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 24th, 2014

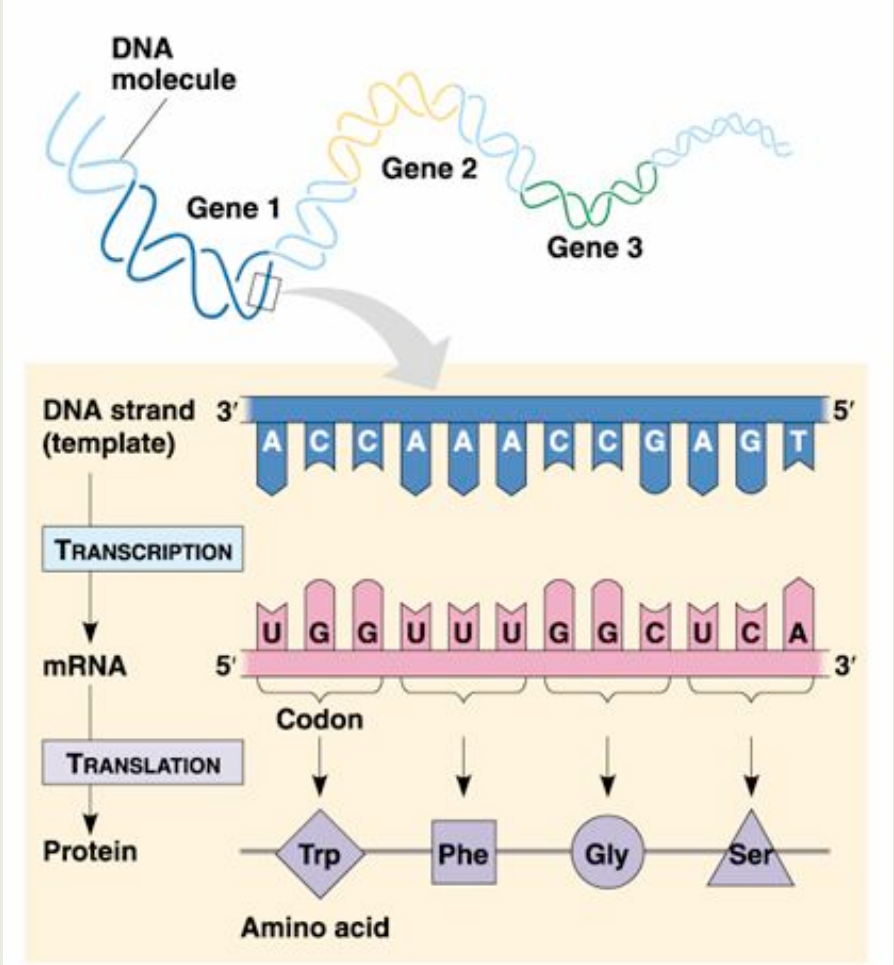
Overview

- 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)



tokresource.org

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


gene-quantification.de

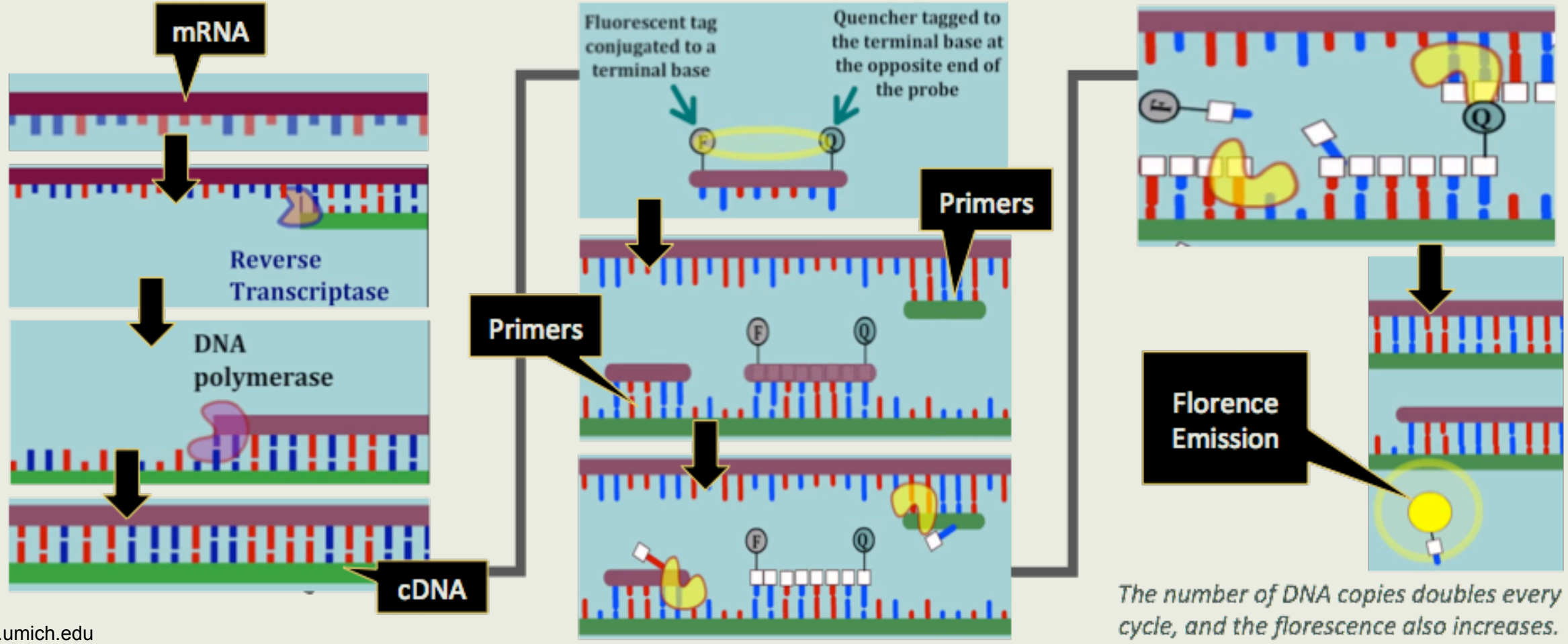
Real Time - PCR (RT-PCR)

- 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

Background: Real Time - PCR



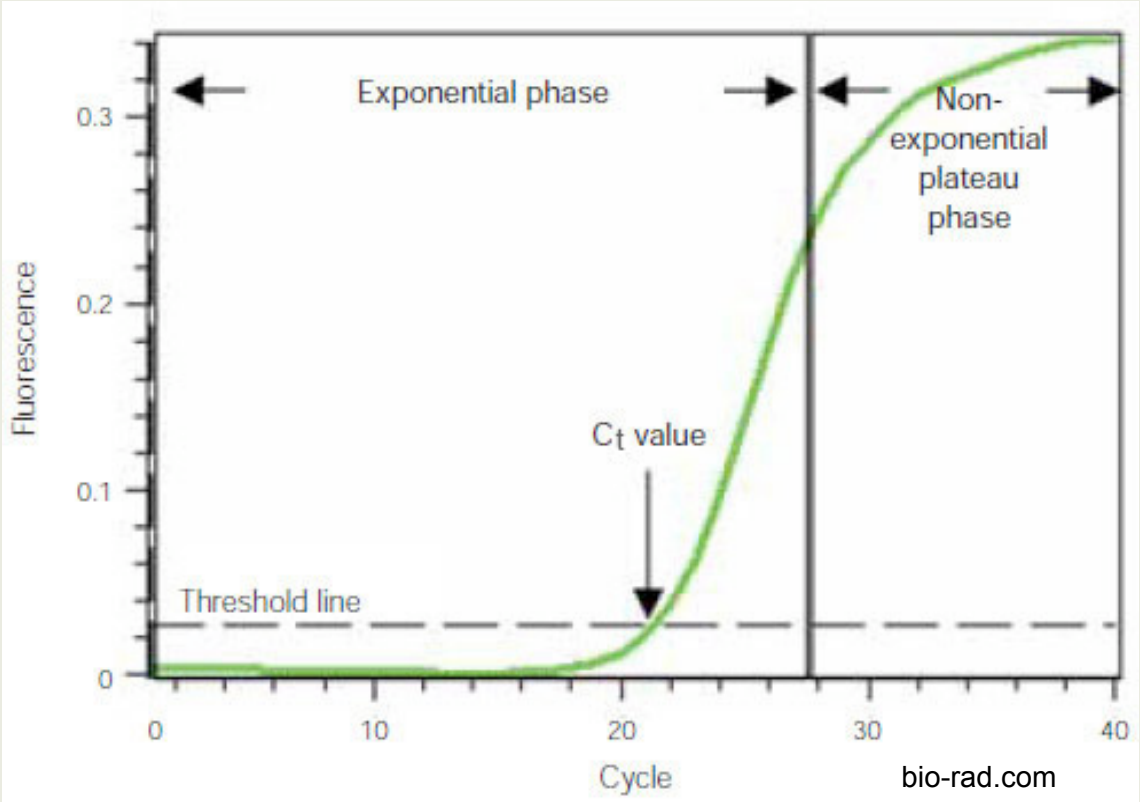
open.umich.edu

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



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

Experime	Experime	Experime	Experime	Experime	Experime	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

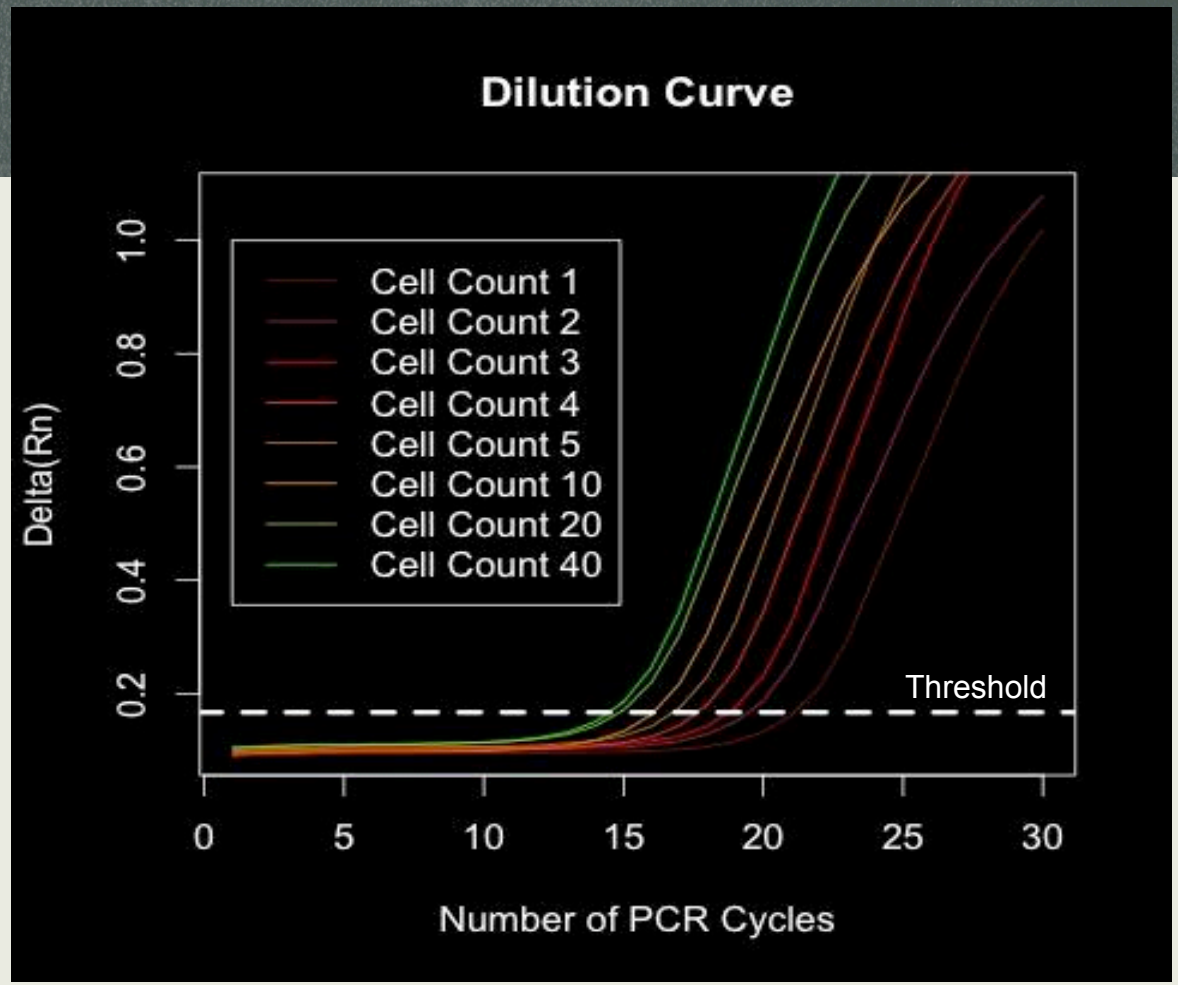
Our Study

Determining Relationship

- Samples:
  - Different sample cell count in sets of 12 for each gene
    - 1, 2, 3, 4, 5, 10, 20, 40

  ↑Sample Cell Count ↓C_t Value

- $X*2^{Ct}=Y$
  - X: Sample Cell Count
  - Y: Relative Fluorescence Threshold



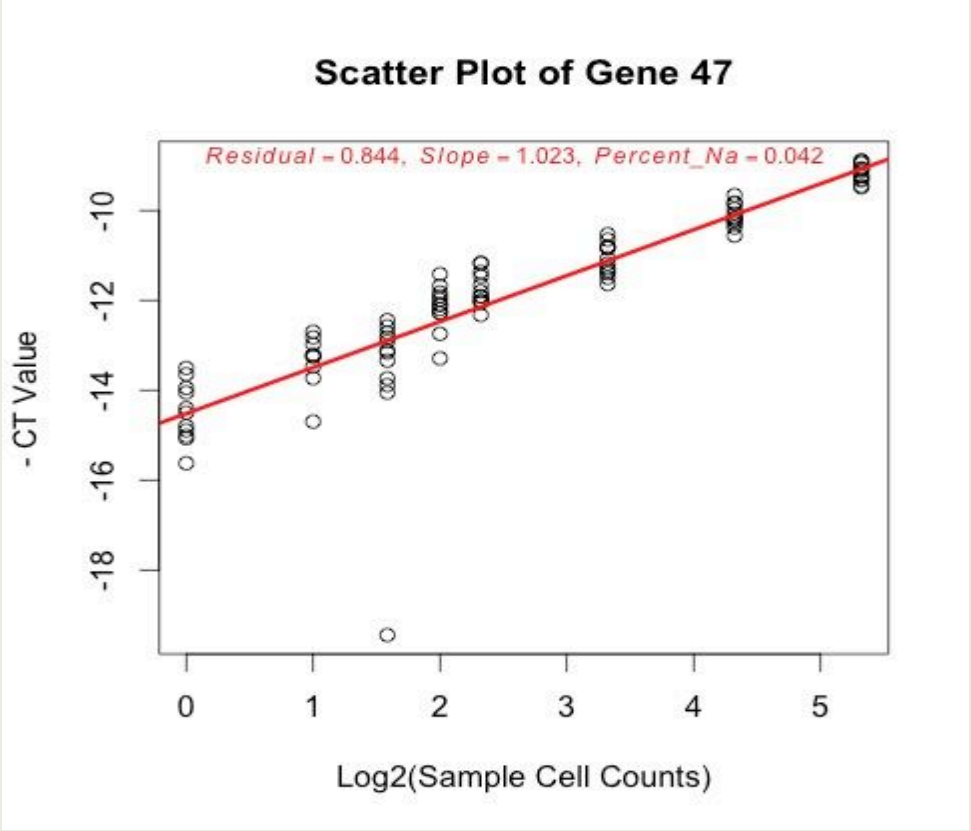
Dilution Curve for Gene 1
Relative Fluorescence (Delta Rn) vs. PCR Cycle Number

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(Sample Cell Count)-a= - C_t
  - "a" is a relative constant



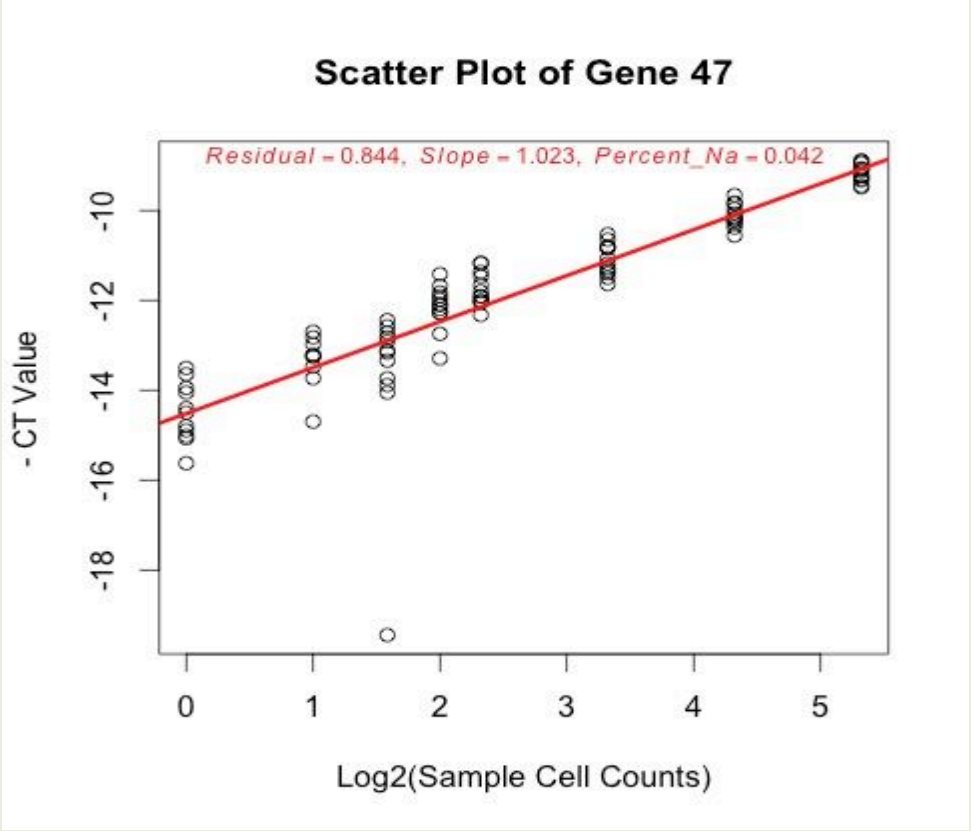
Standard Curve
-C_t Value vs. Log_2(Sample Cell Count)

Our Study

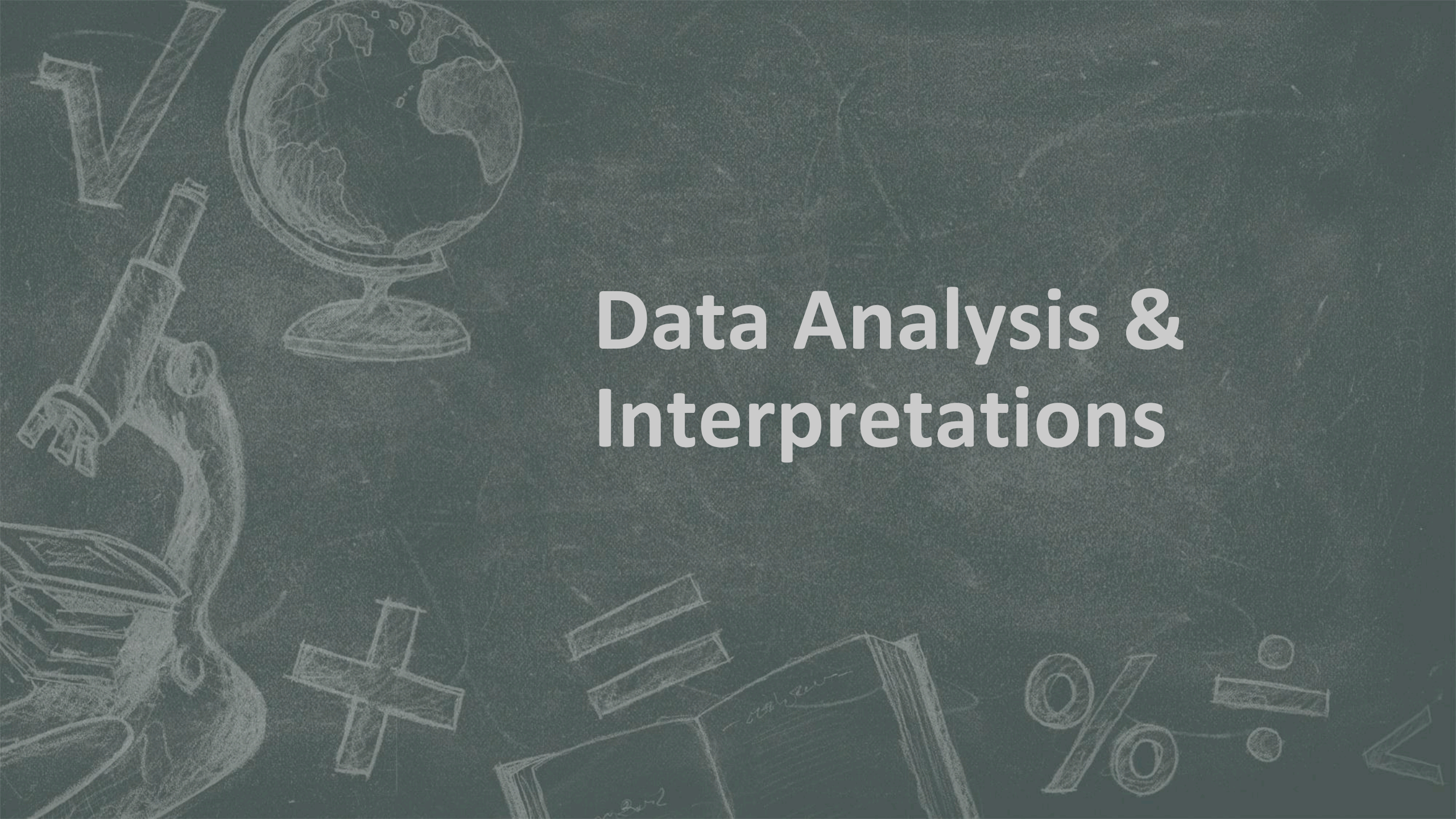
Examining Data

Ideal Plot:

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



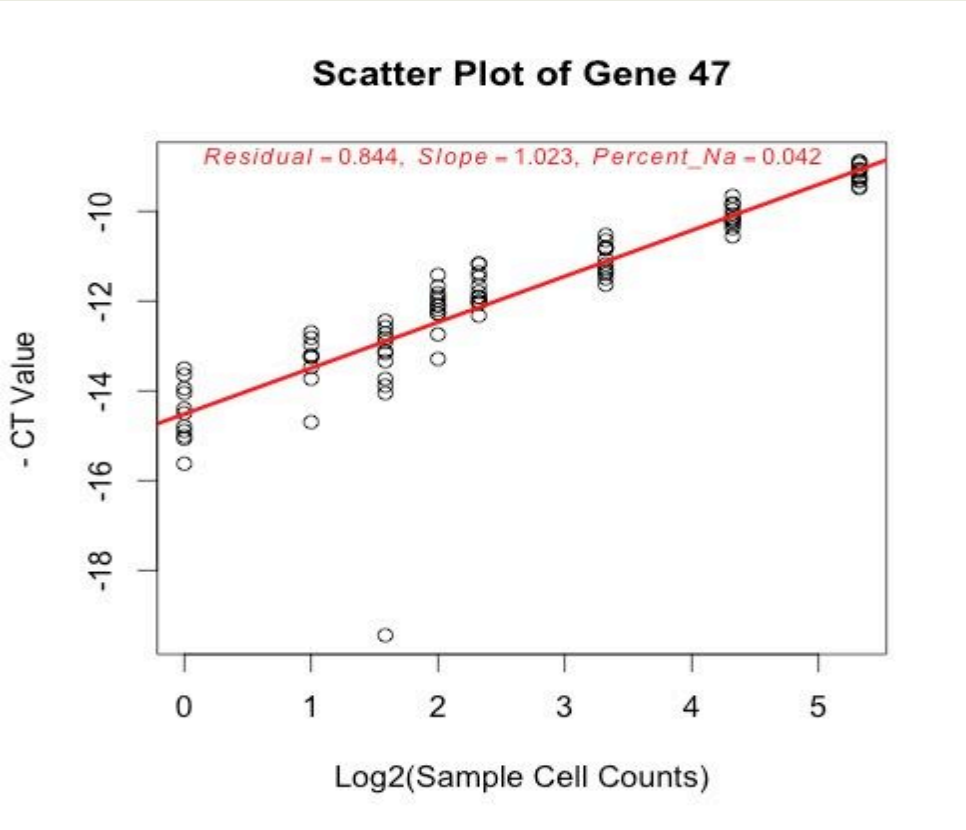
Standard Curve
$-C_t$ Value vs. Log_2(Sample Cell)



Data Analysis & Interpretations

Data Analysis & Interpretations

Observed Results



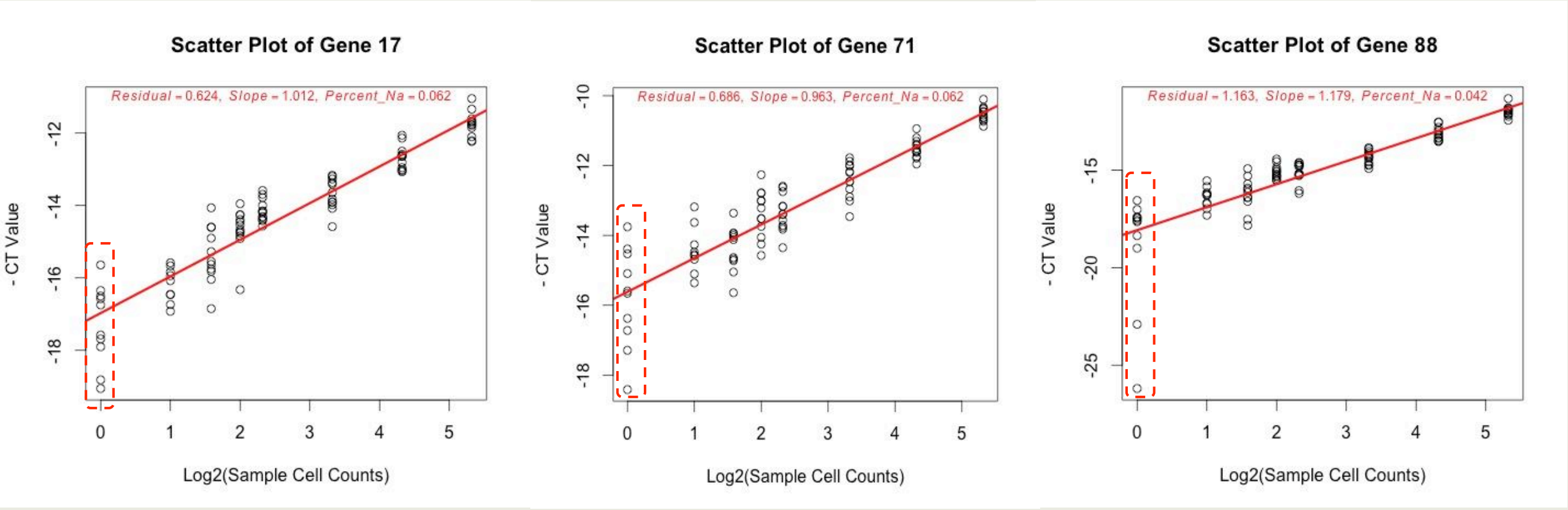
Interpretation

- Slope ≈ 1.023
  - Good accuracy
- Residual error ≈ 0.544
  - Good precision
- % NA ≈ 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

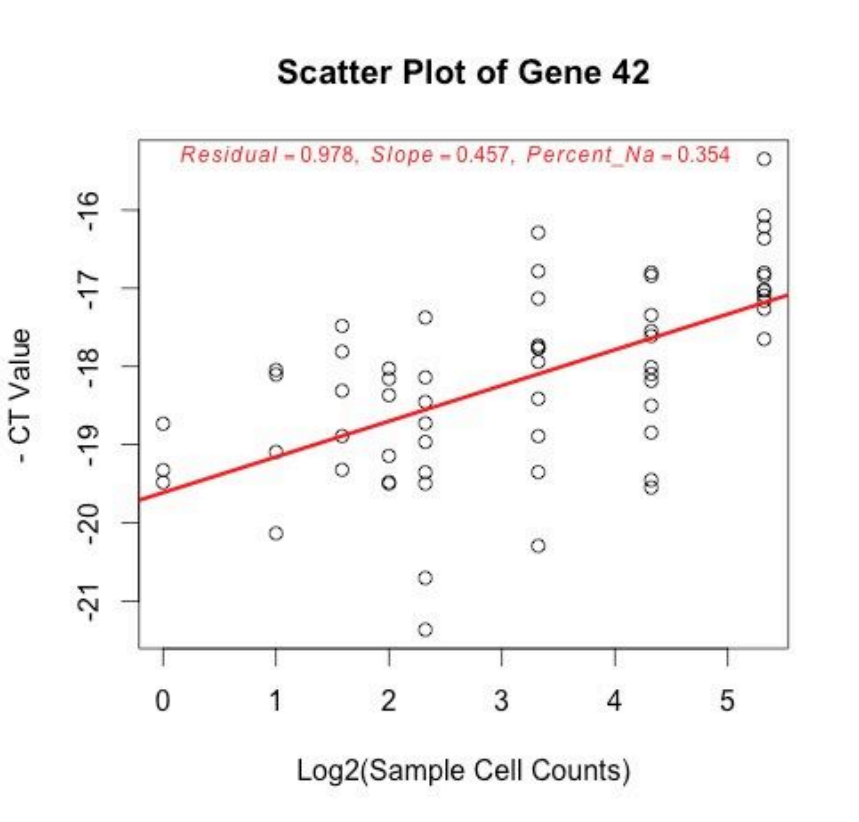
Data Analysis & Interpretations

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



Data Analysis & Interpretations

Observed Results



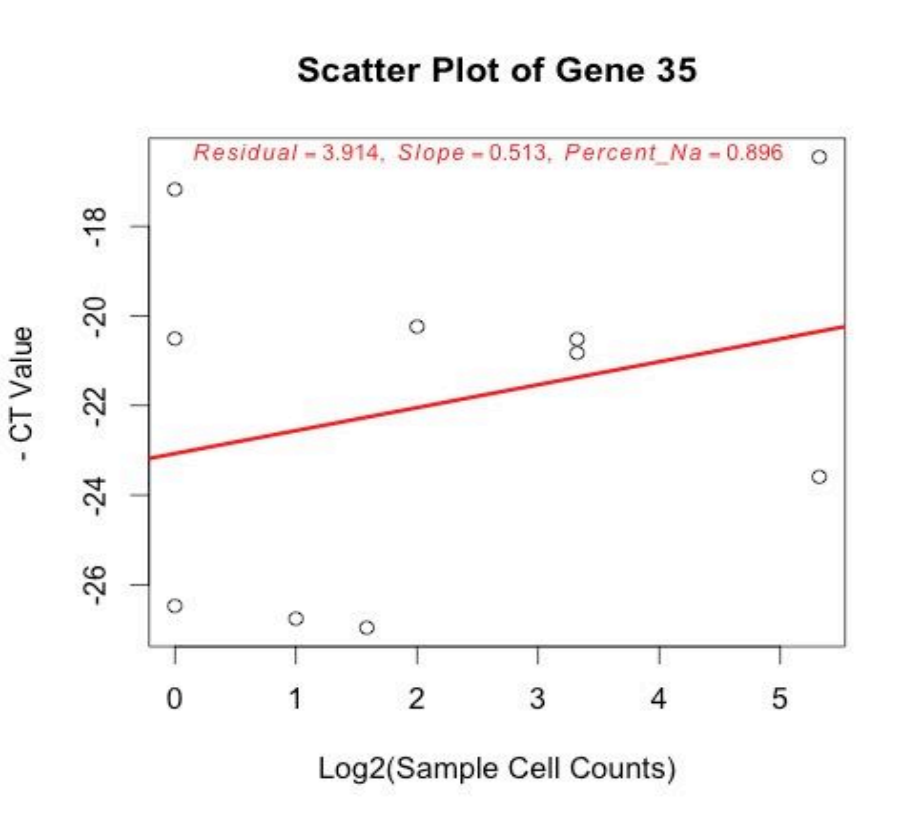
Interpretation

- Slope ≈ 0.457
  - Low accuracy
- High residual error (≈0.978)
  - Lacks precision
- % NA ≈ 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



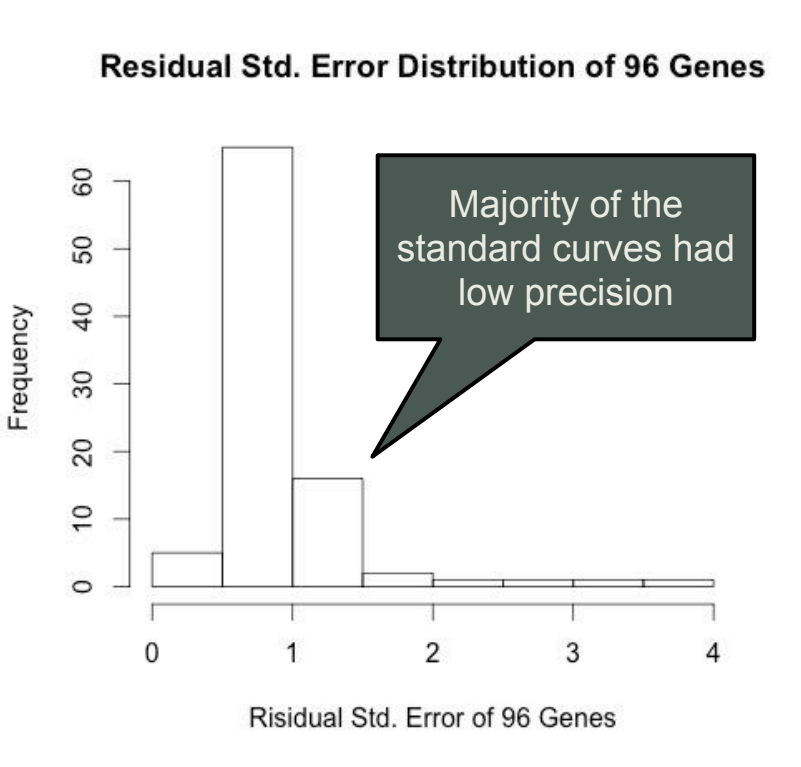
Interpretation

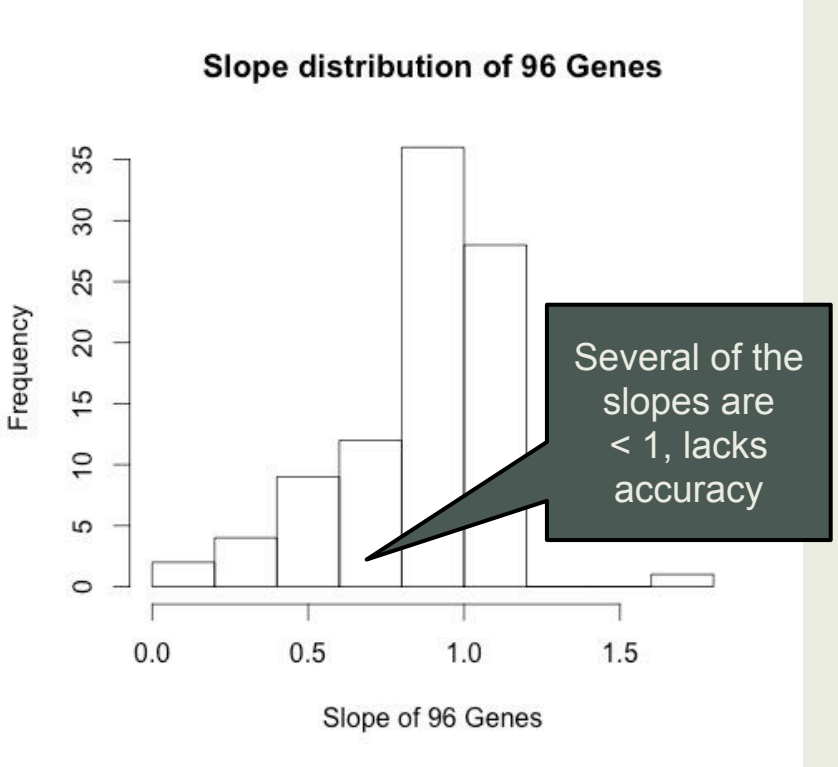
- Slope ≈ 0.513
  - Not close to one
  - Lacks accuracy
- High residual error ≈ 3.914
  - Lacks precision
- Most of the data is NA

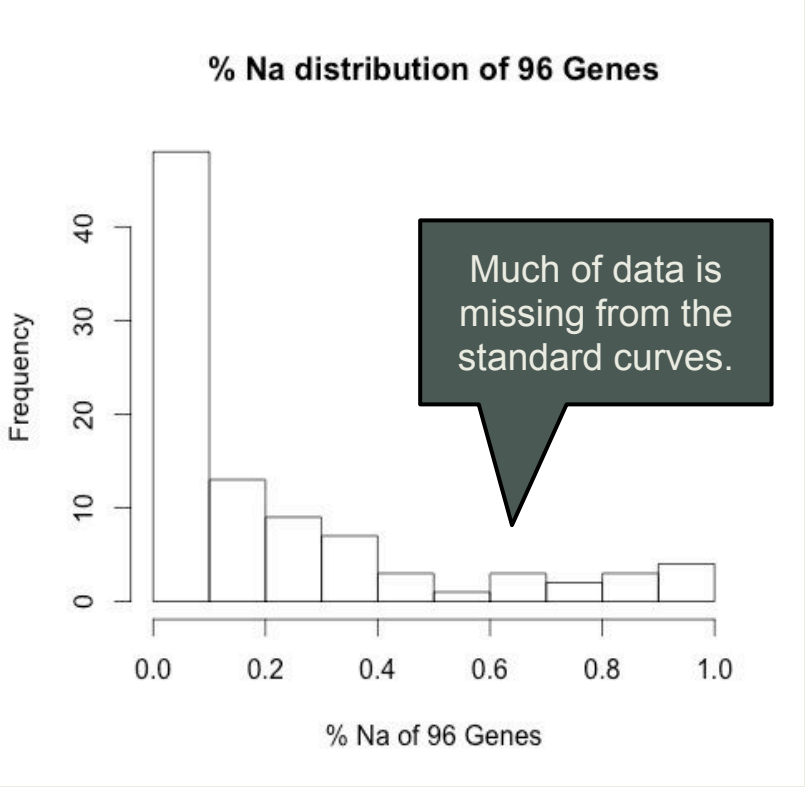
**Note: Gene is not expressed, or gene went undetected by RT-PCR machine when 100% of the data is missing, i.e. Gene 54*

Data Analysis & Interpretations

Distributions of our data's statistics





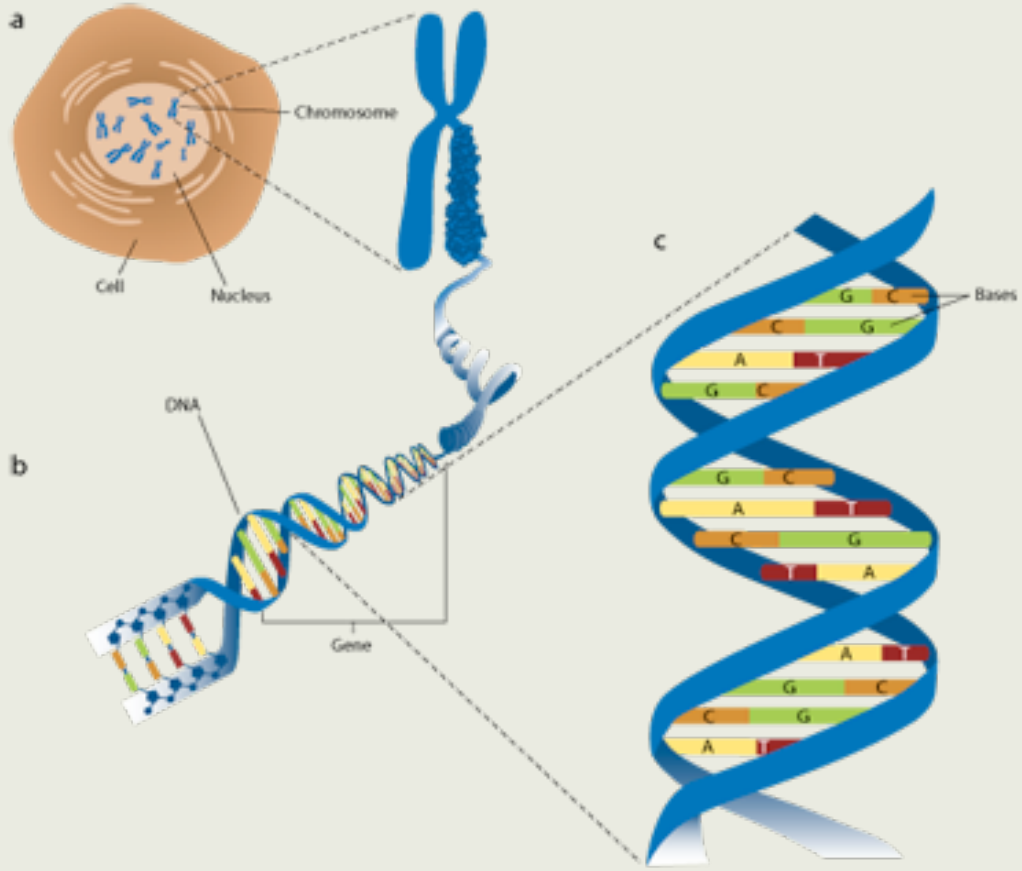


Data Analysis & Interpretations

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

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



down-syndrome.org

Future Works

Future Works

Need for a Statistical Model

- The △ Rn 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

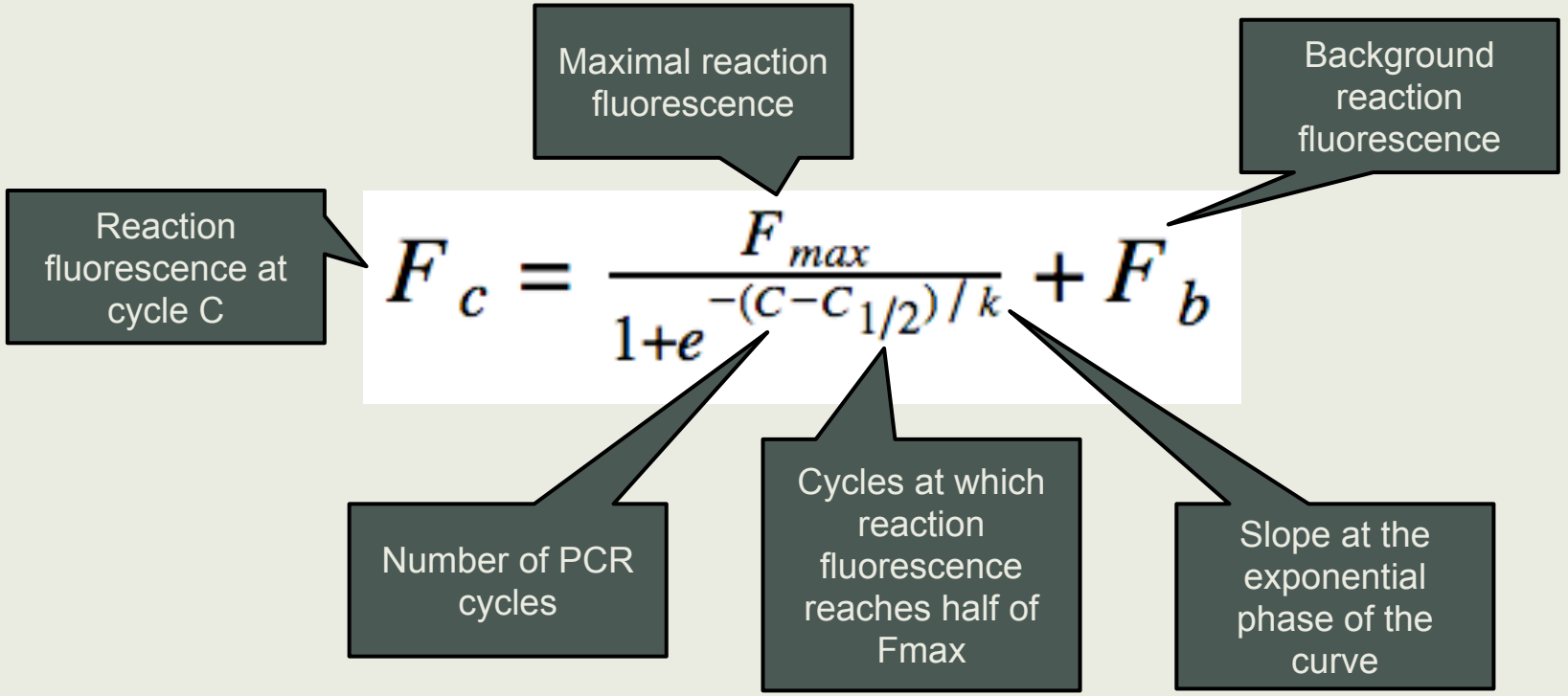
$$X\, a^{Ct} = Y$$

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

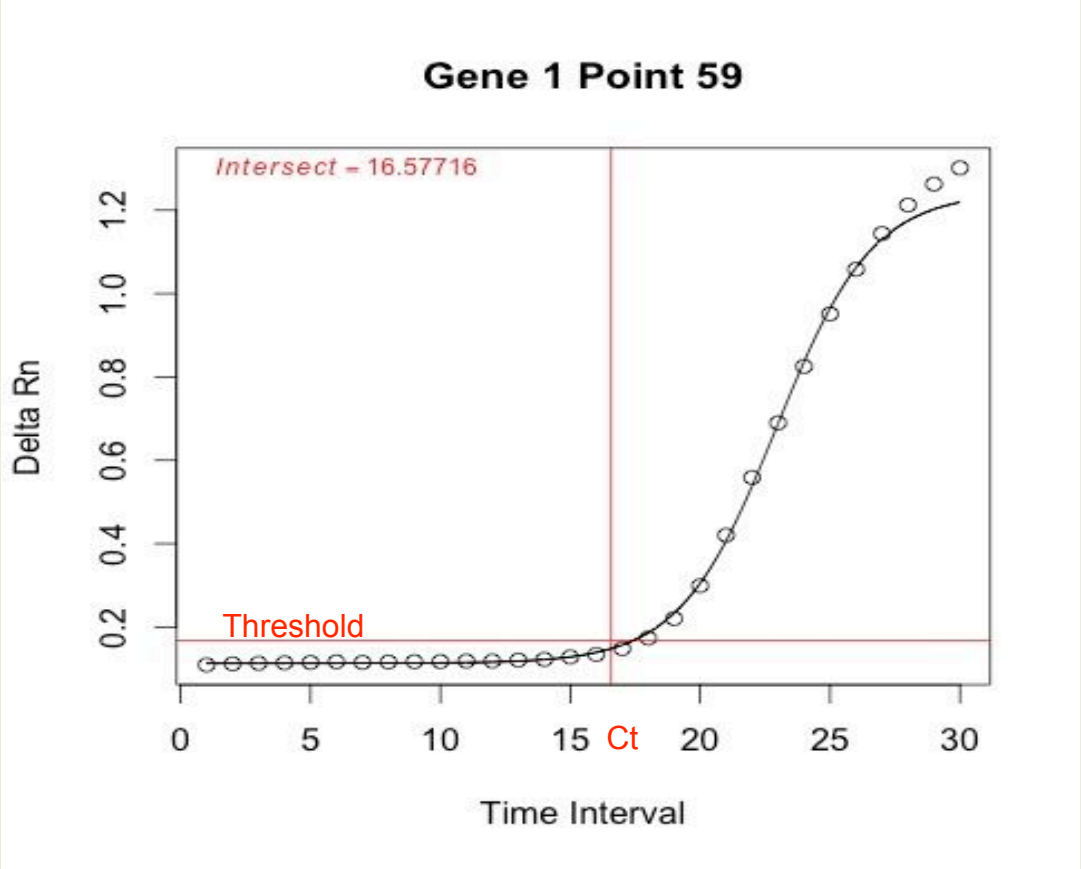
Future Works

Sigmoidal Function



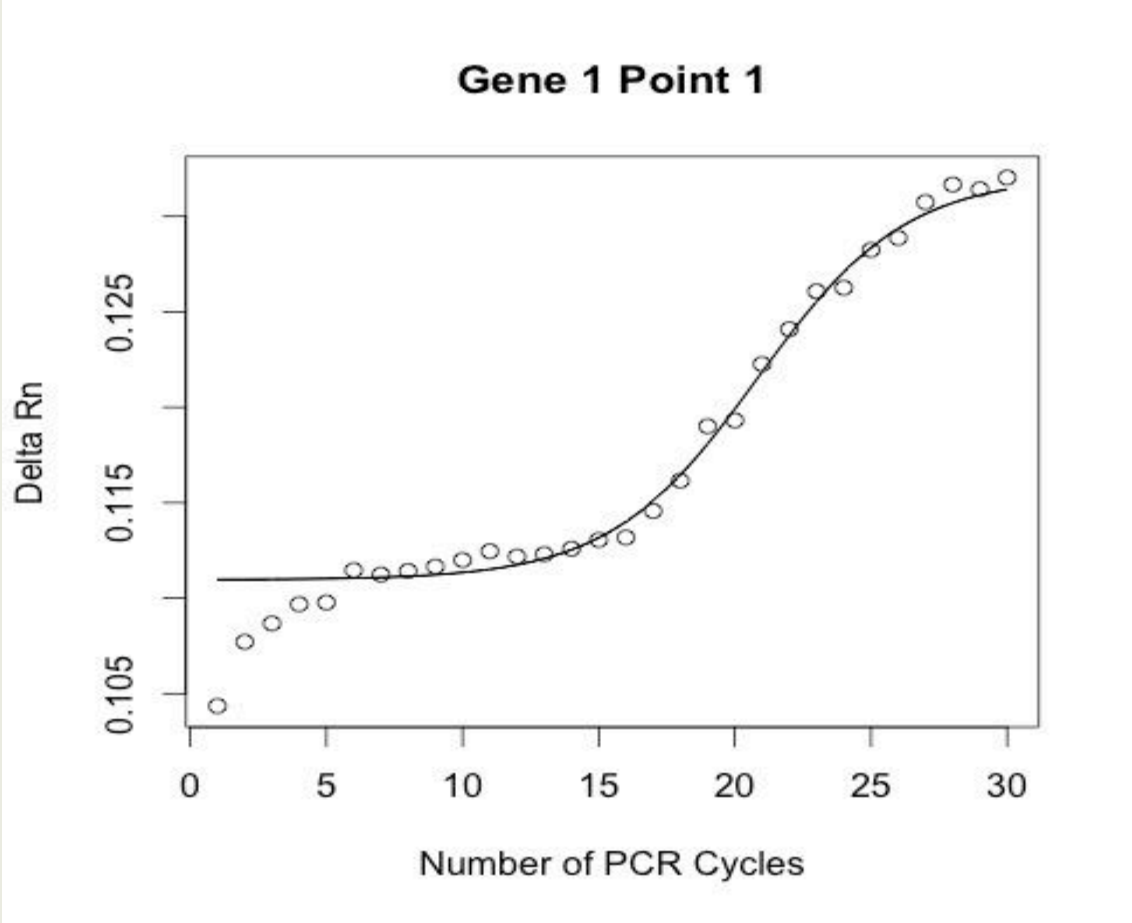
Future Works

Sigmoidal Curve Example



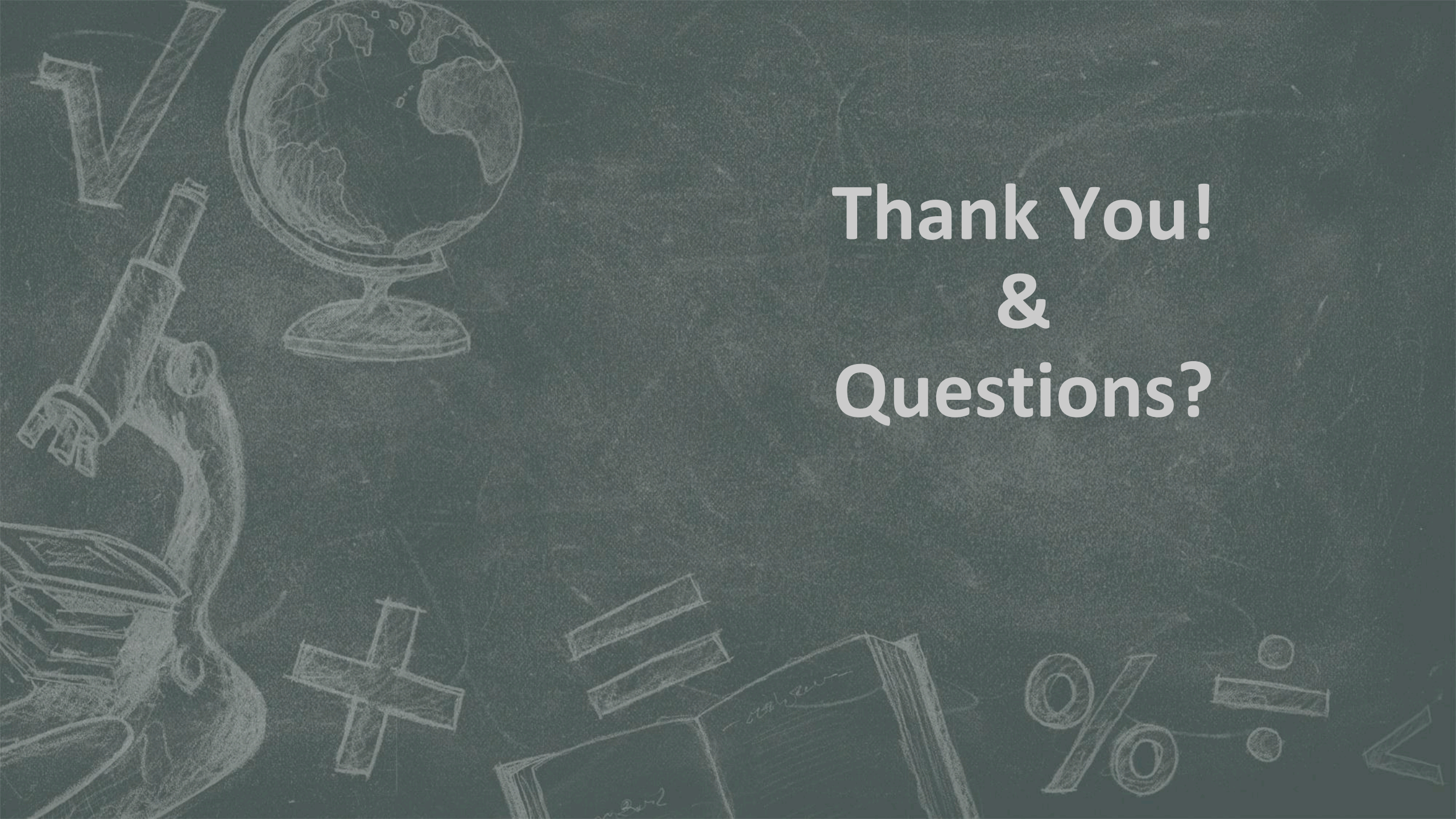
Future Works

Sigmoidal Curve Example



Acknowledgements

Dr. Rafael Irizarry
Stephanie Chan
Dr. Matt Cefalu
Summer Program In Quantitative Sciences at HSPH
PubMed
Makers of R
Jake Conway



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., Chattopadyay, 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.