Effects of Key RNA-sequencing Technical Elements on Toxicity Assessment using Three Biological Replicates

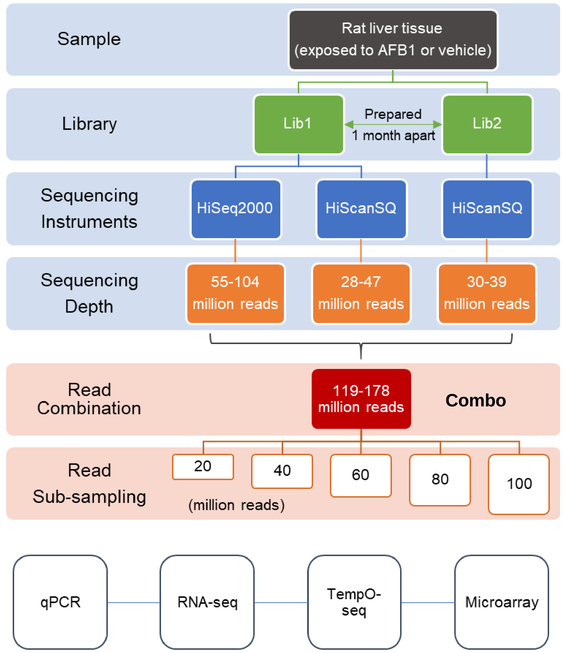
Dongying Li, Binsheng Gong, Joshua Xu, Baitang Ning, and Weida Tong

National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079

Introduction

RNA-Sequencing (RNA-seq) has emerged as a standard approach for toxicogenomics research. Under the 3-R (reduce, refine, and replace) principles for animal welfare protection, three animals per group are commonly used in current toxicogenomics studies. It is critical to understand how toxicological interpretation of RNA-seq data may be affected by key technical elements of RNA-seq, such as sequencing depth and library construction, when only three biological replicates are used. We conducted a comprehensive comparative analysis to address this question using rats treated with aflatoxin b1 (AFB1), a model hepatotoxin and focusing on key mechanisms of AFB1 toxicity. We compared differential gene expression and pathway enrichment in multiple RNA-seq datasets, which were generated from identical samples but with varying sequencing depths and library preparation. A cross-platform analysis was also performed by comparing data from RNA-seq, microarray, TempO-seq, and qPCR using the same samples.

Study Design



Results

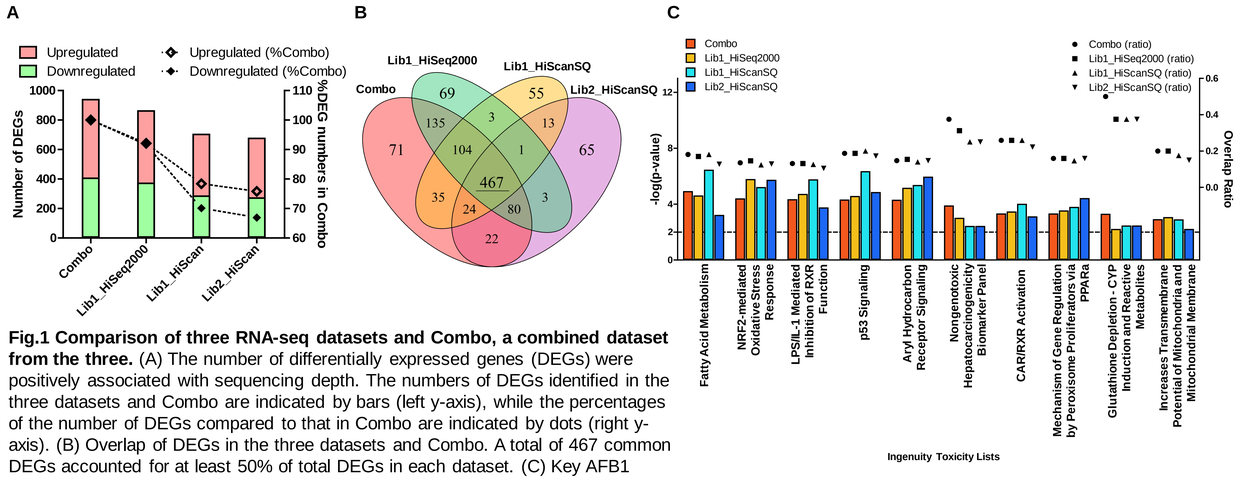


Fig.1 Comparison of three RNA-seq datasets and Combo, a combined dataset from the three. (A) The number of differentially expressed genes (DEGs) were positively associated with sequencing depth. The numbers of DEGs identified in the three datasets and Combo are indicated by bars (left y-axis), while the percentages of the number of DEGs compared to that in Combo are indicated by dots (right y-axis). (B) Overlap of DEGs in the three datasets and Combo. A total of 467 common DEGs accounted for at least 50% of total DEGs in each dataset. (C) Key AFB1 toxicity pathways were enriched by all four datasets in Tox Analysis of Ingenuity Pathway Analysis. Bars indicate -log(p-value) while dots overlap ratios. -log(p-value) > 2 was considered significantly enriched.

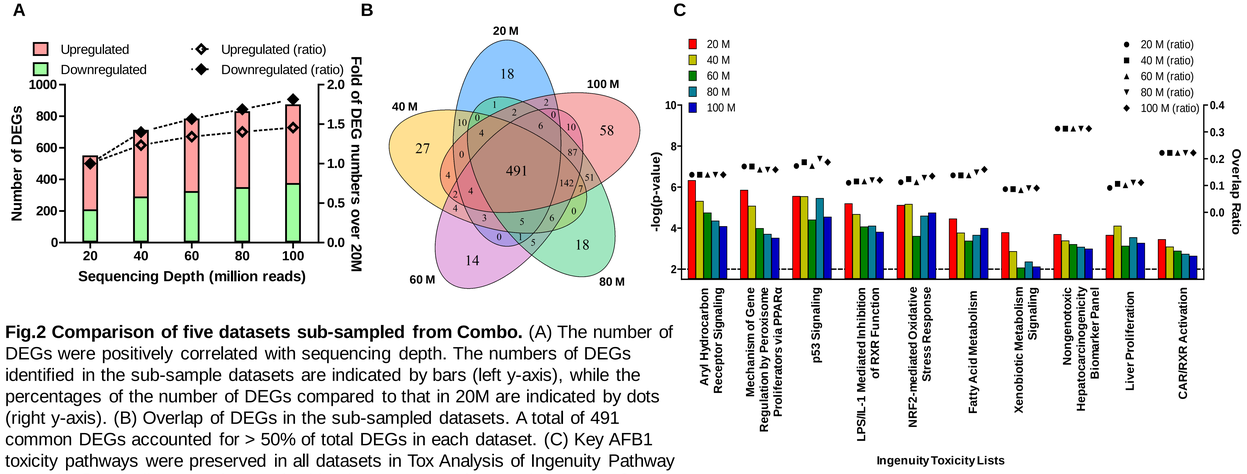


Fig.2 Comparison of five datasets sub-sampled from Combo. (A) The number of DEGs were positively correlated with sequencing depth. The numbers of DEGs identified in the sub-sample datasets are indicated by bars (left y-axis), while the percentages of the number of DEGs compared to that in 20M are indicated by dots (right y-axis). (B) Overlap of DEGs in the sub-sampled datasets. A total of 491 common DEGs accounted for > 50% of total DEGs in each dataset. (C) Key AFB1 toxicity pathways were preserved in all datasets in Tox Analysis of Ingenuity Pathway Analysis. Bars indicate -log(p-value) while dots overlap ratios. -log(p-value) > 2 was considered significantly enriched.

Results

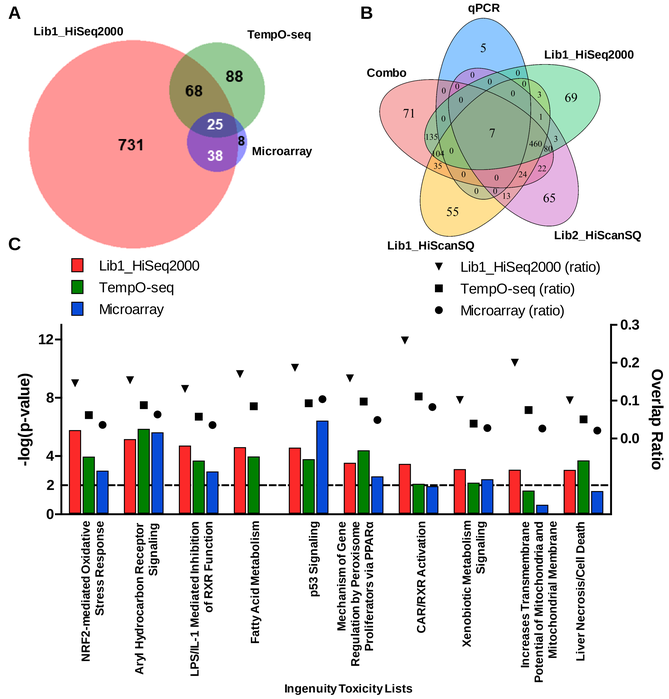


Fig.3 Cross-platform comparison. (A) Overlap of DEGs identified by RNA-seq in the Lib1_HiSeq2000 dataset, TempO-seq, and microarray analysis. Areas in the Venn diagram are proportional to the number of DEGs. (B) Overlap of DEGs identified in qPCR and four RNA-seq datasets with varying sequencing depths. (C) Top 10 enriched toxicity lists using DEGs from RNA-seq, TempO-seq, and microarray analysis. RNA-seq yielded overall higher overlap ratios and significance powers compared to TempO-seq and microarray. Bars indicate -log(p-value) while dots overlap ratios. -log(p-value) > 2 was considered significantly enriched.

Conclusions

- DEG detection power was positively correlated with sequencing depth.
- Key pathways underlying AFB1-induced liver toxicity may be preserved with reduced sequencing depth to a minimum of 20 M.
- RNA-seq had overall better statistical performance than other high-throughput platforms in pathway enrichment to gain toxicity insight.
- Library construction using the same protocol was key to reproducibility in toxicological interpretation of RNA-seq data.