Methodology Section

Genetic Two-Group Comparison

We used a negative binomial model to find genetic targets for treatment of RP. To do so, we analyze analyze gene expression changes between the diseased (Age of 30 Days) and treated groups (Treated 30/Aged to 37 Days). In doing so, we first normalized raw counts across samples by using the transcript per million or TPM method, which normalizes by gene length first and then by the sequencing depth (quantity of reads per sample). Next, we measured dispersion for each gene using maximum likelihood estimation. These gene-specific dispersion values were shrunk to the expected dispersion value using DESeq's empirical bayes method. We utilized DESEQ2 functions in R to fit this model to each gene (Love et al. (2014)):

The model takes the form:

$$\begin{aligned} K_{ij} &\sim NB(\mu_{ij}, \alpha_i) \\ \mu_{ij} &= s_j q_{ij} \\ log_2(q_{ij}) &= x_j \beta_i \end{aligned} \tag{1}$$

where counts K_{ij} for gene i, sample j are modeled using a negative binomial distribution with fitted mean μ_{ij} and a gene-specific dispersion parameter α_i. The fitted mean is composed of a sample-specific size factor s_j and a parameter q_{ij} proportional to the expected true concentration of fragments for sample j. The coefficients β_i give the log2 fold changes for gene i. This modeling framework allows for assessing what genes are expressed significantly differently in a retina that received treatment.

Differentially expressed genes were selected using a 1 percent false discovery rate threshold and those that had a log2fold change of 1. This stringent FDR is meant to limit the number of genes differentially expressed and only select for those with most significant changes. These genes were then loaded into iDEP for pathway and enrichment analysis to observe trends in the nervous system. Genes for interpretation were filtered for those involved with the nervous system and those that had the greatest fold changes. We then checked to see if these genes were also differentially expressed with an empirical bayes modeling framework (Leng et al. (2013)) instead of the above frequentist model.

Genetic Time Course Analysis

We applied an empirical bayes mixture model to analyze bipolar cell transcriptomes of control and diseased retinas across time points. We used this analysis to identify whether or not the genes involved in treatment are changing during degeneration and/or develeopment. A potential concern was that an identified gene that we found to be associated with treatment may have resulted from differences in development. Because diseased and control retinas had ordered age groups (ages 30, 90, and 210), a time-course analysis was appropriate. This approach was not only able to model non-constant expression over time but also group them into expression paths (constantly up, down, sporadic, etc. . .). In addition to identifying differentially expressed genes, this method was also able to evaluate whether some genes expressed equivalent expression. This analysis was done using the EBSEQ-HMM package in R (Leng et al. (2015)).

The model fomulation is as follows:

$$(X_{gt}|X_{g,t-1} = x_{g,t-1}, S_g^{\Delta t} = s, \Theta) \sim \text{Beta} - \text{NB}(\mu_{gt} = \frac{r_{g,t-1}(\beta + \xi_g^s, \bar{x}_{g,t-1})}{\alpha - 1}) \tag{2}$$

where X_{gt} represents the expression values for gene g at time t, $r_{g,t-1}$, $\beta + \xi_g^s, \bar{x}_{g,t-1}$ represent contribution from the beta prior, Θ represents the parameters in the beta prior, and $S_g^{\Delta t}$ is the specified time point in the time course (see Leng et al. (2015) for full derivation and code). The model parameters were fit using method of moments. We identified genes differentially expressed according to a 1 percent false discovery rate.

Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M., Haag, J. D., Gould, M. N., Stewart, R. M., and Kendziorski, C. (2013), "EBSeq: An empirical bayes hierarchical model for inference in rna-seq experiments," *Bioinformatics*, Journal Article„ 29, 1035–43. https://doi.org/10.1093/bioinformatics/btt087.

Leng, N., Li, Y., McIntosh, B. E., Nguyen, B. K., Duffin, B., Tian, S., Thomson, J. A., Dewey, C. N., Stewart, R., and Kendziorski, C. (2015), "EBSeq-hmm: A bayesian approach for identifying gene-expression changes in ordered rna-seq experiments," *Bioinformatics*, Journal Article„ 31, 2614–22. https://doi.org/10.1093/bioinformatics/btv193.

Love, M. I., Huber, W., and Anders, S. (2014), "Moderated estimation of fold change and dispersion for rna-seq data with deseq2," *Genome Biol*, Journal Article„ 15, 550. https://doi.org/10.1186/s13059-014-0550-8.