[BIOINFO 2023]

Comparative RNA-seq Analysis Utilizing Large-Scale Control Datasets in *Oryza sativa*.

Il San Jeong[1], and Yang Jae Kang[1*]

[1]*Department of Bio and Medical Big Data (BK4 Program), Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Gyeongsang National University*

**Corresponding author: kangyangjae@gnu.ac.kr*

The advent of high-throughput sequencing has yielded abundant RNA-seq data for gene expression analysis. Yet, traditional comparative transcriptomic analyses with small control groups may present limitations in capturing gene expression variability, affecting robustness and reproducibility across the experiments.

In this study, we manually selected 73 control and 3 treated samples to establish a normal expression dataset for Oryza sativa root tissue, downloaded via sra-toolkit, and quantified using Kallisto. Using RAP-DB annotation, transcriptome ids were mapped to Locus IDs in the control group, followed by expression checks of housekeeping genes (UBQ5, EF-1 alpha, 18S, 25S) for sample filtering. Post-normalization, normality tests were conducted, selecting 3,146 genes from 37,967, adhering to a normal distribution. A comparison using Locus ids of genes related to zinc metabolism revealed our analysis identified 57 zinc-related Loci while DESeq2 identified 149, with 36 common Loci.

Our analysis revealed differentially expressed genes not identified in DESeq2 analysis, highlighting our method's potential for a more comprehensive understanding of transcriptomic alterations under zinc deficit conditions.

This methodology, employing a large control group for comparative analysis, is a step forward in harnessing RNA-seq data abundance. The approach not only enhances transcriptomic comparisons but suggests further reliability enhancements with more control data inclusion.

Through this, we believe that our analysis method can enable the discovery of genes not found in other analyses or enhance the accuracy of the analysis results.