

## MDSeq

MDSeq: Gene expression mean and variability analysis for RNA-seq counts.

Di Ran and Z. John Daye

MDSeq: Performs analysis of both gene expression mean and variability on RNA-seq counts, as described in Ran and Daye (2017). A zero-inflated mean-dispersion model is applied, based on the coefficient of dispersion. Generalized linear models (GLMs) are implemented that can incorporate treatment effects and additional covariates on both the mean and dispersion. The package provides comprehensive features to model technical excess zeros, identify outliers efficiently, and evaluate differential expressions at biologically interesting levels. Several assistant functions were implemented to help users in filtering and normalization of raw read counts. Options are provided that allow users to incorporate normalization factors as re-scaled counts or offsets in the mean-dispersion GLMs. Additionally, the software allows for parallel processing with multiple threads for efficient computations.

This package corresponds to the following paper, where further details can be found:

- Ran, Di, and Z. John Daye (2017). “Gene Expression Variability and the Analysis of Large-Scale RNA-Seq Studies with the MDSeq.” *Nucleic Acids Research*. doi: 10.1093/nar/gkx456.

**Installation** Install MDSeq from local source with

```
install.packages("MDSeq_1.0.5.tar.gz", repos=NULL, type="source")
```

Install MDSeq from GitHub with

```
library(devtools)
install_github("zjdaye/MDSeq")
```

**Dependencies** The MDSeq depends on the following packages: edgeR, cqn, quadprog, VarianceGamma, parallel, gtools.

**Vignette** A pdf Vignette is available in the inst/doc folder.