Khan Inan
BI-GY 7653
Excercise #5

M5_v1_1) Define transcription and translation

When RNA is copied from a DNA strand, that process is called transcription. When a protein is subsequently synthesized from RNA, it is called translation

M5_v1_2) How can one gene locus produce two different variants of the same gene?

From what I understand, since every person has two parents, we have two genetic variations/alleles per gene locus

M5_v1_3) What kind of questions can we ask about with transcriptomics data?

We can ask if the RNA is a 1 to 1 transcription from the DNA. Additionally we can check for mutations in this step since it is prior to the protein synthesis step.

M5_v2_1) Why is the northern blot not considered “high throughput”?

The main reason it cannot be considered high throughput is due to its main limitation analyzing one gene at a time

M5_v2_2) How are the probes on the microarray designed?

They are designed using various printing/synthesis techniques involving cDNA and oligonucleotides

M5_v2_3) Can microarray technology detect genes that were previously not identified to be on the genome? why or why not?

Yes it can because microarray techniques involve opening a cell and observing the entirety of its genetic components. It doesn’t exclude any material

M5_v3_1) What is competitive hybridization and what is its limitation?

Competitive hybridization is essentially a way to test for cytogenetic variations without a reference sample or cell culture. Its main disadvantage is that you need copy number changes

M5_v3_2) How many probes do Affymetrix have in a gene probeset?

Around 11-20 pairs

M5_v3_3) How is expression intensity measured in RNA-seq?

It mainly compares the levels of expression across different replications of the same expression

M5_v4_1) Why is it important to normalize your data?

Because it allows you to make more accurate extrapolations and also use machine learning more effectively

M5_v4_2) How does MAS5 calculate background noise?

The average of the lowest 2% of intensities within each section

M5_v4_3) How does RMA calculate background noise?

The method that it uses essentially involves assuming that the background noise is normally distributed across the array

M5_v5_1) What are the different steps of quantile normalization?

The steps are that the genes are first ranked by magnitude, then the average value for the genes are calculated, and then finally there is substitution of all the genes that occupy any given rank/average value.

M5_v5_2) Can quantile normalization be used for both, microarrays and RNA-seq? Explain your answer.

Yes because quantile normalization can take from two different data sets in its calculation

M5_v6_1) How do you calculate RPKM and what factors does it account for?

You divide the number of reads by the scaling factor of a million

M5_v6_2) Why is TPM preferred over RPKM?

TPM has a more efficient order of operations for calculation

M5_v7_1) Why do we log our fold change?

It is simpler to work with and it still provides an equivalent value due to its symmetric nature

M5_v7_2) What is the purpose of centering?

It is important because it defines the 0 point and eliminates multiple linearity between variables

M5_v7_3) What is the purpose of scaling?

Scaling allows for larger scale extrapolation and prediction of data outcomes without sacrificing any accuracy that one would obtain from smaller scale arrays

M5_v8_1) Why would genes be detected in RNA-seq but not in microarray?

RNA-seq involves actually sequencing the transcriptome which can result in many alterations/mutations compared to a microarray that is based off predefined reference values.

M5_v8_2) Which type of genes show consistent change in expression across the technologies.

From what I understand protein-coding genes show the most consistent types of change in expression simply due to the nature of selection for these genes and their importance in organism

M5_v8_3) What are some advantages of using RNA-seq over microarray for transcriptomics.

Higher specificity and sensitivity to changes in genes and sequences compared to microarrays