Supporting new massive scale single cell research initiatives through development of automated solutions for sample preparation

Dan Dubinsky, Cole Walsh, Tom Howd, Jim Meldrim, Sheila Dodge, Tim Desmet, Stacey Gabriel, Niall Lennon
Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge MA 02141

Introduction

Single cell and single nuclei sequencing is a continuously growing and powerful method of interrogating the transcriptome of individual cells. Platforms like 10x Genomics, make it possible to encapsulate thousands of cells/nuclei into a single reaction. However, technically complex and highly manual workflows associated with tissue dissociations limit the yield of cells/nuclei being targeted and the amount of reactions that can be batched together and processed at the same time.

We present a customized automated solution making it possible to receive nuclei from multiple sources and sequence at scale. We are able to receive, store, then batch cells/nuclei from multiple sources and run plate based cDNA generation to generate libraries consistently at 450bp with concentrations above 2ng/µl.

Results

Validation data with one library run on our automated platform compared a to comparable sample run manually resulted in the following Cell Ranger outputs:

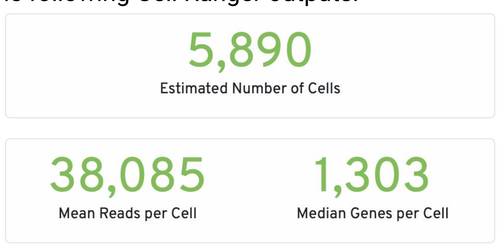
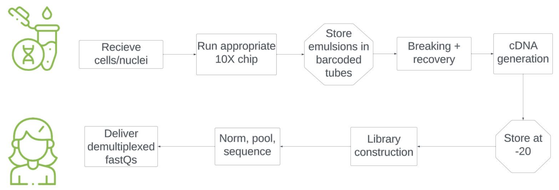


***Figure 1**: Cell calling and sequencing depth from the automated workflow matching the input target.*

Workflow



- Frozen PBMCs are prepared according to the 10x 10X manufacturers cell preparation protocol.
- Two aliquots of 8,000 cells each are loaded into two channels of a single 10X Genomics 3’ V3.1 chip.
- Emulsions are frozen at -20 in barcoded tubes.
- Subsequent steps of cDNA generation and library construction, from breaking and recovery to the final indexed PCR cleanup are completed on customized automated protocols on the Agilent Bravo instrument, with the final normalization and pooling on a Hamilton Star.
- Samples are evenly pooled and sequenced on NextSeq5000 to a read depth of 40,000 reads/cell.

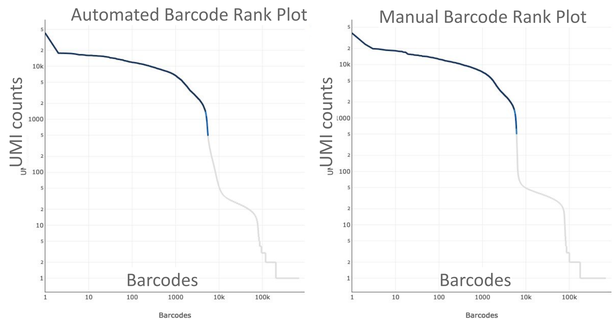


***Figure 2**: Single cell counts are consistent across both automated and manual protocols.*

Automated		Manual	
Number of Reads	224,318,379	Number of Reads	221,269,732
Valid Barcodes	97.2%	Valid Barcodes	95.6%
Valid UMIs	99.9%	Valid UMIs	100.0%
Sequencing Saturation	82.3%	Sequencing Saturation	72.4%
Q30 Bases in Barcode	97.3%	Q30 Bases in Barcode	94.4%
Q30 Bases in UMI	96.6%	Q30 Bases in UMI	93.8%

***Table 1**: Cell Ranger sequencing analysis of the automated protocol (left) compared to a library run manually (right). Minimal difference is observed in valid barcodes and UMIs. Q30 bases in barcode and UMI reads are comparable, and sequencing saturation was slightly improved with the automated platform.*

Scalability

- Sample receipt to completed LC in a simple, two day workflow.
- cDNA generation and library construction can be batched in sets of 96 samples.
- Aims to target each library at 5,000-10,000 cells and achieve data for over a million cells per batch.
- Potential for more than 5,000 libraries sequenced a year.
- Quality control to assess processing success is achieved with reviewing traces using the Agilent TapeStation.
- All sample metadata and QC is tracked via LIMS.

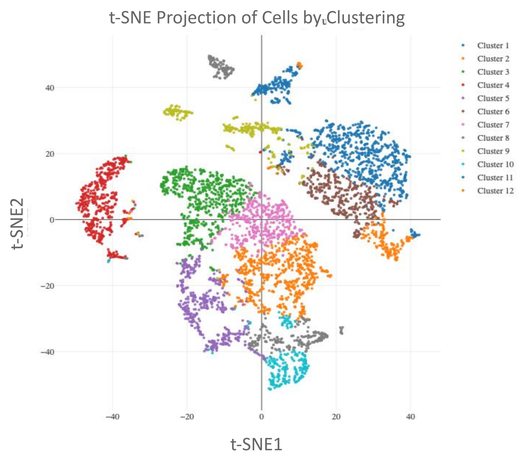


***Figure 3**: Clustering of PBMCs run through the automated workflow.*

Discussion

To achieve consistent results, we have determined that the cellular input should be within a range of 8,000-15,000 cells per batch. This ensures consistent RNA input for even PCR amplification.

Utilizing the Broad Clinical labs WalkUp Sequencing service, individual libraries can be pooled and customized to the projects needs and delivered directly to secure buckets for each collaborator.

While this enables a massive potential scale of both libraries and cell inputs, batching many emulsions from different sources limits the use of custom primers or potential alternate LC methods.

- **Compatible Products:**
  - 3’ Gene Expression
    - Feature Barcoding Optional
  - 5’ Gene Expression
    - Feature Barcoding Optional
    - V(D)J Enrichment Optional
  - Multiome Gene Expression+ATAC

Summary

Using this automated protocol with 10X Genomics 3’, 5’, or Multiome products, combined with increasingly accessible sequencing options, such as the Illumina NovaSeqX platform, we have massively increased the potential scale of single cell projects while not increasing costs. The automation shows little to no detrimental effect on the quality of data, and has the added benefit of reducing human pipetting and handling errors in lab. By running large batches, large scale single cell studies are faster, more accessible, and cost effective.

Acknowledgement

10x GENOMICS illumina® Agilent