Charlee Cobb - Transcriptomics, Exercise 5

Charlee Cobb

2023-02-27

M5_v1_1) Define transcription and translation Transcription is the act mRNA being created from DNA.Translation uses mRNA molecules to construct amino acids sequences that form proteins.

M5_v1_2) How can one gene locus produce two different variants of the same gene? This happens with a splicing event when introns are removed, and exons are combined in a various number of ways.

M5_v1_3) What kind of questions can we ask about with transcriptomics data? With Transcriptomics data, we can measure the RNA levels in cells to identify their states in life (disease state, developmental stage, etc.) And we can see which genes are expressed in different cells compared to other cells. We can also learn about the function of genes based on their expression, and group genes to see if they have similar functions.

M5_v2_1) Why is the northern blot not considered "high throughput"? A northern blot is not high throughput because you couldn't look at multiple genes at a time, and you had to know the gene sequence and it's complimentary ahead of the experiment.

M5_v2_2) How are the probes on the microarray designed? The probes are placed on solid substrate in an array fashion and you check to see how much of the RNA hybridizes.

M5_v2_3) Can microarray technology detect genes that were previously not identified to be on the genome? why or why not? No because the probes are based on what are previously known. You would have to predict a novel gene in the genome and create a probe based on the prediction.

M5_v3_1) What is competitive hybridization and what is its limitation? Competitive hybridization is where you take two differently labeled samples and allow their molcules to compete with each other to hybridize to the probe. It's a limitation because it's only designed to do a comparison between two conditions.

M5_v3_2) How many probes do Affymetrix have in a gene probeset? Affymetrix is not competitive hybrydization. Affymetrix has 10-20 olligonucleotides probes in a gene probeset, including a mismatch probe.

M5_v3_3) How is expression intensity measured in RNA-seq? It measured by quantification of abudance of a an mRNA molecule in relation to other present mRNAs

M5_v4_1) Why is it important to normalize your data? It's important to normalize data so it's scaled for comparison between tissues or experiments.

M5_v4_2) How does MAS5 calculate background noise? MAS5 is calculates background noise by observing the weighted mean from Tukey, and considers outside factors by normalizing each array.

M5_v4_3) How does RMA calculate background noise? RMA uses quantile normalization and results are given on log scale to calculate background noise and also considers GC content.

M5_v5_1) What are the different steps of quantile normalization? Sort the values, then calculate their average by row, then substitute the new average values for the real ones, then restore order to original state

M5_v5_2) Can quantile normalization be used for both, microarrays and RNA-seq? Explain your answer. Yes because this normalization method can allow you to compare results that have different read amounts per experiment.

M5_v6_1) How do you calculate RPKM and what factors does it account for? You first obtain reads per million (RPM), the divide this values by the length of the gene in kilobases.

M5_v6_2) Why is TPM preferred over RPKM? TPM is preferred over RPKM because you use the same values, but calculate reads per kilobase (RPK), and then you count the total counts and scale that to a million.

M5_v7_1) Why do we log our fold change? We log the fold change because the log values make the data more normally distributed. Values are symmetrical around 0. This helps us use statistically tests that expect normally distributed data.

M5_v7_2) What is the purpose of centering? With Centering you subtract the mean of the data from each value. So the average becomes zero and the values will be symmetrical around that. This allows us to find patterns in data and see the range of high and low values.

M5_v7_3) What is the purpose of scaling? Scaling allows you to compare values in the data in a relative manner, so you can see changes in values relative to the data itself.

M5_v8_1) Why would genes be detected in RNA-seq but not in microarray? You are able to build a denovo trancript with RNA-seq that can let you discover sequences that are novel genes. With the microarray, probes were made to catch RNA based on genes that were already know or gene sequences that were predicted. This leads to genes going undetected in a microarray experiment.

M5_v8_2) Which type of genes show consistent change in expression across the technologies. Highly expressed genes show consistent change in expression across the technologies. So if the gene is highly expressed, it will be picked up by both technologies.

M5_v8_3) What are some advantages of using RNA-seq over microarray for transcriptomics. With RNA-seq you can build a transcriptome de novo, and all protein coding genes are represented in the RNA-seq experiment. With RNA-seq we can conduct differential expression between tissue types. Finally, RNA-seq captures different tyes of RNA.