RESEARCH SNAPSHOT

Elevate the Quality of Your Single-Cell Sequencing Data Using Levitation Technology

Overview and Single-Cell RNA Sequencing Pipeline

Single-cell RNA sequencing (scRNA-seq) is a powerful tool that can help researchers uncover rare and complex cell populations, evaluate the role of certain genes and/or screen drugs for their effect on different cell types in the context of numerous disorders. ScRNA-seq enables transcriptional profiling of thousands of single cells, and analysis of variably regulated genes specific to certain cells at an individual level[1]. This powerful tool is rendered useless without the assurance of data accuracy and quality. As a result, scRNA-seq datasets are routinely assessed and optimized using a set of parameters called QC metrics.

In this research snapshot, we demonstrate how the quality of your downstream scRNA-seq data and analysis can be affected by poor quality samples, and how a simple enrichment step can change everything. Single cells were first isolated from different brain subtypes, including neurons, astrocytes, oligodendrocytes, microglia and endothelial cells. A subset of cells were enriched with the LeviCell™ system, while the rest were not enriched. All cells were profiled using 10x Genomics Chromium Controller, and analyzed with the latest Seurat pipeline (v4). In every QC metric that was assessed, the enriched cells outperformed the non-enriched cells, delivering better quality single-cell data and results.

HIGHLIGHTS

- Enriched cells outperformed non-enriched cells in every scRNA-seq QC metric that was assessed.
- Enriched cells lead to ~2.5 fold lower cell loss based on set QC thresholds.
- Enriched cells have 1.7 fold less ambient RNA contamination, preventing undesired barcoding and sequencing of free mRNA versus native mRNA.

Current Challenges In Accurate scRNA-seq Analysis

1. Sub-optimal Single-Cell Isolation and Enrichment

It is of utmost importance to efficiently isolate and enrich cells of the highest quality, without compromising their transcriptomic profile. Widely used techniques include enzymatic or mechanical dissociation followed by Fluorescence Activated Cell Sorting (FACS)[2]. This puts the isolated and enriched cells through a significant amount of stress, which may alter their transcriptional profile, i.e. upregulation of stress or inflammatory genes, increase in mitochondrial genes, among others. We have previously demonstrated that FACS sorting puts undue stress on single cells, resulting in an artificially activated transcriptional profile. In comparison, Levitation Technology enables fast, gentle isolation and enrichment of different cell types while also preserving a more homeostatic prolife of microglial cells (#90-00081A0522, Enrichment of Homeostatic Mouse Brain Subtypes).

2. Cell Viability vs Cell Quantity

During the isolation and enrichment steps of the workflow, some cells may enter an apoptotic state which results in a mixed population of viable, dead and dying cells. These dying and apoptotic cells may be considered live through standard counting or cell enrichment methods. However, they are biologically very different from viable cells and have different gene signatures. Higher numbers of mitochondrial genes are observed, and other stress/inflammatory/apoptotic signaling pathways are significantly upregulated. To accurately distinguish viable cells from dying and apoptotic cells, researchers should utilize three quality control (QC) parameters such as number of counts per cell (nCount), number of genes per cell (nFeature) and percentage of mitochondrial genes per cell (Percent. Mt). Cells with low nCount, low nFeature, and high Percent.Mt are considered low quality and/or dying.

In the field of scRNA-seq, determination of these QC thresholds is one of the most debated topics. There needs to be a trade-off between the quality of data generated against the quantity of cells available in the dataset. Thresholds that are too stringent will yield fewer cells in the final dataset for downstream analysis. Conversely, if the QC thresholds are too lenient, the dataset will have more cells with lower overall quality[3] and potentially less relevant to the biological question they are asking. To eliminate doublets from downstream analysis, remove cells with nCount < 400, nFeature < 200, and the top 1% of all nCount, nFeature and Percent.Mt outliers. These thresholds should be set prior to analysis to achieve the highest quality datasets, which require fewer exclusion of cells (% cells lost in QC). The quality of downstream analysis can be further improved with the addition of a simple enrichment step using Levitation Technology, leading to ~2.5 fold lower cell loss as noted in Figure 1A.

3. Low Sequencing Quality

10x Genomics provides a list of parameters (Table 1) that can be used to assess sequencing quality. These parameters determine the reliability of data and thus the validity of downstream interpretation. In Figures 1B through 1G, we highlight key parameters that saw significant improvements as a result of sample enrichment with our label-free Levitation Technology.

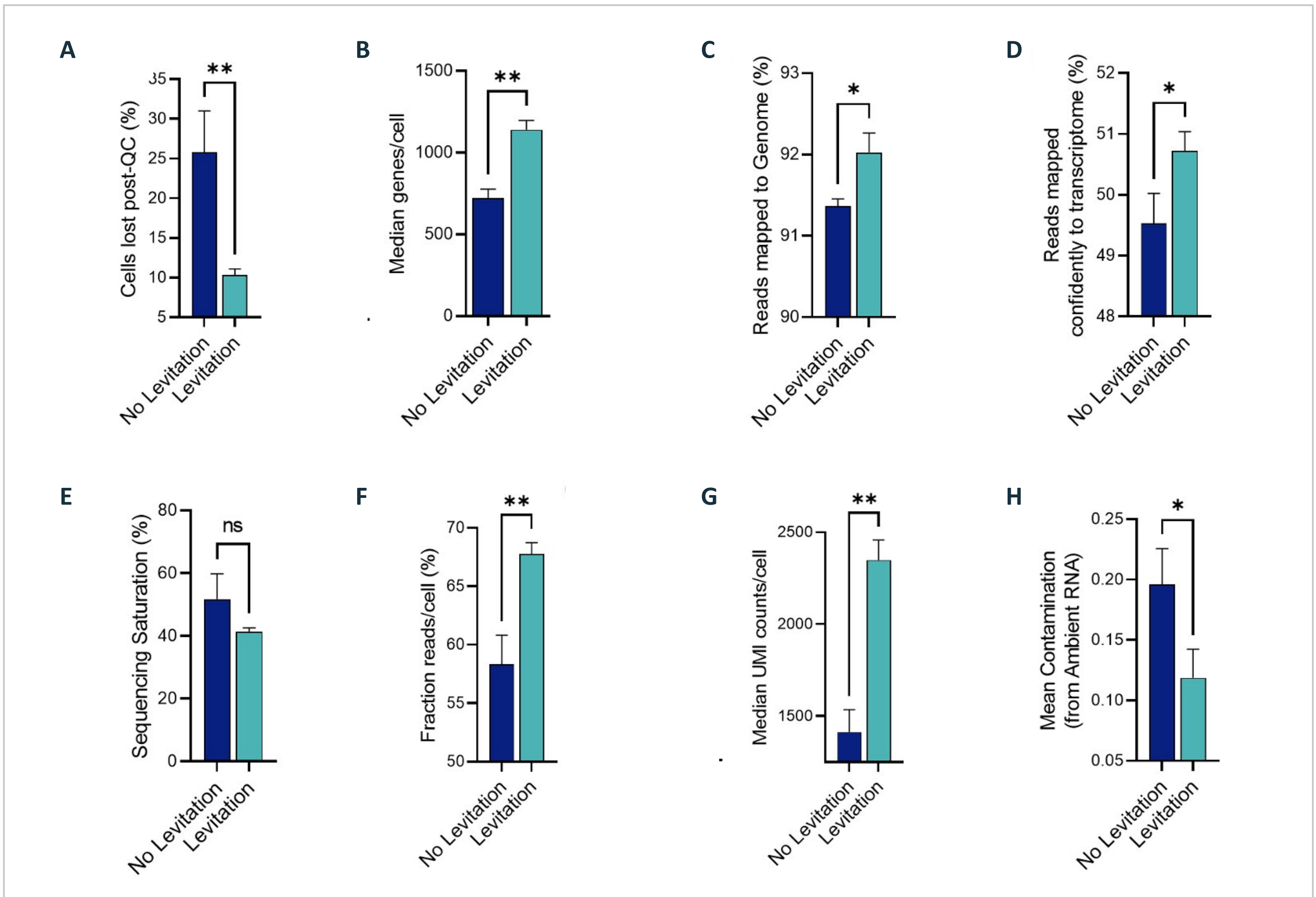

Figure 1. LeviCell enriched samples consistently perform better. *(A) Significantly fewer cells were lost or excluded from levitated samples based on set QC thresholds. (B-G) Samples that were enriched using Levitation Technology demonstrated significantly better readouts when assessed with CellRanger QC metrics. (H) Significantly lower ambient RNA contamination was present in enriched versus non-enriched samples. Error bars indicate mean±SEM from 4 different samples. *p> 0.01; **p>0.001 via Student's t-test.*

METRIC	DESCRIPTION	IDEAL
Cells Lost Post QC	Low quality cells removed from the dataset based on nCount, nFeature and percent mito	Lower the better
Median Genes per Cell	The median number of genes detected (with nonzero UMI counts) across all cell-associated barcodes	Normal: 500-5000 Ideal: 1000-3000
Reads Mapped to Genome	Fraction of reads that mapped to the genome	Lower values indicates poor library quality
Reads Mapped Confidently to Transcriptome	Fraction of reads that mapped to a unique gene in the transcriptome with a high mapping quality score as reported by the aligner	>30% Lower percentages indicates poor library quality
Sequencing Saturation	Sequencing saturation is a measure of the fraction of library complexity that was sequenced in a given experiment. The inverse of the sequencing saturation can be interpreted as the number of additional reads it would take to detect a new transcript	Lower sequencing saturation indicates a high proportion of the library complexity has not been captured by sequencing
Fraction Reads in Cells	The fraction of cell-barcoded, confidently mapped reads with cell-associated barcodes	>70% Lower percentages indicate that a high level of ambient RNA partitioned into all (cell-containing and non-cell-containing) GEMs
Median UMI Counts per Cell	The median number of total UMI counts across all cell-associated barcodes	Lower values suggest shallower sequencing depth and/or poor library quality
Mean Contamination (from Ambient RNA)	Cross-contamination with RNA transcripts from different apoptotic cells	Lower the better

Table 1. *List, description and ideal range of key CellRanger metrics which are evaluated to determine sequencing quality.*

4. Ambient RNA Contamination

mRNA molecules can be released into the cell suspension during isolation and enrichment, usually from stressed or apoptotic cells. When these free ambient mRNA molecules are mixed with single cells and encapsulated into the droplets during 10x GEM formation, they are barcoded and amplified along with the cell's native mRNA. This contamination can be seen when a highly expressed cell-type specific gene is observed at very low levels in different cell populations. Ambient RNA contamination can also be accurately predicted with DecontX, a novel Bayesian method developed by a computational biologist, and subsequetly removed from those individual cells. This pipeline can be incorporated into a scRNA-seq workflow to significantly improve the downstream analysis[4]. As depicted in Figure 1H, we observed significantly lower contamination from ambient RNA in samples that were enriched versus non-levitated controls (~1.7 fold).

In summary, this research snapshot highlights the efficiency and effectiveness of our label-free Levitation Technology, and how it can significantly enhance different stages of scRNA-sequencing from single cell isolation and enrichment, sequencing and data quality assessment to the final transcriptomic profiling of different cell types (#90-00081A0522, Enrichment of Homeostatic Mouse Brain Subtypes).

References

1. Haque, A., Engel, J., Teichmann, S.A. and Lönnberg, T., 2017. "A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications." Genome medicine, 9(1), pp.1-12.

2. Hwang, B., Lee, J.H. and Bang, D., 2018. "Single-cell RNA sequencing technologies and bioinformatics pipelines." Experimental & molecular medicine, 50(8), pp.1-14.

3. Luecken, M.D. and Theis, F.J., 2019. "Current best practices in single-cell RNA-seq analysis: a tutorial." Molecular systems biology, 15(6), p.e8746.

4. Yang, S., Corbett, S.E., Koga, Y., Wang, Z., Johnson, W.E., Yajima, M. and Campbell, J.D., 2020. "Decontamination of ambient RNA in single-cell RNA-seq with DecontX." Genome biology, 21(1), pp.1-15.

For Research Use Only. Not for use in diagnostic procedures. ©2022 LevitasBio, Inc. All rights reserved. LeviCell, LevitasBio, and the LevitasBio logo are trademarks of LevitasBio, Inc. All other tradermarks are properties of their respective owners. support@levitasbio.com
90-00086A0722