

## **Meta-Analysis of RNA sequencing to rank cotton genes with unknown function**

Authors: JONATHAN ZIRKEL, Grant Billings, Amanda R Storm, and Amanda M Hulse-Kemp

There are currently thousands of genes in the cotton genome that code for proteins with minimal or no annotated function, many of which may have agricultural relevance. How do we know which proteins to focus further research efforts on? One way to narrow down the search is to focus on genes showing differential expression across many studies. Citation records for RNA-seq papers citing major genome assemblies in cotton were used to retrieve lists of differentially expressed genes (DEGs) of unknown function, resulting in 82 differential expression group comparisons extracted from 48 publications. The DEGs were sorted by the publication and trait and aggregated into a table counting the number of total papers mentioning each gene. Examples of traits studied in the source publications included stress tolerance, fiber quality, and plant physiology. In total 4882 genes of unknown function were found to be differentially expressed in at least 1 publication, with a small number of genes found in multiple publications. This research stemmed from interaction with a Course-based Undergraduate Research Experience (CURE), developed as a collaboration between USDA-ARS and Western Carolina University, where students learn bioinformatics techniques and apply them to predict the function of their protein. The list of prioritized genes identified here will be used to assign protein sequences to students in upcoming semesters in the CURE. The results will give students new resources to find publications mentioning their protein, as well as provide a focus on proteins of agricultural relevance.