Hannah Trautmann

BIO539: Big Data Analysis

Title:

Exploratory Analysis of Preliminary qPCR Data in the Investigation of Specialized Ribosomes

Introduction:

Specialized ribosomes refer to structural and possibly functional heterogeneity in ribosomes. There has been shown to be compositional differences among ribosomes in a number of different ways. Primarily, in *E. coli*, there are seven rRNA operons with sequence differences in the genes, leading to different ribosomes (Byrgazov et al. 2013). There can also be post-translational modifications of ribosomal proteins, such as acetylation or methylation, post-transcriptional modifications of rRNA, and exchanges of ribosomal proteins. Of this latter group, many replacements occur because of homologous proteins (Byrgazov et al. 2013). In the model organism being used in this experiment, *Francisella tularensis*, three homologs of one ribosomal protein exist. Because of the shown structural heterogeneity in ribosomes, the question arises as to whether these different ribosomes have distinct functions.

In this study, we have attempted to analyze this using a technique called ChIPPAR, which allows for the investigation of many co-transcriptional processes, such as translation in prokaryotes. The idea behind ChIPPAR is a basic chromatin immunoprecipitation experiment to crosslink proteins of interest to nearby DNA, in the presence or absence of rifampicin (Kambara et al. 2018). Rifampicin is an antibiotic that effectively stops transcription, so rifampicin presence should lead to fewer transcription-

associated proteins, like ribosomes, at the promoter regions of genes. Rifampicin thus acts as an effective control to determine that presence of ribosomes at those genes is indeed due to transcription/translation activities and not background presence (Kambara et al. 2018). Additionally, we analyzed both protein-coding genes and non-protein coding genes, as we expect to see transcription machinery but no translation machinery at the non-protein coding genes. We hypothesized that there would be specific enrichment of ribosomal proteins at protein-coding genes when no rifampicin is added.

Methods:

To test this hypothesis, a plasmid was cloned to overexpress the ribosomal protein bS21 with a VSVG tag, and was then introduced into bacterial cells. Three biological replicates were grown in liquid culture and rifampicin was added to half of each culture during the last hour of incubation. An input sample was taken at this stage. A standard chromatin immunoprecipitation was then completed using α-VSVG beads, and the DNA associated with this bS21 protein was then purified from both the input and immunoprecipitation (IP) samples. Quantitative PCR was used to calculate the relative amount of DNA recovered from a protein coding gene (FTL_1364) and a non-protein coding gene (23s rRNA) to indicate specific enrichment of ribosomal proteins at those two genes, in the presence or absence of rifampicin.

Statistical analysis of the qPCR data was completed using R studio (R Core Team 2018). The values collected are termed CP, or crossing point, which represents the PCR cycle number at which each sample reached a certain threshold of fluorescence. Smaller CP values indicate higher amounts of protein at the gene because more DNA was recovered during immunoprecipitation. First, the CP values for

the three technical replicates of each sample were averaged and standard deviation determined. Next, ΔCP was calculated by comparing the immunoprecipitated DNA to the input DNA for each sample, and these values were averaged for the three biological replicates. ΔΔCP and standard deviation (s) were calculated by comparing the rifampicin-treated samples to the control treatment at both genes, normalizing such that the control treatment samples were set to 0. It is also possible at this step to compare based on gene, rather than based on treatment, as explained in more detail in the Rmarkdown file. Finally, these values were normalized for primer efficiency, by assuming an efficiency of 1.8 and raising this to the additive inverse of the calculated ΔΔCP (i.e. normalized value = $1.8^{-\Delta\Delta CP}$). The data were then plotted using a bar plot with standard deviation used for error bars.

Results:

The results of the qPCR analysis are included in Figure 1 below.

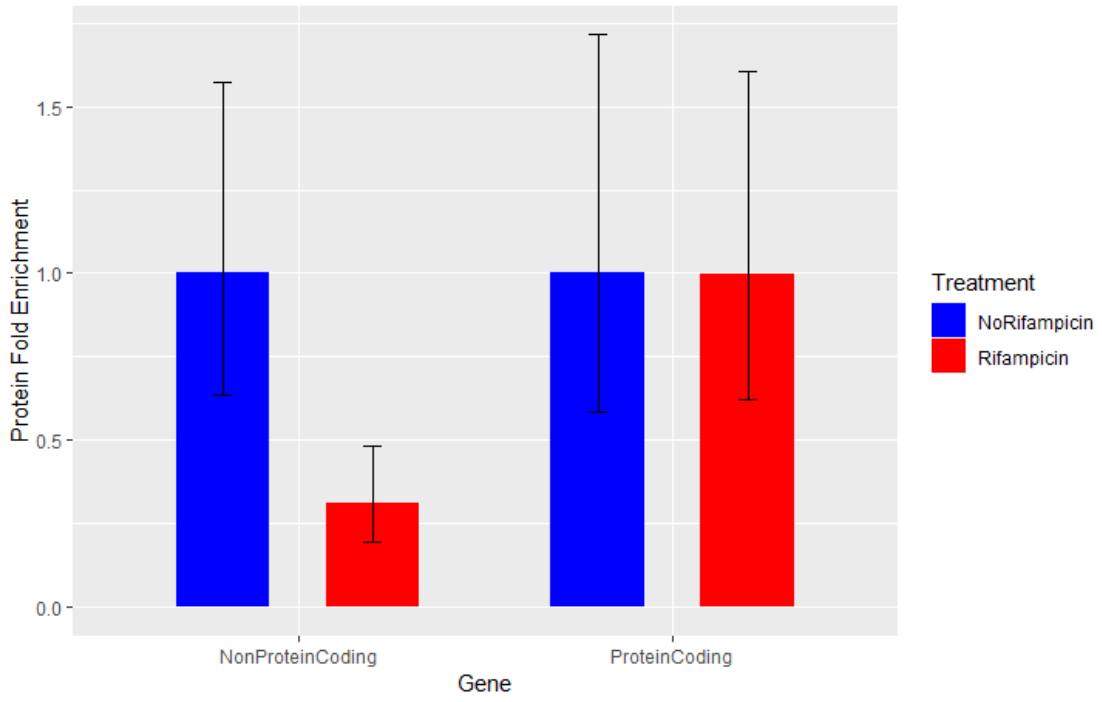

Figure 1: qPCR results normalized to control treatment, with no rifampicin added. Red bars represent those treated with rifampicin, while blue bars represent control treatment. Protein coding refers to amplification of FTL_1364 and non-protein coding refers to 23s rRNA.

Discussion:

There was not a significant difference in protein enrichment based on rifampicin treatment at our protein coding gene, FTL_1364. However, we did see a significant decrease in the amount of immunoprecipitated DNA at the 23s gene in the presence of rifampicin, which was not as expected. This likely reflects that the 23s region is a poor control because we may have nonspecifically immunoprecipitated large amounts of this gene when there were many proteins involved in transcribing the 23s gene, and the nonspecific immunoprecipitation of this gene may be less when there is no transcription of this gene.

To improve this experiment, we will be using only one gene, but controlling for ribosome presence by modifying the ribosomal binding site of the gene. Furthermore, to

reduce the high amount of nonspecific immunoprecipitation, we will attempt a two-step purification process using TAP-tagged B’ RNA polymerase then VSVG-tagged bS21. This will hopefully reduce the number of ribosomes to only those that are actively translating nascent mRNA. Finally, the large error bars indicate inconsistency between biological and technical replicates. Steps will be taken to optimize the ChIPPAR protocol in order to minimize variation.

References:

Byrgazov, Konstantin, Oliver Vesper, and Isabella Moll. 2013. “Ribosome Heterogeneity : Another Level of Complexity in Bacterial Translation Regulation.” *Current Opinion in Microbiology* 16 (2): 133–39. https://doi.org/10.1016/j.mib.2013.01.009.

Kambara, Tracy K, Kathryn M Ramsey, Simon L Dove, Tracy K Kambara, Kathryn M Ramsey, and Simon L Dove. 2018. “Pervasive Targeting of Nascent Transcripts by Hfq Article Pervasive Targeting of Nascent Transcripts by Hfq.” *CellReports* 23 (5): 1543–52. https://doi.org/10.1016/j.celrep.2018.03.134.

R Core Team. 2018. “R: A language and environment for statistical computing.” R Foun dation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.