Supplementary material
Resistant Fit Normalization for Single-cell RNA-seq Data

Da Kuang
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, USA
kuangda@seas.upenn.edu

Junhyong Kim
Department of Biology
University of Pennsylvania
Philadelphia, USA
junhyong@sas.upenn.edu

We use Splatter[S1] to generate two scenarios (SIM I and II) and demonstrate that resistant fit normalization works under extreme differential expression and extreme drop-out rate as examples of biological variations and technical variations.

S.I. Resistant Fit Normalization Works with Extremely Expressed Genes

In simulation SIM I, condition B with 100 cells and 100 genes is simulated from Splatter, where parameters are estimated from Tungs' data [S2]. For each cell in B, select one gene (DE gene) and scale 100 times up as the fold change of differential expression then make it a cell in condition A. For each pair of cells, resist fit normalization and log-scale normalization are used to reduce the biological variation between two cells, where condition B is the responsible variable and condition A is the explanatory variable.

Cell.1 and Cell.2 are two typical cells picked from condition A and B. are a set of visualizations representing before normalization as well as normalized by log-scale and resistant fit. We apply linear regression on each pair of cells after normalization. The slope of the linear model demonstrates the linearity between cells. The slope should be 1 to have a biological comparable normalization.

We found that log-scale normalization introduces bias while resist fit ignores the extreme expression and keeps the ratio among non-differentially expressed genes (common genes). It is because the size factor of log-scale normalization is affected by the extreme expression, hence the counts of all the genes in cell.1 become relatively smaller after normalization. Therefore we constantly have larger scopes in fig 2d. For resistant fit normalization, the DE gene has no effect because the linear regression is only based on the biological feature set which excludes the DE gene. normalization.

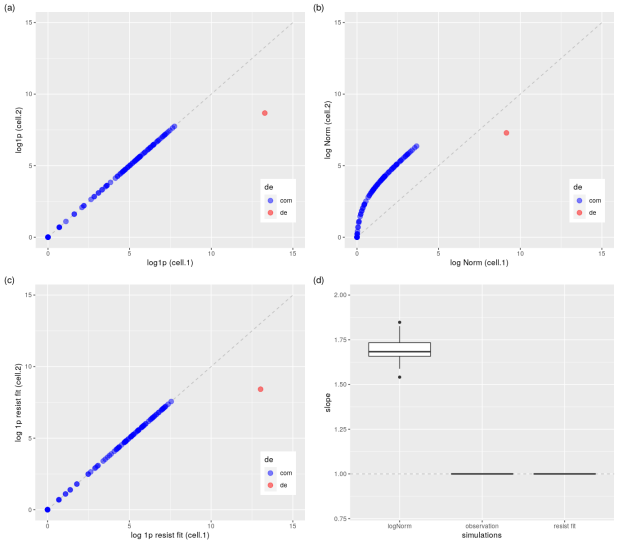

Fig. S1. **(a-c)** Biplots of one typical pair of cells where cell.1 is in condition A and cell.2 is in condition B. (a) before normalization, (b) after log-scale normalization and (c) after resist fit normalization. **(d)**: the barplot for the slope of 100 pairs of such cells under the same scenario in a,b and c. It shows that the log-scale normalization increase the expression levels in condition A which contains the DE gene..

S.II. Resistant Fit Normalization Works with Extreme Drop-out Rate

In simulation SIM II, condition B is the same as SMI I. For each cell in B, select half of genes (drop-out genes) and set the counts to 0 as the drop-out event then make it a cell in condition A. For each pair of cells, resistant fit and log-scale normalization are used to reduce the technical variation between two cells.

Similar to SMI I, we found that log-scale normalization also introduces bias under extreme drop-out rate because the size factor is reduced by the drop-out genes in condition A so that the scopes after log-scale normalization become smaller(Fig 3d). It can be observed that resist-fit still keep two cells comparable after

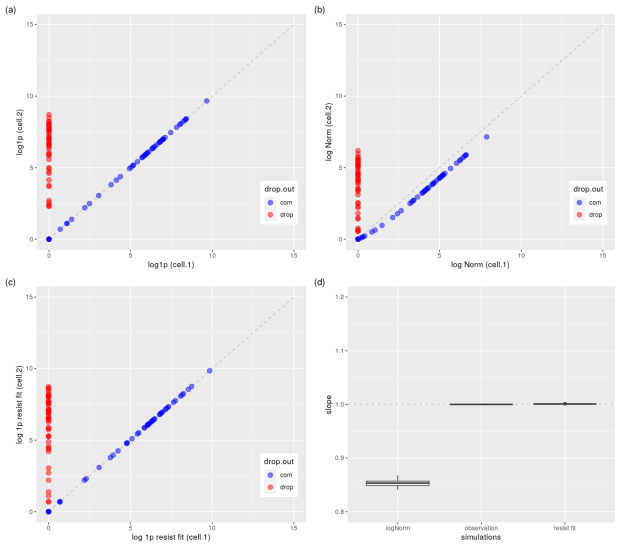

Fig. S2. **(a-c)** Biplots of one typical pair of cells where cell.1 is in condition A and cell.2 is in condition B. (a) before normalization, (b) after log-scale normalization and (c) after resist fit normalization. **(d)**: the barplot for the slope of 100 pairs of such cells under the same scenario in a,b and c. It shows that the log-scale normalization reduce the expression levels in condition A which contains the drop-out genes..

REFERENCES

[S1] L. Zappia, B. Phipson, and A. Oshlack, "0-Splatter: simulation of single-cell RNA sequencing data," Genome Biology, vol. 18, no. 1, p. 174, Sep. 2017, doi: 10.1186/s13059-017-1305-0.

[S2] P.-Y. Tung et al., "Batch effects and the effective design of single-cell gene expression studies," Scientific Reports, vol. 7, no. 1, Art. no. 1, Jan. 2017, doi: 10.1038/srep39921.