An Evaluation of RNA-seq Differential Analysis Methods

Dongmei Li[1], Martin S. Zand[1], Timothy D. Dye[2], Maciej L. Goniewicz[3], Irfan Rahman[4], Zidian Xie[1]

[1]Department of Clinical and Translational Research, University of Rochester Medical Center, Rochester, NY, USA; [2]Department of Obstetrics and Gynecology, School of Medicine and Dentistry, University of Rochester, Rochester, NY, USA; [3]Department of Health Behavior, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA; [4]Department of Environmental Medicine, School of Medicine and Dentistry, University of Rochester, Rochester, NY, USA

Introduction

❖ High-throughput transcriptome sequencing technologies have profound impact on our ability to address diverse range of biomedical problems and improve our understanding of human diseases.

❖ Gene differential analysis is an important downstream analysis for selecting candidate genes for further experimental validations.

❖ Current methods for RNA-seq differential analysis include both parametric methods and nonparametric methods such as edgeR, DEseq, DEseq2, baySeq, EBSeq, Voom, SAMSeq and NOISeq, etc.

Aims/Objectives

❖ We aim to systematically evaluate the performance of eight popular RNA-seq differential analysis methods including edgeR, DEseq, DEseq2, baySeq, EBSeq, Voom, SAMSeq and NOISeq.

Methods

❖ The performance of the eight RNA-seq differential analysis methods were evaluated by the false discovery rate (FDR), power, and stability.

❖ We conducted both simulation studies and real RNA-seq data examples for the evaluation.

❖ The simulation data were generated from both negative binomial distributions and log-normal distributions.

❖ The real data are gene expression levels on 36536 genes from 10 B6 mouse samples and 11 D2 mouse samples in neuroscience research.

Conclusions

❖ The significant genes lists are different when apply different gene differential analysis methods.

❖ For RNA-seq data analysis, the EBSeq method is recommended for studies with sample size as small as 3 in each group.

❖ The DESeq2 method is recommended for sample size of 6 or higher in each group.
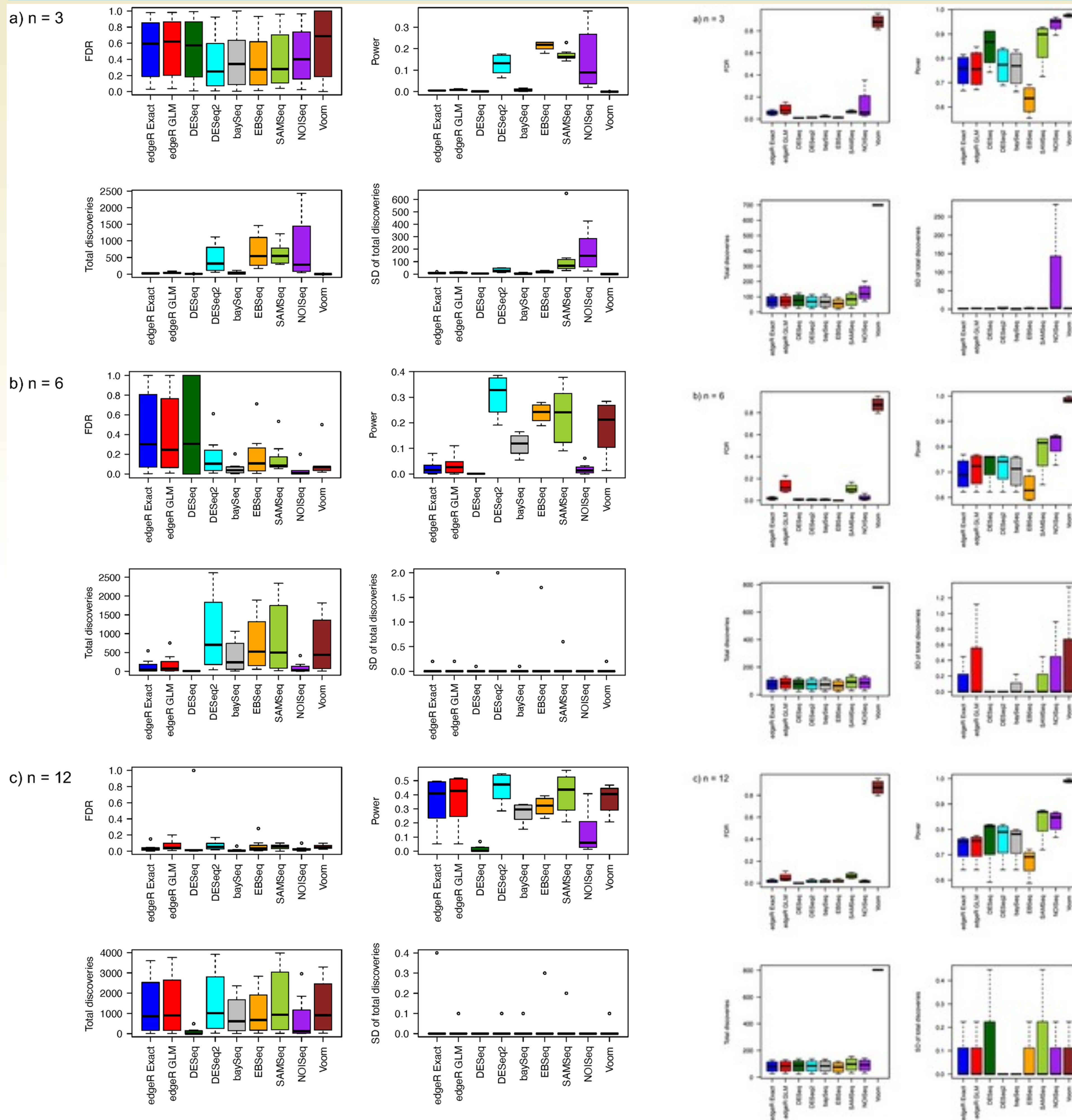
Results

Figure 1. Simulation results from data generated from negative binomial distributions.
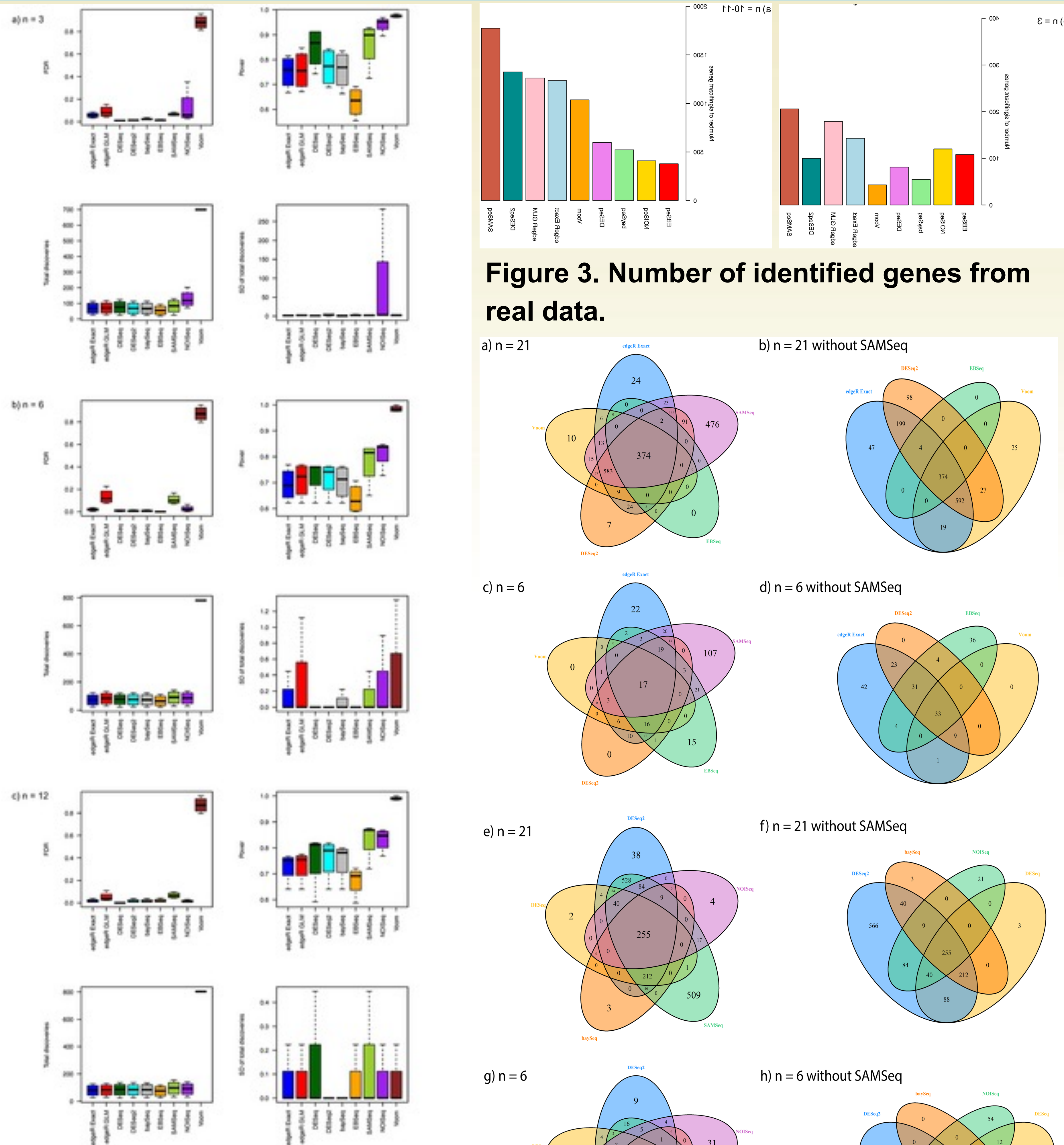

Figure 2. Simulation results from data generated from log-normal distributions.
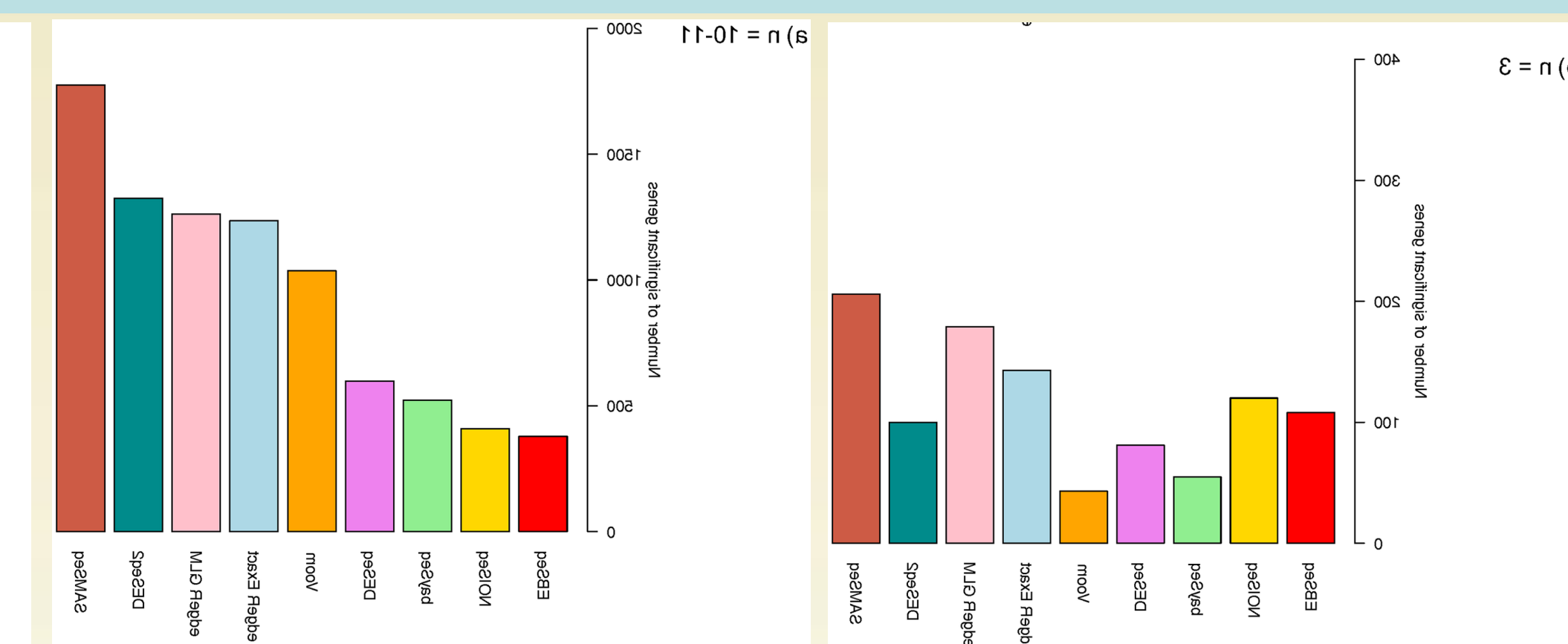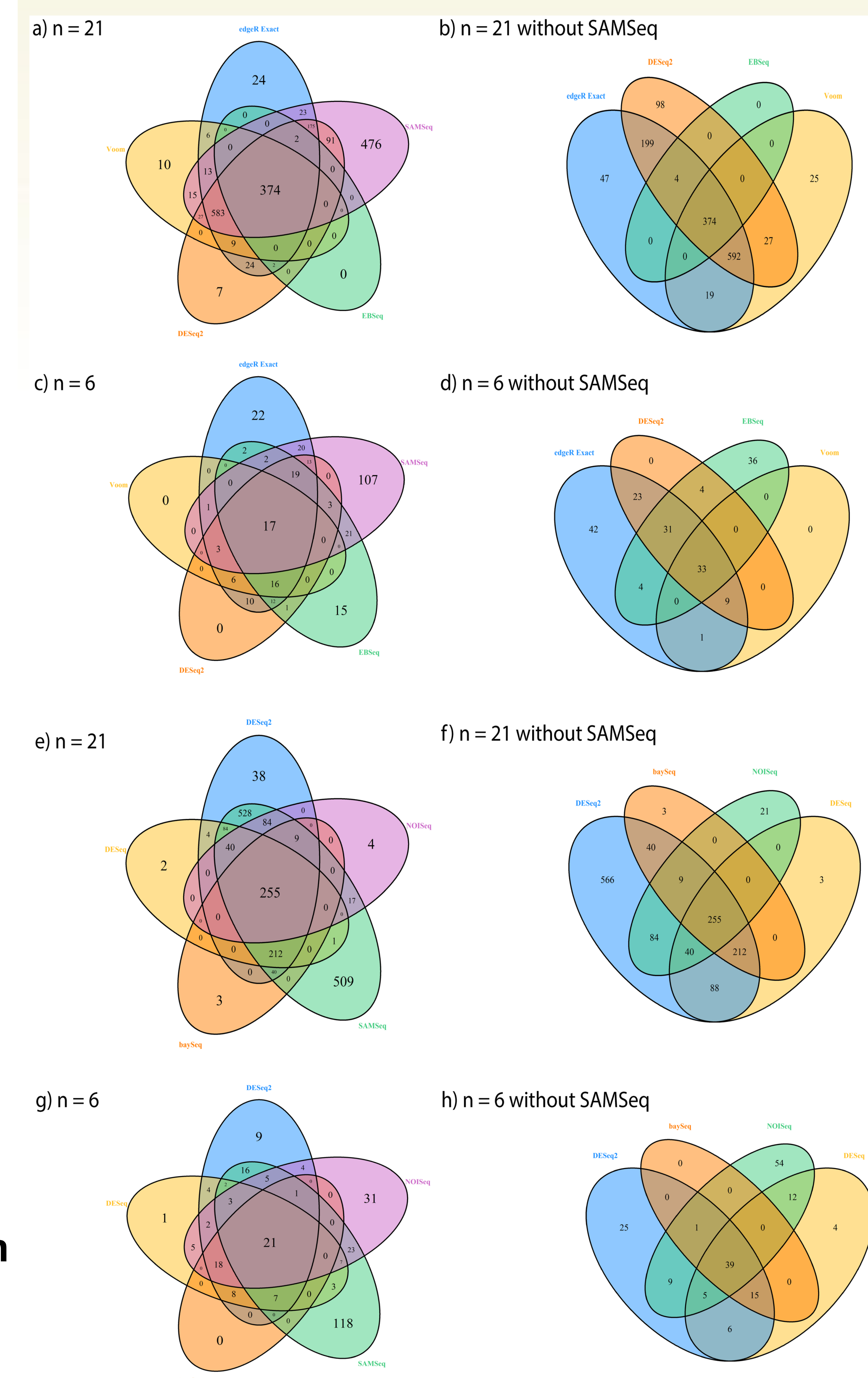

Figure 3. Number of identified genes from real data.

Figure 4. Venn diagram of significant genes from different methods.

Acknowledgements

Citation: Li D, Zand MS, Dye TD, Goniewicz ML, Rahman I, Xie Z. An evaluation of RNA-seq differential analysis methods. PLoS One. 2022 Sep 16;17(9):e0264246. doi: 10.1371/journal.pone.0264246. PMID: 36112652; PMCID: PMC9480998.

CLINICAL&TRANSLATIONAL SCIENCE INSTITUTE

CROFT — WNY CENTER FOR RESEARCH ON FLAVORED TOBACCO

UNIVERSITY of ROCHESTER MEDICAL CENTER