Correcting and classifying SARS-CoV-2 RNA expression in single cells

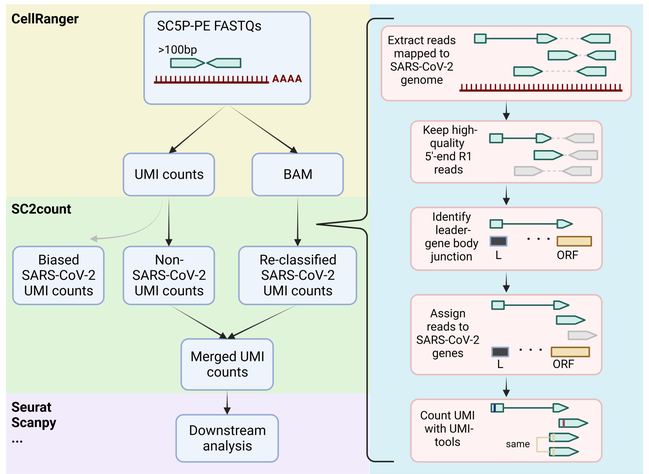
Wendao Liu, Zhongming Zhao
The University of Texas MD Anderson Cancer Center UTHealth Houston Graduate School of Biomedical Sciences, Houston, TX
Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX

INTRODUCTION

SARS-CoV-2 genes encoding structural proteins (S, E, M, N) and accessory proteins (ORF3a, 6, 7a, 8, 10) are transcribed into subgenomic RNAs (sgRNAs). Canonical sgRNAs contain a common 5' leader of around 70 nt, a gene body, and all downstream sequence of the genome. Only the gene body sequence is translated into the protein.

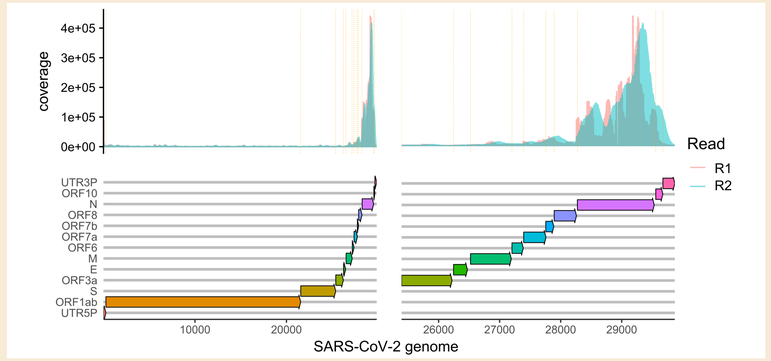
scRNA-seq is used to profile the gene expression of both host and viral genes in COVID-19 patients. However, the transcription biology of coronaviruses differs from eukaryotes, making it challenging to accurately measure the expression levels of individual SARS-CoV-2 genes using current bioinformatic methods.

METHODS

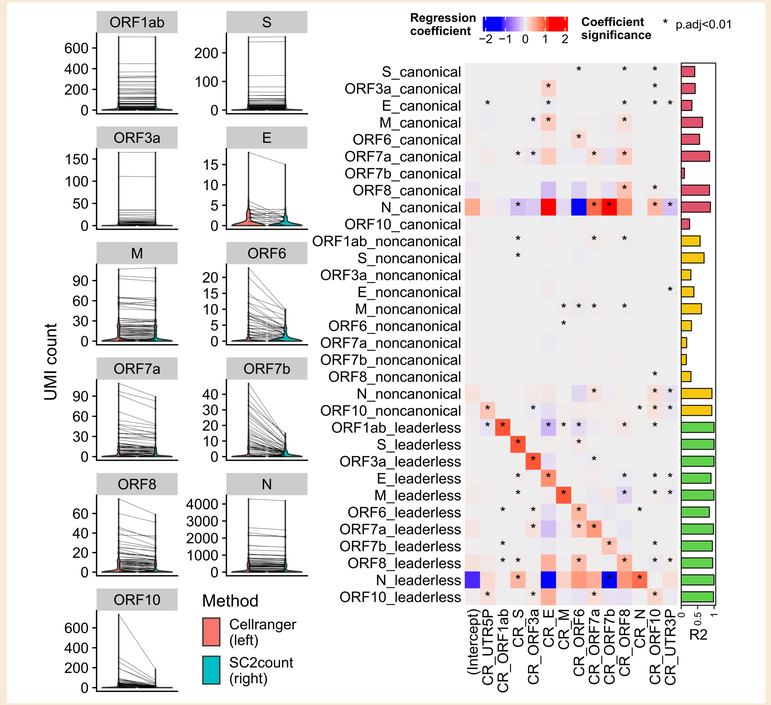


SC2count assigns SARS-CoV-2 reads to individual genes and categorizes viral RNAs into different groups using paired-end 5' library scRNA-seq data. It output corrected re-classified UMI counts for downstream analyses.

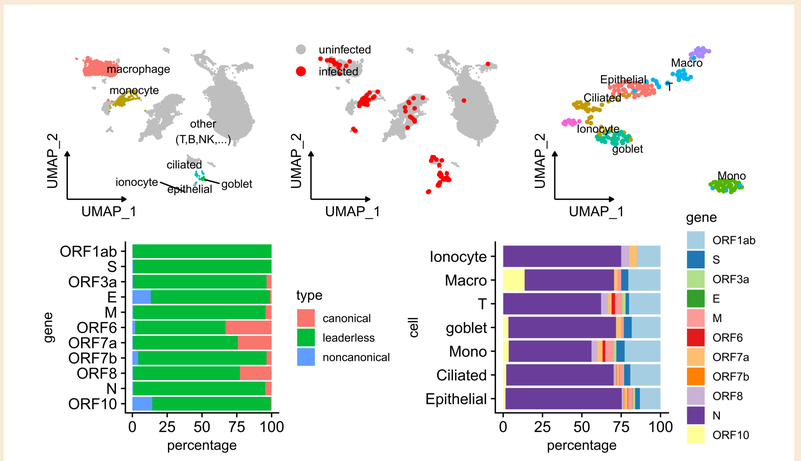
RESULTS



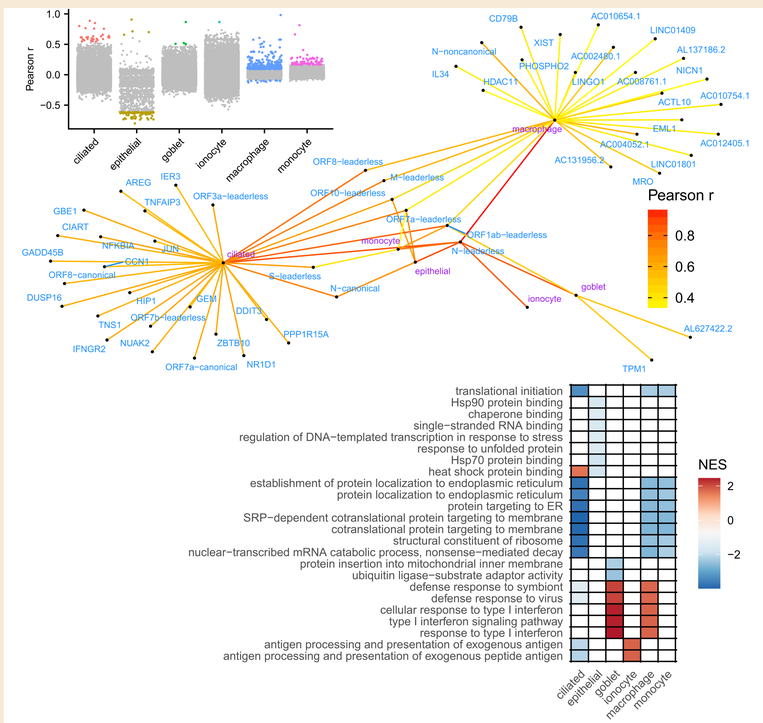
SARS-CoV-2 gene expression quantification based on read 2 from single-cell 3'/5' library is biased.



SC2count corrects individual viral gene expression quantification and categorizes RNAs.



In different infected cell populations, some viral genes are selectively transcribed. (GSE167118)



Different host genes are correlated with SARS-CoV-2 total RNA expression in several cell types, involved in multiple virus-induced pathways.

CONCLUSION

We introduce SC2count to correct biased quantification of SARS-CoV-2 gene expression, which can be used for identifying additional features of SARS-CoV-2 transcription in COVID-19 scRNA-seq data. It may facilitate downstream analyses in understanding cell type-specific viral gene expression and host immune responses.

ACKNOWLEDGEMENTS

Wendao Liu is supported by Cancer Prevention and Research Institute of Texas Biomedical Informatics, Genomics and Translational Cancer Research Training Program (CPRIT RP210045).

Our COVID-19 scRNA-seq review article