Identification of Differentially Expressed Genes with CiBER-Seq

David Noble
UC Berkeley

Spring 2021

Abstract

CiBER-Seq is a development on the CRISPR interference method which quantifies changes in expression across genome-wide perturbations. This novel method yields higher granularity with the expense of higher levels of noise. Early iterations of this method relied on pooled sequencing of barcode DNA for control in differential expression analysis. The goal of my project was to perform differential expression analysis on CiBER-Seq data with the constraint of not using barcode DNA counts, and to assess the effect on sensitivity this constraint causes. I performed this analysis using DESeq2, a method which uses negative binomial regression and the empirical Bayes method to shrink dispersion and log-fold change estimates. I attempted to leverage experimental redundancy to increase sensitivity with meta-analysis approaches like Fisher's method and Stouffer's Z-score method, which did not improve sensitivity. The results of my analysis had lower sensitivity than analysis performed without the constraint, but a comparable positive predictive value, highlighting a tradeoff between statistical sensitivity and experimental efficiency.

Acknowledgements

I'd like to thank Dr. Ingolia for the opportunity to work on some truly incredible projects, as well as for challenging me and expanding the way I think. I want to thank Dr. Kendra Reynaud for her constant support and encouragement. Congratulations on the completion of your doctorate! Last but not least, a warm thank you to Dr. Eric Van Dusen for guiding me through the completion of this thesis and for fostering an amazing, diverse community among the Honors cohort.

This thesis is dedicated to my grandparents, Ben and Julia Noble, who financially supported my education and taught me the meaning of perseverance and grace.

Reproducibility

This research project was conducted under the open-source model. All visualizations and results are reproducible with the data and code available in my GitHub repository: https://github.com/dvdnobl/CiBER-Seq.

Table of Contents

I. Introduction

The quantification of specific, molecular phenotypes across a genome-wide set of genetic perturbations is an important approach for understanding gene function and dissecting complex networks of genetic regulation. CiBER-Seq is a novel technique to profile changes in expression for a gene of interest when other genes are perturbed [1]. It accomplishes this by combining CRISPR interference with barcoded reporter expression. CRISPR interference (CRISPRi) uses a catalytically dead Cas9 lacking endonuclease activity with a guide RNA to repress expression of the target gene corresponding to the complementary sequence of the guide [2]. CiBER-Seq utilizes a genome-wide library of guides which are individually transformed into *Saccharomyces cerevisiae* cells. This creates a heterogenous population of cells that each express one distinct guide. Barcoded expressed reporter sequencing is a method adapted for CiBER-Seq that connects guides from the library to their phenotypic consequences. Each guide is given a unique barcode sequence whose expression is driven by the query promoter. High-throughput sequencing of the barcodes reveals the phenotypic profile that reflects the effect of the corresponding guide. A schematic adapted from the CiBER-Seq paper [1] is shown in Figure 1.

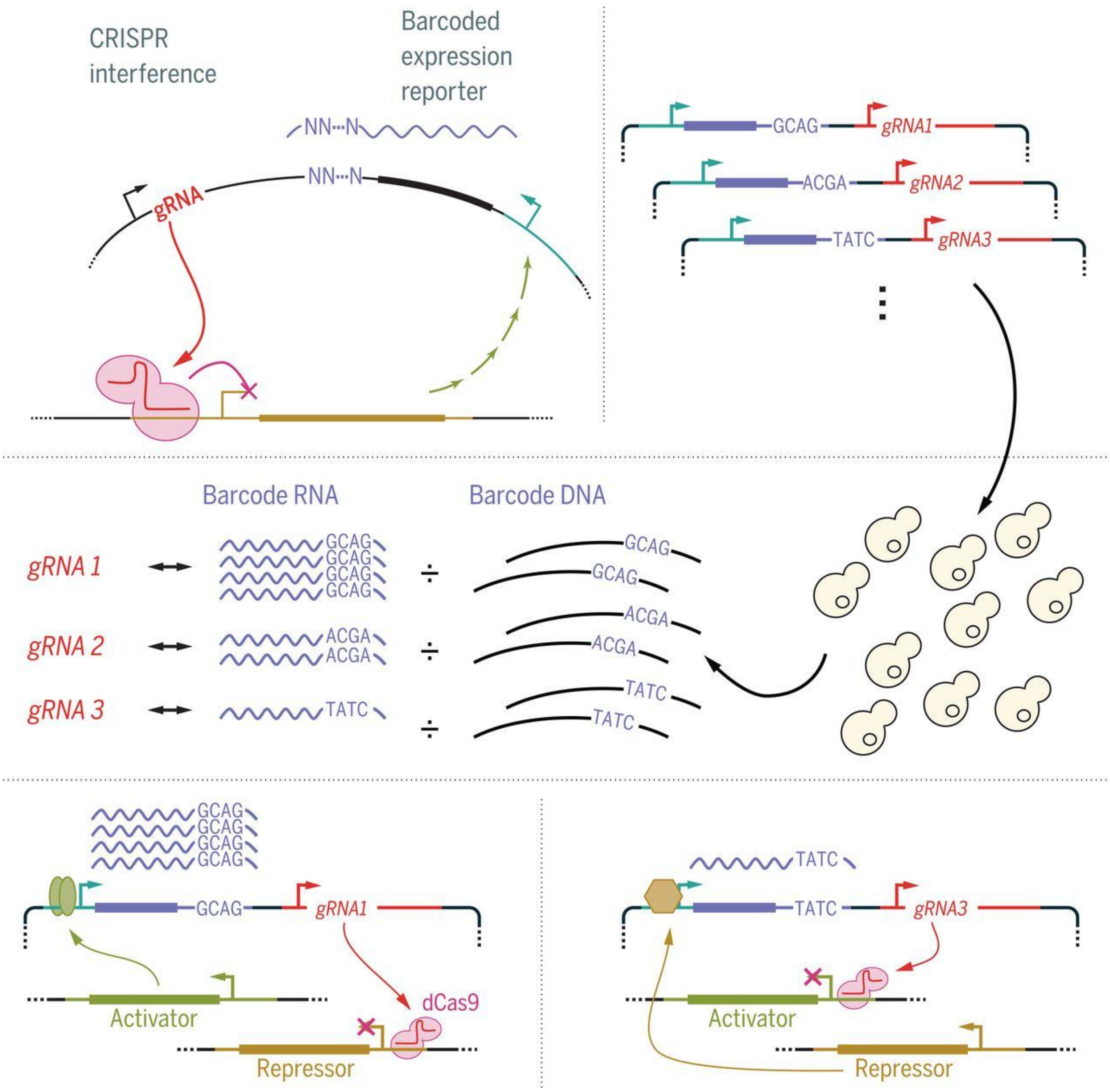


Fig. 1: *Visual schematic of CiBER-Seq* [1]

The phenotypic effect associated with a guide is measured quantitatively by the read count of the associated barcode mRNA. Barcodes are quantified through high-throughput sequencing of library populations. The barcode mRNAs are counted before and after guide induction, and are normalized against the count of barcode DNA present. The log of the ratio of the normalized counts from post-guide induction to pre-guide induction is indicative of the regulatory link between a specific guide and the query promoter. A high log-ratio indicates that the targeted gene is likely a repressor for the gene associated with the query promoter, whereas a low log-ratio indicates that the targeted gene is likely an activator. The identification of likely repressors and activators is done statistically. This type of statistical analysis is known more generally as differential gene expression analysis.

One statistical approach that works well with CiBER-Seq is massively parallel reporter assay linear model (mpralm). This method models log-ratios of count data from a massively parallel reporter assay with a linear model to estimate the change in reporter expression caused by guide induction. The weights for the linear model are estimated from the smoothed empirical variance of the log-ratios and the log-DNA counts [3], as previously described [1]. The use of mpralm analysis for CiBER-Seq experiments on multiple promoters yielded results which could be corroborated by each promoters' known function and regulation. This makes mpralm a good standard for comparing other statistical methods.

The setup outlined above and in Figure 1 requires sequencing of the barcode DNA present in the cell population to normalize counts of the sequenced barcode RNA. Count normalization is required to build accurate phenotypic profiles for the guides, which may have varying levels of barcode DNA present. However, the process for sequencing all of the barcode DNA is expensive and cumbersome. A possible workaround to this is to perform CiBER-Seq with the PGK1 promoter as an additional query promoter. PGK1 is abundantly expressed and has stable mRNA, so it is feasible to normalize the barcode counts for a given guide associated with the primary query promoter to the barcode counts for the same guide associated with the PGK1 promoter. This method is easier to perform, but introduces an additional source of noise.

The goal of this thesis is to develop a statistical approach to model the sequencing count data from CiBER-Seq for the HIS4 promoter using count data for PGK1 for normalization. The HIS4 gene encodes an enzyme involved in the synthesis of histidine, an essential amino acid. The function and regulation for HIS4 is well-known and documented, allowing us to compare the results of our analysis to what is known from other experiments. Finding such an approach requires a model that is more flexible than mpralm and has high statistical power even with the additional noise. Developing a method for this workaround that has comparable power and false-discovery rate (FDR) to mpralm with barcode DNA counts for normalization would expedite CiBER-Seq while reducing its cost. Conducting multiple runs of CiBER-Seq for many promoters would be more feasible, allowing us to produce a large amount of data from which we could find patterns to elucidate regulatory activity on a genome-wide scale.

II. Data

Data description

The primary data for this thesis are the RNA count matrices from two separate CiBER-Seq experiments, one with HIS4 as the query promoter and the other with PGK1. Each of these experiments were conducted in biological duplicate in two turbidostats. Turbidostats are continuous-culturing devices that enable phenotypic analysis of large libraries [4]. Each run-through of CiBER-Seq was conducted with the same library of 53,061 guides, 788 of them being empty guides as a control. There are between approximately 5,300 to 5,400 protein-coding genes in the yeast genome [5], so on average there are 10 guides targeting each gene. Each of the guides targeting a given gene does so at different genomic coordinates relative to the gene's transcription start site, and therefore may have varying levels of repression efficacy. The entire library of guides consists of 195,759 barcodes, so on average each guide has 3-4 barcodes in the library. The redundancy in the guide library helps to reduce noise in our statistical analysis.

The count matrices for HIS4 and PGK1 each have 5 columns. The first column in each is the barcode sequence and acts as the primary key for the table. The other four columns represent the sample from which the sequencing count was taken: left turbidostat before guide induction, left turbidostat after guide induction, right turbidostat before guide induction, right turbidostat after guide induction. Therefore, each row represents the sequencing counts in each sample for a barcode. The HIS4 count matrix has 625,102 rows while the PGK1 count matrix has 610,091 rows. There are more rows in these tables than there are unique guides in the library because single nucleotide sequencing errors can introduce sequence artifacts into the library. Errors in these sequencing reads are easy to filter out, however, by joining the count matrices to a table containing the barcode sequences and guides from the library. The count matrices also contain uncertain sequencing reads. These result when the signal for a specific nucleotide in the sequence is not strong enough to distinguish, and are marked by an "N" in the barcode sequence.

Exploratory data analysis

The bulk of my analysis was done in RStudio. The code for my exploratory data analysis can be found in the GitHub repository. The results shown below are reproducible with the latest version of R and all the relevant dependencies.

I began by joining each of the two count matrices with a table of barcodes and guides from the library. This attached the guide names to the count matrices, allowing me to aggregate the tables by guide. This aggregation was done by summing the sequencing counts of barcodes associated with the same guide. I then joined the two count matrices together by matching guides. To get a preliminary overview of the distribution of counts in the matrix, I plotted a histogram for the sequencing counts in the left turbidostat after guide induction under HIS4, as shown in Figure 2.

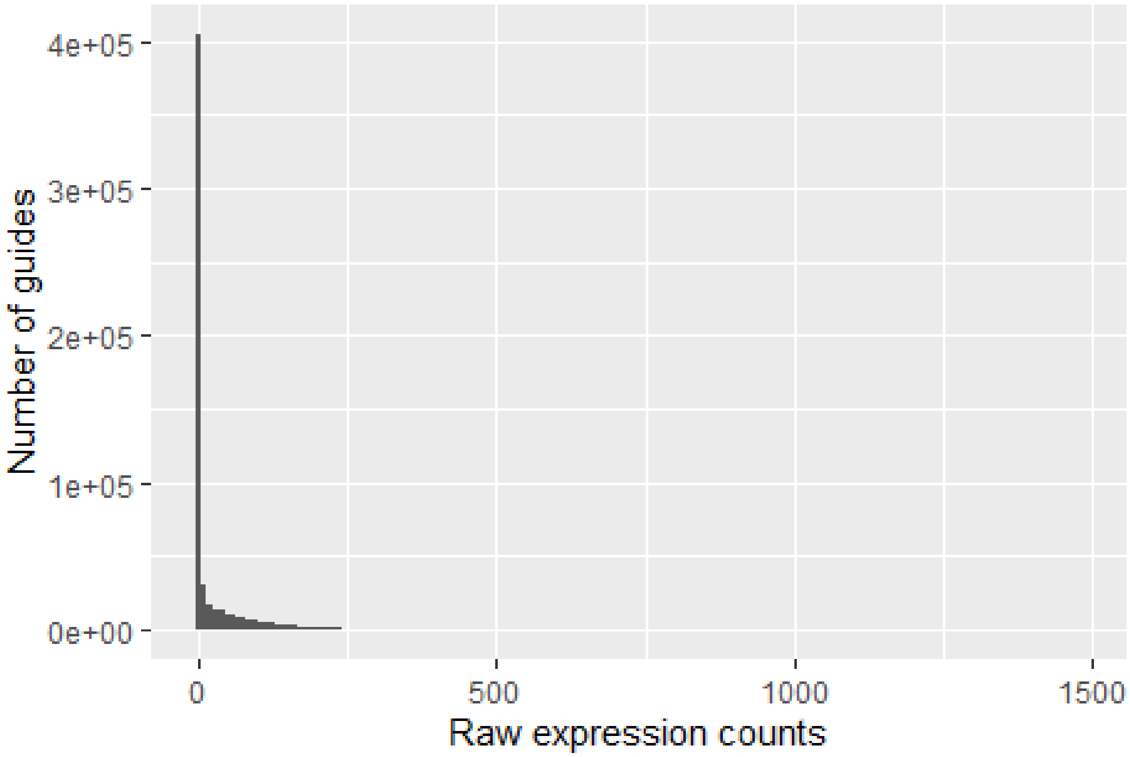


Fig. 2: *Distribution of sequencing counts in CiBER-Seq sample*

The distribution of sequencing counts for the other samples looked very similar, with high enrichment of zeroes. Low sequencing counts, and especially zeroes, are unuseful and add noise to the analysis. I filtered out rows where the sequencing count was less than 32 for any of the samples before guide induction. This filtering of low counts is common practice in differential gene analysis, since they provide little insight and add a lot of noise. The threshold of 32 was chosen empirically from experience in past experiments. The distribution of sequencing counts for the same sample as the figure above is shown in Figure 3. The effect filtering on the data had is clearly seen in the new histogram, where we can get a better sense of the variance and magnitude of sequencing counts.

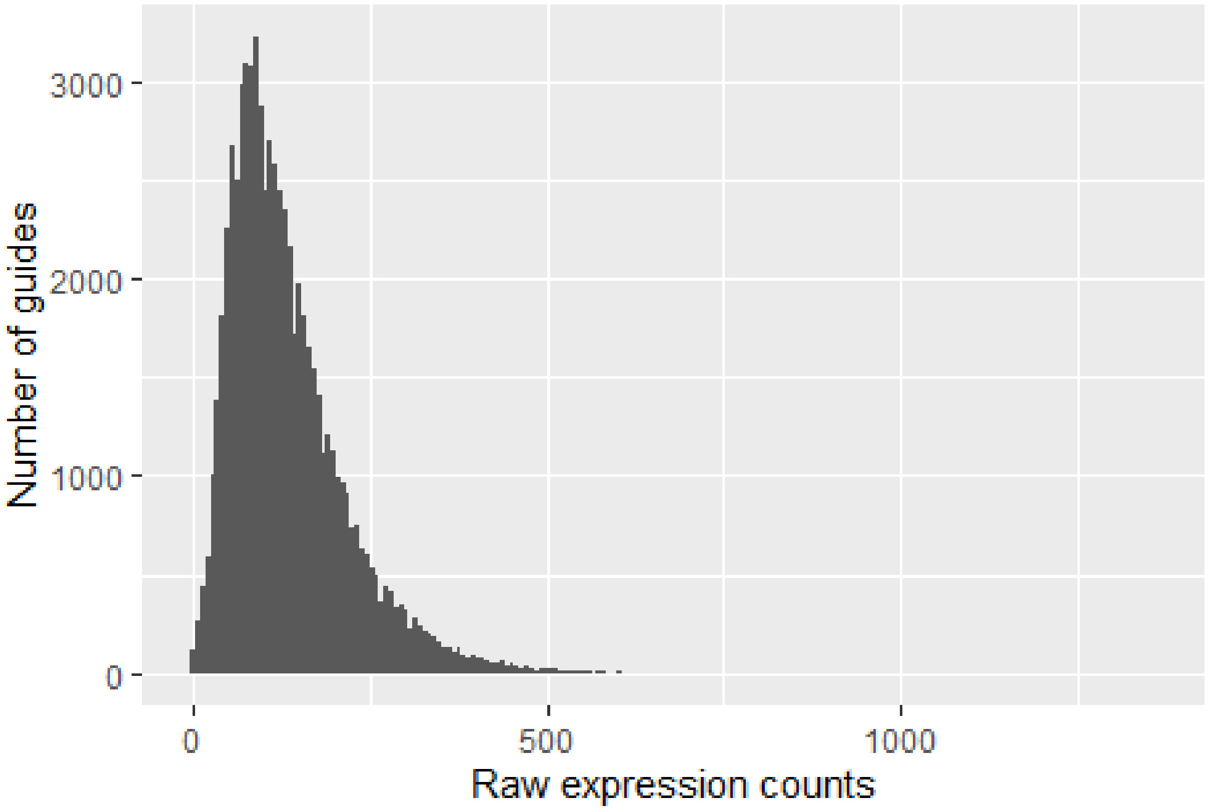


Fig. 3: *Distribution of sequencing counts in CiBER-Seq sample after filtering*

There are many variations of regression methods used for differential gene expression analysis. In order to pick the most suitable one for this dataset, it is crucial to understand the relationship between the means and variances within each sample. Below in Figure 4, I have plotted the mean versus the variance among samples for HIS4 and PGK1 respectively. The blue lines in each plot represent the constant line for the means, while the red lines represent a simple linear regression between the mean and variance. In both, the red lines are above the blue and have sharper slopes, indicating that the variance is greater than the mean in our samples. The prevalence of overdispersion rules out using Poisson regression to model our data. Rather, a negative binomial distribution would be a better model, since it mixes the Poisson and gamma distributions to accommodate for heterogeneity in the data.

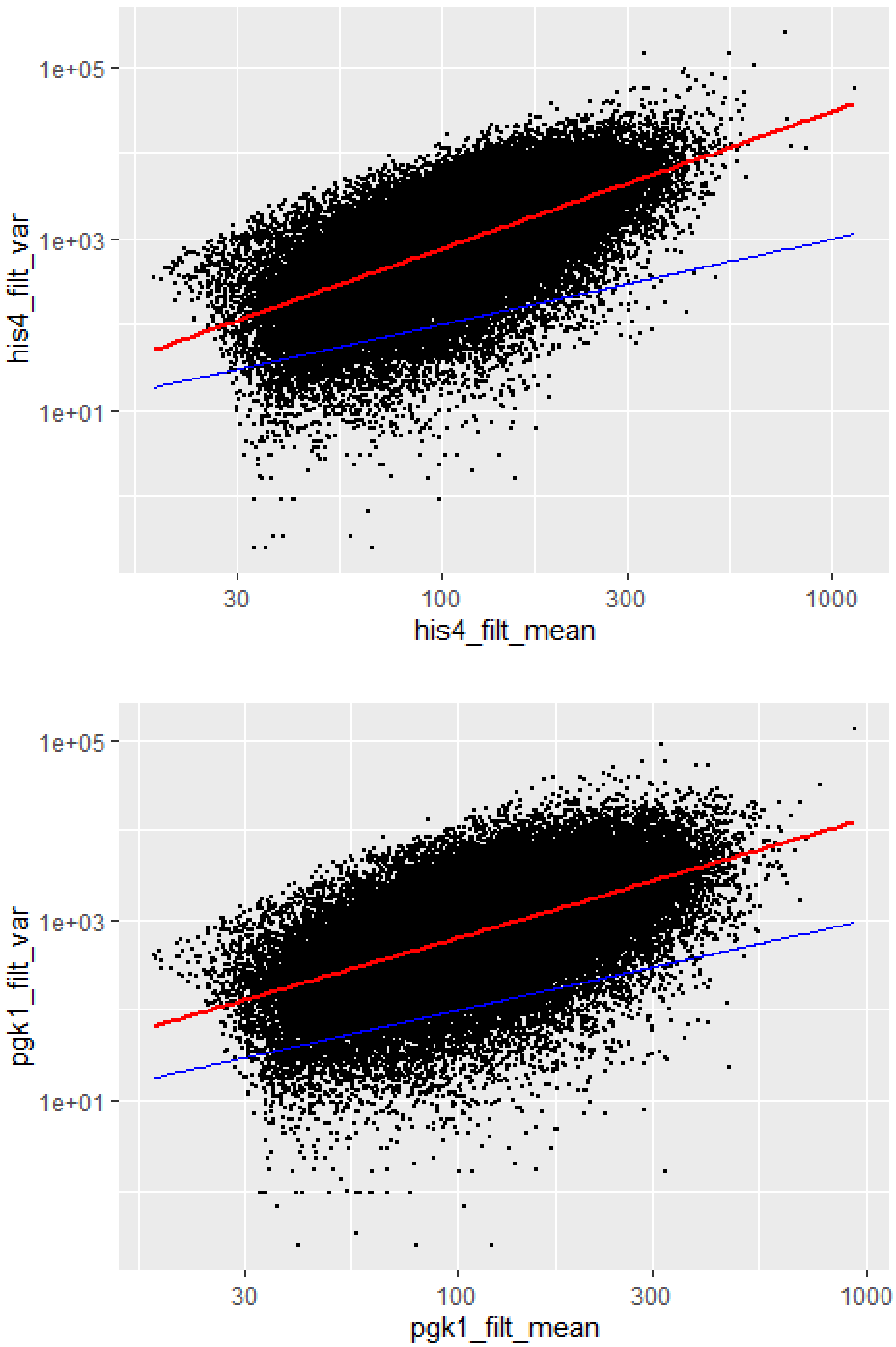


Fig 5: *Variance vs. Mean in HIS4 and PGK1 samples*

To further understand the variance of the data, I performed principle component analysis (PCA). This also served as data quality control by allowing me to observe if the main sources of variance were represented by the experimental conditions. Figure 6 below shows the result of this PCA colored three ways, one for each condition. The first principal component is the query promoter condition (HIS4 vs. PGK1), while the second principal component is the guide induction condition (pre-guide induction vs. post-guide induction). As the PCA shows, the left vs. right turbidostat condition was not a significant source of variance. This is expected and helps confirm the quality of our data, since the two turbidostats are experimental replicates. The first two principal components are also as expected, since the goal of the experiment is to distinguish guides corresponding to differential expression after being induced between HIS4 and PGK1.

The results of this exploratory data analysis provide crucial insights into the quality of the data as well as the assumptions that can be made fairly regarding the underlying distributions. PCA confirmed that the experimental conditions from these runs of CiBER-Seq were indeed the primary sources of variance in the data, and that the replicates are not a significant source of variance. The scatter plots shown in Figure 5 illustrate the unequal relationship between sample mean and variance, suggesting that the more flexible negative binomial model is more appropriate than the Poisson distribution for modelling the count data.

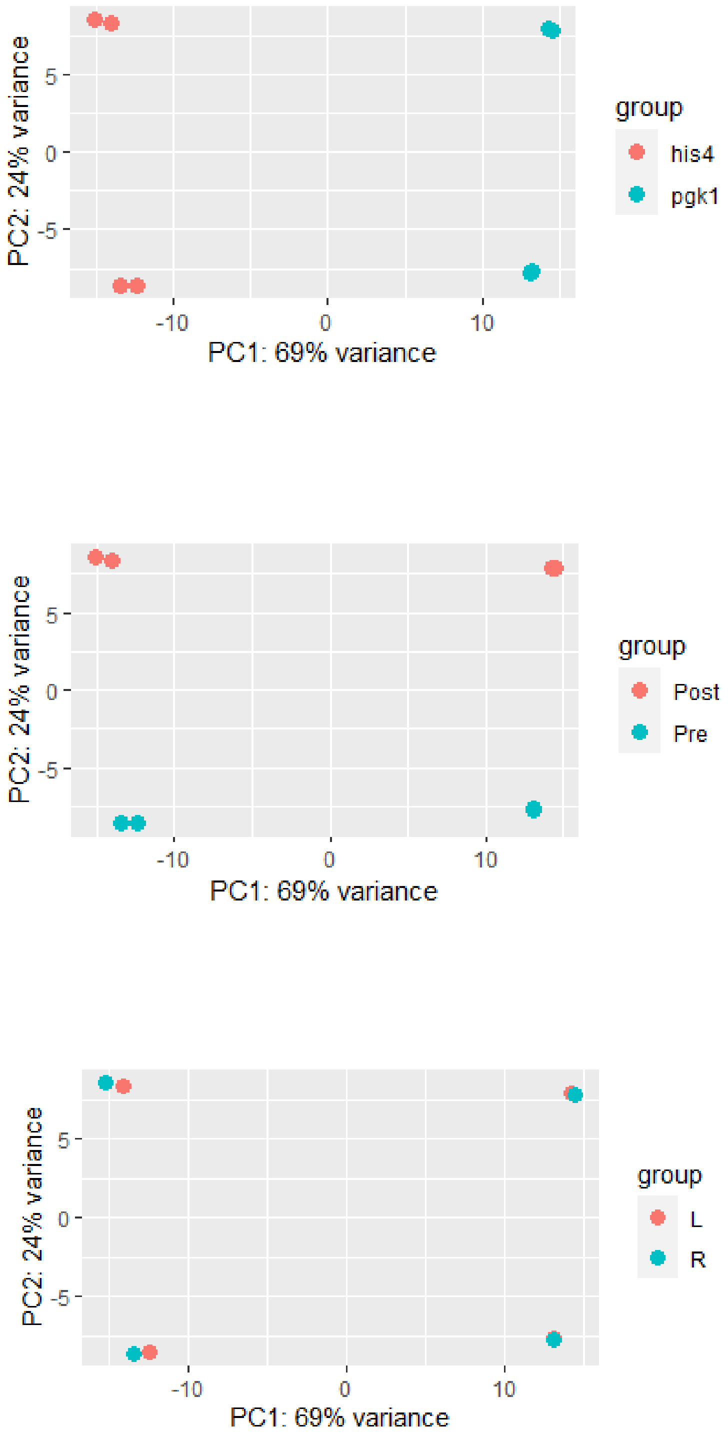


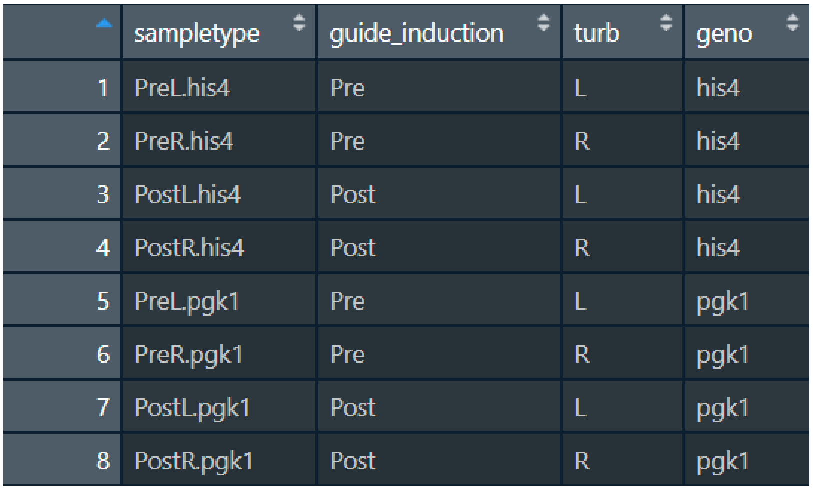
Figure 6: *PCA of count data*

III. Methods

DESeq2

The insights gained from my exploratory data analysis suggested DESeq2 as a strong contender for modelling the data. DESeq2 is a method for differential analysis of count data designed for RNA-Seq that is available as an R package [6]. This method uses negative binomial regression and estimated shrinkage for dispersion and fold change to handle noisy data. To account for low counts, high variance, and non-normality, DESeq2 pools information across genes (in our case, the guides) with similar counts to make assumptions about similarity of variance. These features of DESeq2 give it high precision and sensitivity even with noisy data like ours.

Using DESeq2 begins with creating a metadata table which includes the sample information for the count matrix. The metadata table is included below in Figure 7. This table provides the statistical environment we will create with DESeq2 with context about the sample type for each variable, and will be used to extract results after performing the analysis at a later step. To tell DESeq2 how we would like to compare the factors in the metadata table, we input the design formula as follows: design = ~ guide_induction * geno + turb. This means we want to model changes in expression for HIS4 and PGK1 before and after guide induction, and to account for the experimental replicates.



	sampletype	guide_induction	turb	geno
1	PreL.his4	Pre	L	his4
2	PreR.his4	Pre	R	his4
3	PostL.his4	Post	L	his4
4	PostR.his4	Post	R	his4
5	PreL.pgk1	Pre	L	pgk1
6	PreR.pgk1	Pre	R	pgk1
7	PostL.pgk1	Post	L	pgk1
8	PostR.pgk1	Post	R	pgk1

Figure 7: *Metadata table for DESeq2*

One feature of DESeq2 that is particularly useful for these data is the flexibility to use factor crossing in the model. This is crucial for our schema since it allows for comparing counts across HIS4 and PGK1 for pre-induction and post-induction conditions. Other tools developed for differential analysis like mpralm rely on DNA counts to model the relationship between copy number and the variability of the outcome measure. Since we have constrained ourselves not to use DNA counts, this tool is unsuitable for our schema.

The DESeq2 package for R creates a statistical environment which automates each step of the analysis. The first step of this analysis is to normalize the sequencing counts to account for factors such as sequencing depth and RNA composition. The next step is the estimation of guide dispersions. DESeq2 will then use the empirical Bayes method to shrink those dispersion estimates. Next, the raw counts will be modeled with the negative binomial distribution and fitted to estimate coefficients. These coefficients are the estimates for the log2 fold change for each sample group, and are shrunken using empirical Bayes again. Finally, hypothesis testing is performed to test for differential expression.

Estimating size factors

Various factors such as sequencing depth and RNA composition can affect sequencing counts obtained from CIBER-Seq in uninteresting ways that we need to account for. To accomplish this, DESeq2 normalizes sequencing counts using the median of ratios method [7]. This method divides the raw counts by the estimated size factor, calculated as the median ratio of guide counts relative to the geometric mean per guide.

Dispersion estimation and shrinkage

DESeq2 uses a measure of dispersion which relates the variance and the mean with the following equation:

$$\text{Var} = \mu + \alpha * \mu^2$$

The dispersion value α is directly related to variance and inversely related to the mean, so guides with low count means have high dispersion and counts with high means have low dispersion. DESeq2 estimates dispersion for each guide using maximum likelihood estimation.

Our data have small sample sizes for each guide (3 to 4 as mentioned in the Data section), so dispersion estimates are likely to be noisy. To accommodate for this noise, DESeq2 uses the empirical Bayes method to share dispersion estimates across guides with similar counts to shrink them. This is done by calculating the expected dispersion for guides given an expression strength. This results in a curve that DESeq2 will fit the dispersion estimates to, depending on how close the estimate is to the curve and the sample size of the guide. This shrinkage helps leverage the redundancy in our dataset to reduce noise and false positive rate. Figure 8 below depicts a plot of the dispersion estimates along with the fitted curve for our data. The estimates follow a trend of decreasing as the mean of normalized counts increases, which is expected. Additionally, the estimates cluster together around the fitted curve, indicating that shrinkage was applied appropriately.

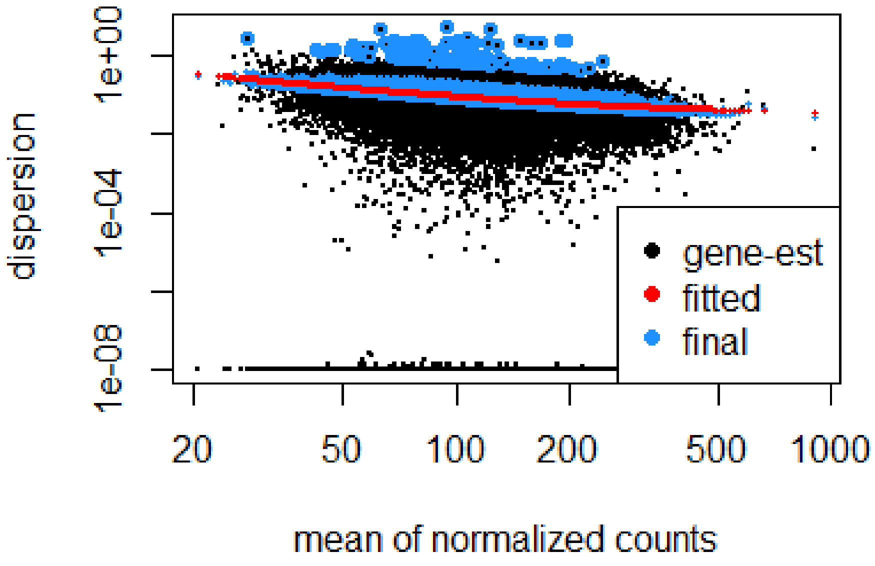


Figure 8: *DESeq2 dispersion estimates and fitted curve*

GLM fitting

DESeq2 then takes the normalized counts and shrunken dispersion estimates to fit a negative binomial distribution to each guide. This modeling step is described by the following equation:

$$K_{ij} \sim NB(s_{ij}q_{ij}, \alpha_i)$$

The letter *i* indexes the guide and *j* indexes the sample group. *K* represents the raw count, *s* is the estimated size factor, *q* is a quantity proportional to the expected true concentration of fragments, and α is the shrunken dispersion estimate. After the model is fit, DESeq2 will estimate coefficients using the following formula:

$$\log_2 q_{ij} = \Sigma_r x_{jr}\beta_{ir}$$

where x is the design matrix and β is the coefficient representing the log2 fold-change for the sample group.

Shrinkage of LFC estimates

A common issue in dealing with sequencing count data is heteroskedasticity. Guides with low counts tend to exhibit high dispersion, which gives the LFC (logarithmic fold-change) estimates high variance. Similarly to the shrinkage of dispersion estimates, DESeq2 uses the empirical Bayes method to shrink the fold-change estimates to zero when the information for a guide is low. The distribution of LFC estimates for all guides is used as the prior which each LFC estimate is shrunken to. The effect of this is that LFC estimates for guides with higher dispersion are shrunken towards zero, solving the issue of exaggerated LFC estimates for guides with low counts [8].

The purpose of this shrinkage step can be expressed in terms of a bias-variance tradeoff. LFC estimates for guides with higher counts have low variance and low bias, so this method will not shrink them very much. On the other hand, this shrinkage will reduce the variance for LFC estimates on low counts at the cost of moving the bias to zero. This is not done by DESeq2 automatically and does not affect the number of guides identified as causing significant differential expression. This step is only needed if more accurate LFC estimates are needed in downstream analysis.

Hypothesis testing

DESeq2 uses the Wald test to identify guides causing differential expression. We take the null hypothesis that there was no differential expression across sample groups, or an LFC of zero. This is tested against the alternative hypothesis that there was differential expression across sample groups. We use a significance level of 0.05. P-values obtained from the Wald test are adjusted for multiple testing using the Benajimi-Hochberg method [9].

Figure 9 below shows the result of fitting our data and performing the Wald test as an MA plot. Each point represents a guide. The horizontal axis represents the mean counts for a guide, and the vertical axis represents the estimated LFC for that guide. The plot below depicts the non-shrunken LFC estimates. Points on the plot that are highlighted blue represent guides found to cause differential expression. Points that lie beyond the vertical range of the plot are depicted as triangles on the upper or lower bounds.

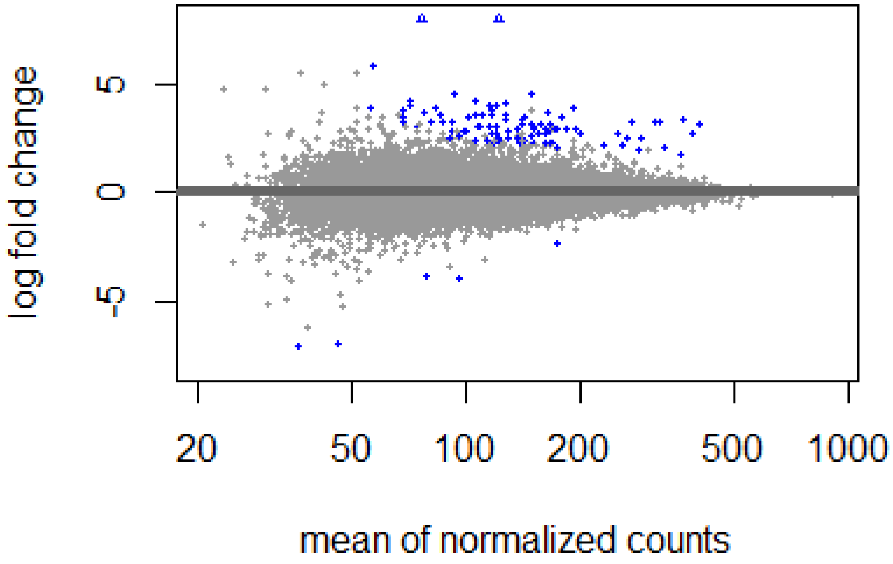


Figure 9: MA plot of results

Meta-analysis

The results from our analysis with DESeq2 gives us LFC estimates for each guide. As mentioned previously, our data contains multiple guides for each gene. The question arises: how do we aggregate the results for each guide to draw conclusions about differential expression caused by perturbation of each gene? The goal of this meta-analysis is to answer this question while maximizing the sensitivity of our overall analysis. We measure sensitivity by the number of genes found to cause differential expression.

One simple approach to this question is to aggregate guides by minimum adjusted p-value. That is, we conclude a gene was differentially expressed if at least one of the guides targeting it had an adjusted p-value below our significance level of 0.05. This method has the benefit of being the simplest to interpret. However, it would be ideal to share information across all the guides for a given gene and further leverage the redundancy in our experimental design. This method serves as our baseline to compare sensitivity with other meta-analysis methods.

A situation that could arise from our analysis is that one gene had multiple guides each with a low adjusted p-value, but none of them below 0.05. Since the significance level is a hard cutoff, we would classify that gene as insignificant using the method above. To increase our sensitivity, it may help to combine the results of each guide's test. One method for doing this is Fisher's method. Fisher's method combines the result of multiple independent tests under the same hypothesis. The test statistic is calculated as such:

$$X_{2k}^{2} \sim -2\sum_{i=1}^{k}\ln(p_i)$$

The key assumption is that each test is independent, which may not be the case for our data. When applying this method to tests with positive dependence, evidence against the null hypothesis is overstated. After applying this method, however, I did not find any new genes that were neglected by the baseline minimum method.

Another meta-analysis method related to Fisher's method is Stouffer's Z-score method. This method has nearly identical power to Fisher's method. However, the formulation it uses makes it easy to incorporate weights, allowing us to give certain guides more importance than others. This method is formulated as:

$$\mathrm{Z} \sim k^{-1/2} \sum_{i=1}^{k} Z_i$$

where $Z_i = \phi^{-1}(1 - p_i)$ and$\phi$is the standard normal CDF.

The ability to differentially prioritize specific guides seemed promising because biologically, guides have varying efficacy. The efficacy of a guide can depend on its location in a gene it targets. Thankfully, I had access to estimates for guide efficacy for each guide in the library. These efficacy estimates are the probability the guide results in successful suppression of the gene. However, after incorporating the efficacy estimates as weights when applying Stouffer's method, we again did not discover new genes neglected by the baseline minimum method.

The failure of the meta-analysis methods to increase our sensitivity is not conclusive proof that they would not work again with another run of CiBER-Seq. It merely indicates that the simplest approach of aggregating guides targeting one gene by their minimum adjusted p-value provided the highest sensitivity for our particular data. In the future, these methods may prove useful for identifying genes whose guides individually had weak signals, but in total have a stronger signal.

IV. Results and Discussion

The results of DESeq2 identified 99 guides causing differential expression. After aggregating these by the criterion that a gene knock-down causes differential expression if at least one guide targeting it causes differential expression, we are left with 49 genes. To put this result into context, I will compare it to a similar analysis done on the HIS4 count data without the constraint of not sequencing barcode DNAs. When these CiBER-Seq experiments were conducted, members of the lab also sequenced barcode DNAs so that we could compare results of our analysis with the constraint to results of an analysis done without the constraint. Analysis using the barcode DNA counts was performed with mpralm, which models log-ratio activity with linear models and estimates weights in the model by smoothing the relationship between the variance of the log-ratios and estimated log-DNA copy number.

Due to the lower amount of noise in the data when including barcode DNA counts, we would expect that mpralm would have higher sensitivity and identify more differentially expressed guides. Since my goal is to determine whether it is possible to distinguish differentially expressed guides with the experimental constraint using DESeq2, it is useful to compare the results from both in terms of sensitivity. However, due to the uncertainty inherent in both analyses, it is worth mentioning that it is impossible to truly know the quantified sensitivity produced by our results. We can merely use the number of differentially expressed guides and genes identified as significant to be a measure for the analyses' ability to distinguish signal from noise, and use that as a proxy for sensitivity.

That being said, the mpralm analysis identified 1,364 guides causing differential expression. When aggregating with our minimum method, mpralm identified 961 genes. This is clearly a much greater magnitude than the results obtained from DESeq2. What is promising, however, is that the set of genes identified by DESeq2 is nearly a subset of the genes identified by mpralm. Of the 49 genes identified by DESeq2, each except for 4 of them belonged in the set of genes found by mpralm. Therefore, although DESeq2 with the constraint has less sensitivity than mpralm without the constraint, they have comparable positive predictive values.

The constraint of using PGK1 counts for control rather than barcode DNA counts resulted in the predictable outcome of higher noise and reduced sensitivity. What my analysis shows, though, is that using DESeq2 provides similar positive predictive value to mpralm analysis done without the constraint. Though meta-analysis methods like Fisher's method or Stouffer's Z-score method did not improve sensitivity on our particular dataset, it may prove useful in future iterations of CiBER-Seq to leverage experimental redundancy and increase sensitivity. Fundamentally, my analysis highlights the tradeoff between statistical sensitivity and experimental efficiency. For future studies wishing to reduce cost and prioritize collecting a large magnitude of data quickly, following my analysis methods may provide a pragmatic solution.

V. References

[1] Muller, Ryan, Zuriah A. Meacham, Lucas Ferguson, and Nicholas T. Ingolia. “CiBER-Seq Dissects Genetic Networks by Quantitative CRISPRi Profiling of Expression Phenotypes.” *Science* 370, no. 6522 (2020). https://doi.org/10.1126/science.abb9662.

[2] Qi, Lei S., Matthew H. Larson, Luke A. Gilbert, Jennifer A. Doudna, Jonathan S. Weissman, Adam P. Arkin, and Wendell A. Lim. “Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression.” *Cell* 152, no. 5 (2013): 1173–83. https://doi.org/10.1016/j.cell.2013.02.022.

[3] Myint, L., Avramopoulos, D.G., Goff, L.A. *et al.* “Linear models enable powerful differential activity analysis in massively parallel reporter assays.” *BMC Genomics* 20**,** 209 (2019). https://doi.org/10.1186/s12864-019-5556-x

[4] McGeachy, Anna, Zuriah Meacham, and Nicholas Ingolia. “An Accessible Continuous-Culture Turbidostat for Pooled Analysis of Complex Libraries.” *ACS Synthetic Biology* 8, no. 4 (2019): 844–56. https://doi.org/10.1021/acssynbio.8b00529.

[5] Mackiewicz, Pawel, Maria Kowalczuk, Dorota Mackiewicz, Aleksandra Nowicka, Malgorzata Dudkiewicz, Agnieszka Laszkiewicz, Miroslaw R. Dudek, and Stanislaw Cebrat. "How Many Protein-Coding Genes Are There in The Saccharomyces Cerevisiae Genome?" *Yeast* 19, no. 7 (2002): 619–29. https://doi.org/10.1002/yea.865.

[6] Love, Michael I, Wolfgang Huber, and Simon Anders. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15, no. 12 (2014). https://doi.org/10.1186/s13059-014-0550-8.

[7] Anders, Simon, and Wolfgang Huber. "Differential Expression Analysis for Sequence Count Data." *Genome Biology* 11, no. 10 (2010). https://doi.org/10.1186/gb-2010-11-10-r106.

[8] Zhu, Anqi, Joseph G Ibrahim, and Michael I Love. "Heavy-Tailed Prior Distributions for Sequence Count Data: Removing the Noise and Preserving Large Differences." *Bioinformatics* 35, no. 12 (2018): 2084–92. https://doi.org/10.1093/bioinformatics/bty895.

[9] Benjamini, Yoav, and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57, no. 1 (1995): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.