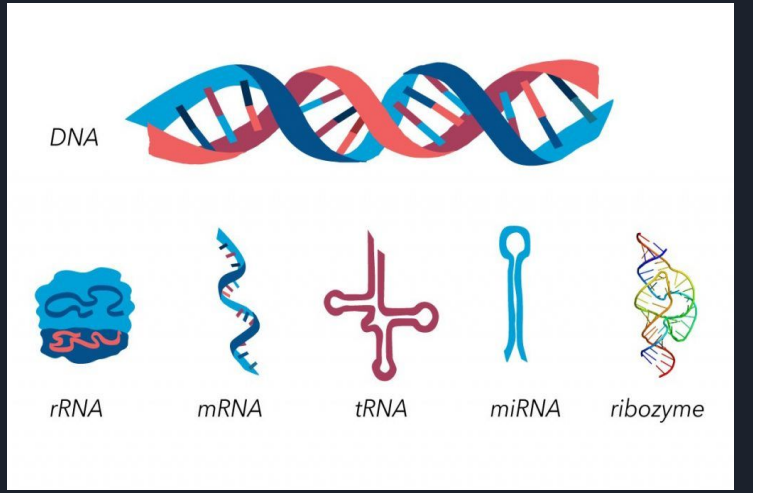
WEEKLY UPDATE 4

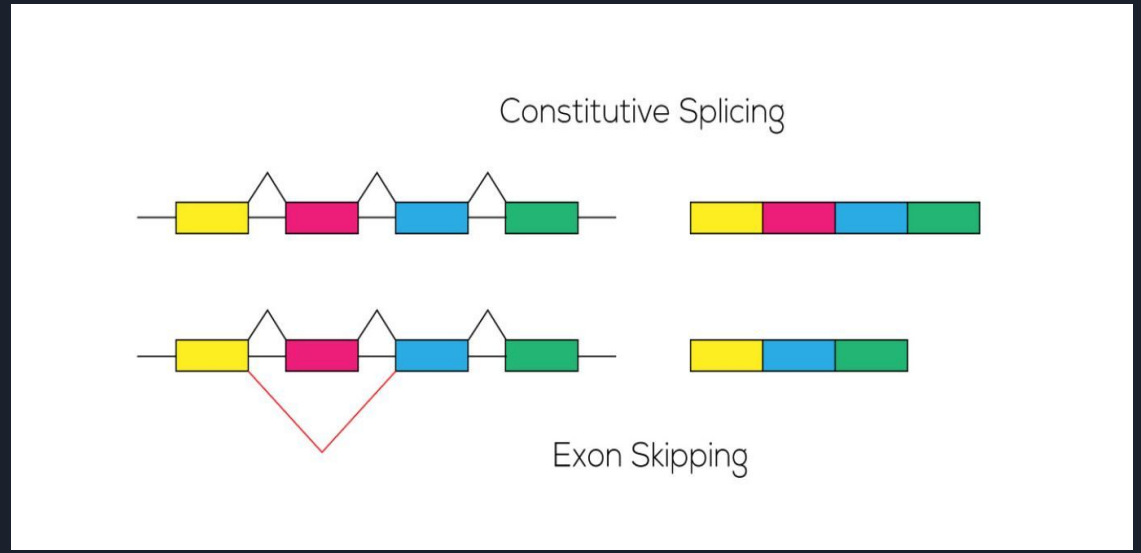
OMICS LOGIC RESEARCH FELLOWSHIP

TRANSCRIPTOMICS - PART 1

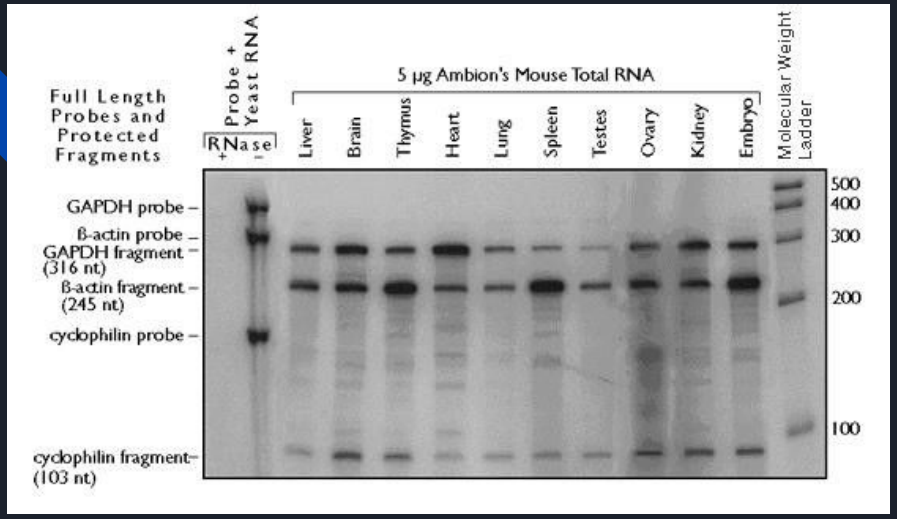
- T1
- T2

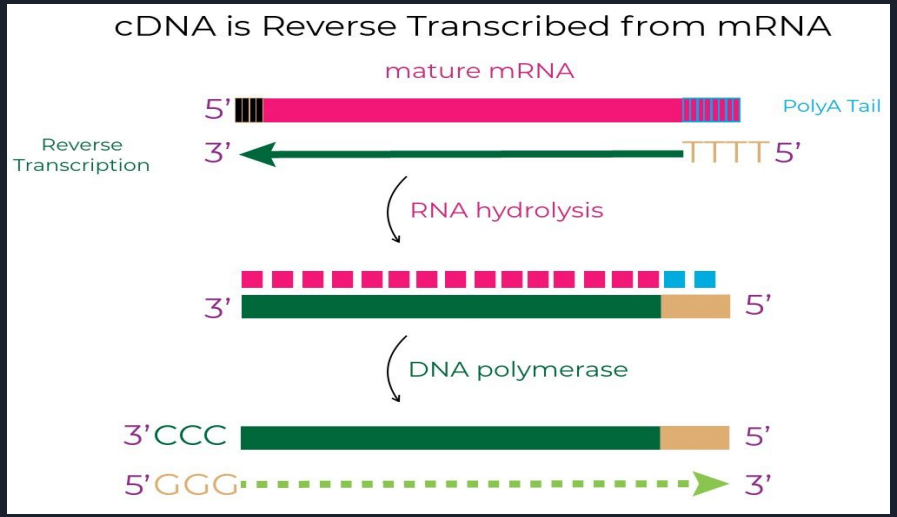
Biology of transcription

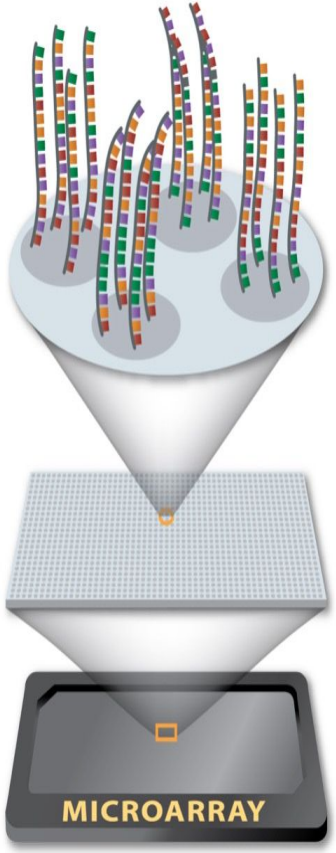


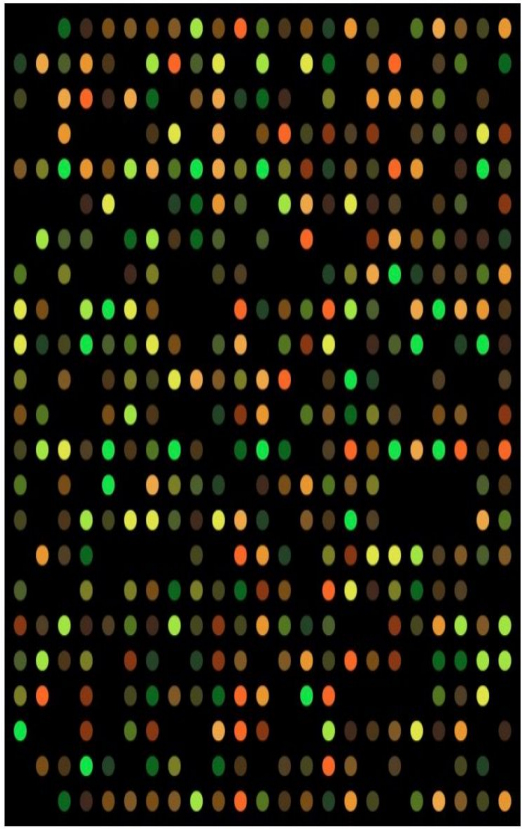


RNA QUANTIFICATION

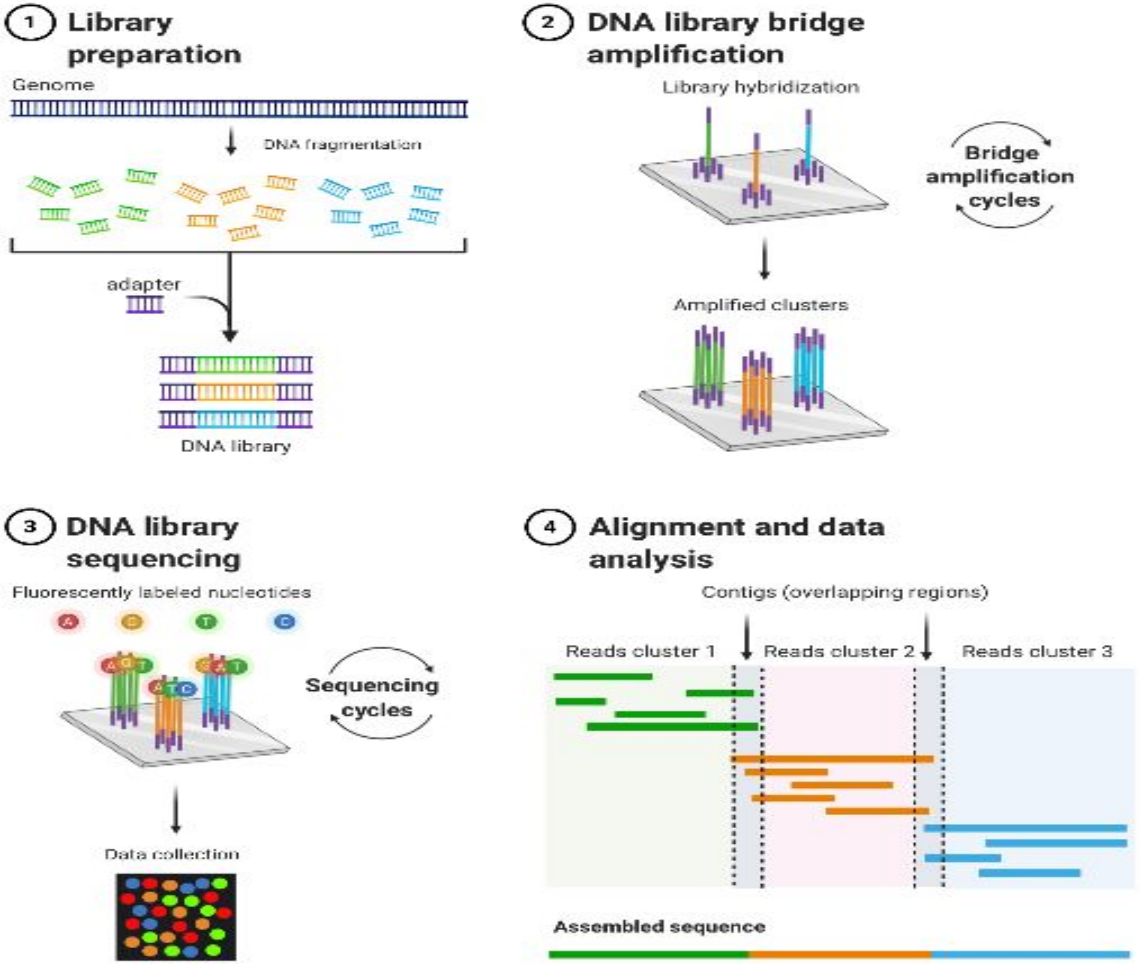


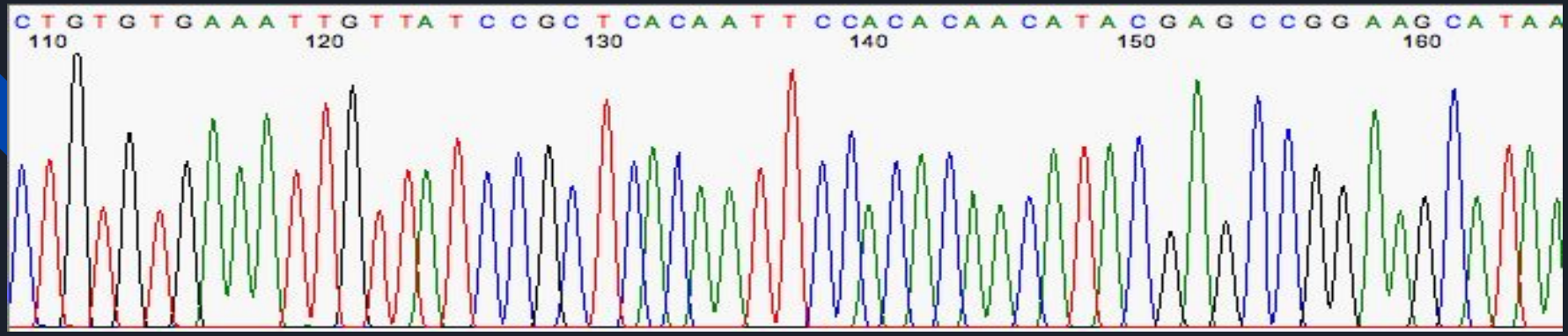


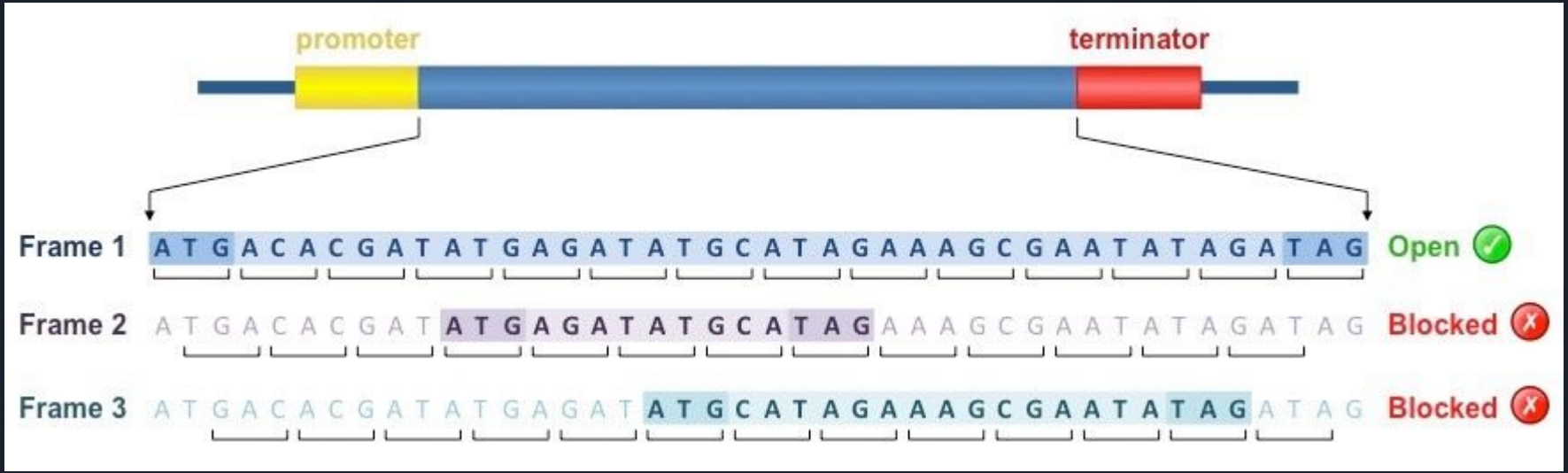




NGS


1 Library preparation
Genome
DNA fragmentation
adapter
DNA library
2 DNA library bridge amplification
Library hybridization
Bridge amplification cycles
Amplified clusters
3 DNA library sequencing
Fluorescently labeled nucleotides
Sequencing cycles
Data collection
4 Alignment and data analysis
Contigs (overlapping regions)
Reads cluster 1
Reads cluster 2
Reads cluster 3
Assembled sequence


C T G T G T G A A A T T G T T A T C C G C T C A C A A T T C C A C A C A A C A T A C G A G C C G G A A G C A T A A
110
120
130
140
150
160


promoter
terminator
Frame 1
A T G A C A C G A T A T G A G A T A T G C A T A G A A A G C G A A T A T A G A T A G
Open
Frame 2
A T G A C A C G A T A T G A G A T A T G C A T A G A A A G C G A A T A T A G A T A G
Blocked
Frame 3
A T G A C A C G A T A T G A G A T A T G C A T A G A A A G C G A A T A T A G A T A G
Blocked

Analysis of Raw RNA-seq Data

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

Step 2: Mapping on a reference genome

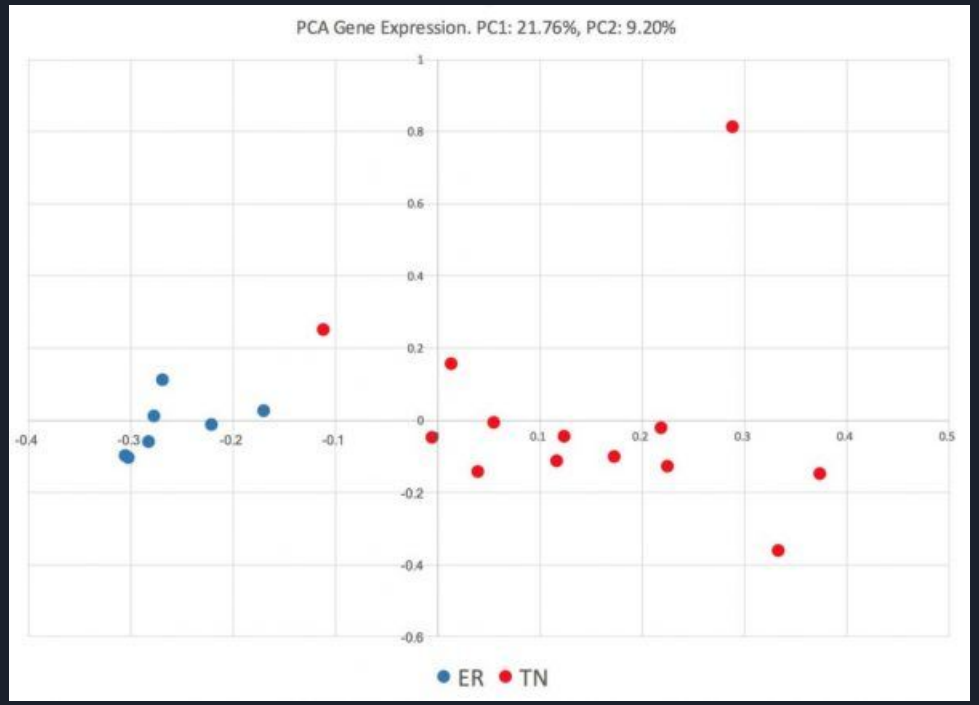
Step 3: Calculating the abundance of reads aligned to the reference genome

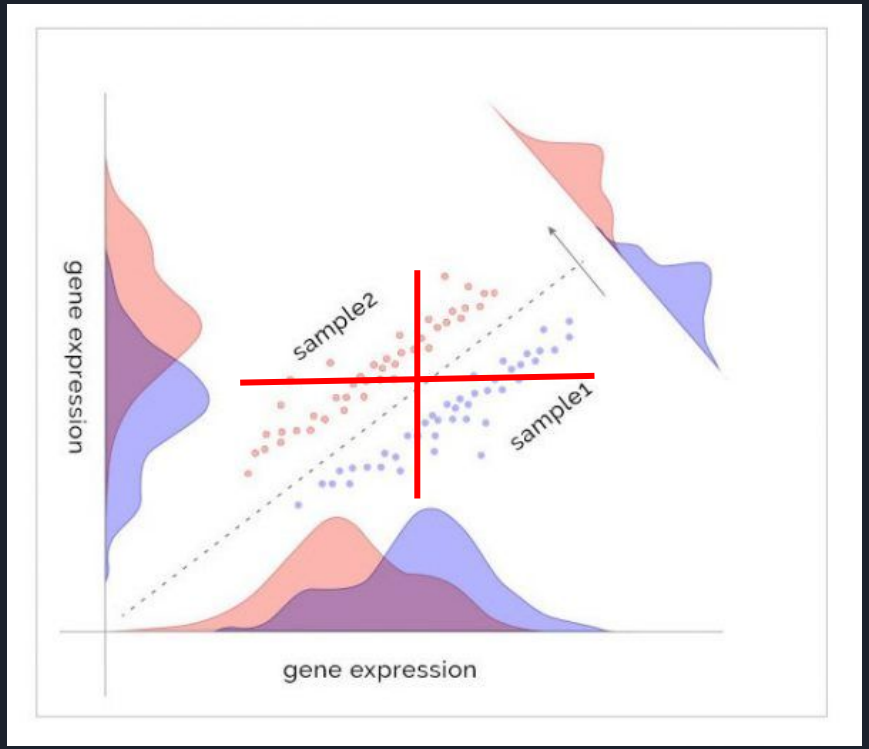
Step 4: Post-processing

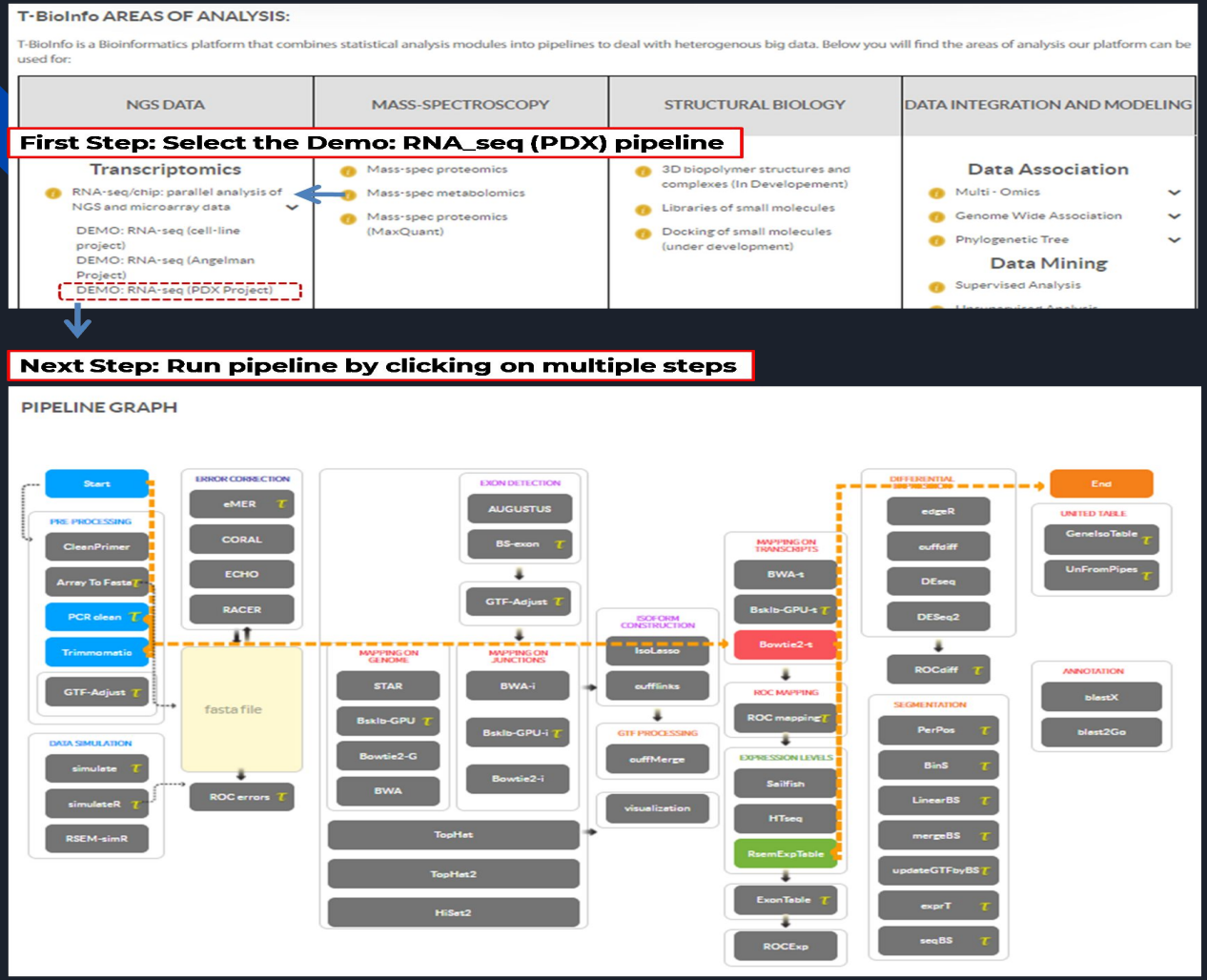
Step 5: Exploratory Analysis

Step 6: Biological Interpretation

PCA


PCA Gene Expression. PC1: 21.76%, PC2: 9.20%
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.4
-0.3
-0.2
-0.1
0.1
0.2
0.3
0.4
0.5
ER
TN

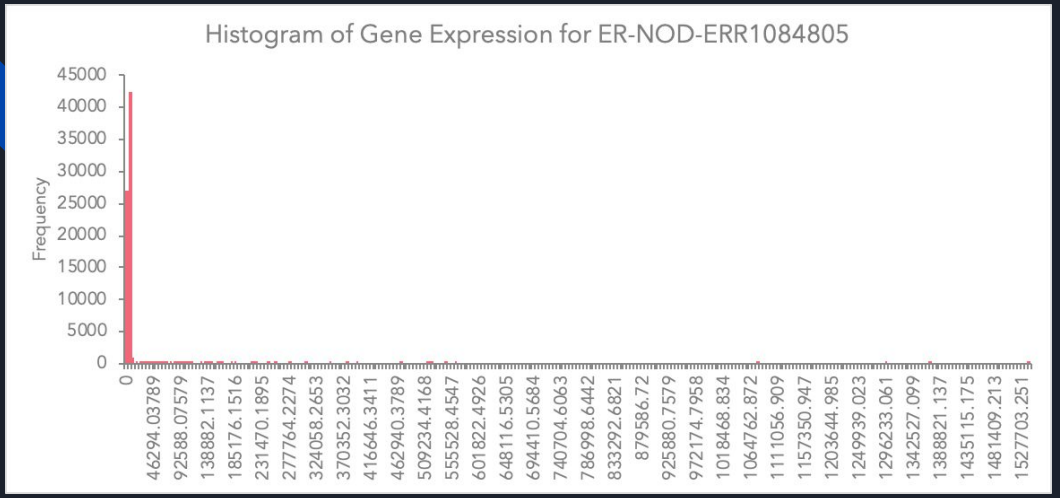

gene expression
sample2
sample1
gene expression

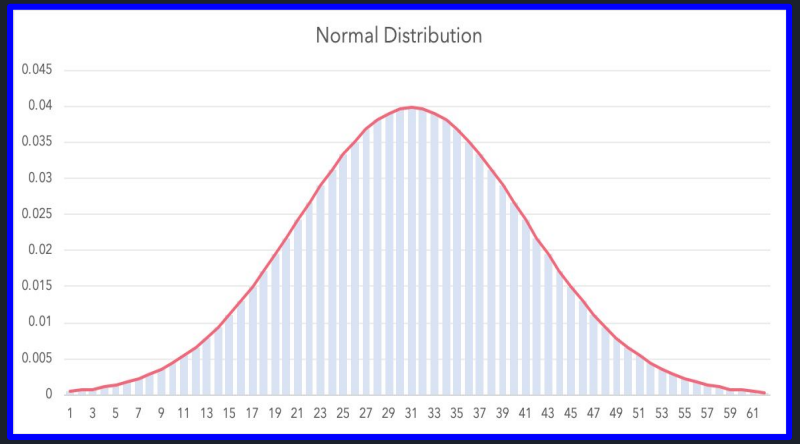


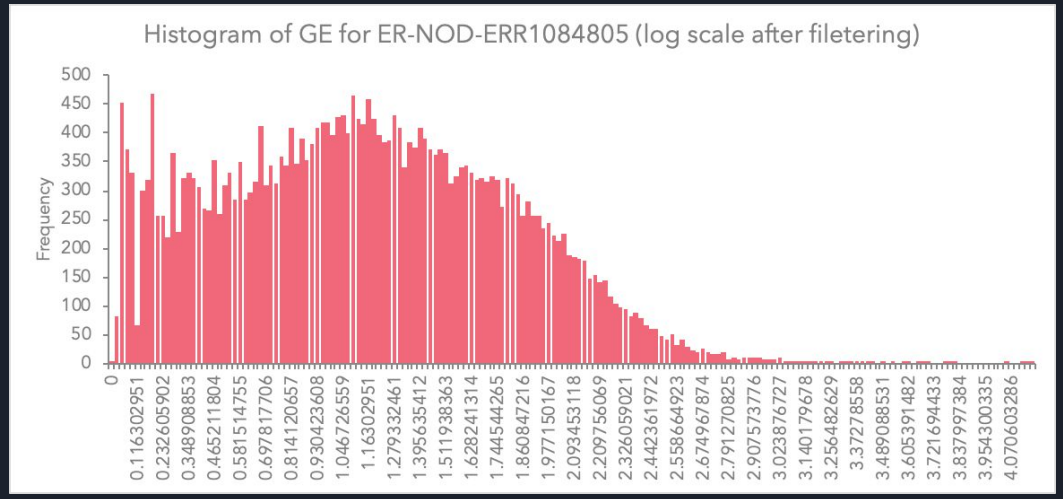
CYP4V2 gene

ER+ tumor sample CYP4V2 genes downregulated

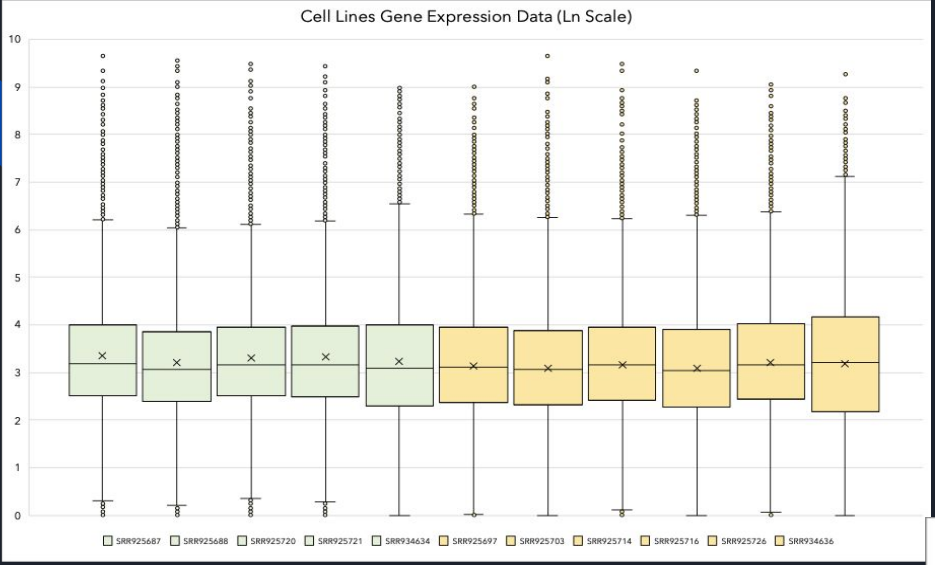
TN- tumor sample CYP4V2 genes upregulated

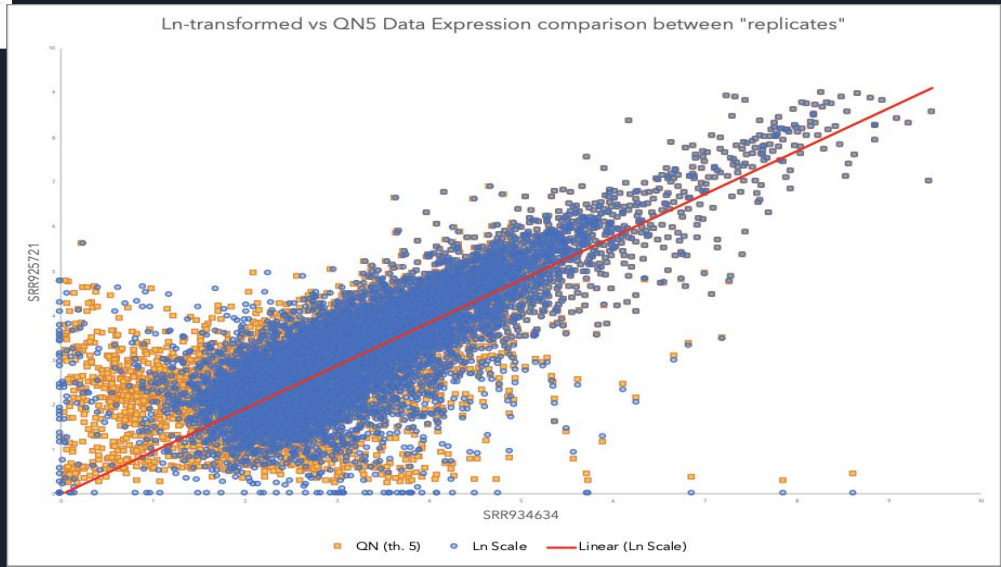






Data Preparation for Downstream Analysis


Cell Lines Gene Expression Data (Ln Scale)
10
9
8
7
6
5
4
3
2
1
0
SRR925687
SRR925688
SRR925720
SRR925721
SRR934634
SRR925697
SRR925703
SRR925714
SRR925716
SRR925726
SRR934636

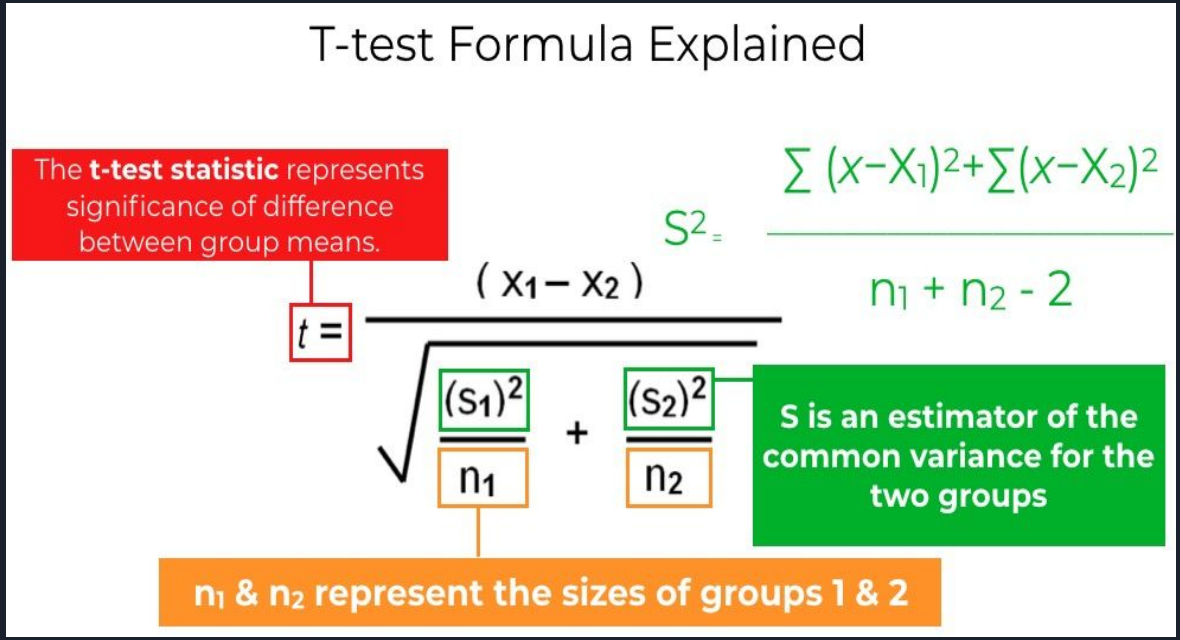

Ln-transformed vs QN5 Data Expression comparison between "replicates"
SRR925721
SRR934634
QN (th. 5)
Ln Scale
Linear (Ln Scale)

Quantile normalization and PCA

Variable	17.79 %	7.36 %	6.45 %	5.75 %	5.46 %
ER-ERR1084763_PE	0.351646	-0.33384	-0.16223	-0.15119	0.053663
ER-ERR1084764_PE	0.342116	-0.24836	-0.18201	-0.18139	0.092981
TN-ERR1084766_PE	-0.21109	-0.54782	0.670529	0.063556	-0.07372
TN-ERR1084768_PE	-0.32142	-0.05837	-0.29367	-0.39249	-0.03926

T-TEST

Ran a demo differential expression pipeline



Differential expression algorithms in Tbioinfo platform

- DEseq
- DEseq2
- Cuffdiff
- EdgeR

TRANSCRIPTOMICS - PART 2

- T3
- Python
- R