

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

# WEEKLY UPDATE 4

OMICS LOGIC RESEARCH FELLOWSHIP



# TRANSCRIPTOMICS - PART 1

- T1
- T2

# Biology of transcription



DNA



rRNA



mRNA



tRNA



miRNA

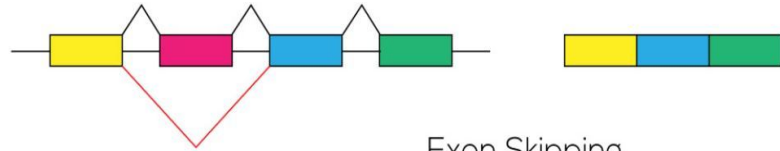


ribozyme

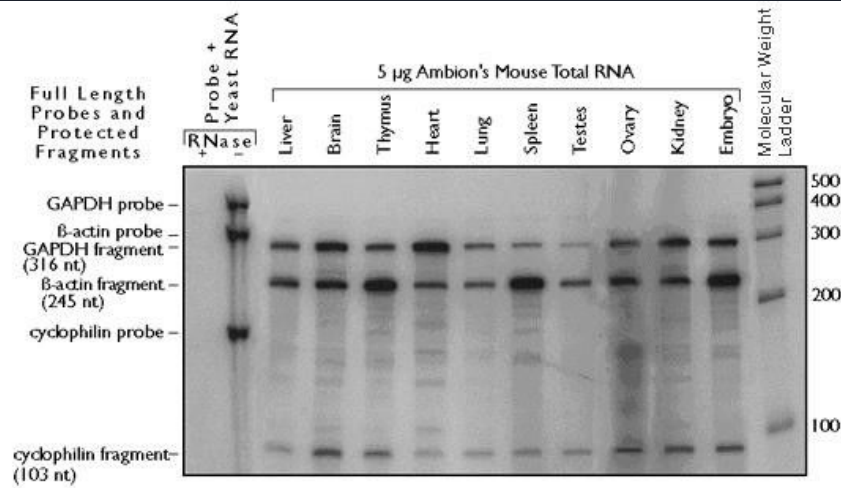
## Constitutive Splicing



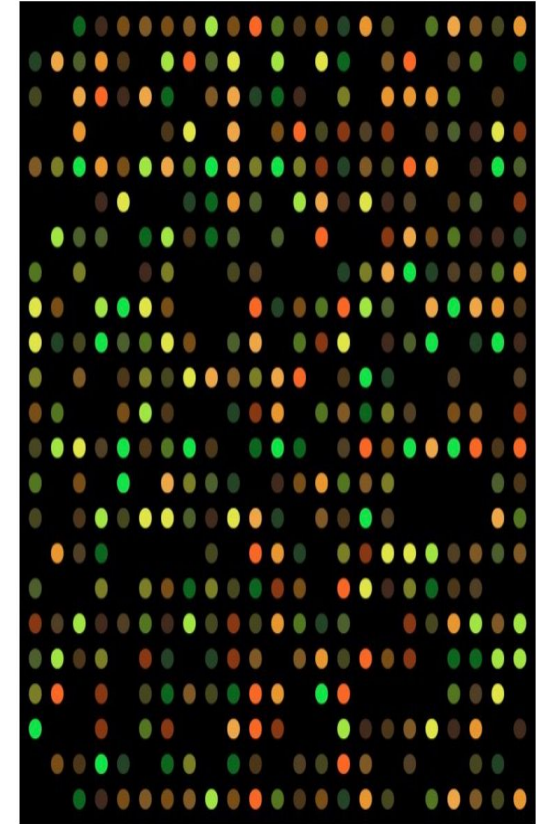
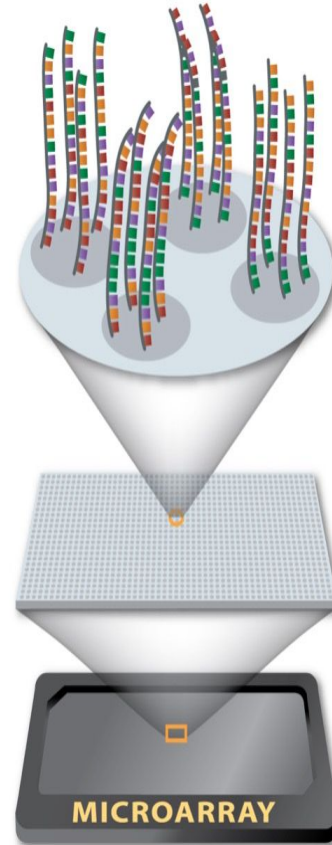
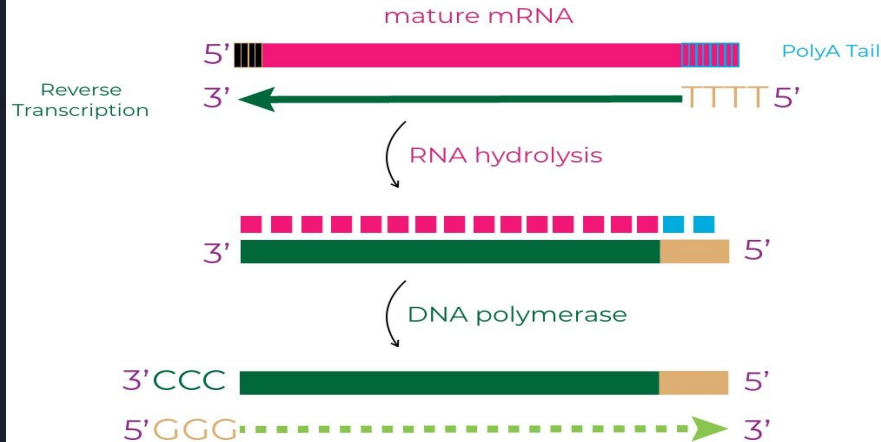
## Exon Skipping



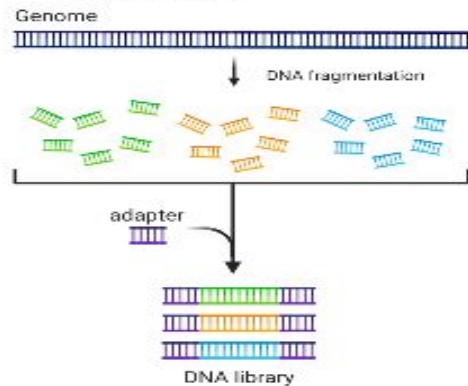
# RNA QUANTIFICATION



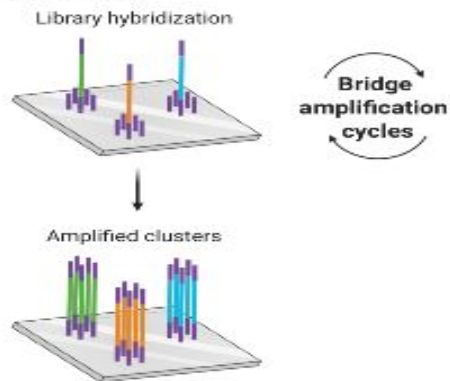
## cDNA is Reverse Transcribed from mRNA



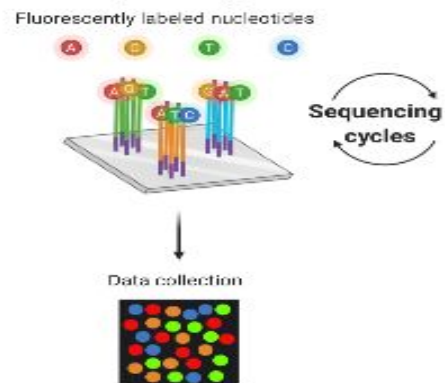
# 1 Library preparation



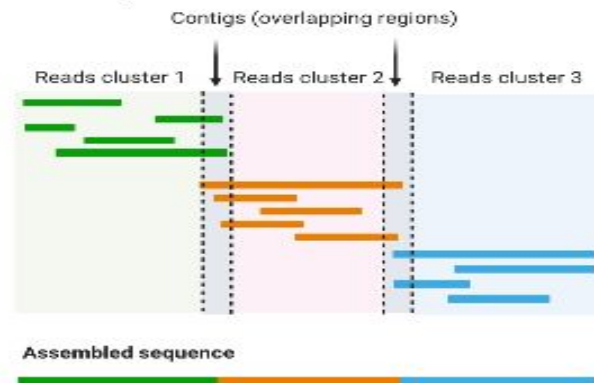
# 2 DNA library bridge amplification



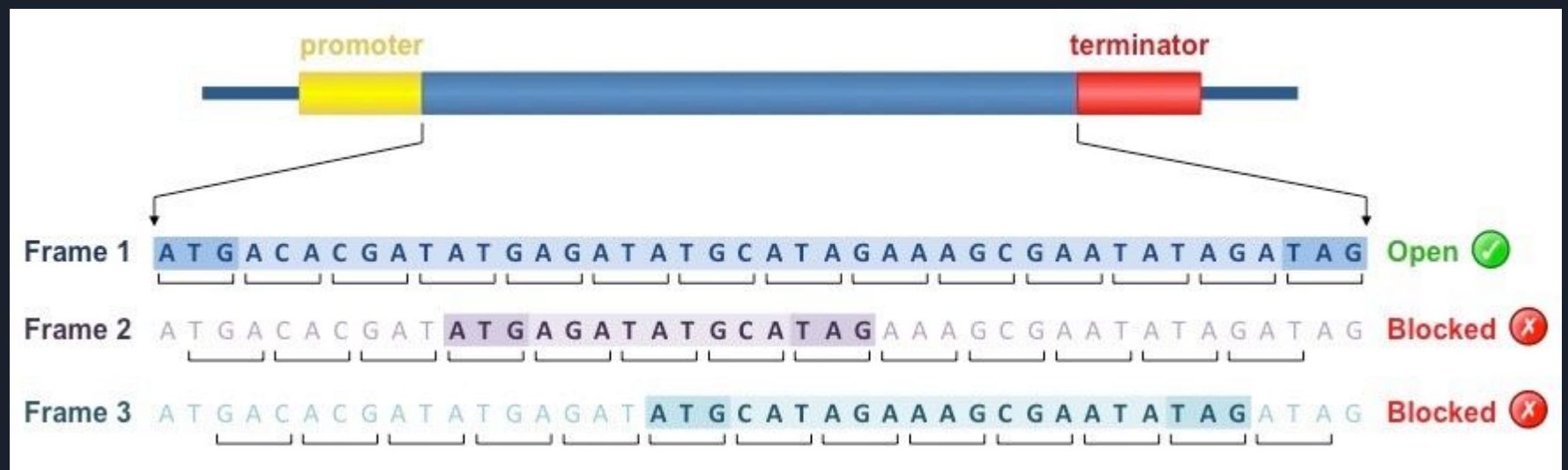
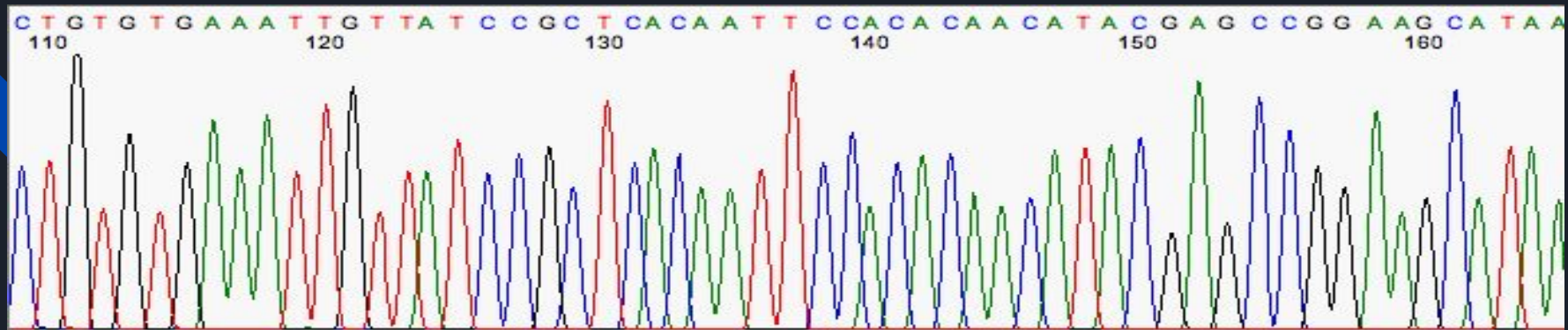
# 3 DNA library sequencing



# 4 Alignment and data analysis



NCS





# Analysis of Raw RNA-seq Data

Step 1: Pre-processing, data simulation, and error correction

Step 2: Mapping on a reference genome

Step 3: Calculating the abundance of reads aligned to the reference genome

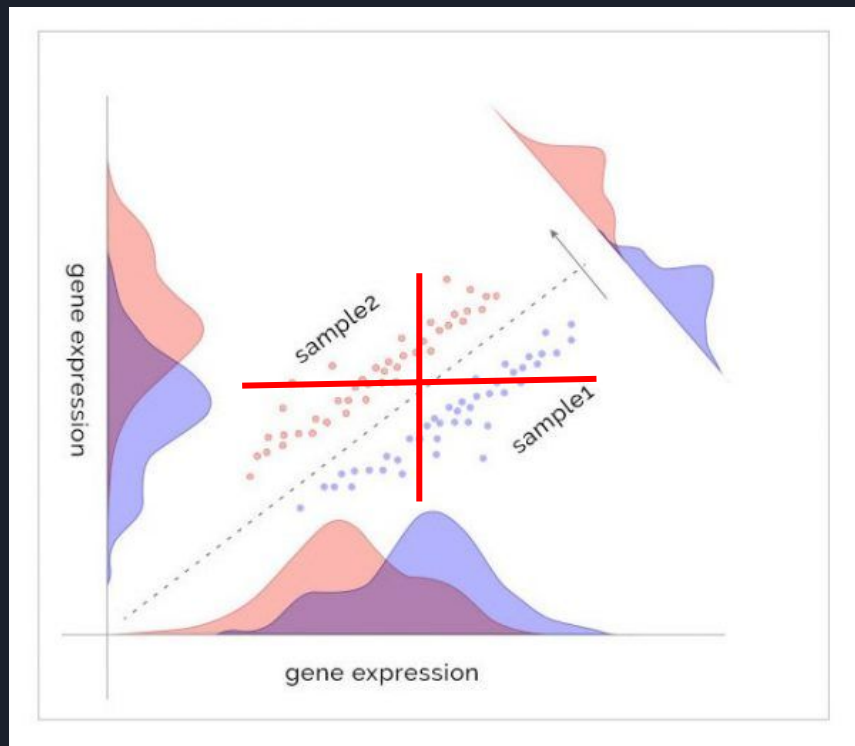
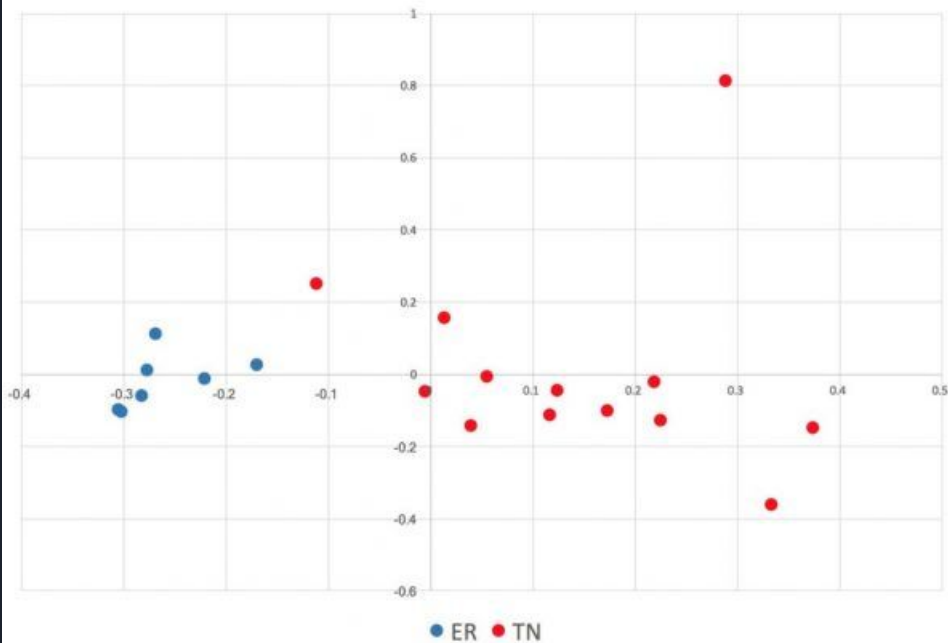
Step 4: Post-processing

Step 5: Exploratory Analysis

Step 6: Biological Interpretation

# PCA

PCA Gene Expression. PC1: 21.76%, PC2: 9.20%





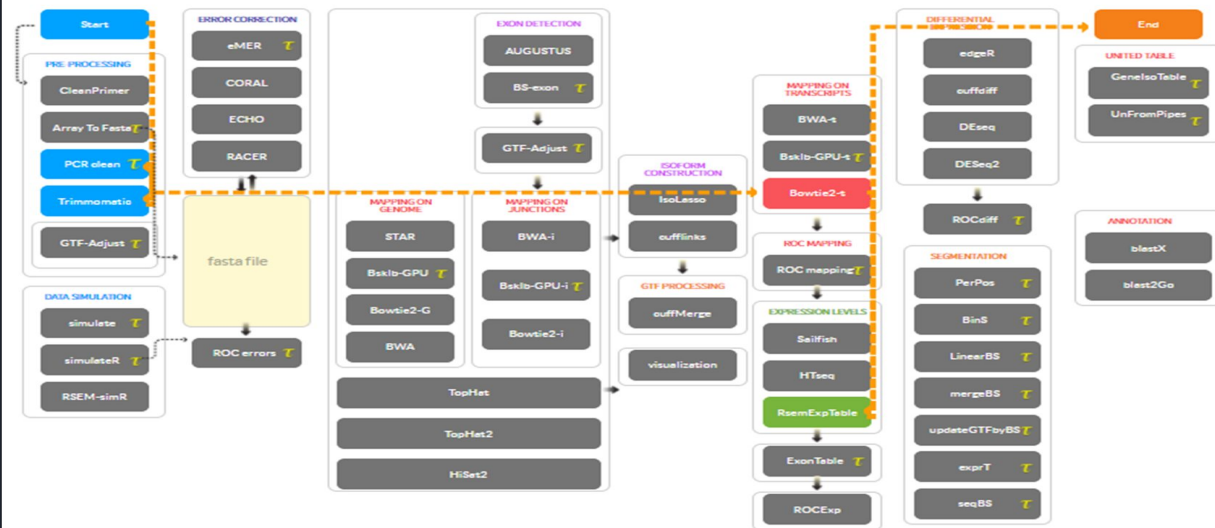
## T-BioInfo AREAS OF ANALYSIS:

T-BioInfo is a Bioinformatics platform that combines statistical analysis modules into pipelines to deal with heterogenous big data. Below you will find the areas of analysis our platform can be used for:

NGS DATA	MASS-SPECTROSCOPY	STRUCTURAL BIOLOGY	DATA INTEGRATION AND MODELING
<b>First Step: Select the Demo: RNA_seq (PDX) pipeline</b>			
<b>Transcriptomics</b> <ul style="list-style-type: none"> <li>RNA-seq/chip: parallel analysis of NGS and microarray data</li> <li>DEMO: RNA-seq (cell-line project)</li> <li>DEMO: RNA-seq (Angelman Project)</li> <li>DEMO: RNA-seq (PDX Project)</li> </ul>	<ul style="list-style-type: none"> <li>Mass-spec proteomics</li> <li>Mass-spec metabolomics</li> <li>Mass-spec proteomics (MaxQuant)</li> </ul>	<ul style="list-style-type: none"> <li>3D biopolymer structures and complexes (In Development)</li> <li>Libraries of small molecules</li> <li>Docking of small molecules (under development)</li> </ul>	<b>Data Association</b> <ul style="list-style-type: none"> <li>Multi - Omics</li> <li>Genome Wide Association</li> <li>Phylogenetic Tree</li> </ul> <b>Data Mining</b> <ul style="list-style-type: none"> <li>Supervised Analysis</li> </ul>

## Next Step: Run pipeline by clicking on multiple steps

### PIPELINE GRAPH



CYP4V2 gene

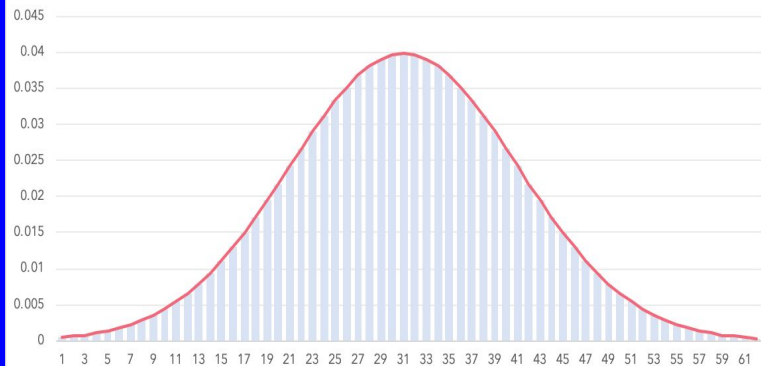
ER+ tumor sample CYP4V2  
genes downregulated

TN- tumor sample CYP4V2  
genes upregulated

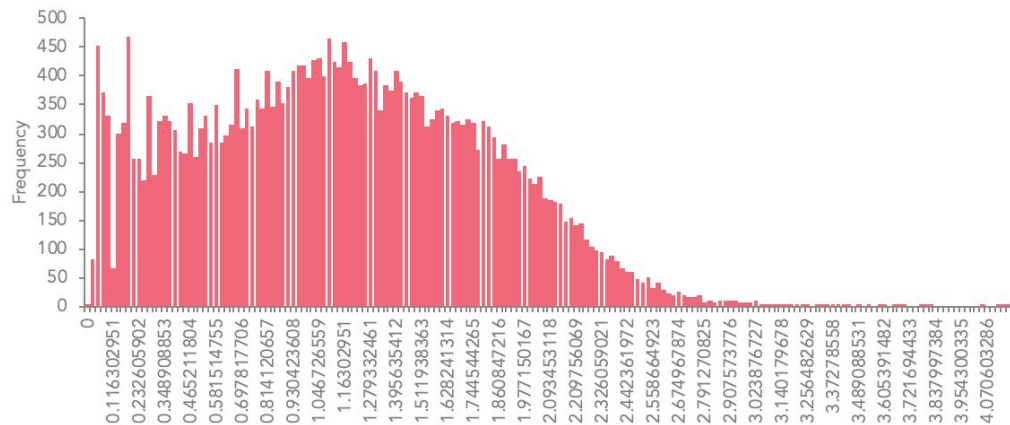
Histogram of Gene Expression for ER-NOD-ERR1084805



Normal Distribution

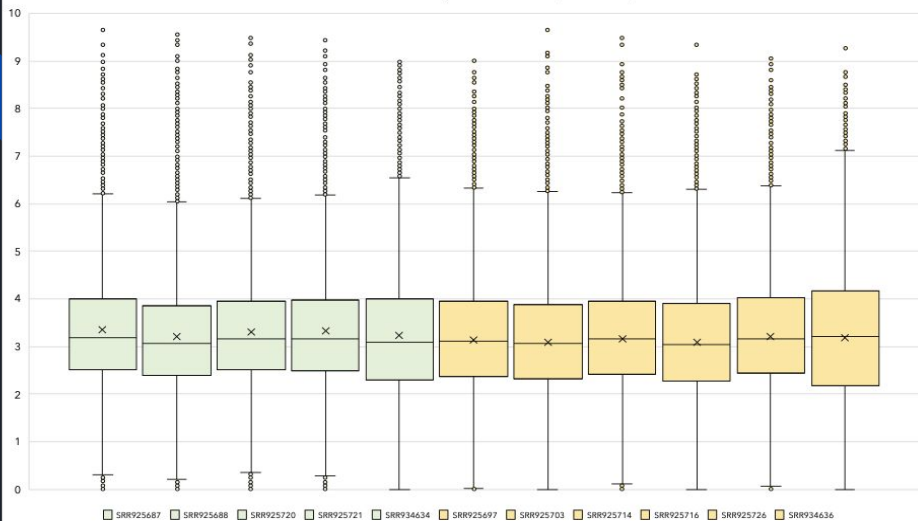


Histogram of GE for ER-NOD-ERR1084805 (log scale after filetering)

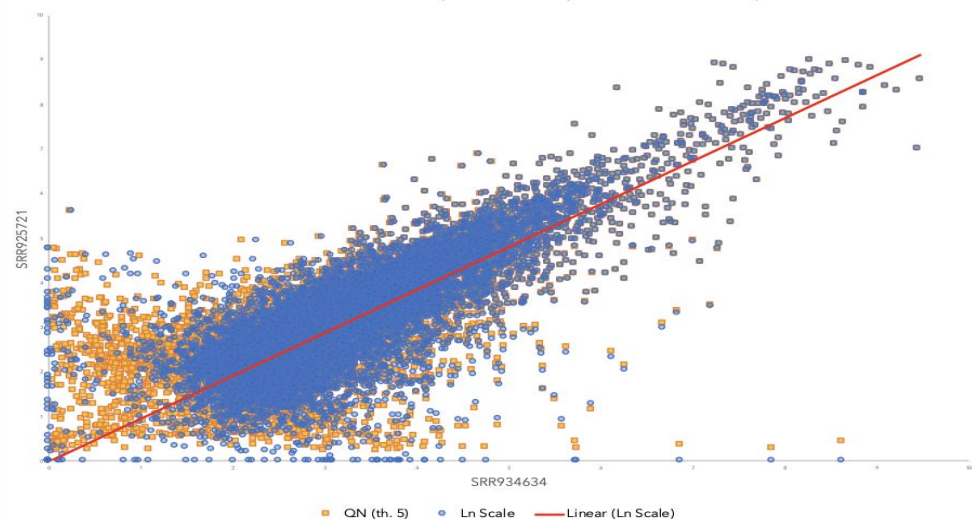



**Data Preparation for  
Downstream Analysis**

Cell Lines Gene Expression Data (Ln Scale)



Ln-transformed vs QN5 Data Expression comparison between "replicates"





## Quantile normalization and PCA

Variable	17.79 %	7.36 %	6.45 %	5.75 %	5.46 %
ER-ERR108 4763_PE	0.351 646	-0.33 384	-0.16 223	-0.15 119	0.053 663
ER-ERR108 4764_PE	0.342 116	-0.24 836	-0.18 201	-0.18 139	0.092 981
TN-ERR108 4766_PE	-0.21 109	-0.54 782	0.670 529	0.063 556	-0.07 372
TN-ERR108 4768_PE	-0.32 142	-0.05 837	-0.29 367	-0.39 249	-0.03 926

# T-TEST

Ran a demo differential expression pipeline

## T-test Formula Explained

The **t-test statistic** represents significance of difference between group means.

$t =$

$(X_1 - X_2)$

$$\sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}$$

$S^2 =$

$$\frac{\sum (x - X_1)^2 + \sum (x - X_2)^2}{n_1 + n_2 - 2}$$

S is an estimator of the common variance for the two groups

$n_1$  &  $n_2$  represent the sizes of groups 1 & 2

Differential expression algorithms in Tbioinfo platform

- DEseq
- DEseq2
- Cuffdiff
- EdgeR



# TRANSCRIPTOMICS - PART 2

- T3
- Python
- R