TECHNICAL NOTE

Biological & Technical Variation in Single Cell Gene Expression Experiments

Introduction

The Chromium Single Cell 3' v2 Reagent Kits protocol (Document CG00052) produces Single Cell 3' short-read sequencer compatible libraries. Technical and biological variation may be present in the experiment design, and may impact data interpretation. Potential sources of technical variation include running a sample on two separate microfluidic chips or at different well positions on the same chip, and/or technical variation introduced by sequencing libraries on separate Illumina flowcells or sequencing lanes. This Technical Note examines the potential sources of technical and biological variation and their effects on single cell gene expression. These factors need to be considered when designing an experiment to minimize bias and generate reliable single cell gene expression data.

Method

Single cell gene expression profiles from embryonic day E18 mouse (Document LIT000015) were compared. Briefly, combined cortex, hippocampus, and ventricular zone from C57EHCV E18 mice (BrainBits) were processed according to the Demonstrated Protocol – Dissociation of Mouse Embryonic Neural Tissue for Single Cell RNA Sequencing (Document CG00055) to obtain single cell suspensions. Single Cell 3' v2 libraries were generated following the Single Cell 3' v2 Reagent Kits User Guide (Document CG00052). The libraries were sequenced on the Illumina HiSeq 4000 platform. Data were processed with Cell Ranger (v1.3.0) and the "cellranger-aggr" pipeline was applied to combine the individual datasets and normalize read depth across samples to approximately 18,500 reads per cell. The study design to investigate different sources of biological and technical variation is shown in Table 1. The presence of potential batch effects on gene expression introduced by the experimental conditions were assessed using tSNE clustering plots, R^2-values, and Pearson correlation between replicate samples (Figures 1-4).

Results

Biological Replicates

To assess gene expression variation between biological replicates, neuronal tissue from two different mice, dissociated on different days (Replicates 1 & 2), were used to prepare single cell suspensions. Two libraries were generated by loading the cell suspensions on the same sample input well but on two separate microfluidic chips. Both libraries were sequenced on two separate lanes of the same Illumina flowcell (Table 1).

Gene expression profiles between the biological replicates, were highly correlated with a Pearson correlation of 0.968 and R^2 of 0.938 (Figure 1A). tSNE clustering showed a high degree of concordance of the cell distribution between the two biological replicates (Figure 1B). In addition, expression signatures for known neuronal marker genes were similarly distributed across the cell clusters in the tSNE plots (Figure 1C).

Table 1: Study design to investigate different sources of technical variation.

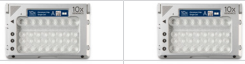
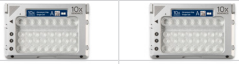
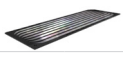
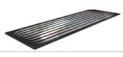
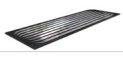
	Biological Replicate	Technical Replicate A	Technical Replicate B	Technical Replicate C
Sample				
Chip				
Channel				
Flowcell				
Differences	Different Samples	Different Flowcells	Different Single Cell A Chips	Different Channels

Figure 1. Gene expression variation between biological replicates.

A. Scatter plot showing the correlation of average gene expression levels between the replicates on a log10 scale.

B. Merged (left panel) and split (right panel) tSNE clustering for the biological replicates.

C. Expression profiles for known neuronal marker genes Fut9 (Inhibitory Neurons), Stmn2 (Excitatory Neurons), Aldoc (Astrocytes), and Slc1a2 (Radial Glia).

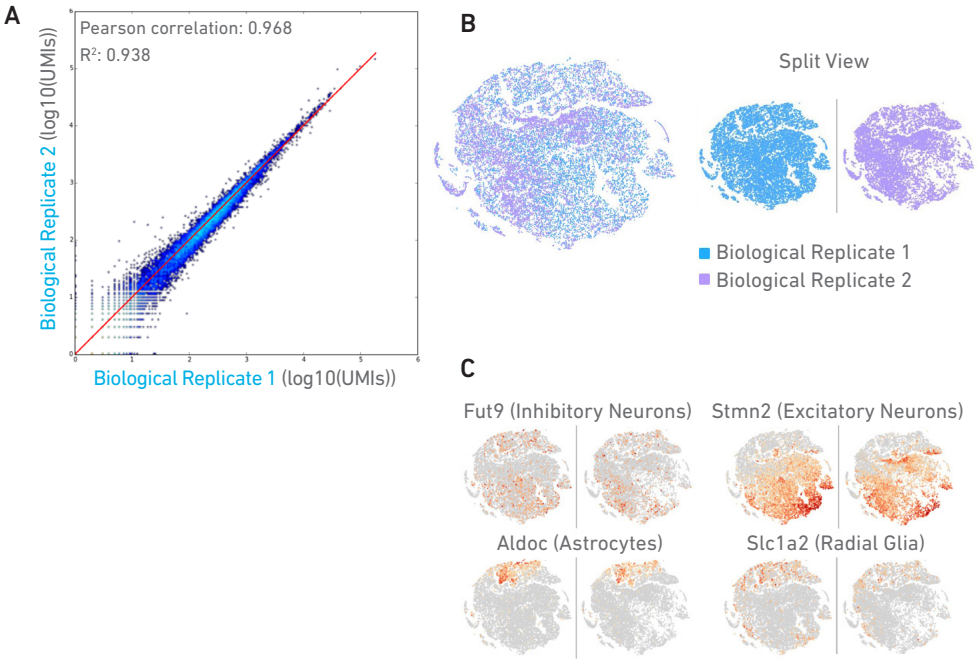


Figure 2. Gene expression variation when the same library is sequenced on two separate flowcells.

A. Scatter plot showing the correlation of average gene expression levels between the replicates on a log10 scale.

B. Merged (left panel) and split (right panel) tSNE clustering for the technical replicates.

C. Expression profiles for known neuronal marker genes Fut9 (Inhibitory Neurons), Stmn2 (Excitatory Neurons), Aldoc (Astrocytes), and Slc1a2 (Radial Glia).

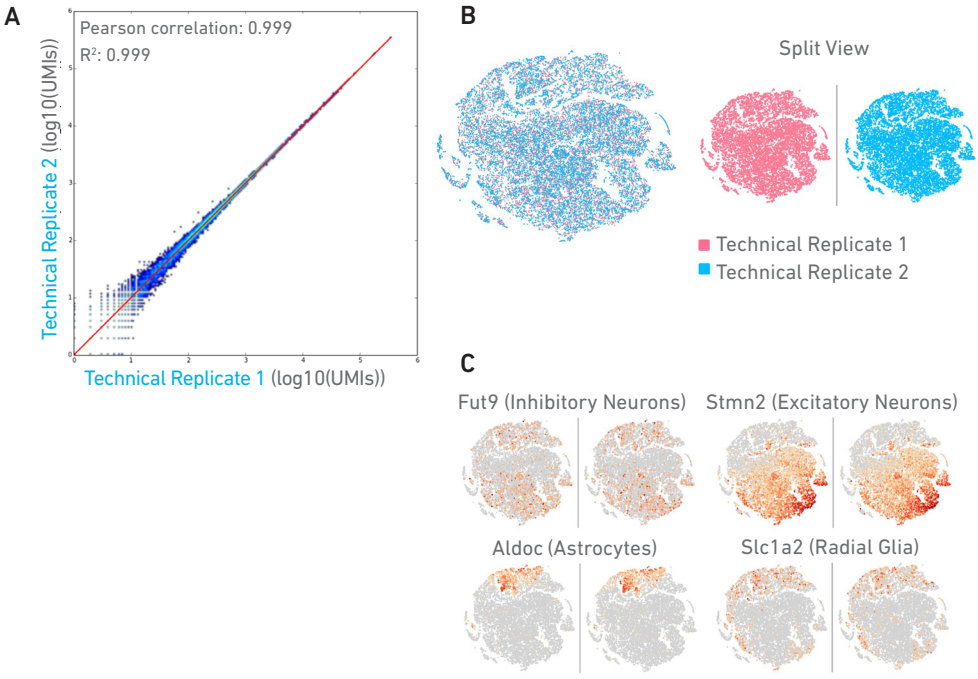


Figure 3. Gene expression variation when the same sample is run on two separate chips.

A. Scatter plot showing the correlation of average gene expression levels between the replicates on a log10 scale.

B. Merged (left panel) and split (right panel) tSNE clustering for the technical replicates.

C. Expression profiles for known neuronal marker genes Fut9 (Inhibitory Neurons), Stmn2 (Excitatory Neurons), Aldoc (Astrocytes), and Slc1a2 (Radial Glia).

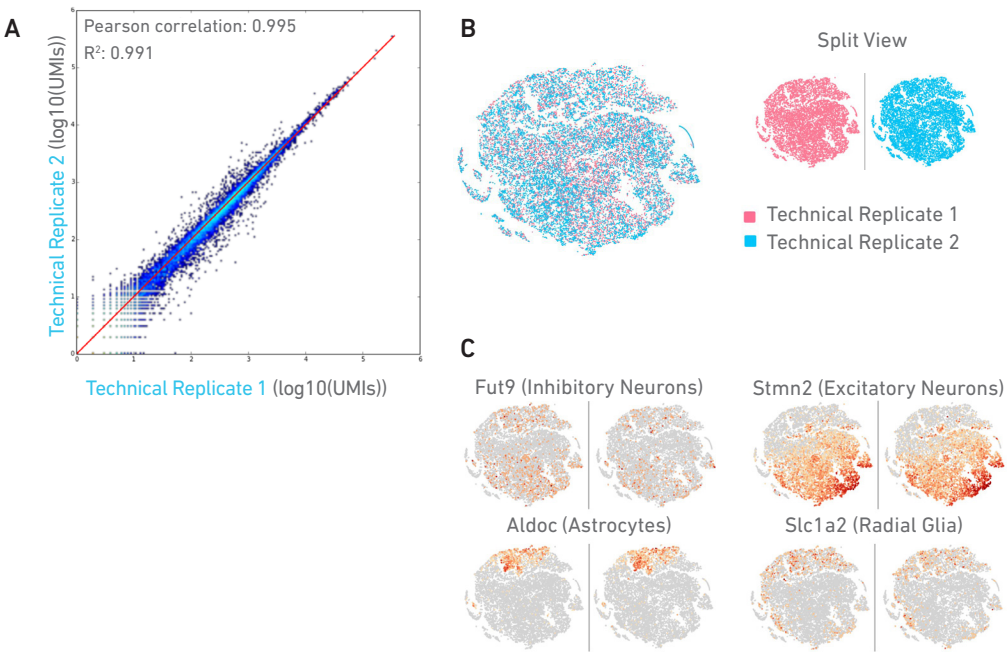
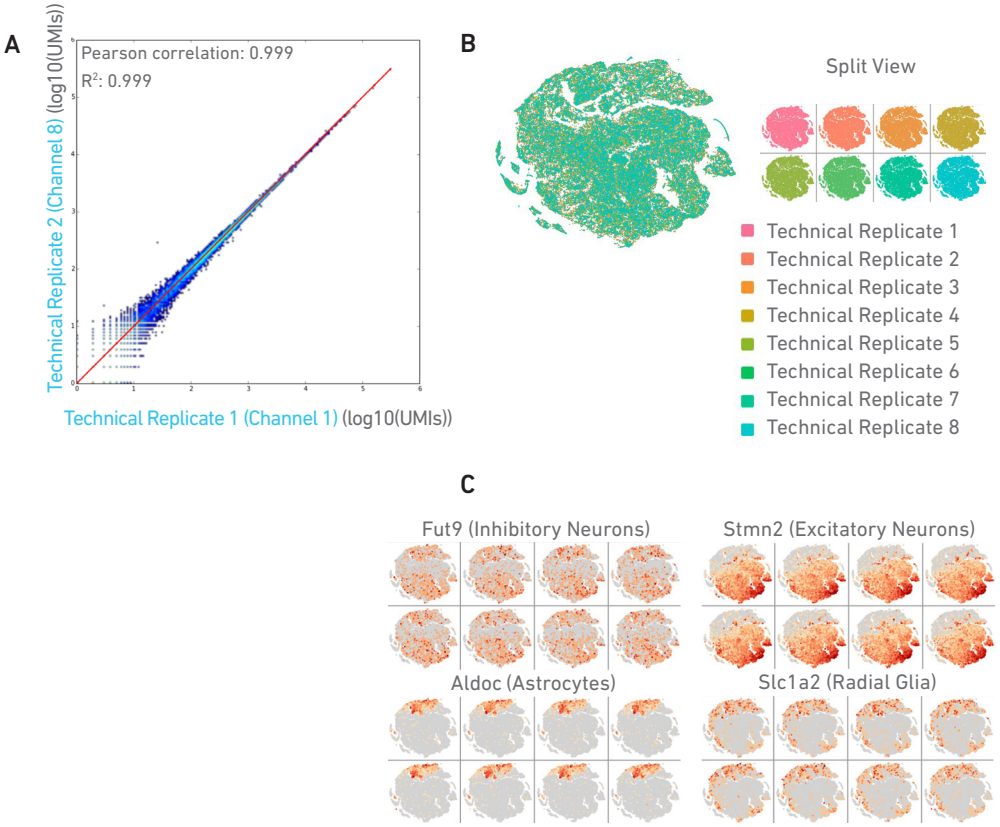


Figure 4. Gene expression variation when the same sample is run on different sample input wells of the same chip.

A. Scatter plot showing the correlation of average gene expression levels between wells 1 and 8 on a log10 scale.

B. Merged (left panel) and split (right panel) tSNE clustering for the 8 technical replicates.

C. Expression profiles for known neuronal marker genes Fut9 (Inhibitory Neurons), Stmn2 (Excitatory Neurons), Aldoc (Astrocytes), and Slc1a2 (Radial Glia).



Technical Replicate A

To assess the potential contribution of flowcell variability on gene expression analysis, neuronal tissue from one mouse was used to prepare a single cell suspension. A single library was generated by loading the cell suspension on a microfluidic chip. The library was sequenced on two separate flowcells (Replicates 1 & 2) on the same HiSeq 4000 sequencer (Table 1).

The technical replicates showed highly correlated gene expression profiles, evidenced by a Pearson correlation of 0.999 and R^2 of 0.999 (Figure 2A). tSNE clustering showed a high degree of concordance of cell distribution (Figure 2B) and nearly identical signatures for known neuronal marker gene expression across the cells (Figure 2C). These data indicate a minor contribution of flowcell to the overall gene expression differences between two samples assessed using the Chromium Single Cell 3' v2 workflow.

Technical Replicate B

To assess the potential contribution of the microfluidic chip on gene expression analysis, neuronal tissue from one mouse was used to prepare a single cell suspension. Two libraries (Replicates 1 & 2) were generated by loading the cell suspension on equivalent sample input well on two separate chips. Both libraries were sequenced on the same flowcell (Table 1).

The correlation between the gene expression profiles was high with Pearson correlation of 0.995 and R^2 of 0.991 (Figure 3A). High concordance of cell distribution was observed for the two separate libraries after tSNE clustering (Figure 3B) and known neuronal gene expression markers showed identical clustering across the cells (Figure 3C).

These data illustrate that use of multiple chips leads to minor variation in the overall gene expression profiles obtained with the Chromium Single Cell 3' v2 workflow.

Technical Replicate C

To assess the potential contribution of well-to-well variability with in a chip, neuronal tissue from one mouse was used to prepare a single cell suspension. Eight libraries (Replicates 1-8) were generated by loading the same cell suspension in each of the eight sample input wells of a chip. All 8 libraries were sequenced on one flowcell (Table 1).

Scatter plot analysis comparing average expression levels between libraries generated from well 1-8 indicate both a high Pearson correlation and R^2 of 0.999 (Figure 4A). Highly concordant cell distribution patterns were seen for the tSNE plots across all 8 wells (Figure 4B) and nearly identical expression signatures for four known neuronal marker genes were observed across wells (Figure 4C). This indicates that variability among wells within a chip is a negligible factor in the Chromium Single Cell 3' v2 workflow.

Conclusion

In this Technical Note, technical and biological sources of variation were evaluated for their effects on single cell gene expression data. Pearson correlation coefficients and R^2 values indicate highly correlated expression profiles between replicates, with biological variation exceeding all investigated sources of technical biases. In addition, near identical cell clustering and neuronal gene-marker expression signatures were observed across all conditions. Other sources of technical variation not investigated here include differences in sample handling (e.g. sample preparation, sample storage conditions), differences in user experience (e.g. handling samples or performing the Single Cell 3' protocol), and differences in sequencing libraries across Illumina platforms. These factors could also impact variability between experiments.

References

- Application Note – Transcription profiling of 1.3 Million Brain Cells with the Chromium Single Cell 3' Solution (Document LIT000015)
- Demonstrated Protocol – Dissociation of Mouse Embryonic Neural Tissue for Single Cell RNA Sequencing (Document CG00055)
- Chromium Single Cell 3' Reagent Kits v2 User Guide (Document CG00052)

Contact:
support@10xgenomics.com
10x Genomics
7068 Koll Center Parkway
Suite 401
Pleasanton, CA 94566 USA

10x GENOMICS