Balan Ramesh and TJ Firneno

Methods

We demultiplexed the raw Ilumina reads and removed the barcodes using the fastx\_barcode\_splitter and the fastx\_trimmer modules of the FASTX-TOOLKIT (WEBSITE), respectively. We examined sequence quality on a per sample basis using FASTQC v0.11.3 and across all twelve samples using MULTIQC v1.8(WEBSITE). We indexed the *Rana pipiens* reference transcriptome (www.davislab.net/rana/) and then quantified the expression of our reads to the indexed reference transcriptome using the default parameters and the --gcBIAS flag in the program Salmon v0.11.3 (CITATION). We used the DESeq2 package (CITATION) in a custom R script find the differential expression between the treated test group and the control group. In brief, this script takes the counts for each transcript for each sample and then, using the given design, it calculates the size factors for each transcript, estimates the dispersion of each transcript, and tries to fit a negative binomial GLM overall. The p-value is then calculated and adjusted using the negative binomial Wald’s test---. All of the transcripts with an adjusted p-value < 0.05 are considered differentially expressed. Then transcripts with read counts </= 10 across all twelve samples are removed from further analysis. Finally, the script normalizes the read counts using variance stabilizing transformations (VST), removes the batch effects due to developmental stages, and creates a principle components analysis biplot using PC1 and PC2. We then identified and blasted differentially expressed transcripts against the *Xenopus laevis* and *Xenopus tropicalis* transcriptomes to annotate the function of our differentially expressed transcripts. We submitted our differentially expressed gene list to the bioinformatics database DAVID (CITATION) to find if these transcripts are involved in any biological pathways or processes, and to cluster the transcripts into biological functions using the default parameters.

Results

A total of 48,017,723 single-end reads were retained across the six control samples and 45,777,387 single-end reads were retained across the six treatment samples (93,795,110 across all samples). FastQC and MultiQC verified that only high-quality reads with a Phred score >30 were retained for further analysis. We identified 471 significantly differentially expressed transcripts before and after filtering, of which 84 were upregulated and 387 were downregulated (Figure #). The PCA plot shows that the treatment has an effect comparative to the control group based on the covariation of gene in PC1, PC2, and PC3 (Figure #). Approximately 308 of our 471 differentially expressed transcripts had BLAST hits to the *X. laevis* and *X. tropicalis* transcriptomes. Of these 308 transcripts, 228 represented genes that are downregulated and 47 represented genes that were upregulated, and had a functional annotation (Table #). The 32 transcripts that did not have functional annotation were uncharacterized proteins in the *Xenopus* transcriptomes. DAVID successfully clustered the following: 42 of our differentially expressed transcripts into separate biological processes, with 180 genes not meeting the enrichment threshold (EASE < 0.1) to be clustered into a biological process; 120 of our differentially expressed transcripts into separate cellular components, with 102 genes not meeting the EASE threshold; and 86 of our differentially expressed transcripts into molecular functions, with 136 genes not meeting the EASE threshold (Figure #).

Molecular Function

Biological Process

Cellular Component

Graph of enriched GO ontology terms.