

# WHAT WOULD TREX DO?

From Experimental Design to Analysis, the TREX Approach



# EXPERIMENTAL DESIGN

- What are my research goals?
- Where are my samples coming from?
  - FFPE
  - Clinical
  - Fresh Tissue/Cells
  - FACS Sorted Cells
- How much RNA will I have?
- What is the expected quality of that RNA?
- How many replicates do I need?
- What is the data analysis going to look like?



Cornell University

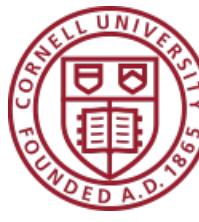
# GENE EXPRESSION ANALYSIS

Control

Treatment

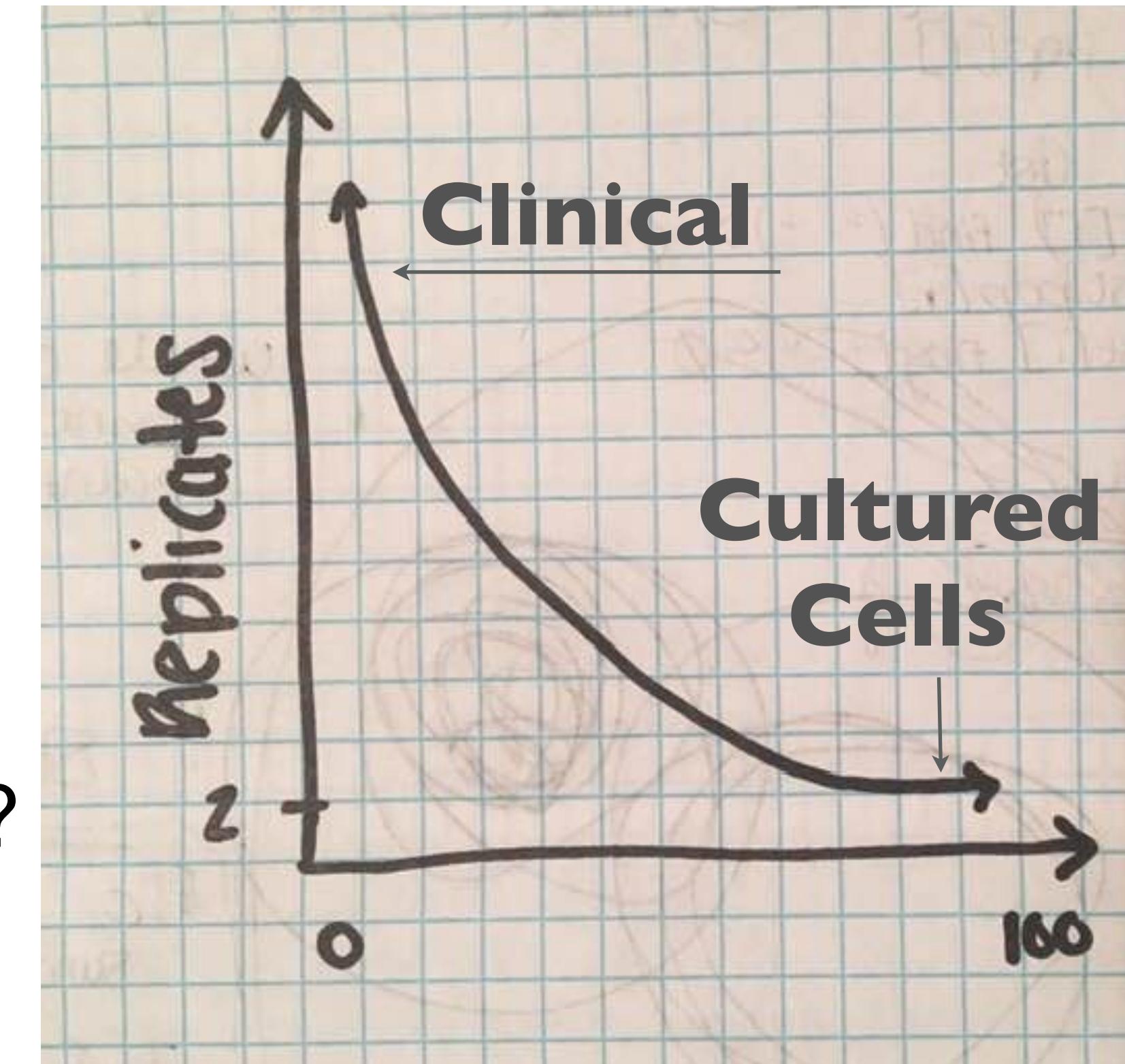
Gene Expression  
Variation

Gene Expression  
Variation



# EXPERIMENTAL DESIGN

- What are my research goals?
- Where are my samples coming from?
  - How much RNA will I have?
  - What is the expected quality of that RNA?
- How many replicates do I need?
  - What is the data analysis going to look like?
- How many samples do I need?
- How much money do I have?

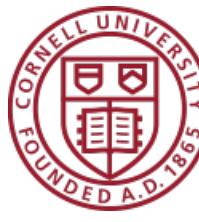


**Biological > Technical**



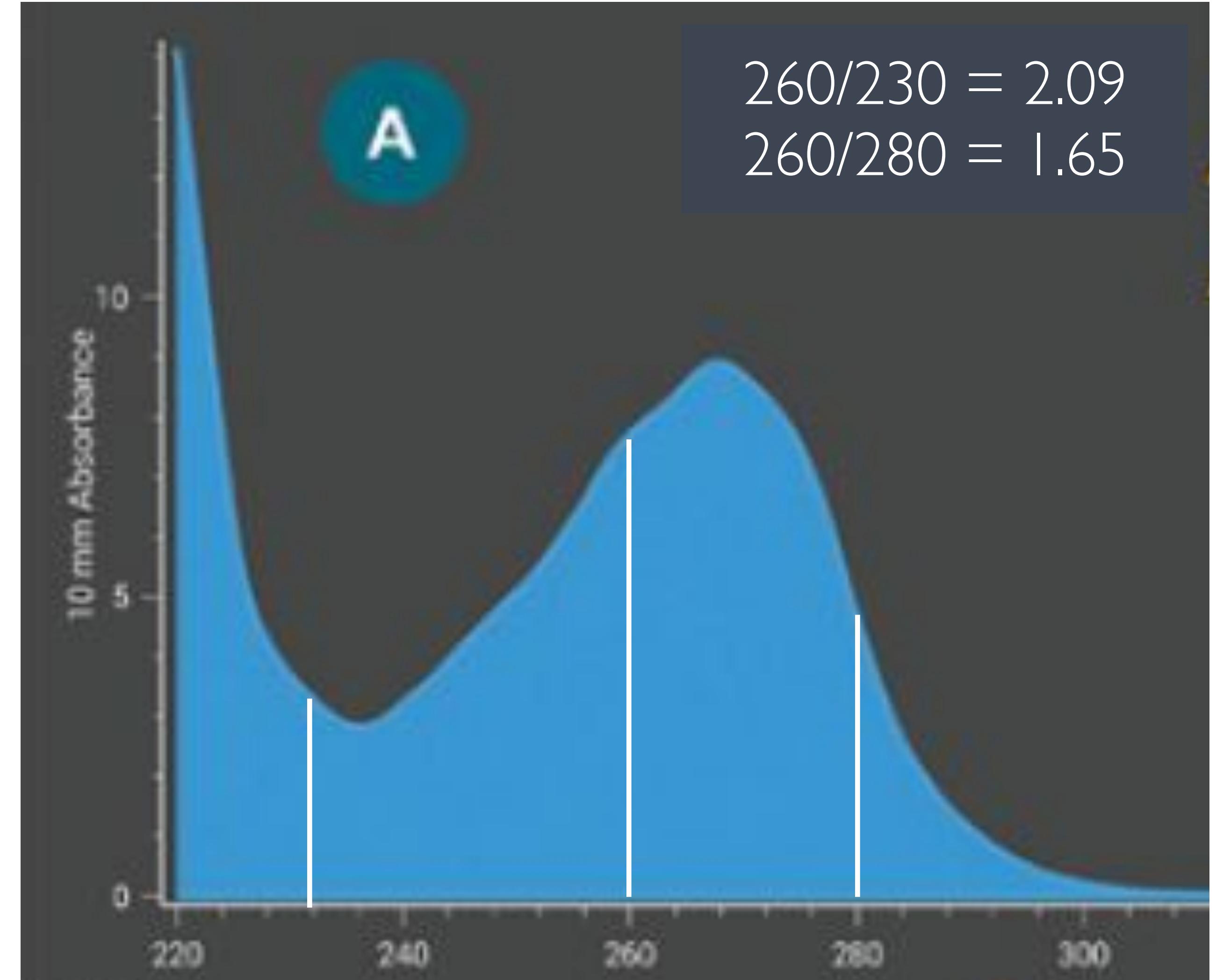
# SAMPLE PREPARATION/ EXTRACTION

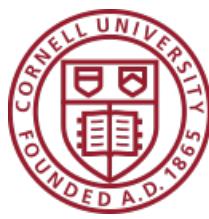
- How did you extract your RNA?
- What does the RNA QC look like?
  - **Nanodrop:** looks at chemical impurities and quantity
  - **Fragment Analyzer:** Looks at RNA integrity
  - **Qubit:** quantity of material



# NANODROP

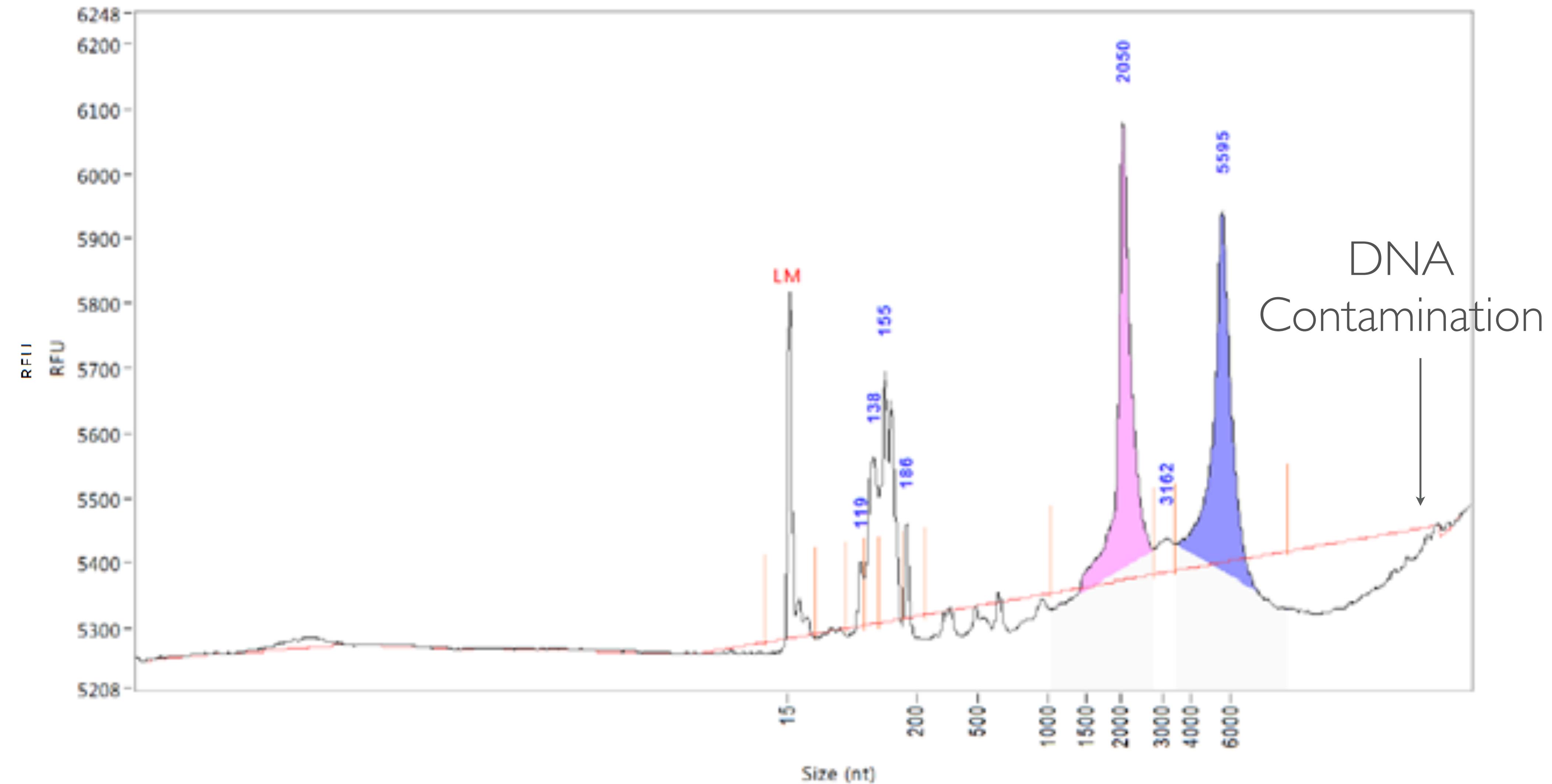
- Spectrophotometer
  - Quantity > 20ng/uL
  - Contaminants
    - Salts
    - Proteins
    - Phenolics
    - Carbohydrates/Sugars
  - 260/230 Ratio: ~2-2.2
  - 260/280 Ratio: ~1.8-2





Cornell University

# FRAGMENT ANALYZER





# QUBIT FLUOROMETER

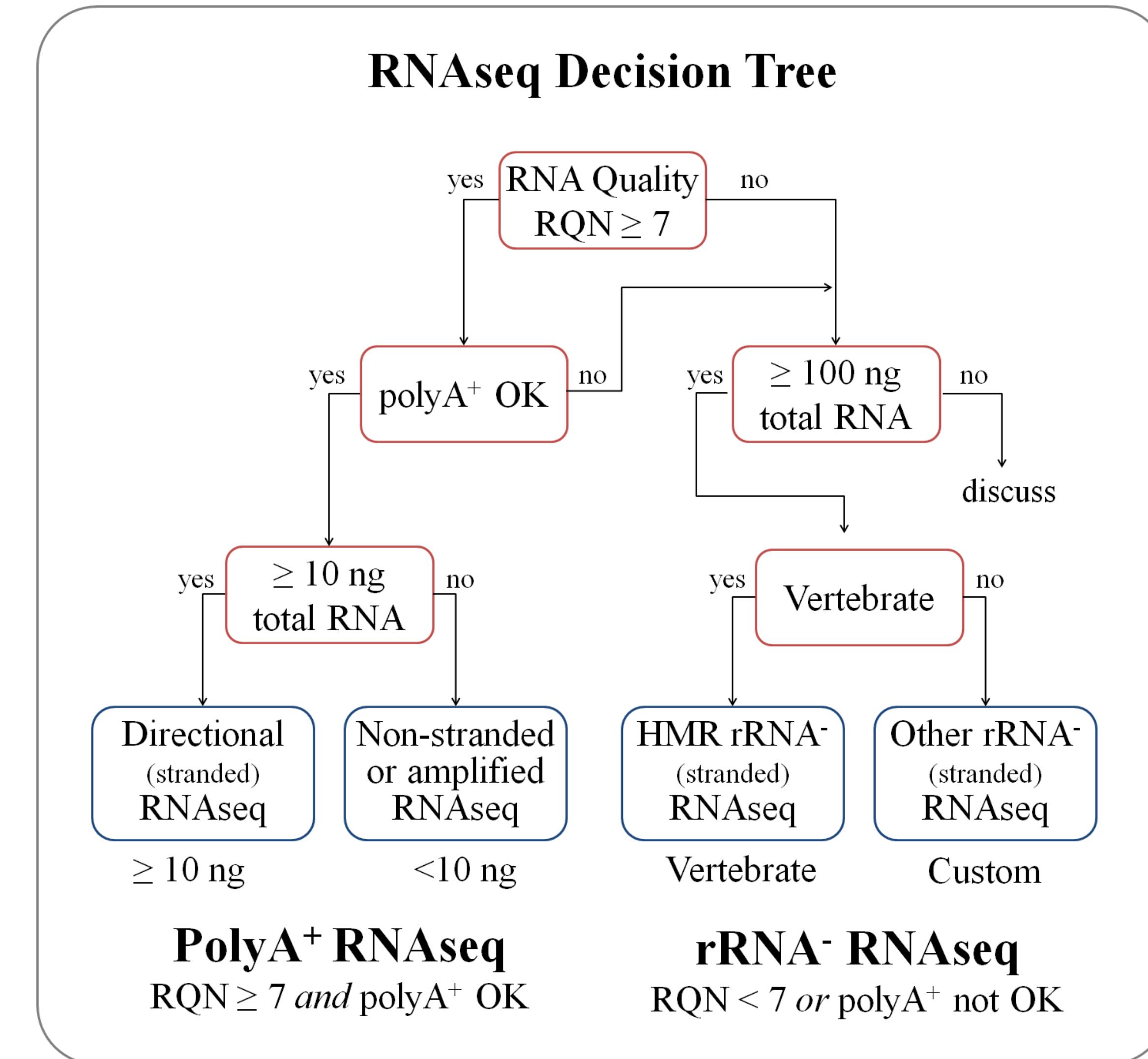
- RNA Quantity < 20
- Can also detect DNA contamination
- Free to use in Genomics Core

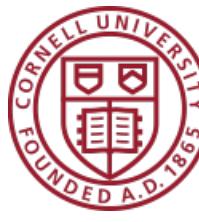




# RNASEQ LIBRARY PREP

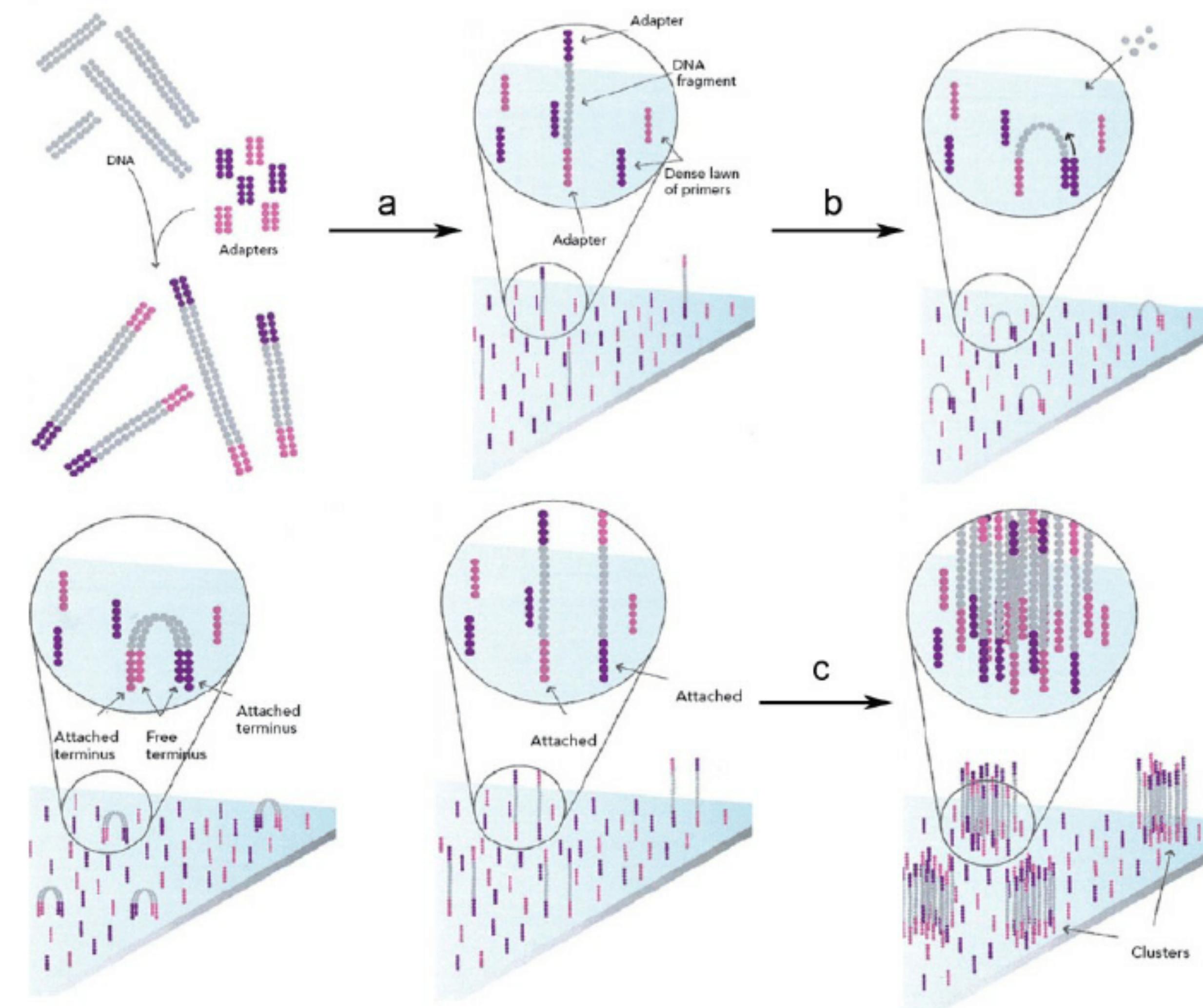
## NEB Next Ultra II RNASeq

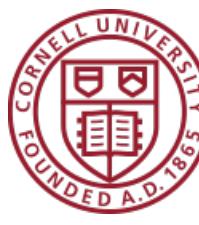




Cornell University

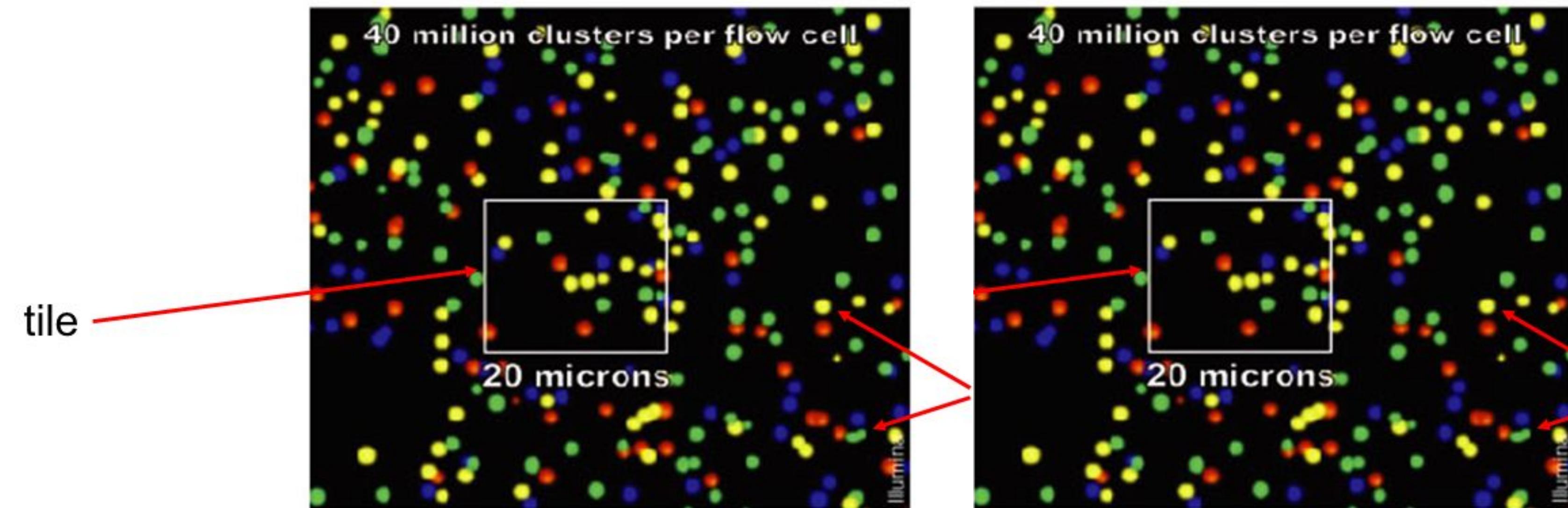
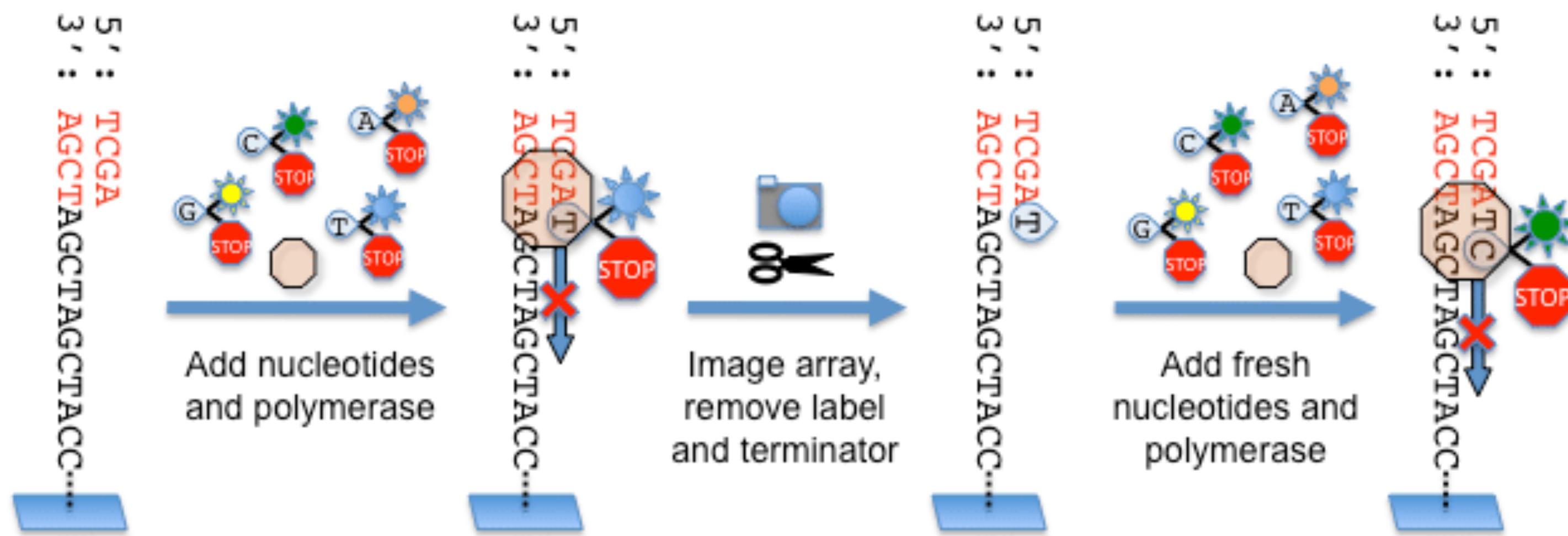
# ILLUMINA SEQUENCING

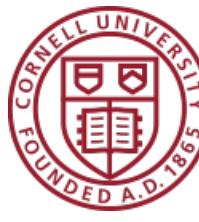




# ILLUMINA SEQUENCING

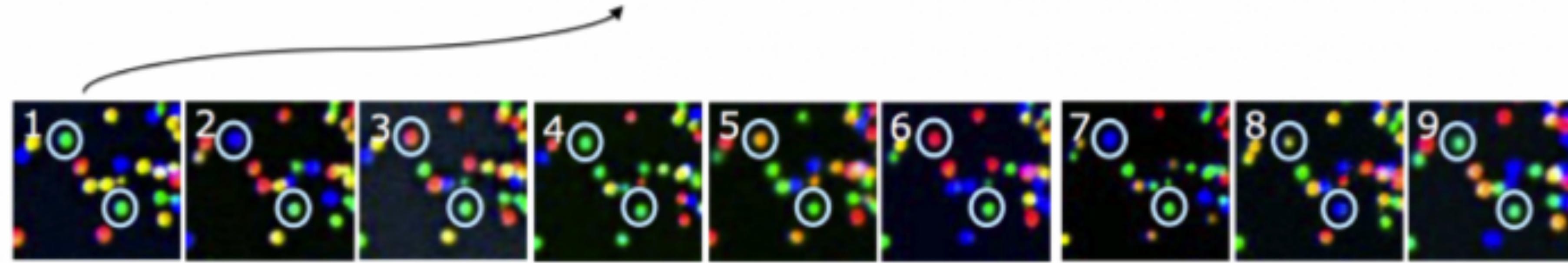
A





# ILLUMINA SEQUENCING

T G C T A C G A T ...



T T T T T T G T ...



# SEQUENCING

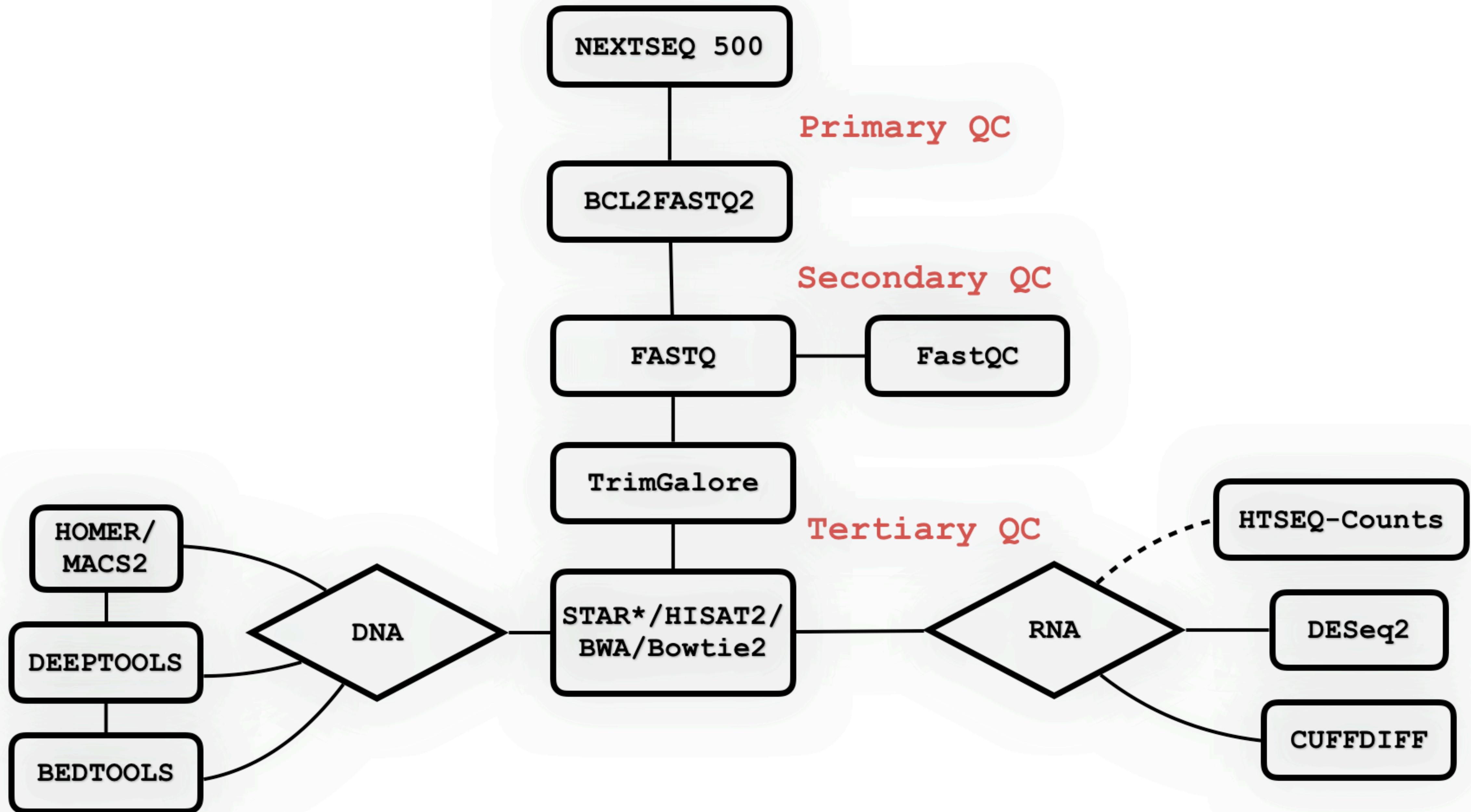
- What Read Length do you want?
  - Our standard is 75bp
  - smRNA ideal is 50bp
  - SNP detection: Paired End
  - Transcriptome Assembly Longest read possible
- How many reads do you need?
  - Our standard: 20m reads
  - smRNA standard: 10m reads
  - Isoform detection: much higher



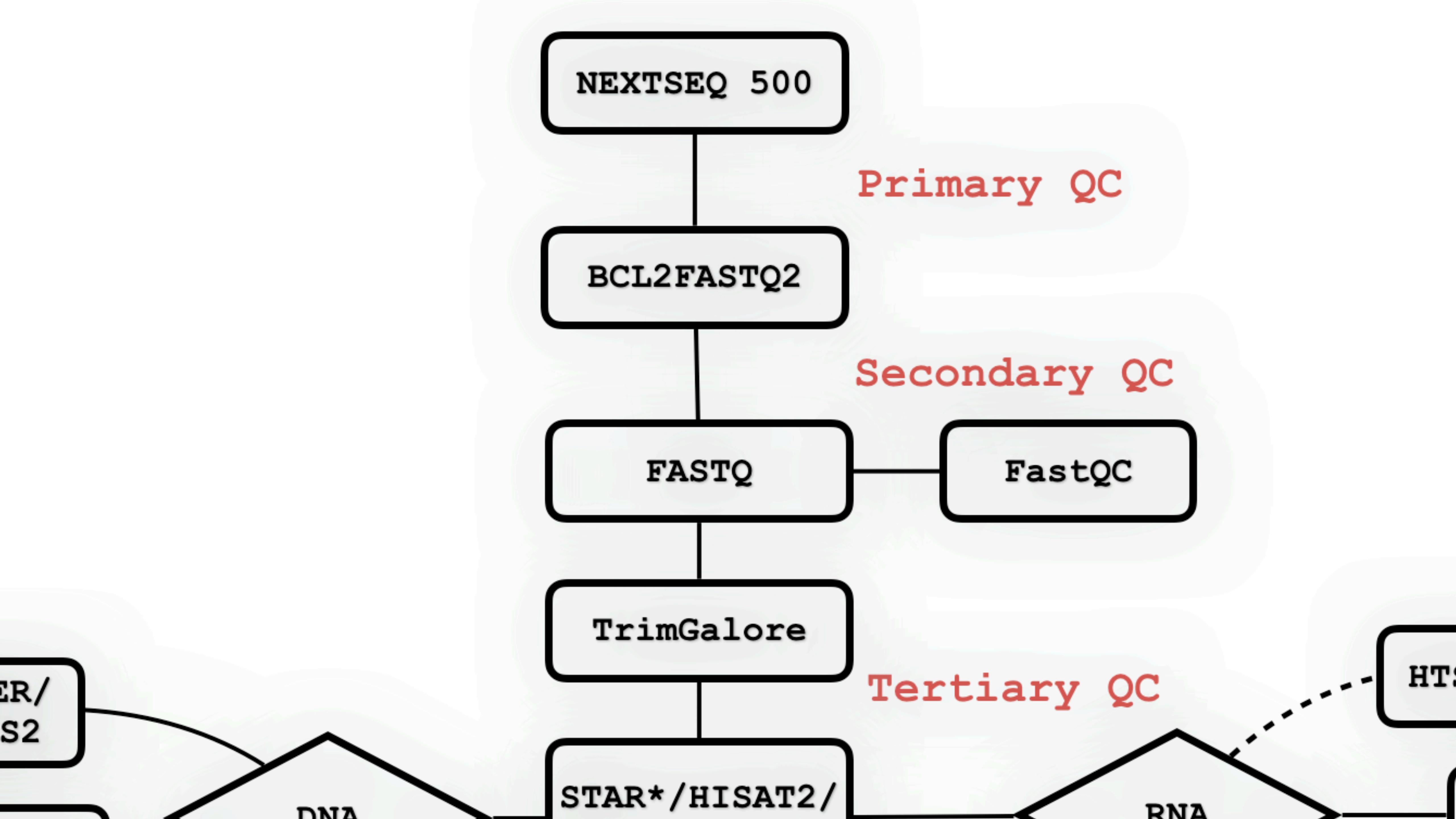
Cornell University

# Data QC

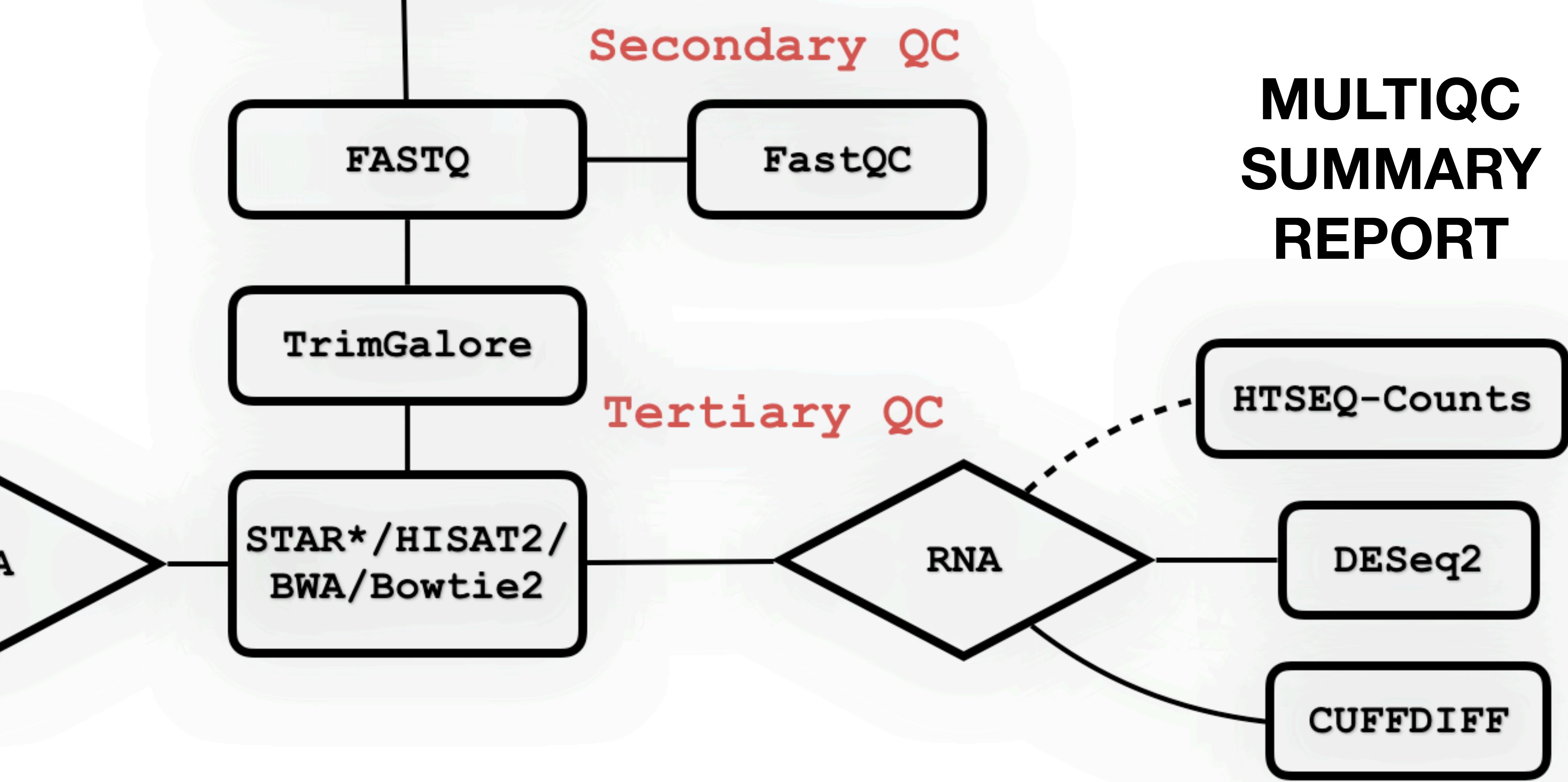




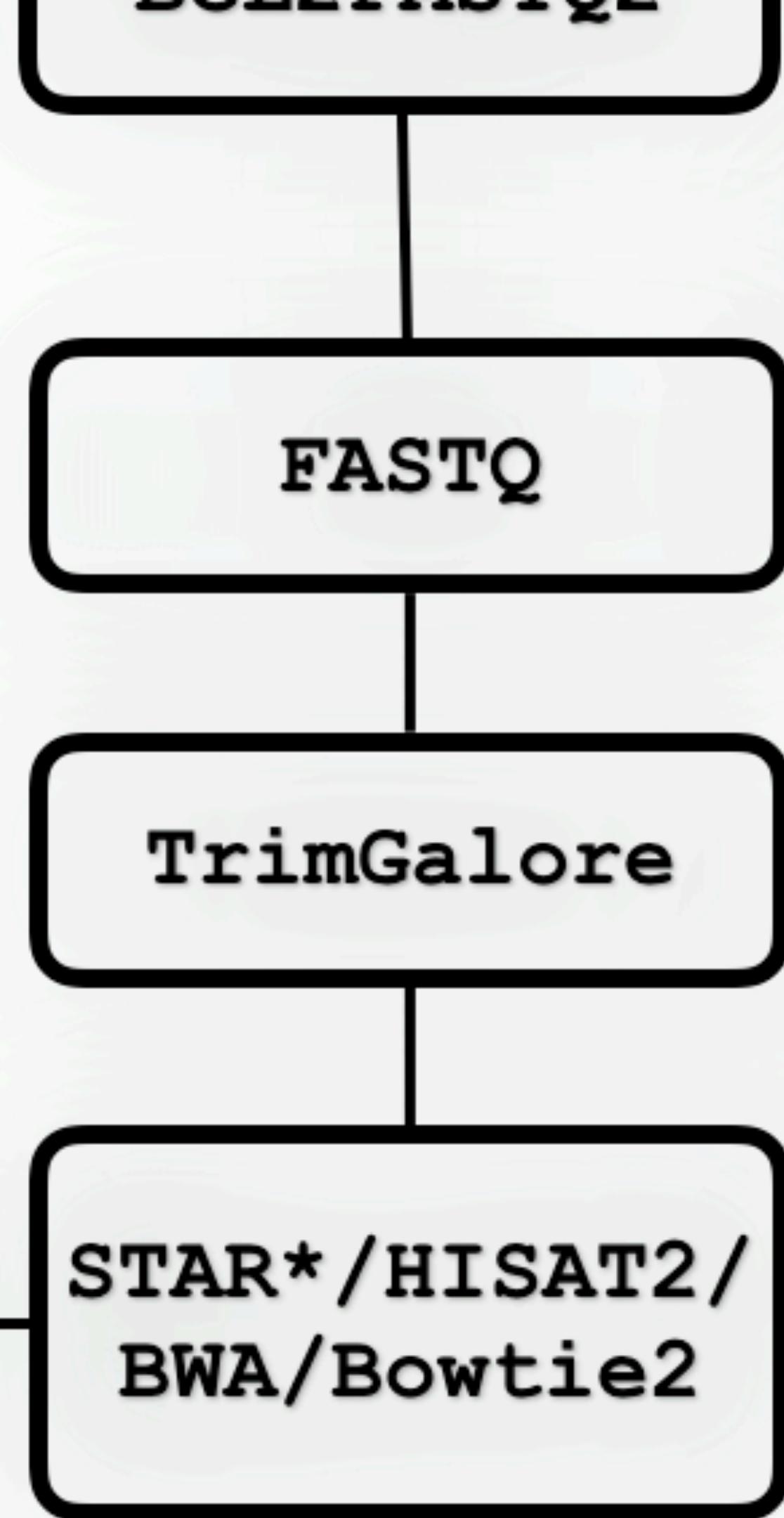
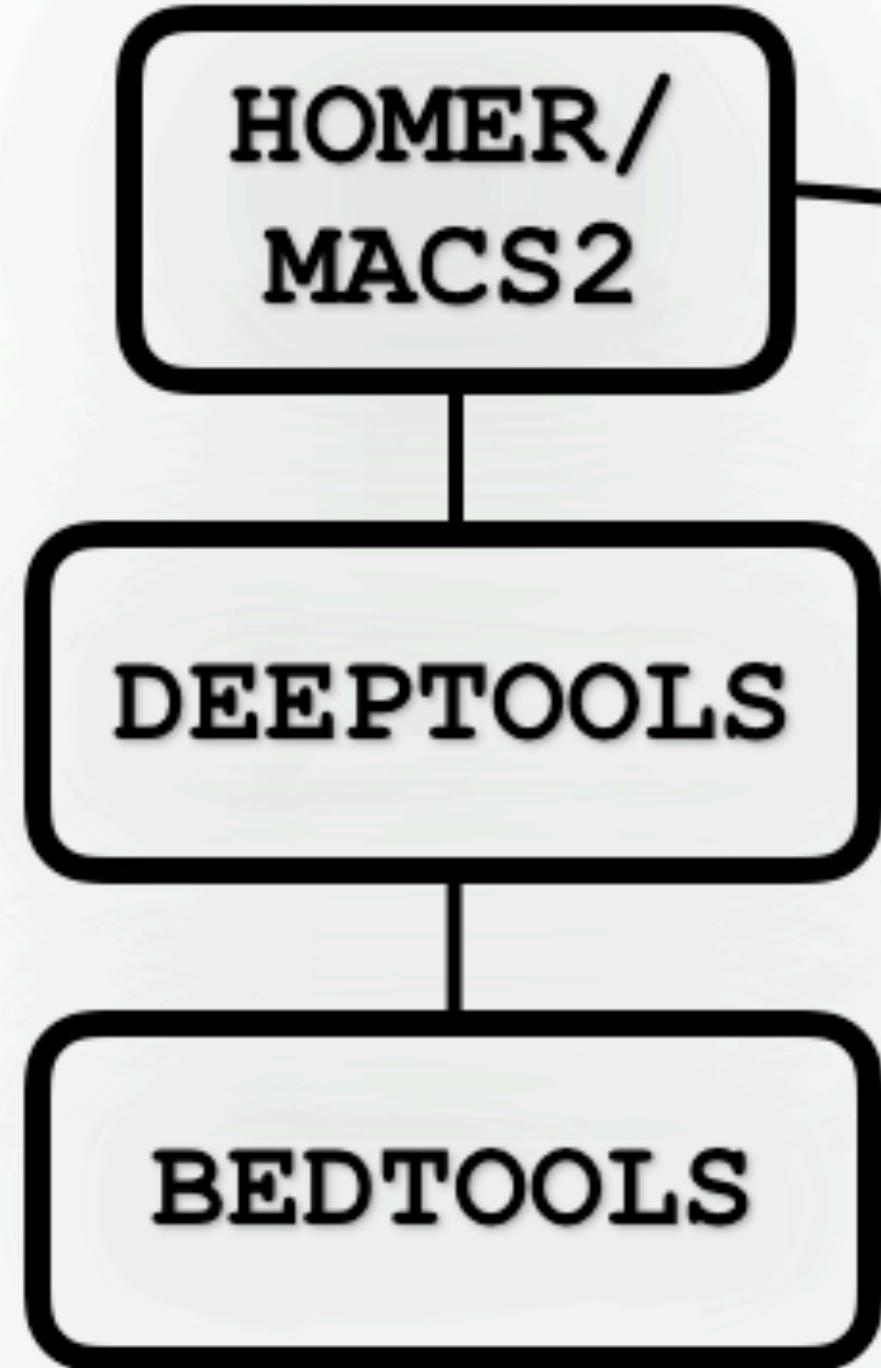
\*---quantMode FOR RNA SEQ READS



# MULTIQC SUMMARY REPORT

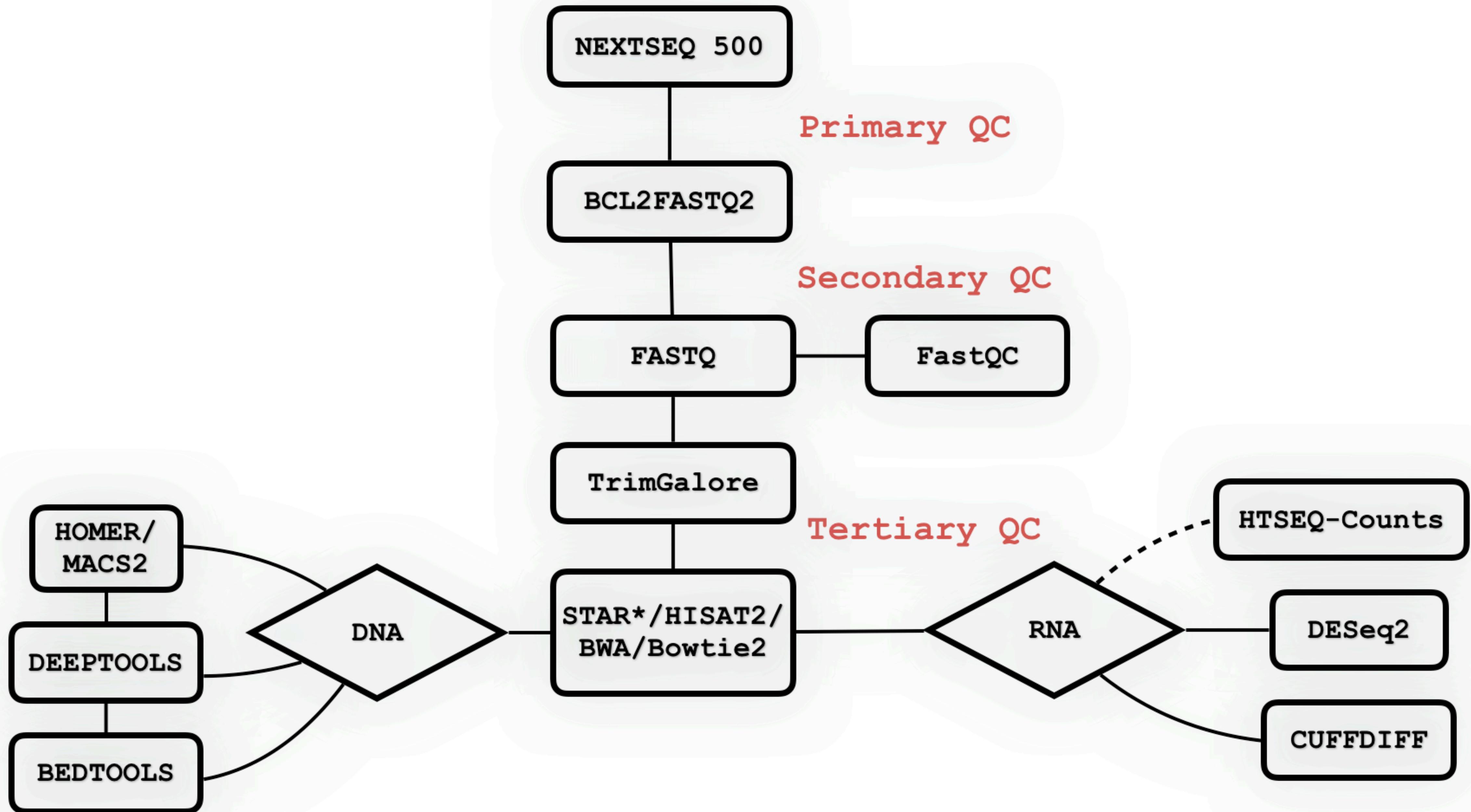


---quantMode FOR RNA SEQ READS

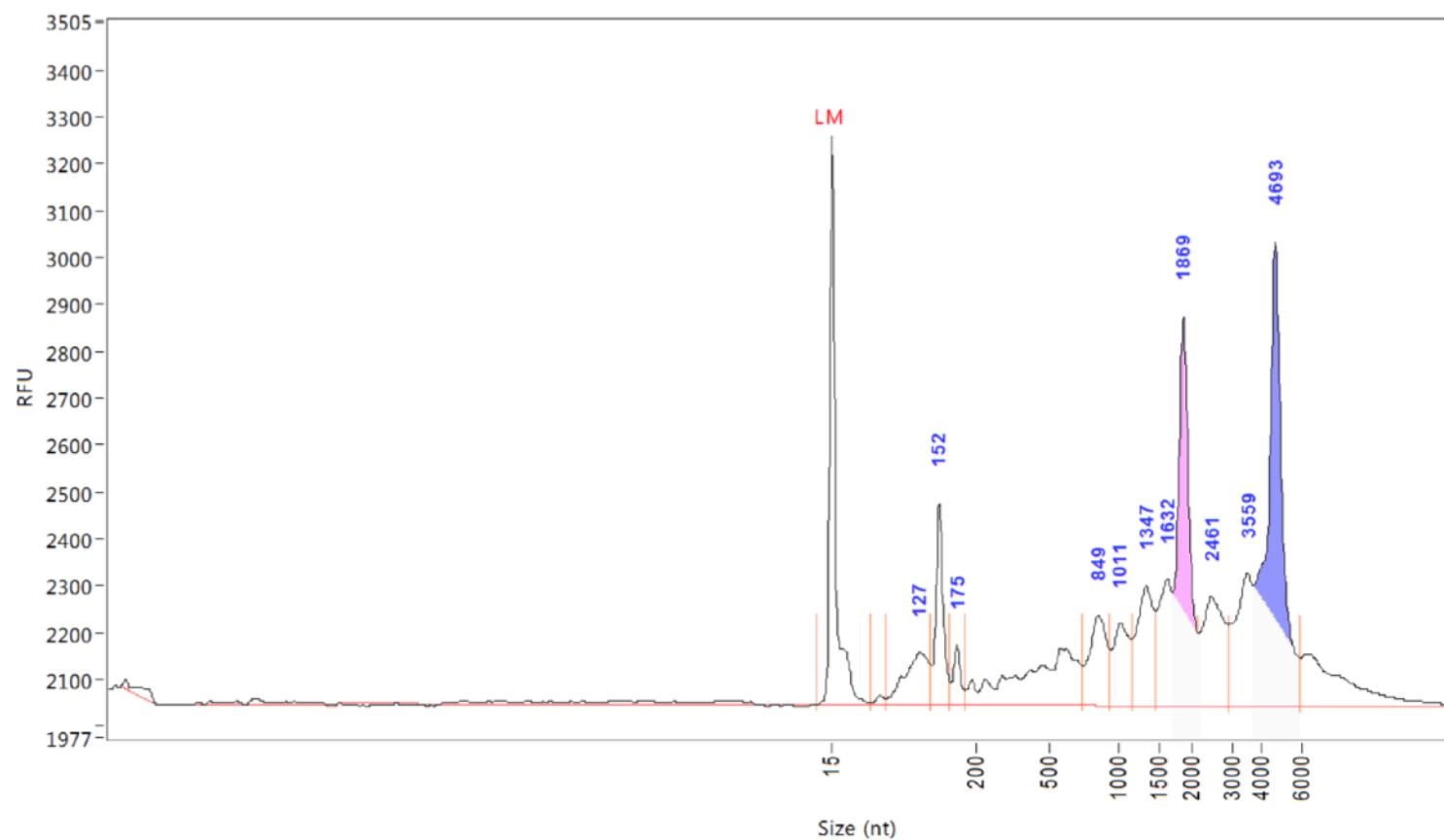


Secondary QC

Tertiary QC



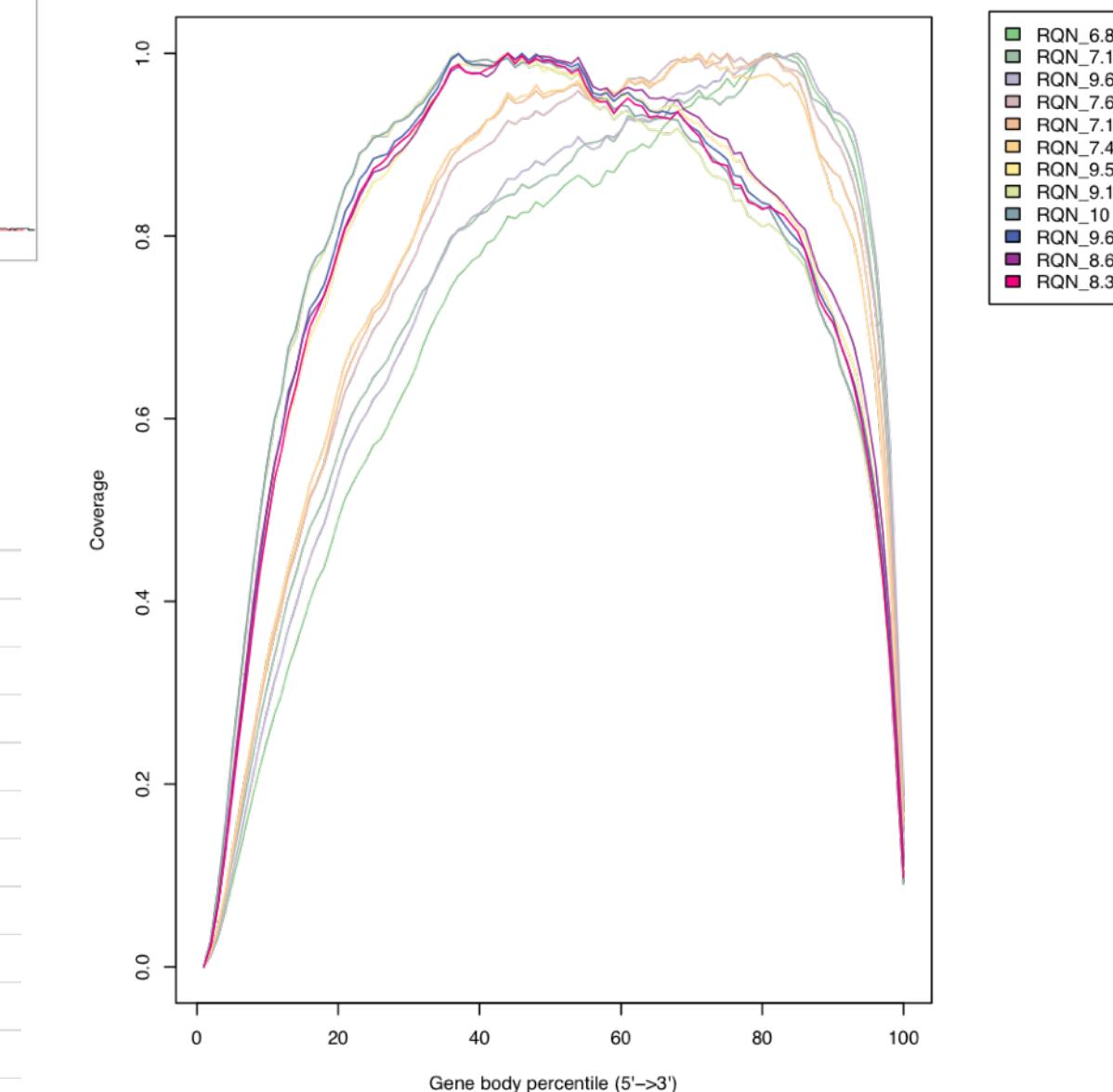
\*---quantMode FOR RNA SEQ READS



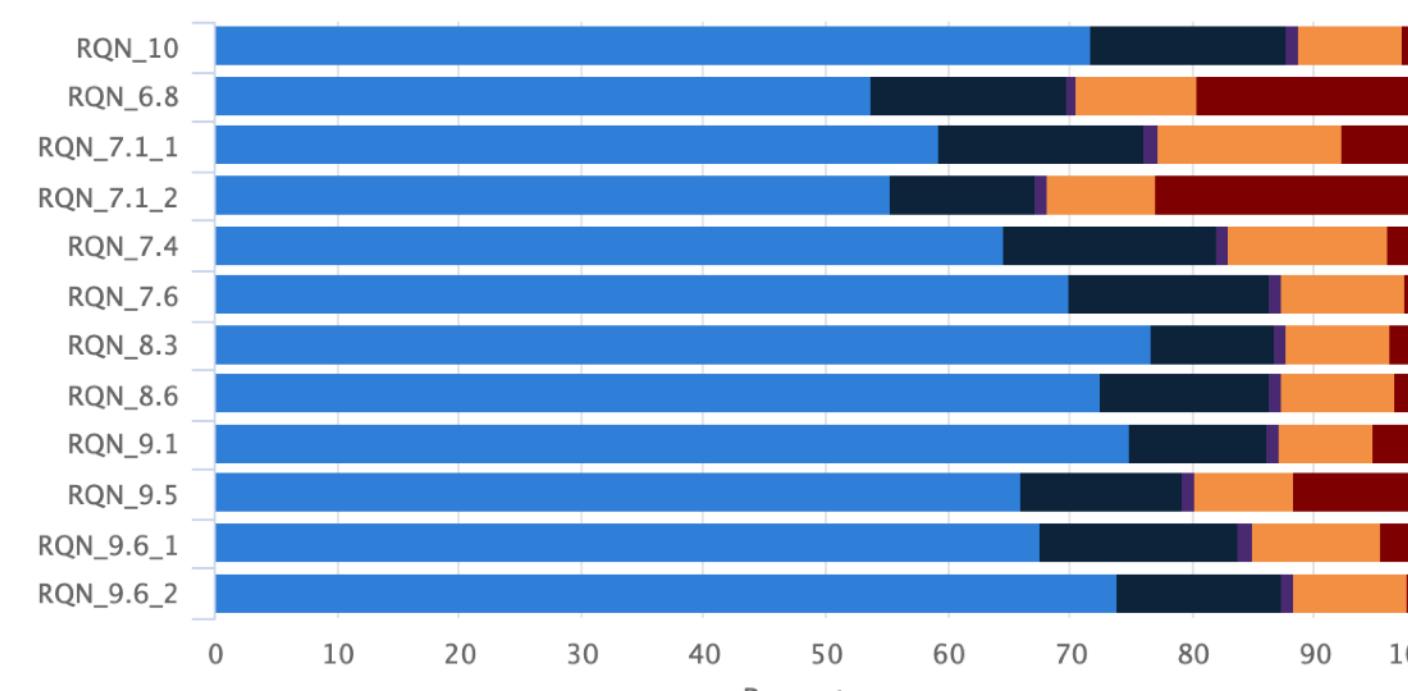
## General Statistics

 Copy table

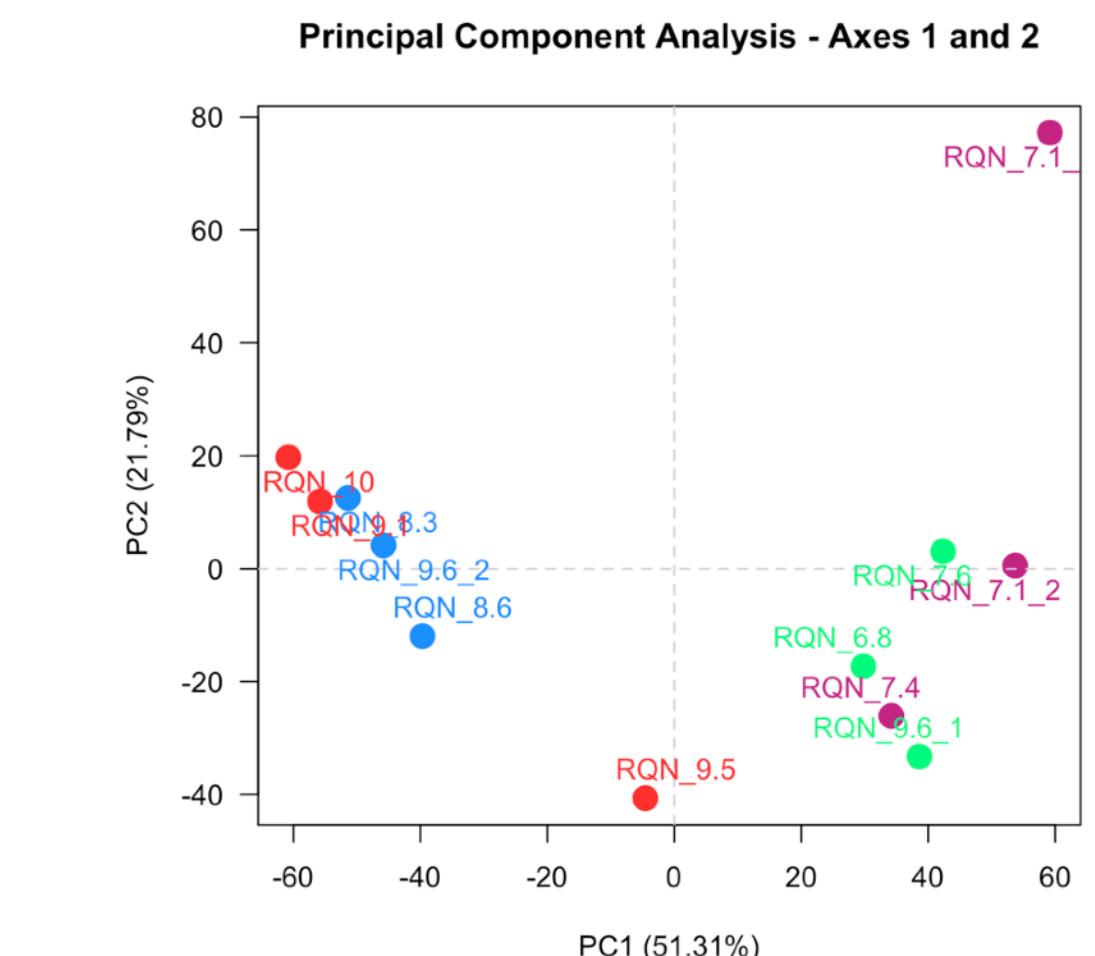
Sample Name	% Aligned ▾	M Aligned
RQN_10	88.7%	32.9
RQN_9.6_2	88.3%	29.6
RQN_8.3	87.7%	31.6
RQN_7.6	87.4%	32.1
RQN_8.6	87.3%	20.9
RQN_9.1	87.1%	28.8
RQN_9.6_1	85.0%	32.0
RQN_7.4	83.0%	29.9
RQN_9.5	80.2%	27.7
RQN_7.1_1	77.3%	26.6
RQN_6.8	70.7%	24.8
RQN_7.1_2	68.1%	23.9



## STAR Gene Counts



● Overlapping Genes    ● No Feature    ● Ambiguous Features    ● Multimapping  
● Unmapped





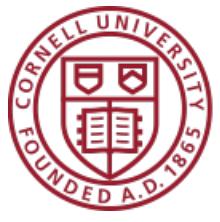
# Diagnostic Plots

- |---|
    - | BioAnalyzer Trace;
    - | MultiQC Report (Alignment Statistics);
    - | GeneBody Coverage;
  
  - |---|
    - | Principal Components Analysis;
    - | Hierarchical Clustering;
- Overall quality**
- Biological Signal**

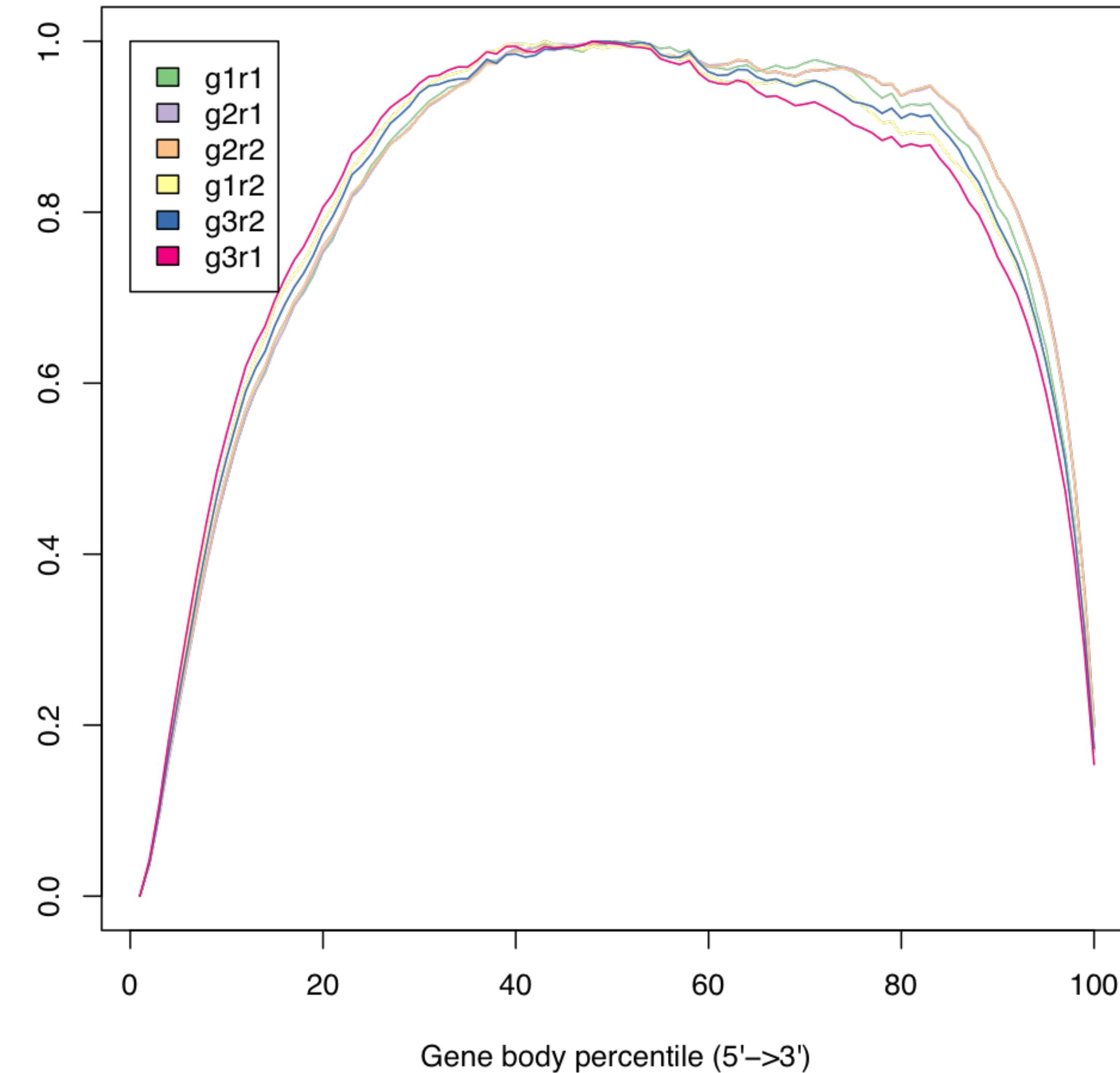
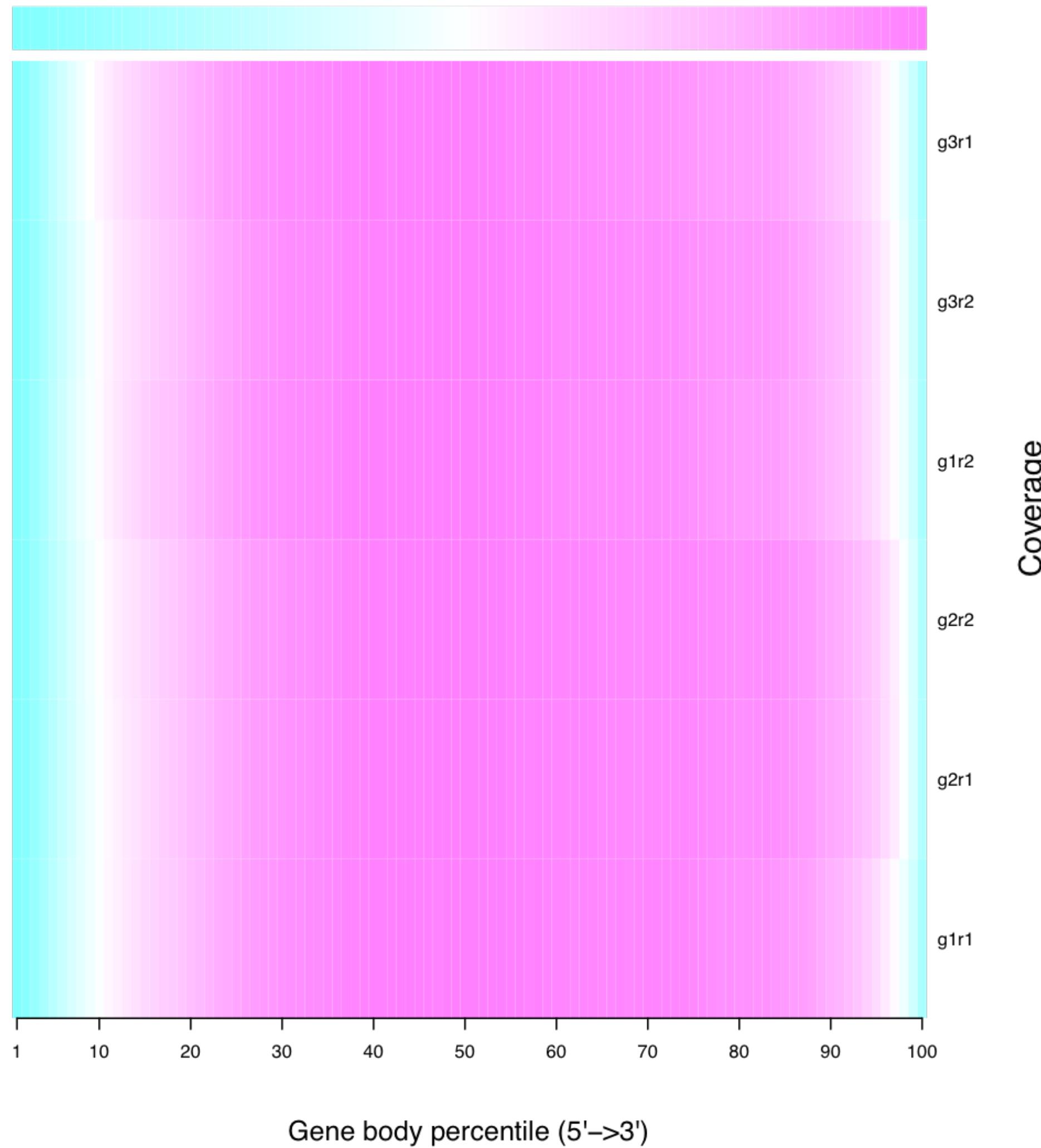


# Diagnostic Plots

- BioAnalyzer Trace;
- GeneBody Coverage;
- Principal Components Analysis;
- Hierarchical Clustering;



- GeneBody Coverage;





# Diagnostic Plots

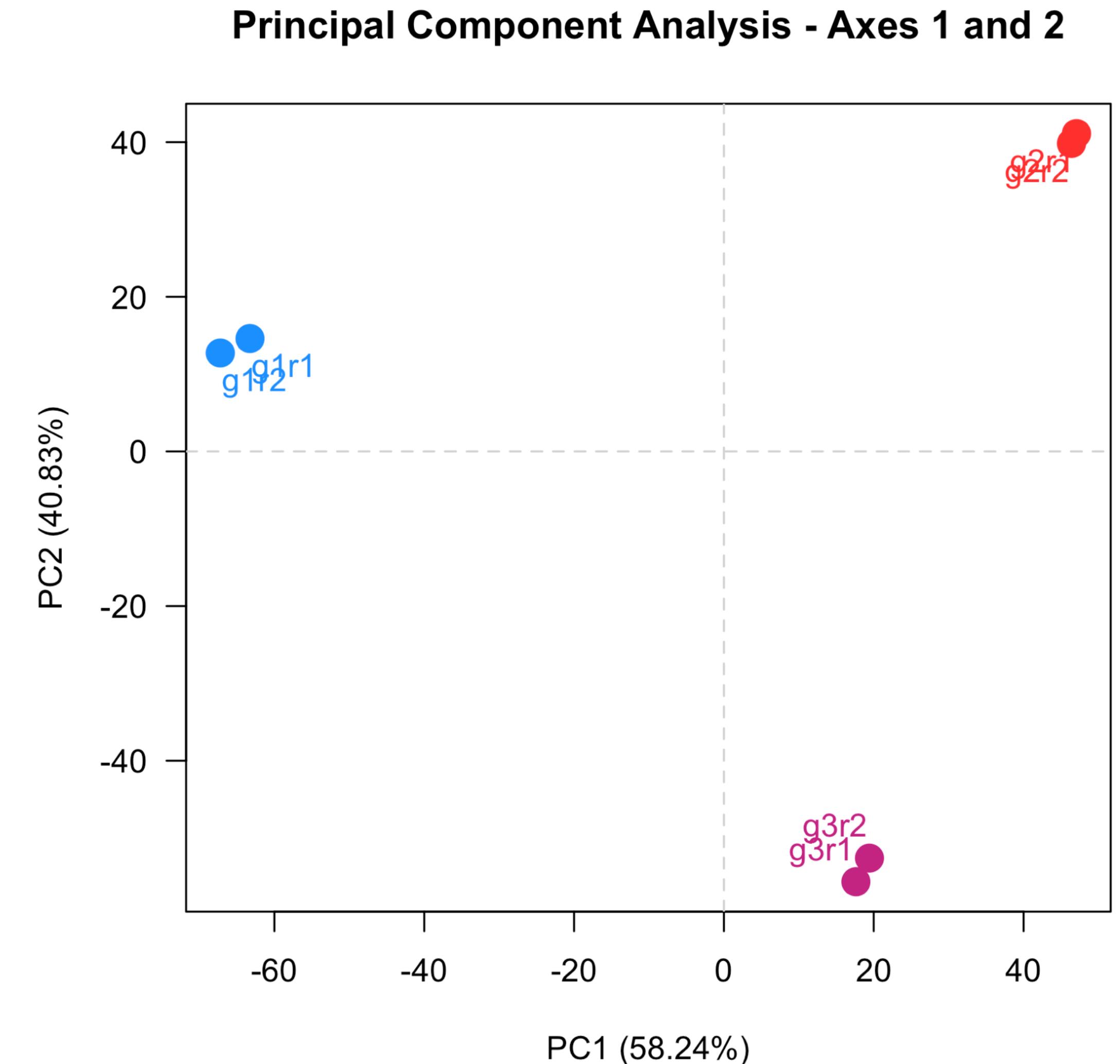
- BioAnalyzer Trace;
- GeneBody Coverage;
- Principal Components Analysis;
- Hierarchical Clustering;



- Principal Components Analysis;

**Color = Treatment Group**

**Multiple data points with  
same color indicate  
biological reps**





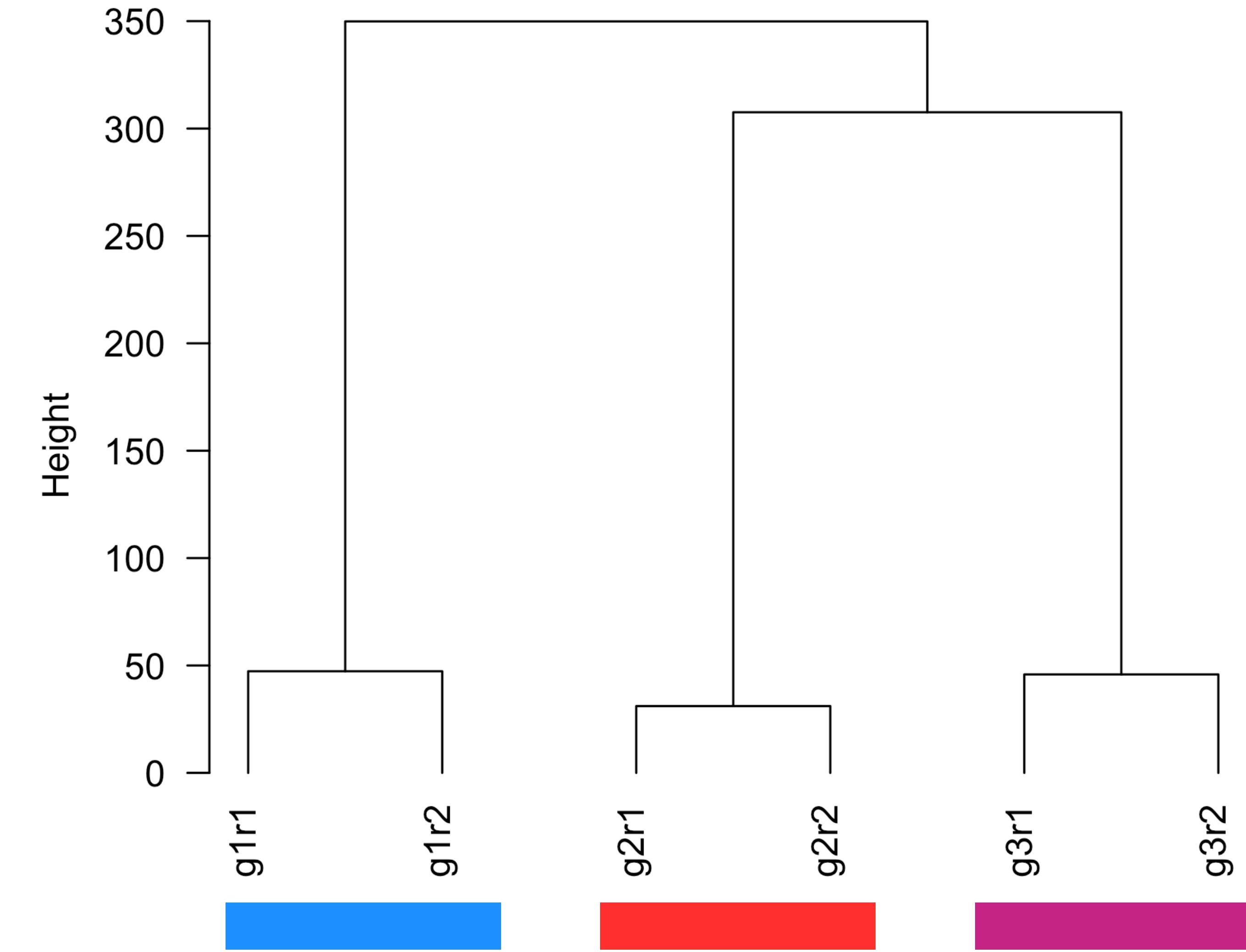
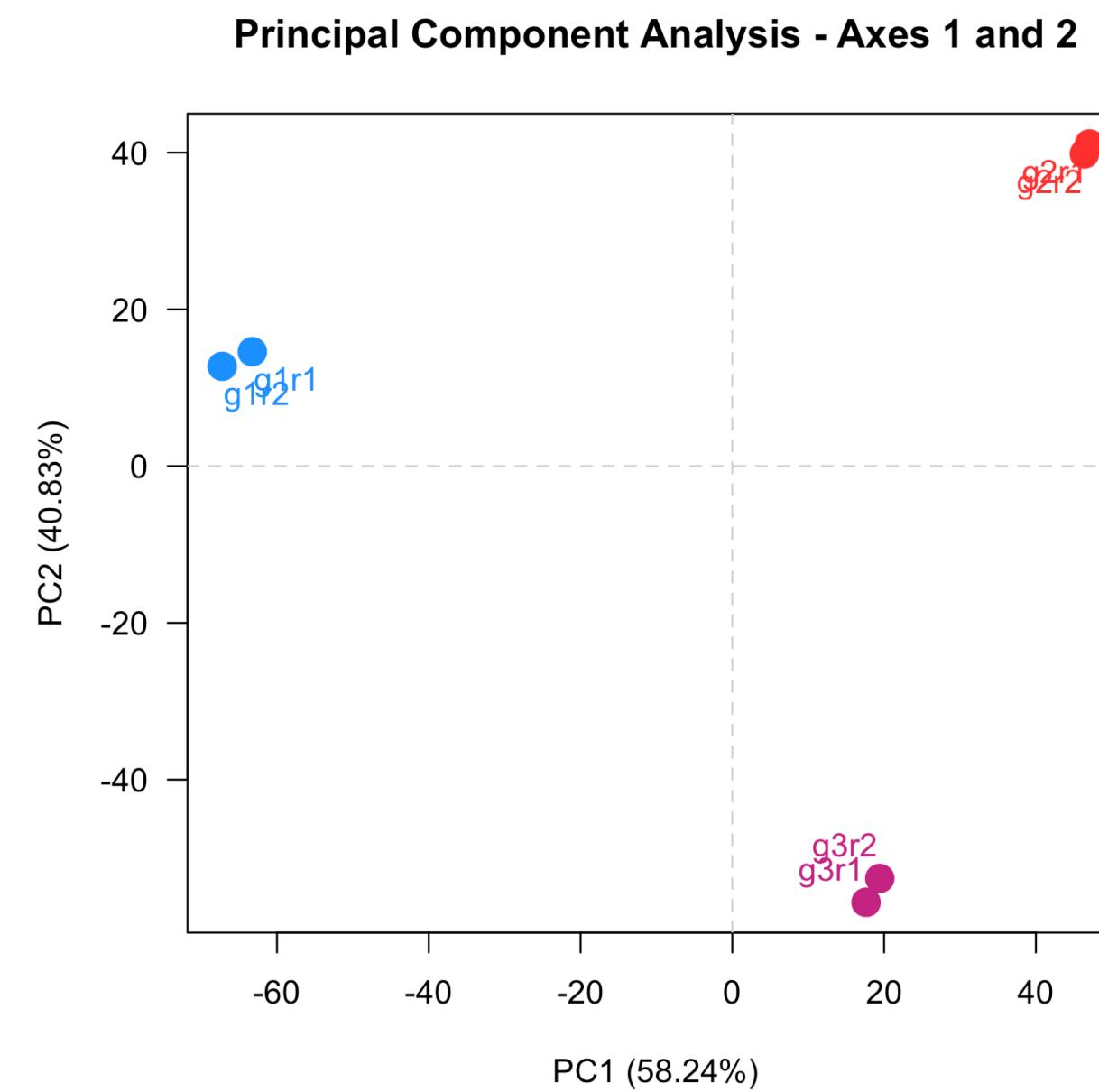
# Diagnostic Plots

- BioAnalyzer Trace;
- GeneBody Coverage;
- Principal Components Analysis;
- Hierarchical Clustering;

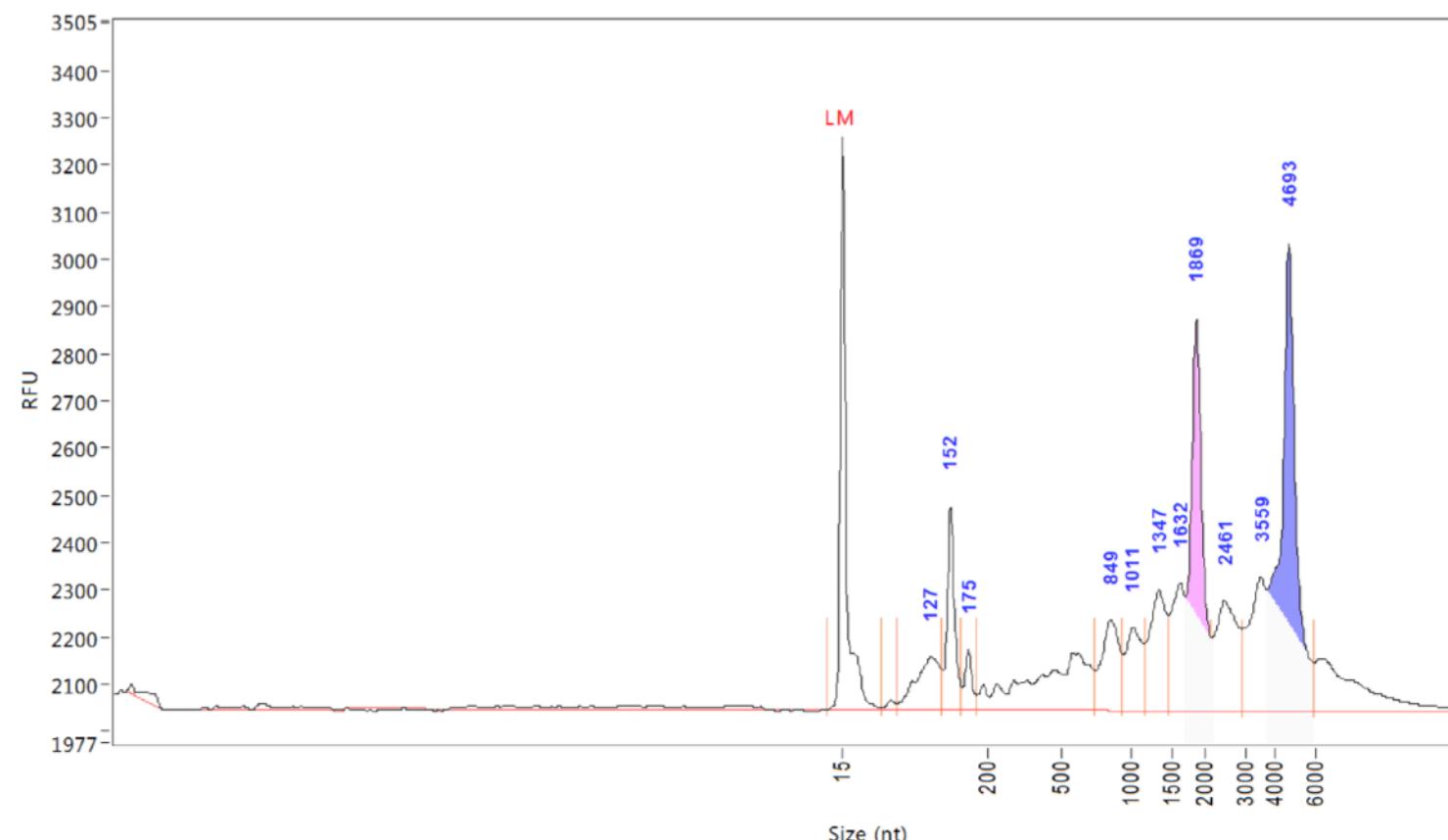
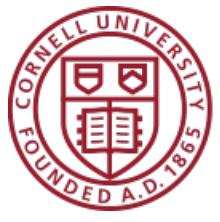


- Hierarchical Clustering;

## Bottom-up approach

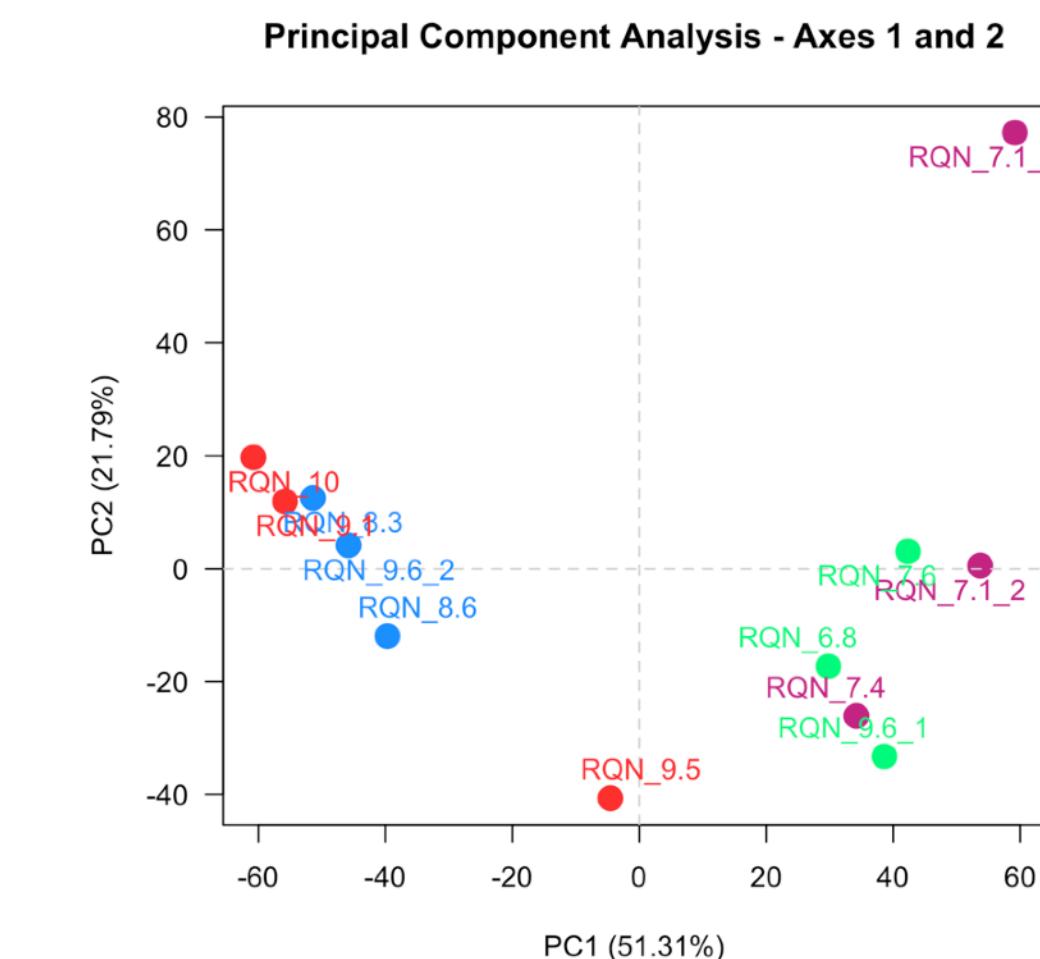
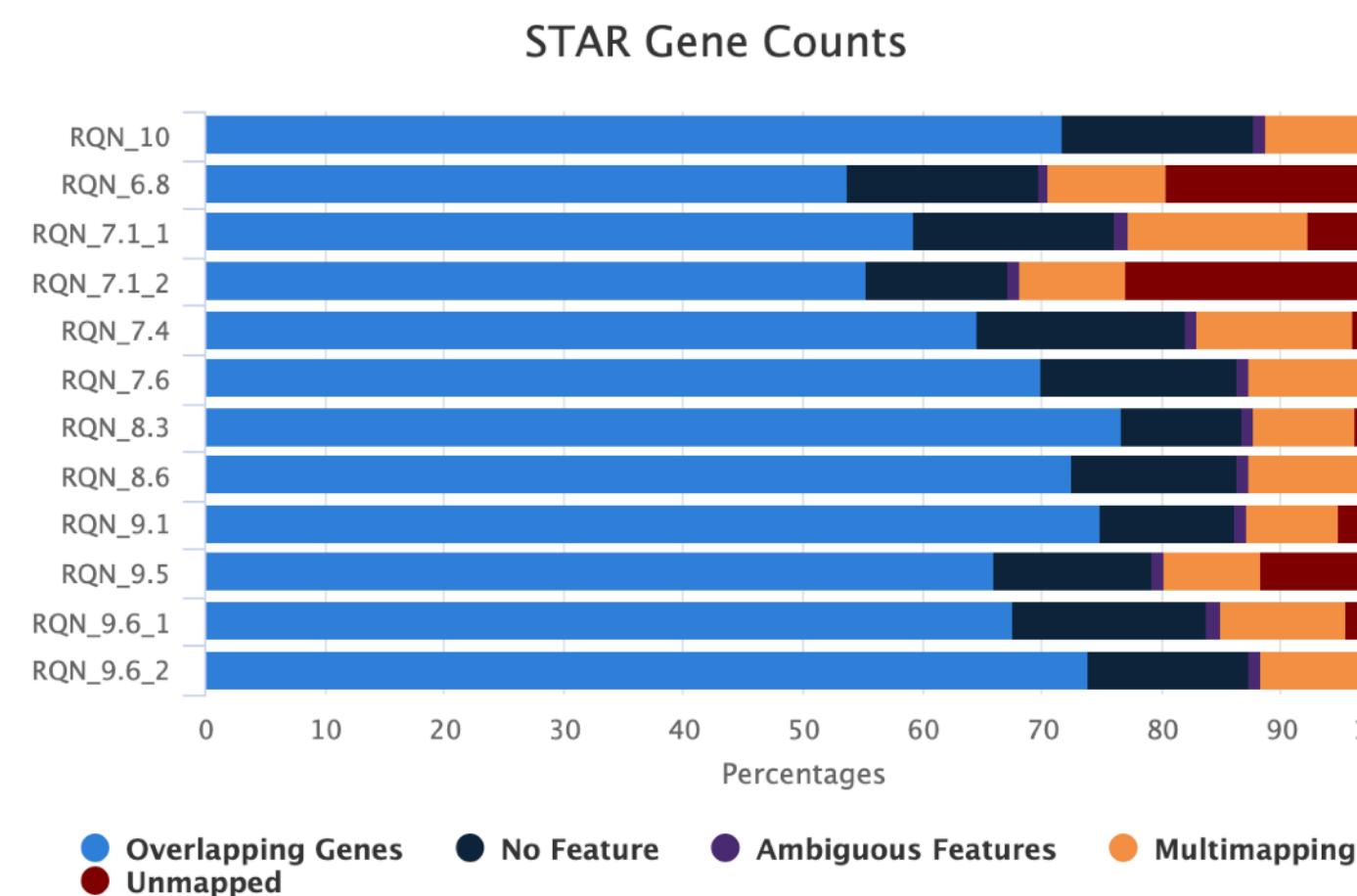
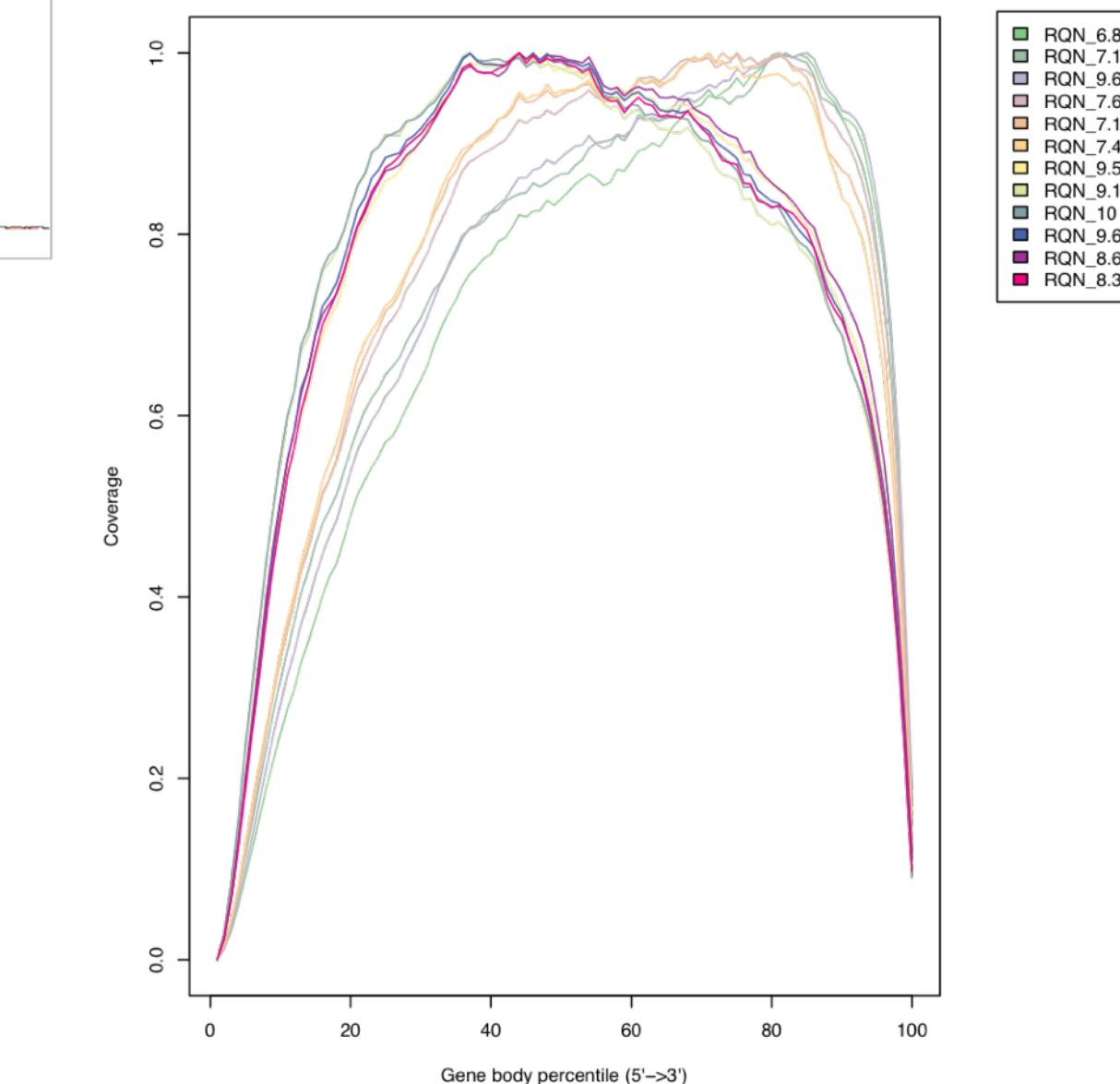


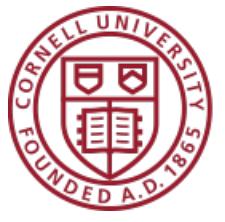
Method: Euclidean distance - Ward criterion  
hclust (\*, "ward.D")



Showing 12/12 rows and 2/2 columns.

Sample Name	% Aligned	M Aligned
RQN_10	88.7%	32.9
RQN_9.6_2	88.3%	29.6
RQN_8.3	87.7%	31.6
RQN_7.6	87.4%	32.1
RQN_8.6	87.3%	20.9
RQN_9.1	87.1%	28.8
RQN_9.6_1	85.0%	32.0
RQN_7.4	83.0%	29.9
RQN_9.5	80.2%	27.7
RQN_7.1_1	77.3%	26.6
RQN_6.8	70.7%	24.8
RQN_7.1_2	68.1%	23.9

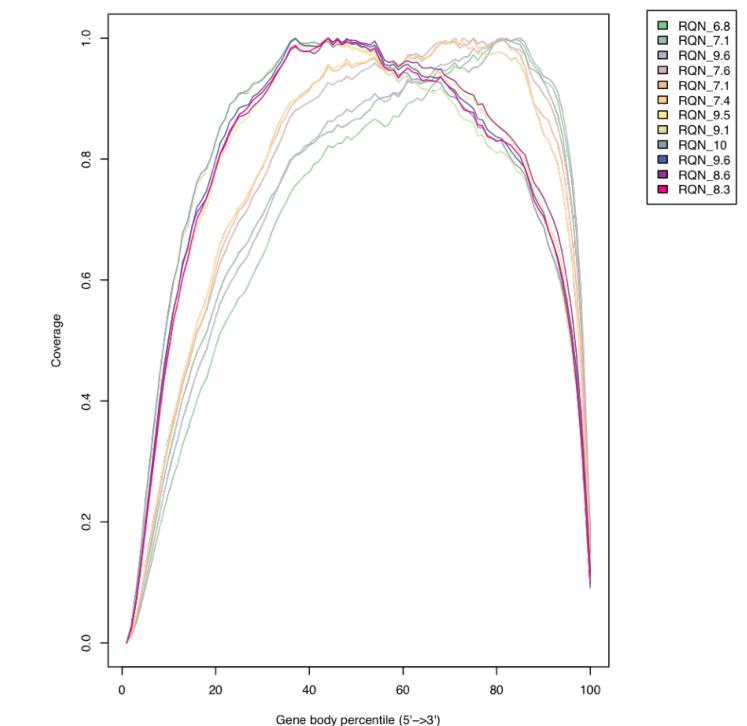
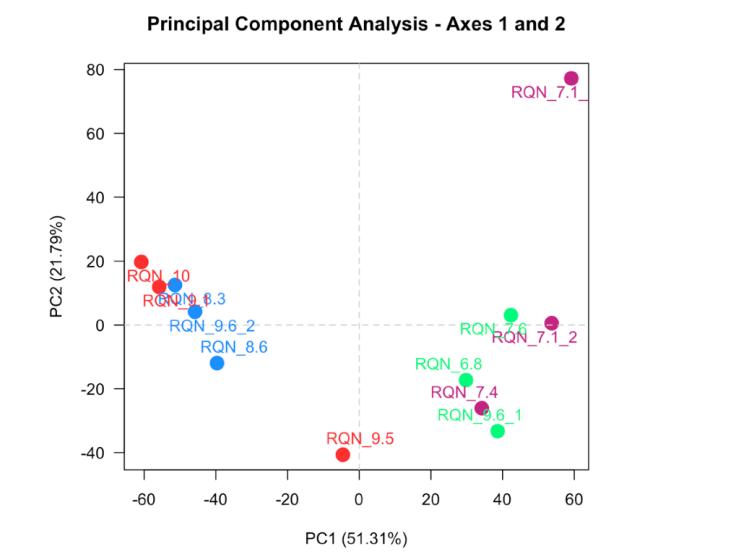
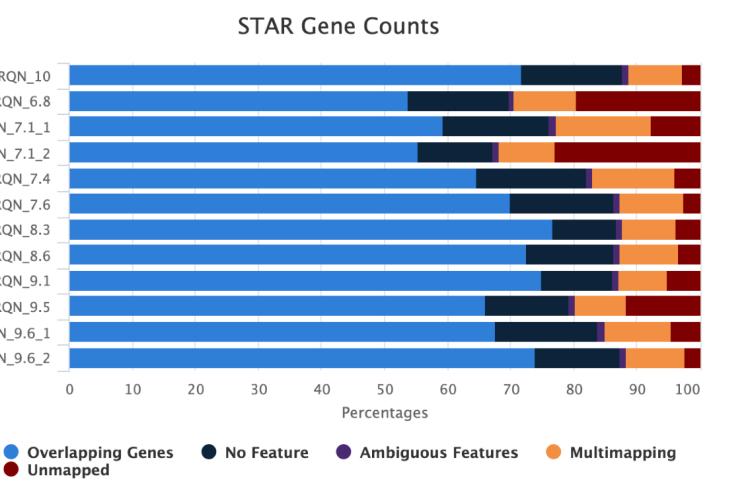
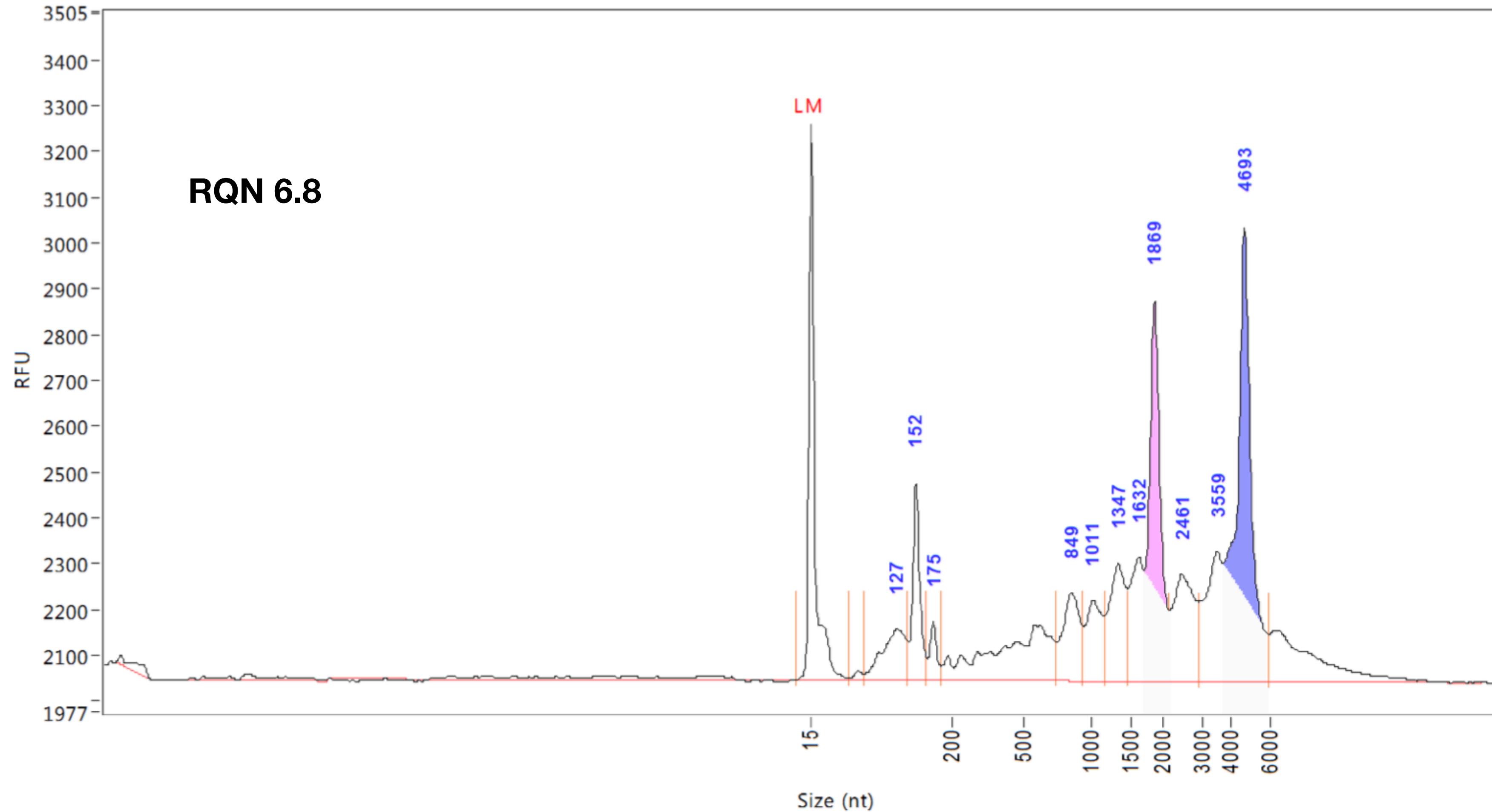




Cornell University

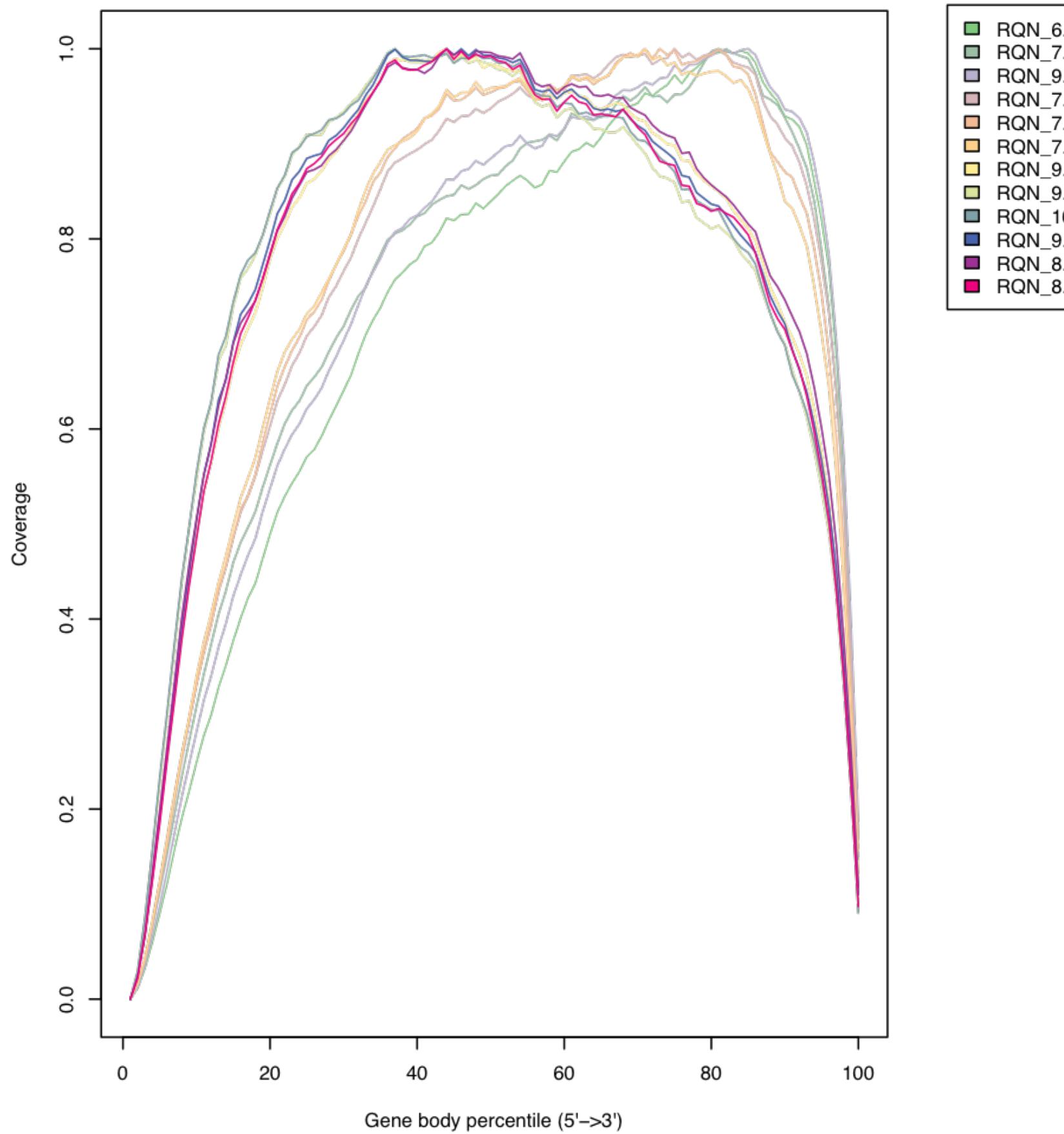
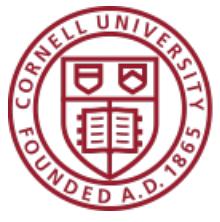
# Case Study 1

## Borderline OKAY sample



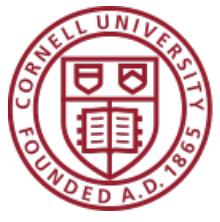
Showing 12/12 rows and 2/2 columns.

Sample Name	% Aligned	M Aligned
RQN_10	88.7%	32.9
RQN_9.6_2	88.3%	29.6
RQN_8.3	87.7%	31.6
RQN_7.6	87.4%	32.1
RQN_8.6	87.3%	20.9
RQN_9.1	87.1%	28.8
RQN_9.6_1	85.0%	32.0
RQN_7.4	83.0%	29.9
RQN_9.5	80.2%	27.7
RQN_7.1_1	77.3%	26.6
RQN_6.8	70.7%	24.8
RQN_7.1_2	68.1%	23.9

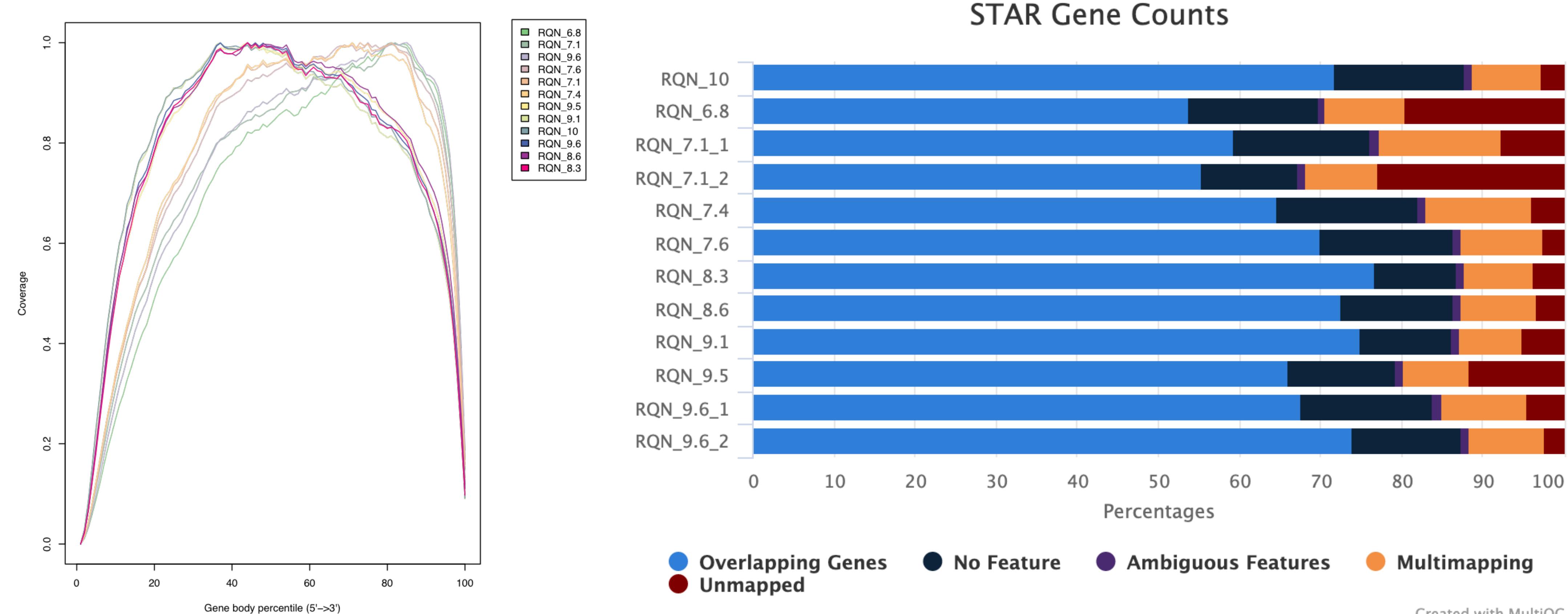


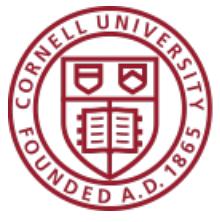
Showing 12/12 rows and 2/2 columns.

Sample Name	% Aligned	M Aligned
RQN_10	88.7%	32.9
RQN_9.6_2	88.3%	29.6
RQN_8.3	87.7%	31.6
RQN_7.6	87.4%	32.1
RQN_8.6	87.3%	20.9
RQN_9.1	87.1%	28.8
RQN_9.6_1	85.0%	32.0
RQN_7.4	83.0%	29.9
RQN_9.5	80.2%	27.7
RQN_7.1_1	77.3%	26.6
RQN_6.8	70.7%	24.8
RQN_7.1_2	68.1%	23.9



Cornell University

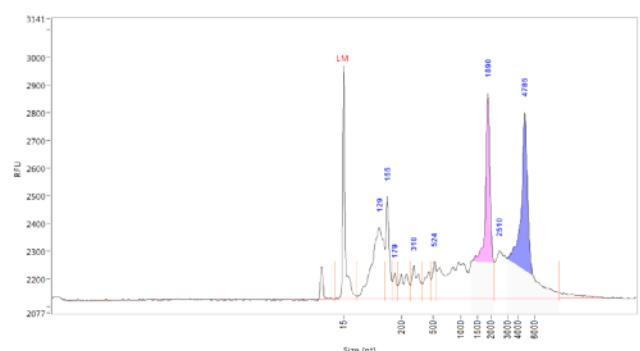
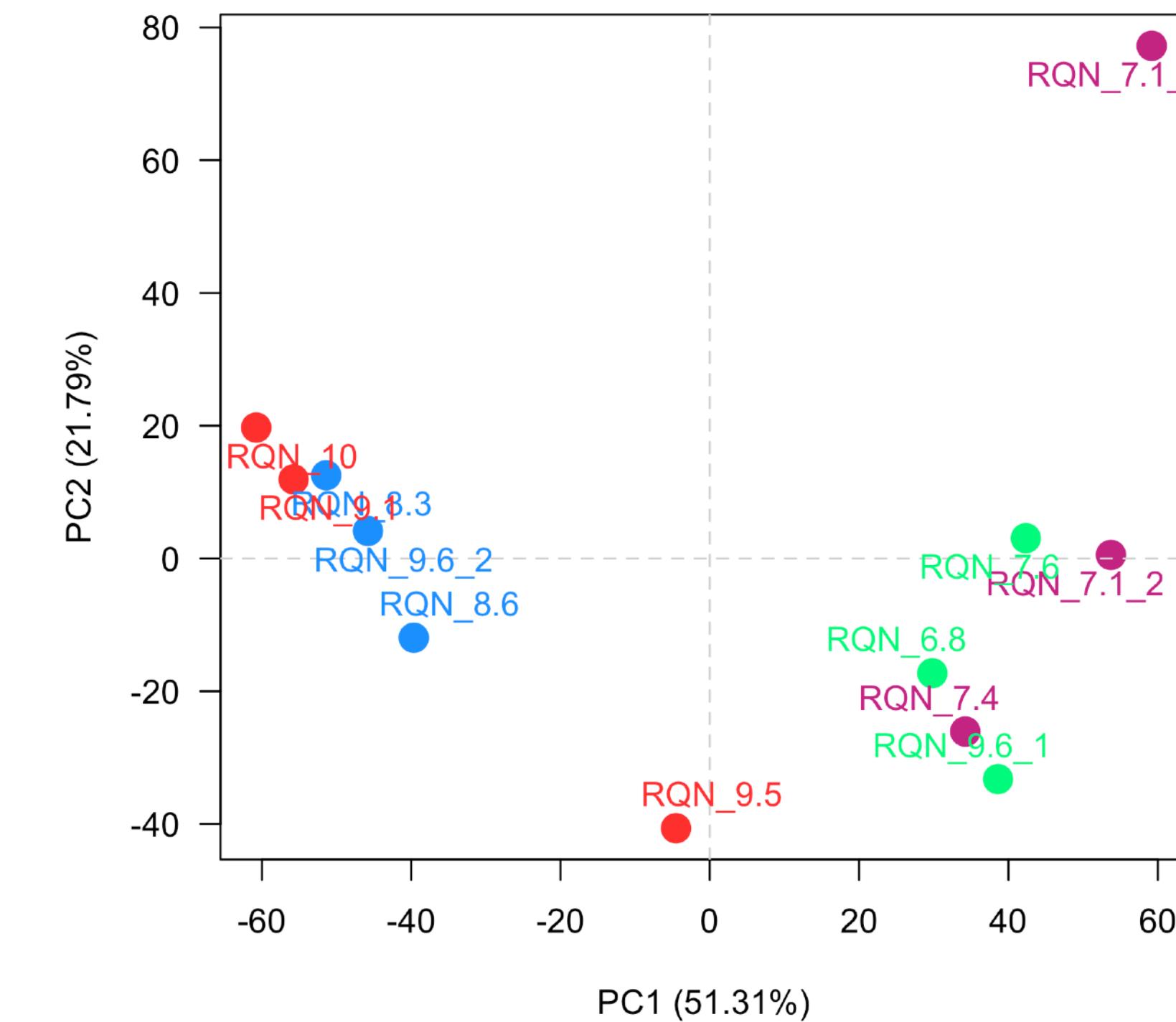


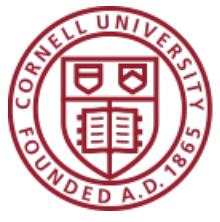


Copy table Showing 12/12 rows and 2/2 columns.

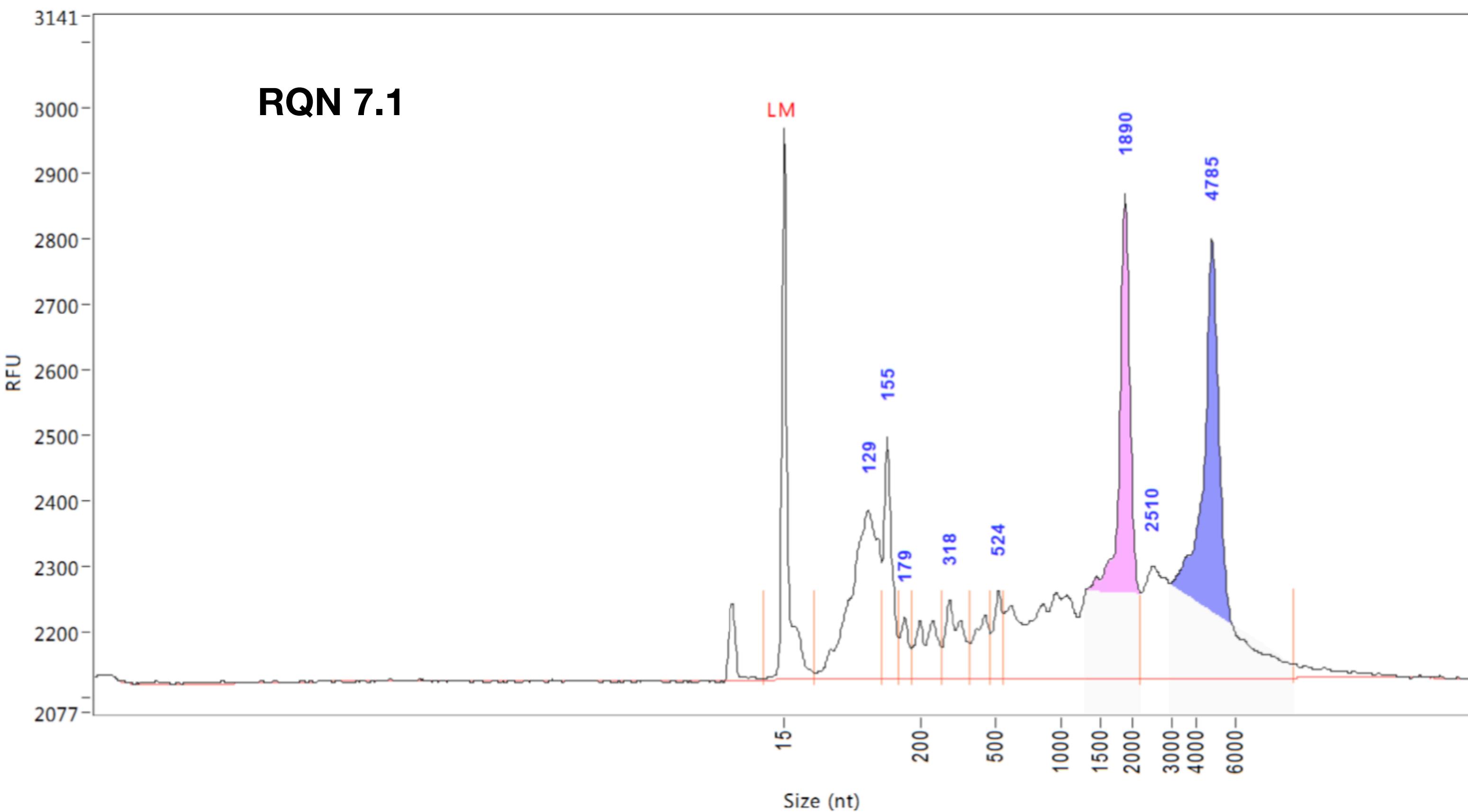
Sample Name	% Aligned ▾	M Aligned
RQN_10	88.7%	32.9
RQN_9.6_2	88.3%	29.6
RQN_8.3	87.7%	31.6
RQN_7.6	87.4%	32.1
RQN_8.6	87.3%	20.9
RQN_9.1	87.1%	28.8
RQN_9.6_1	85.0%	32.0
RQN_7.4	83.0%	29.9
RQN_9.5	80.2%	27.7
RQN_7.1_1	77.3%	26.6
RQN_6.8	70.7%	24.8
RQN_7.1_2	68.1%	23.9

Principal Component Analysis - Axes 1 and 2

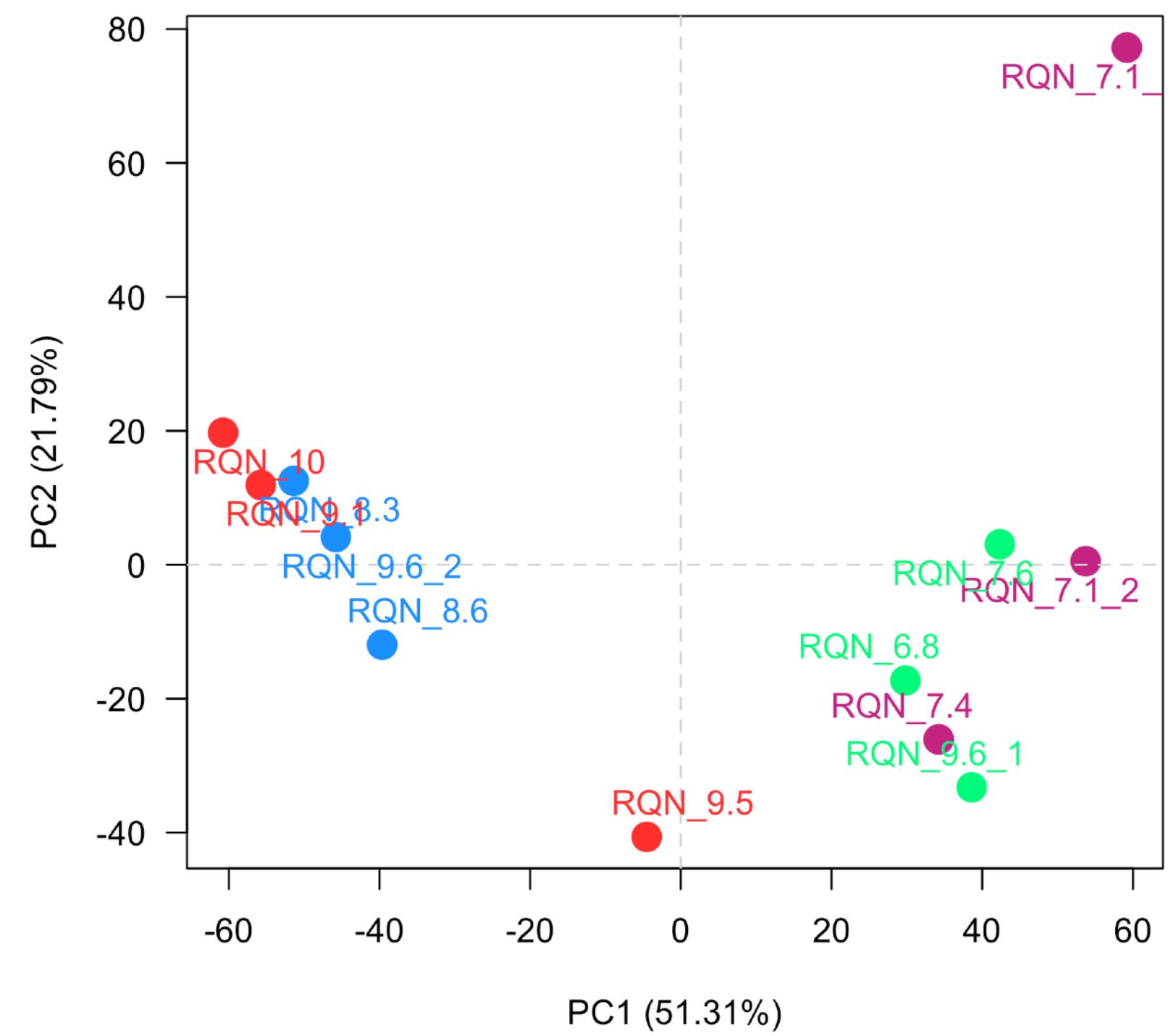


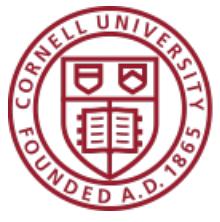


Cornell University

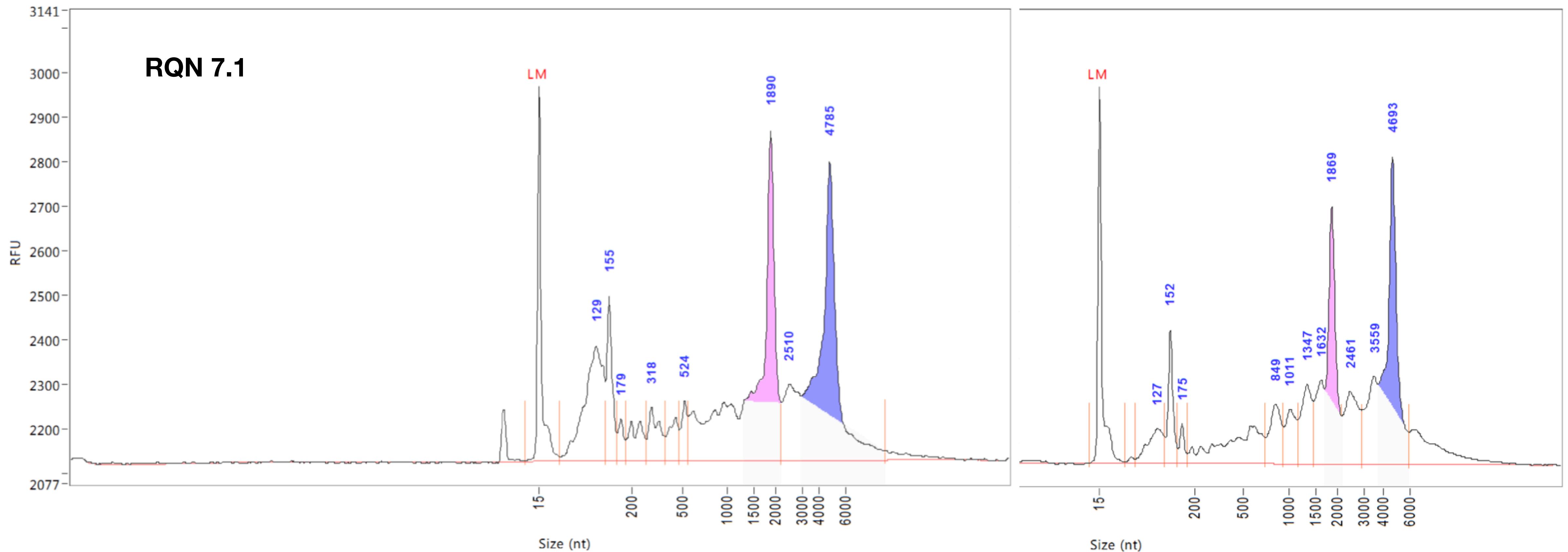


Principal Component Analysis - Axes 1 and 2





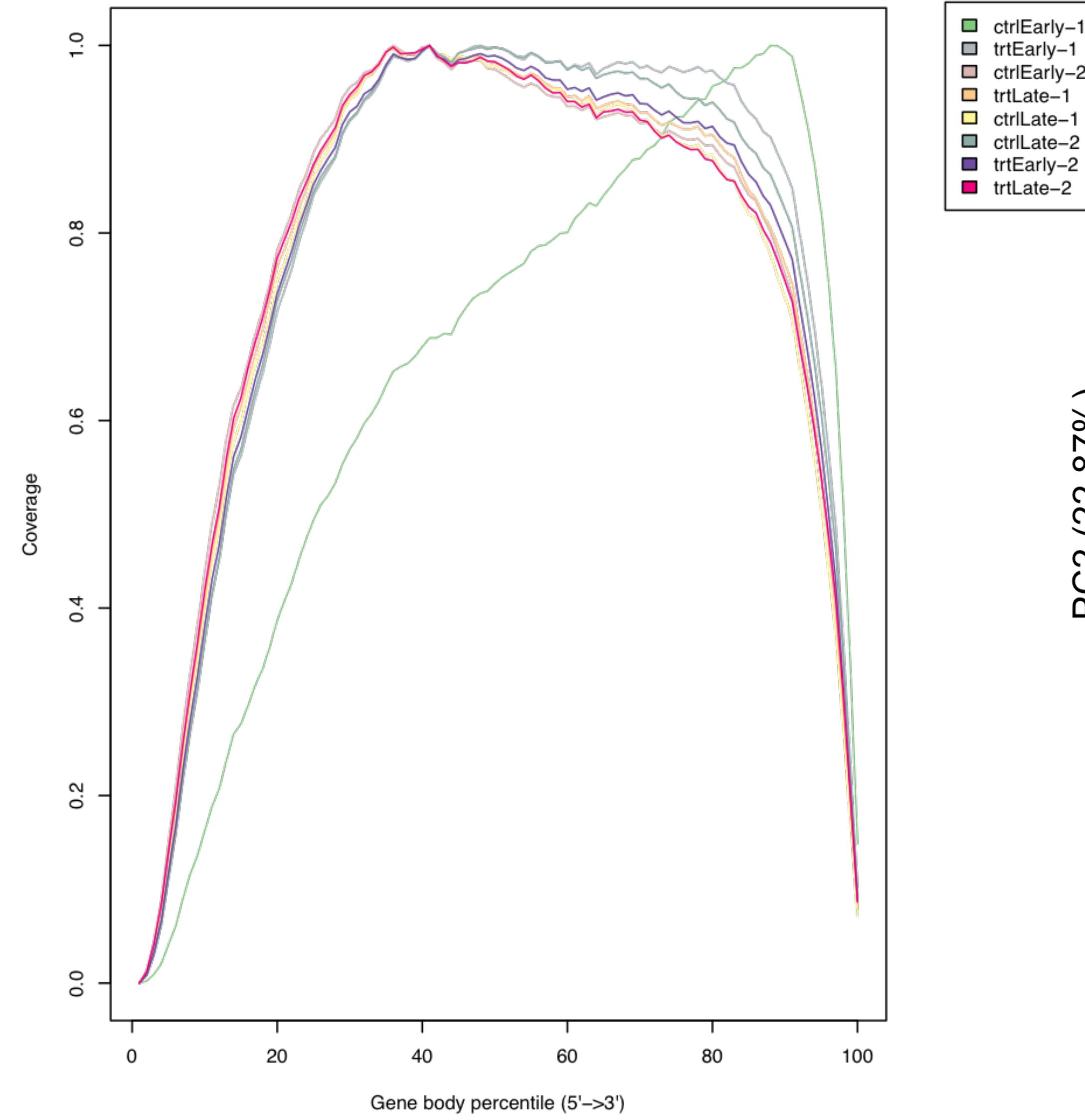
Cornell University



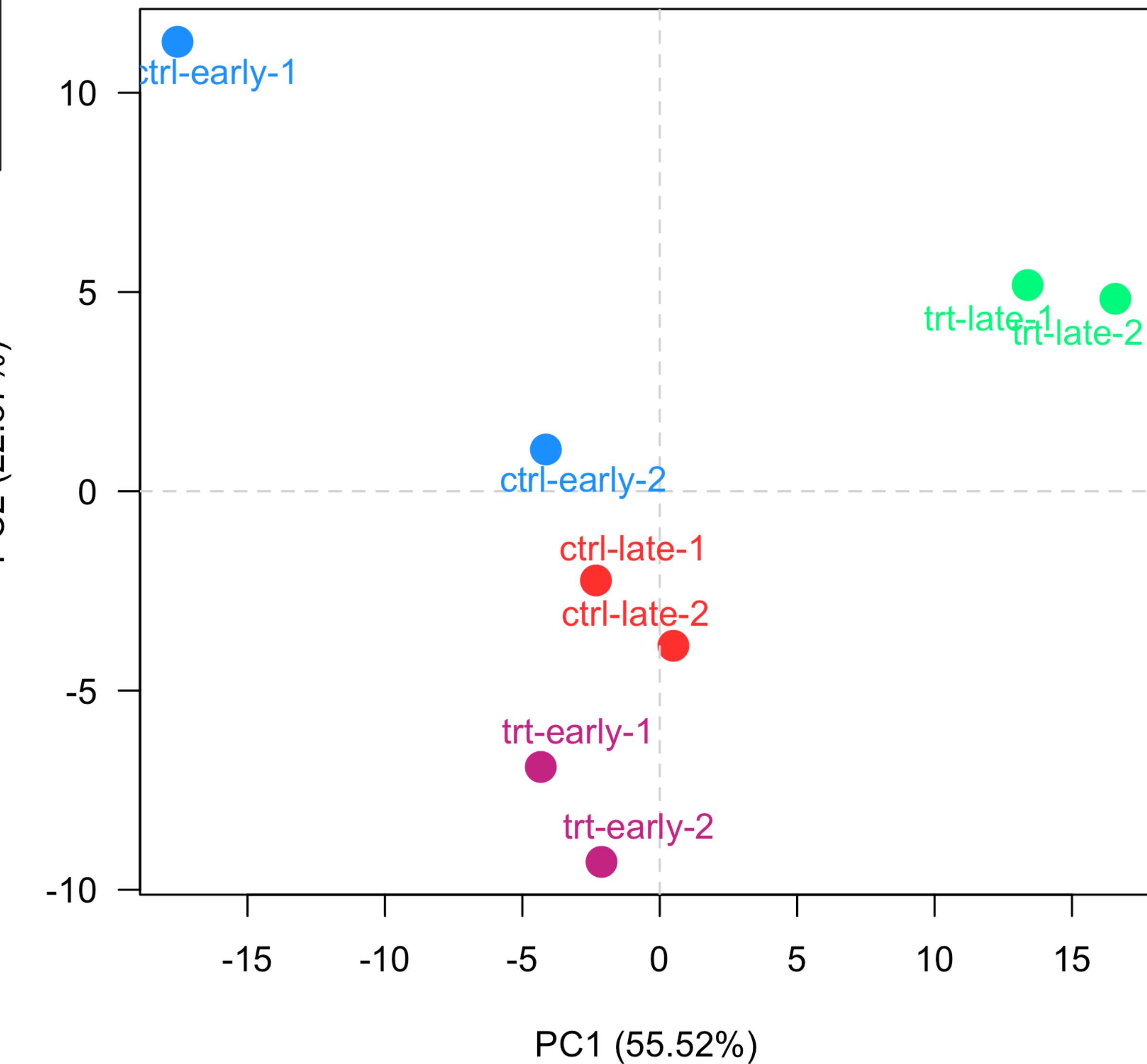


Cornell University

## Case Study 2



Principal Component Analysis - Axes 1 and 2





Cornell University

# Summary Report



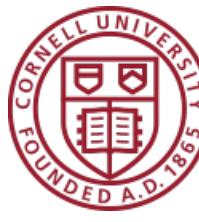
# TREX Analysis Reports

- MultiQC Alignment Summary;
- Data QC Report;
- Raw Count Table;
- DE Genes Analysis Excel File;



# Next Steps...

- DE Assessment of candidate genes;
- DE genes (Panther/DAVID); GO-Term and Pathway Enrichment;
- Gene Set Enrichment Analysis (Broad);
- Ingenuity Pathway Analysis (IPA);



Cornell University

# Thank You for Listening !

Contact Info: <http://rnaseqcore.vet.cornell.edu/>

E-mail List Serve: TREX-GENEREG-L

Jen Grenier: jgrenier@cornell.edu

Christine Butler: cab18@cornell.edu

Ann Tate: aef93@cornell.edu

Faraz Ahmed: fahmed@cornell.edu

