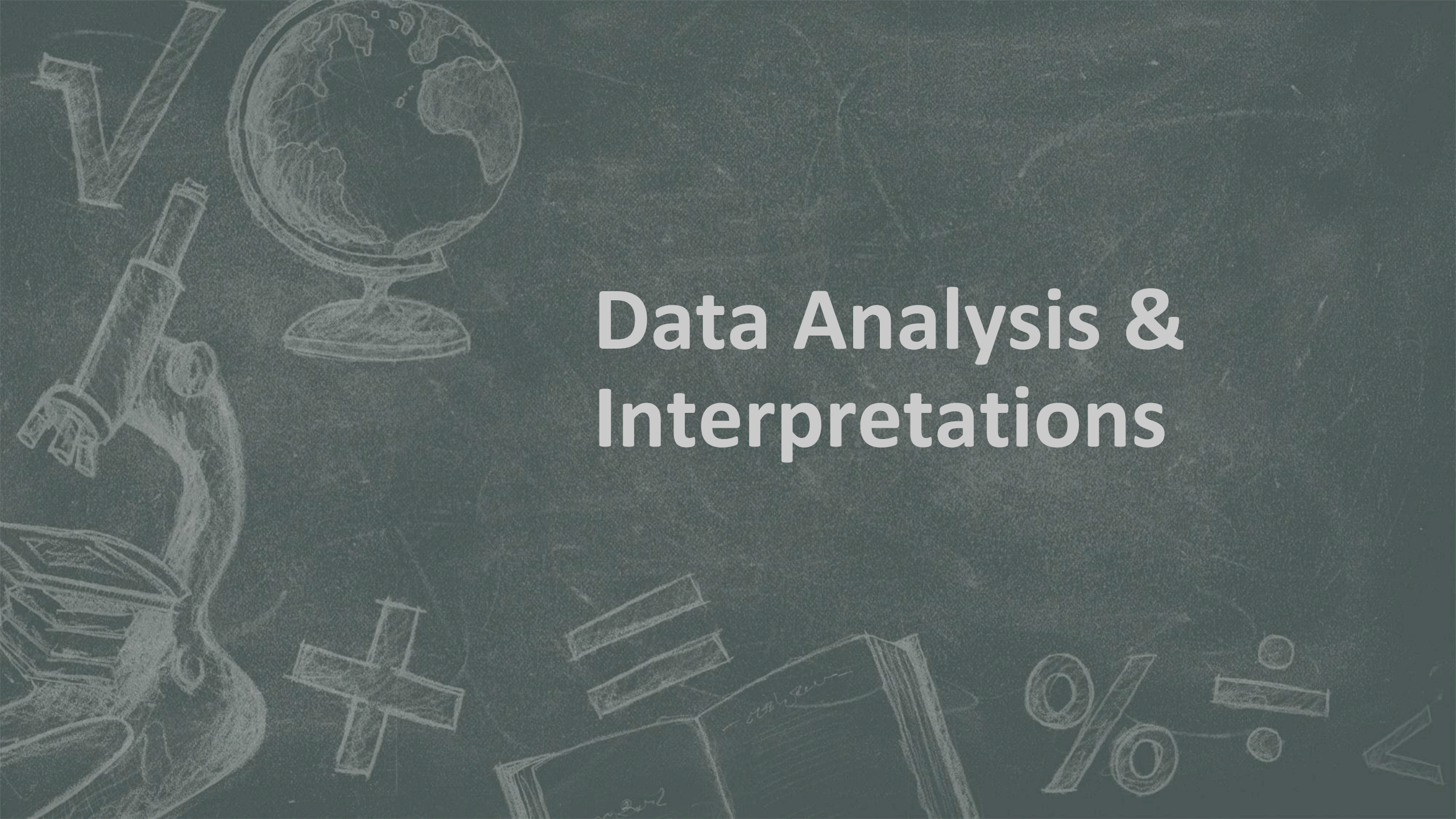
Ideal Plot:

- Slope
  - $\approx 1$
- Residual
  - $\approx 0$
- Recorded Percent N/A
  - Not enough cycles
  - Gene is not expressed
- Variance of Single-Cell
  - Comparable to that of greater cell count



**Standard Curve**  
-C<sub>t</sub> Value vs. Log<sub>2</sub>(Sample Cell)



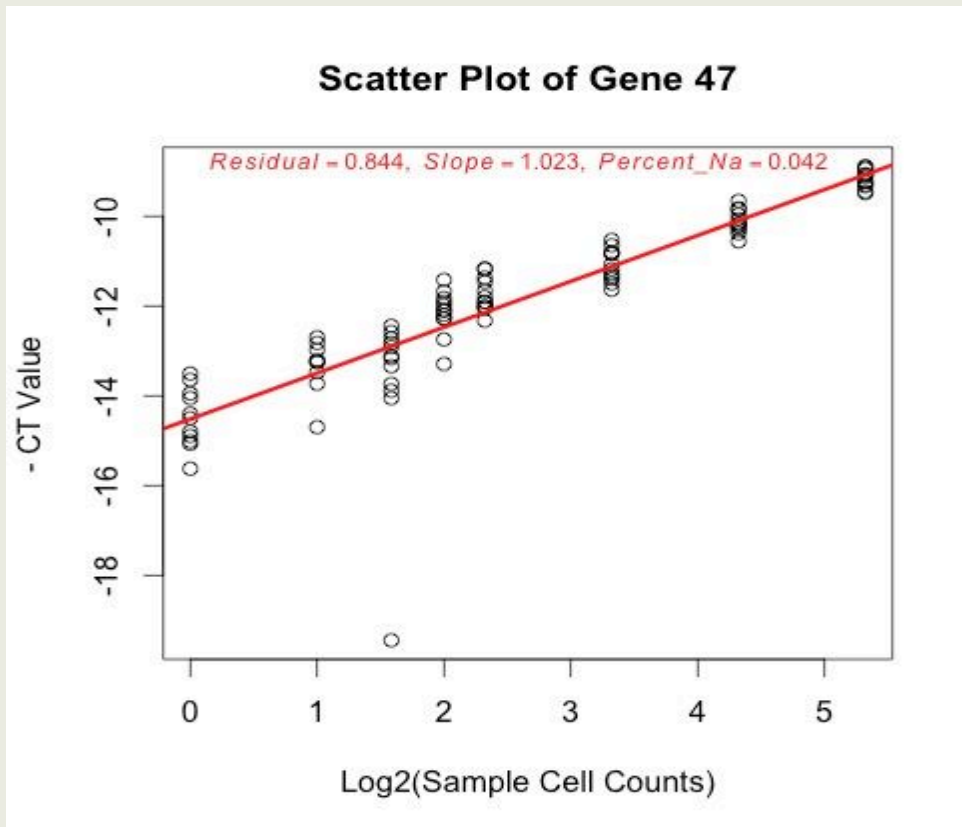
A chalkboard background with various scientific and mathematical sketches in white chalk. On the left, there is a large 'V' shape, a telescope, and a stack of books. In the top left, a globe is shown. In the bottom left, there is a plus sign and an equals sign. In the bottom center, there is an open book with some handwritten text. In the bottom right, there is a percentage sign, a division sign, and a less-than sign.

# Data Analysis & Interpretations



# Data Analysis & Interpretations

## Observed Results



## Interpretation

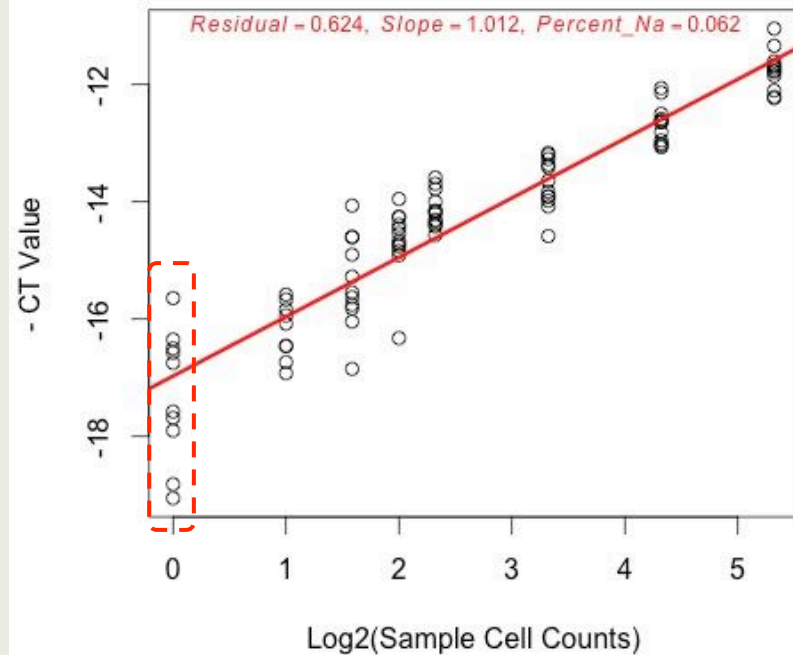
- Slope  $\approx 1.023$ 
  - Good accuracy
- Residual error  $\approx 0.544$ 
  - Good precision
- % NA  $\approx 0.042$ 
  - Majority of  $C_t$  values were captured by RT-PCR machine
- Opens up questions about how precise RT-PCR technology is in the case of single cells



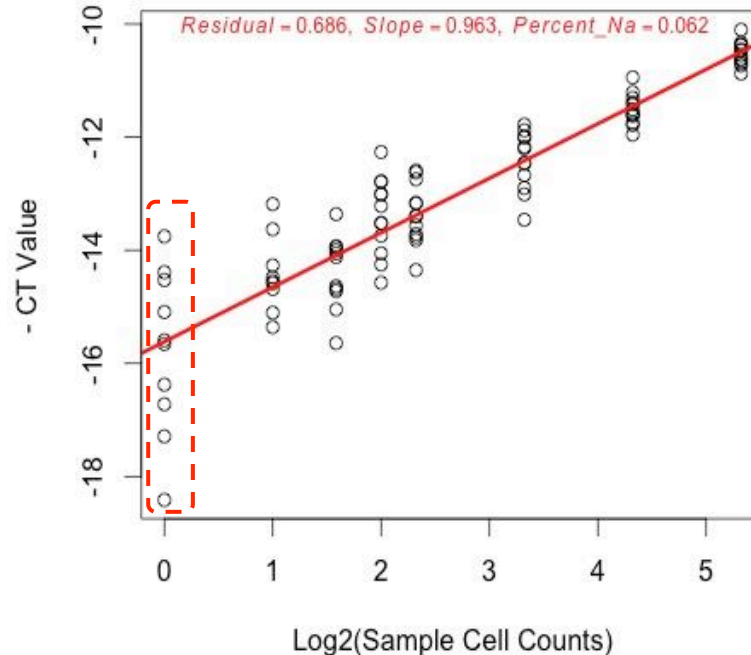
# Data Analysis & Interpretations

Other genes showing the trend of high variance at lower sample cell counts

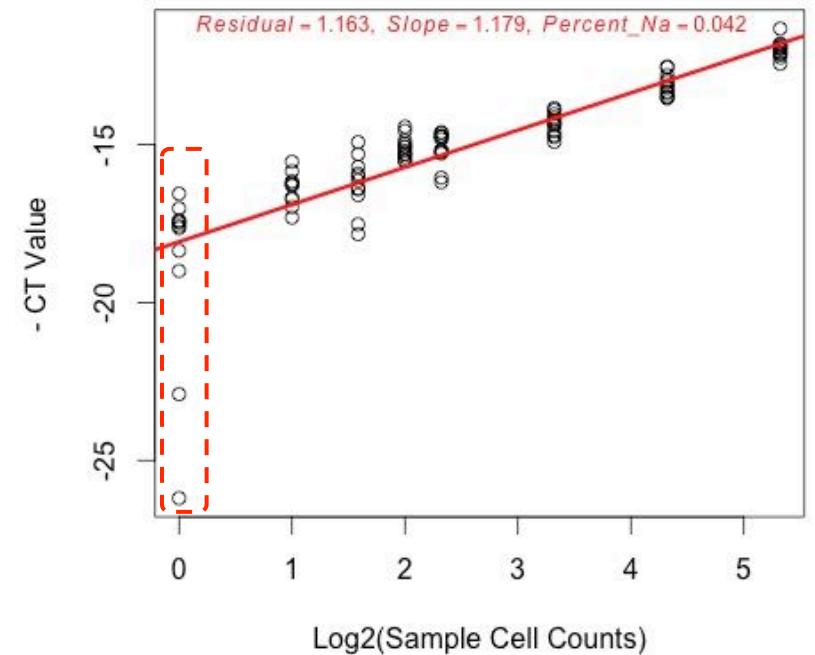
Scatter Plot of Gene 17



Scatter Plot of Gene 71

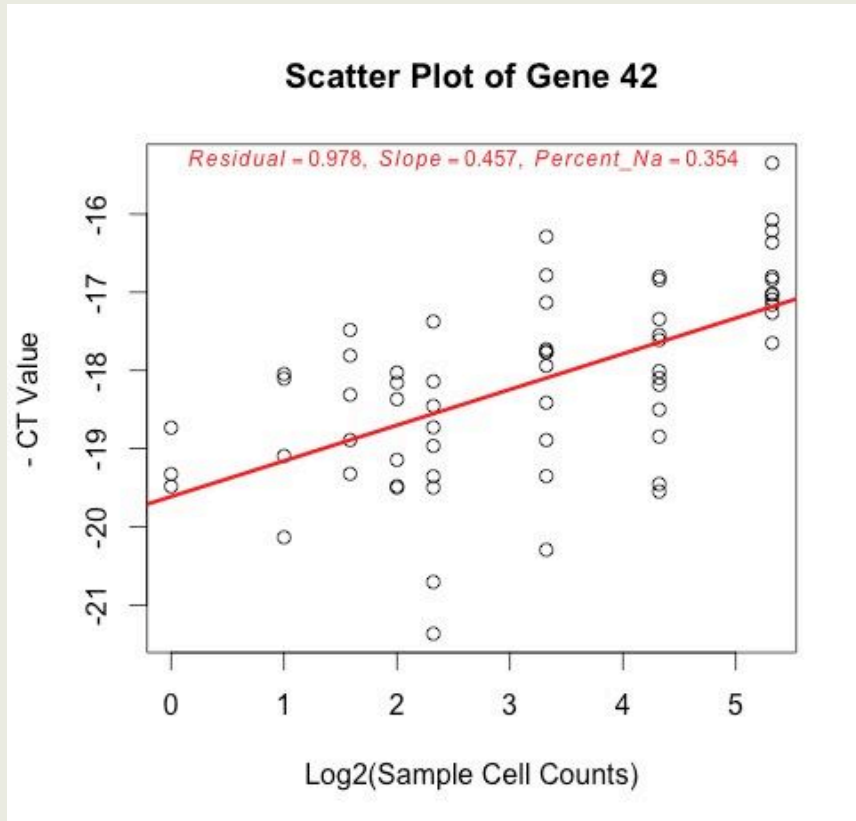


Scatter Plot of Gene 88



# Data Analysis & Interpretations

## Observed Results



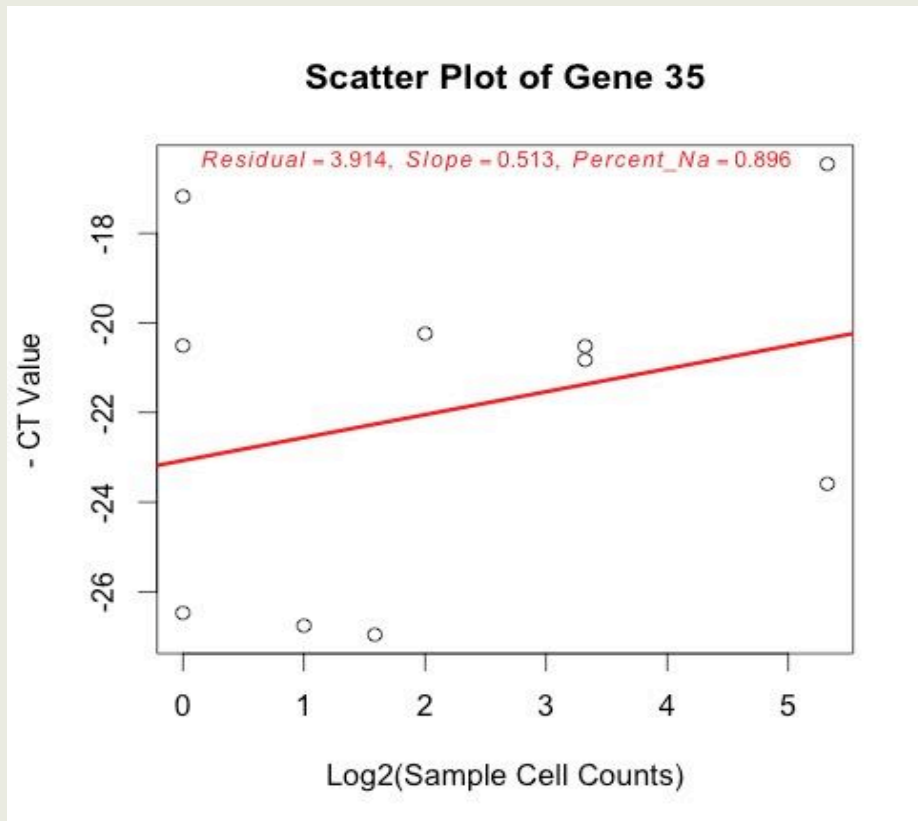
## Interpretation

- Slope  $\approx 0.457$ 
  - Low accuracy
- High residual error ( $\approx 0.978$ )
  - Lacks precision
- % NA  $\approx 35\%$ 
  - RT-PCR captured 65% of  $C_t$  values
- Raised further questions on how well RT-PCR was able to capture the expression levels in our samples



# Data Analysis & Interpretations

## Observed Results



## Interpretation

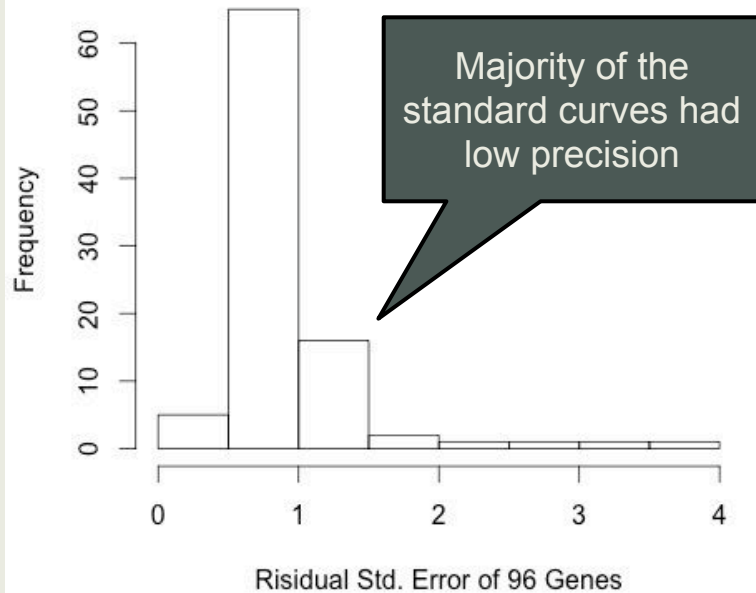
- Slope  $\approx 0.513$ 
  - Not close to one
  - Lacks accuracy
- High residual error  $\approx 3.914$ 
  - Lacks precision
- Most of the data is NA

*\*Note: Gene is not expressed, or gene went undetected by RT-PCR machine when 100% of the data is missing, i.e. Gene 54*

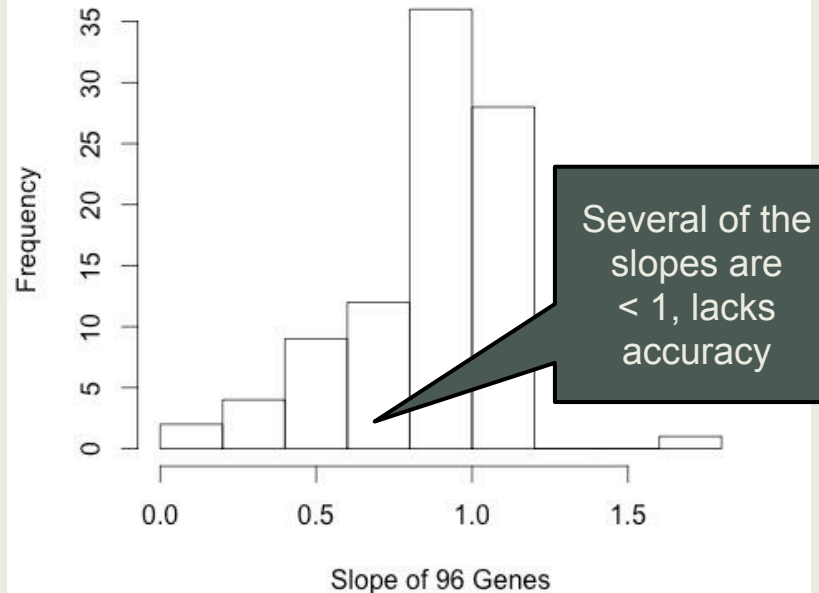
# Data Analysis & Interpretations

## Distributions of our data's statistics

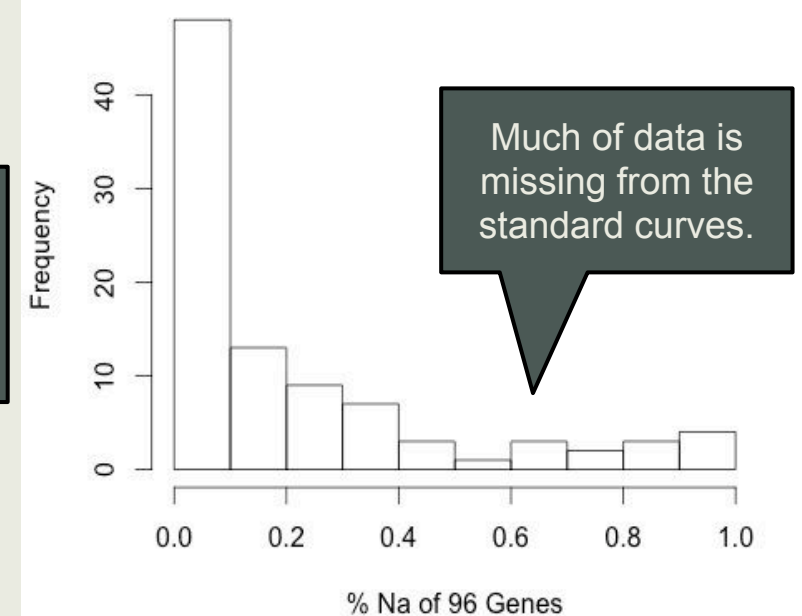
Residual Std. Error Distribution of 96 Genes



Slope distribution of 96 Genes



% Na distribution of 96 Genes



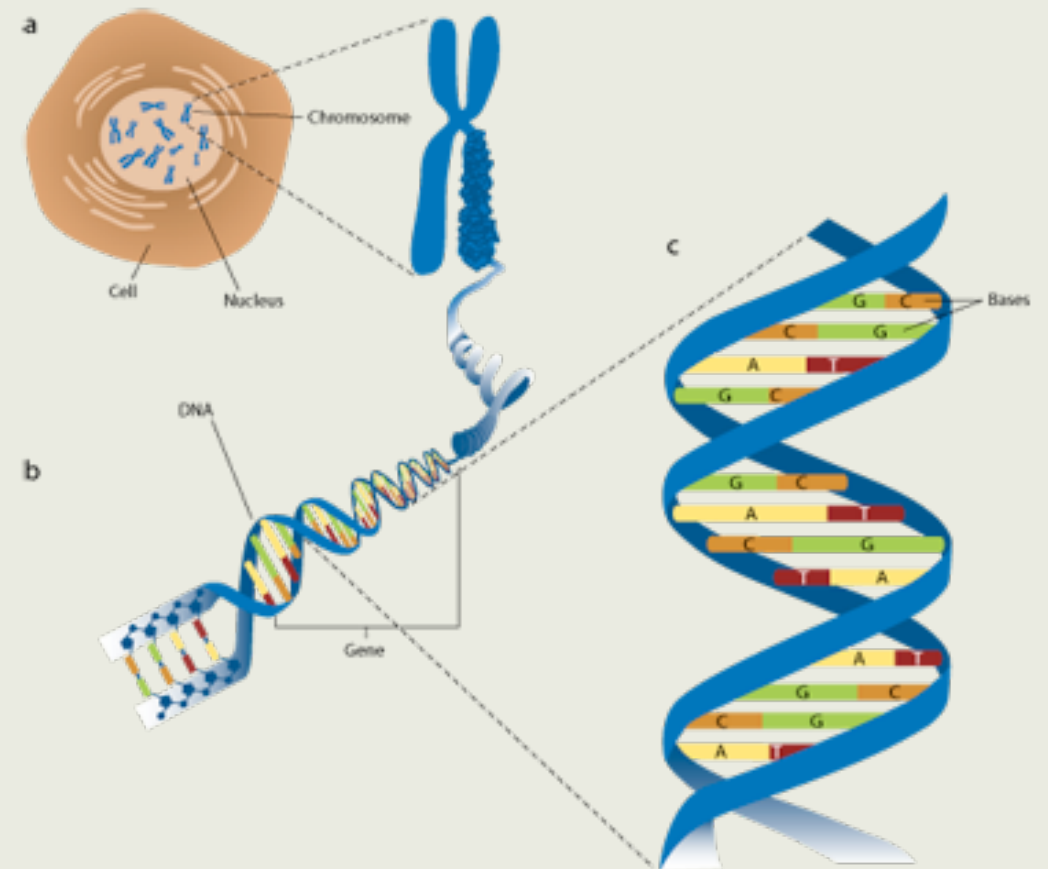


# Data Analysis & Interpretations

## Conclusions

- Not all of the observed outcomes match our expected outcome
  - Precise, but not accurate
  - Accurate, but not precise
- Reproducibility → Consistent results
- Biological variability
- More efficient on multicellular Level

For Single Cell analysis, the Ct-threshold method of RT-PCR is inefficient for estimating initial starting amounts of genetic transcripts





A chalkboard background with various educational sketches in white chalk. In the top left, there is a large letter 'V'. Below it is a drawing of a microscope. To the right of the microscope is a globe showing the continents. In the bottom left, there is a drawing of a stack of books. In the bottom center, there is a drawing of an open book with some text on its pages. To the right of the open book is a drawing of a plus sign. In the bottom right, there is a drawing of a percentage sign and a division sign. The text 'Future Works' is written in the center of the board in a large, white, sans-serif font.

# Future Works



# Future Works

## Need for a Statistical Model

- The  $\Delta R_n$  vs. PCR Cycles graphs are a sigmoidal curve, but RT-PCR fits an exponential model
- The model used by RT-PCR assumes all the values of reproduction rate “a” in the sigmoidal curve is always 2 for every gene, when it might not be.

Model Used by RT-PCR

$$X a^{Ct} = Y$$

## Next Step: Sigmoidal Function

- Independently fit a sigmoidal function to each curve for a better fit
- Distinguish between NAs due to small values of “a” and NAs due to gene non-expression

# Future Works

## Sigmoidal Function

The diagram illustrates the sigmoidal function equation for PCR fluorescence, with callouts identifying each variable:

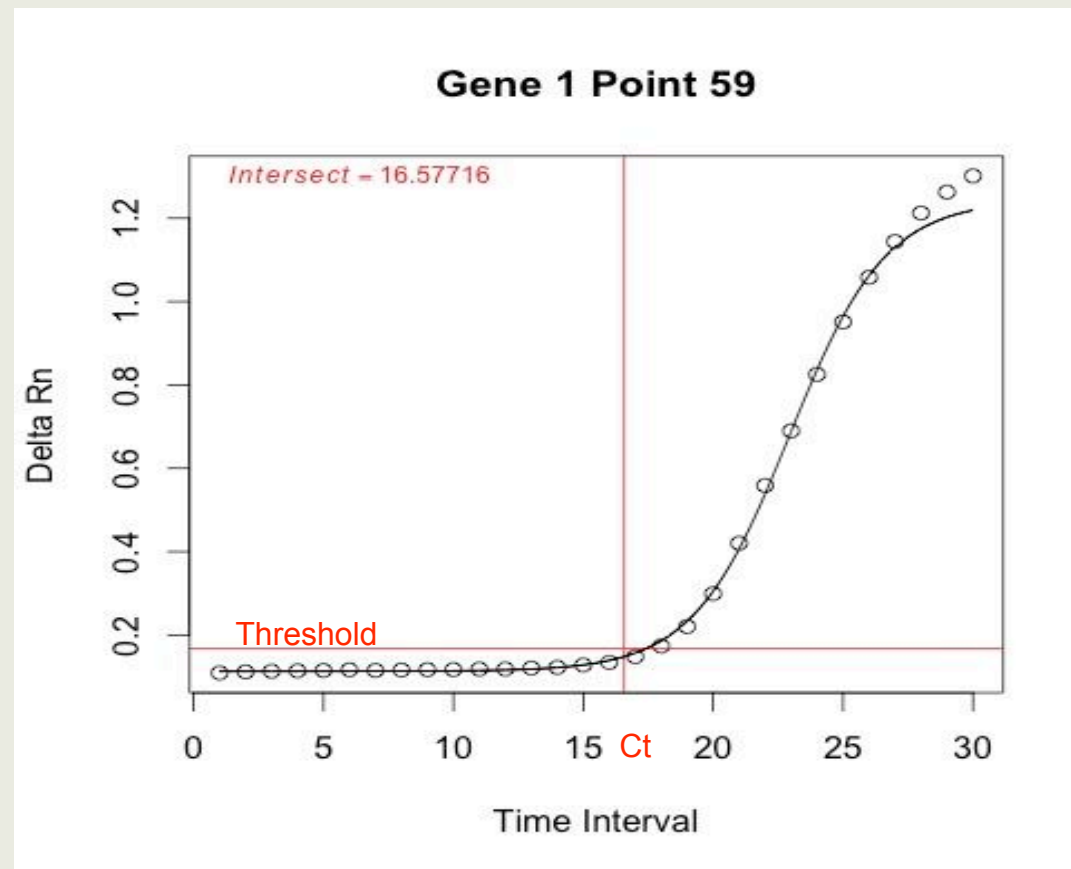
- Reaction fluorescence at cycle C**: Points to the variable  $F_c$  on the left side of the equation.
- Maximal reaction fluorescence**: Points to the variable  $F_{max}$  in the numerator of the fraction.
- Background reaction fluorescence**: Points to the variable  $F_b$  on the right side of the equation.
- Number of PCR cycles**: Points to the variable  $C$  in the denominator's exponent.
- Cycles at which reaction fluorescence reaches half of  $F_{max}$** : Points to the variable  $C_{1/2}$  in the denominator's exponent.
- Slope at the exponential phase of the curve**: Points to the variable  $k$  in the denominator's exponent.

$$F_c = \frac{F_{max}}{1 + e^{-(C - C_{1/2})/k}} + F_b$$



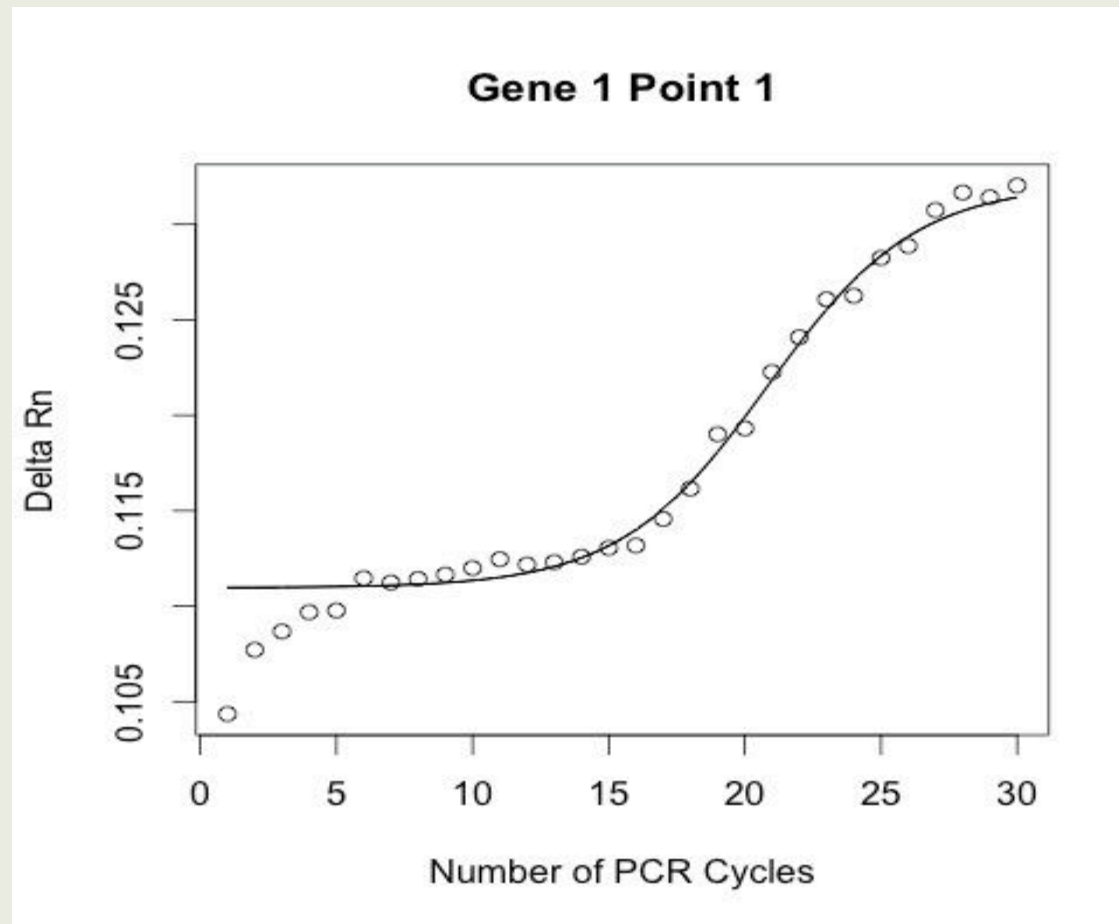
# Future Works

## Sigmoidal Curve Example



# Future Works

## Sigmoidal Curve Example





The background is a dark green chalkboard with various white chalk sketches. On the left, there is a large sketch of a microscope. Above it is a globe showing the continents. Below the globe is a large plus sign. At the bottom, there are sketches of an open book, a stack of books, a percent sign, a division sign, and a less-than sign.

# Acknowledgements

Dr. Rafael Irizarry

Stephanie Chan

Dr. Matt Cefalu

Summer Program In Quantitative Sciences at HSPH

PubMed

Makers of R

Jake Conway



The background is a dark green chalkboard with various white chalk sketches. In the top left, there is a large 'V' and a globe on a stand. To the left of the globe is a microscope. Below the microscope is a stack of books. In the bottom left, there is a plus sign. In the bottom center, there is an equals sign and an open book with some illegible writing. In the bottom right, there is a percent sign, a division sign, and a less-than sign.

**Thank You!  
&  
Questions?**



# Citations

Cui, X., Hwang, J. G., Qiu, J., Blades, N. J., & Churchill, G. A. (2005). Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics*, 6(1), 59-75.

Irizarry, R. A., Parmigiani, G., Guo, M., Dracheva, T., & Jen, J. (2001). A statistical analysis of radiolabeled gene expression data. *Proceedings of Interface 2001*, 1-4.

Liu, W., & Saint, D. A. (2002). Validation of a quantitative method for real time PCR kinetics. *Biochemical and biophysical research communications*, 294(2), 347-353.

McDavid, A., Finak, G., Chattopadhyay, P. K., Dominguez, M., Lamoreaux, L., Ma, S. S., ... & Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics*, 29(4), 461-467.

Ritz, C., & Spiess, A. N. (2008). qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*, 24(13), 1549-1551.

Rutledge, R. G. (2004). Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. *Nucleic acids research*, 32(22), e178-e178.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 3(7), research0034.