A scalable high-throughput method for RNA-Seq analysis of thousands of single cells

Kevin Taylor[1], Lisa Watson[1], Lucas Frenz[2], Doug Greiner[2], Ronald Lebofsky[2], Duc Do[2], Pallavi Shah[2], Pengchin Chen[2], Mary Ma[2], Bin Zhang[2], Preeti Pattamatta[2], Leanne Javier[2], Joshua Mopas[2], Jennifer Chew[2], Sean Cater[2], Carolyn Reifsnyder[2], Felix Schlesinger[1], Irina Khrebukova[1], Jay Patel[1], Charles Lin[1], Jeff Tsai[1], Valerie Montel[1], Mariko Kellogg[1], Andrej Hartnett[1], Allison Yunghans[1], Gary P. Schroth[1], Jeremy Agresti[2]

Abstract

Complex biological systems are fundamentally determined by the coordinated functions of individual cells. The transcriptional heterogeneity that drives this complexity is often masked by conventional technologies that only provide bulk transcriptome data. Although high-dimensional gene expression analysis has been enabled by RNA-Seq, it is currently still a challenge to generate thousands of single-cell NGS libraries in an affordable, high-throughput, and user-friendly manner. To truly deliver on the promise of single-cell biology, a robust technology is required that enables controlled experiments with multiple samples, treatment conditions, and time points.

Here, we present the Illumina Bio-Rad Single-Cell Sequencing Solution. This new platform pairs Bio-Rad's Droplet Digital™ Technology with Illumina NGS library preparation and analysis technology to provide a comprehensive workflow for single-cell analysis. Single cells are individually partitioned into subnanoliter droplets using a disposable cartridge on the one-touch ddSEQ™ Single-Cell Isolator. The cartridge can accommodate multiple samples, and multiple cartridges can be processed in parallel to isolate tens of thousands of cells in a matter of minutes. Cell lysis and cell barcoding occur inside individual droplets, and single-cell-barcoded RNA-Seq libraries are subsequently prepared using Nextera® Technology. Data analysis is conducted via BaseSpace Sequence Hub®, the Illumina cloud-based genomics computing environment.

This droplet-based method is agnostic to mammalian cell size, enabling unbiased profiling of diverse cell populations. Additionally, because the time from culture to lysis is on the order of a few minutes, transcriptional signatures are not affected by lengthy experimental workflows allowing for acute transcriptional responses to be detected and tracked by time course. This combination of a cost-effective, simple, and fast workflow enables new types of single cell information to be revealed by allowing users to analyze multiple samples in parallel, under multiple treatment conditions and at multiple time-points. We demonstrate reproducible interrogation of single cell transcriptomes from multiple cell types. This scalable, robust single-cell NGS sample prep methodology will enable more researchers to apply the sensitivity and precision of RNA-Seq to questions in single cell biology.

1. Illumina, Inc.
2. Digital Biology Center, Bio-Rad Laboratories, Inc.

Illumina Bio-Rad Single-Cell Sequencing Solution

Fig. 1. Sample to answer workflow. The workflow leverages proven cell isolation using Droplet Digital technology with the ddSEQ Single-Cell Isolator, SureCell™ WTA 3' Library Prep Kit with Nextera technology, Illumina sequencing, and BaseSpace NGS analysis.

Overview of 3' RNA-Seq Assay

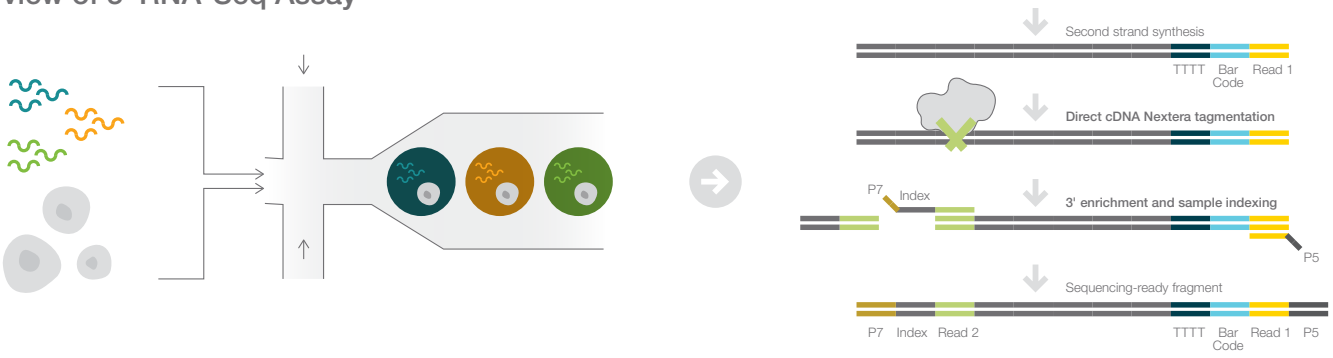

Fig. 2. Overview of 3' RNA-Seq assay. The SureCell WTA 3' Library Prep Kit. Lysis and cell barcoding takes place in each droplet. Droplets are disrupted and cDNA pooled for second strand synthesis in bulk. Libraries are generated with direct cDNA tagmentation followed by 3' enrichment and sample indexing.

Methods

HEK293 cells and NIH3T3 cells (unless otherwise noted) were mixed at a 1:1 ratio, loaded across 4 sample chambers of a single ddSEQ M Cartridge and encapsulated and barcoded by the ddSEQ Single-Cell Isolator. Barcoded transcripts were processed for single cell sequencing using the SureCell™ 3' WTA Library Prep Kit for the ddSEQ System and sequenced on the Illumina NextSeq® 550 sequencer. Sequencing results were analyzed using the SureCell™ RNA Single-Cell App.

Detection of Genes in a Heterogeneous Population of Cells

RNA-Seq Analysis of 1,384 Cells Using BaseSpace Single Cell App

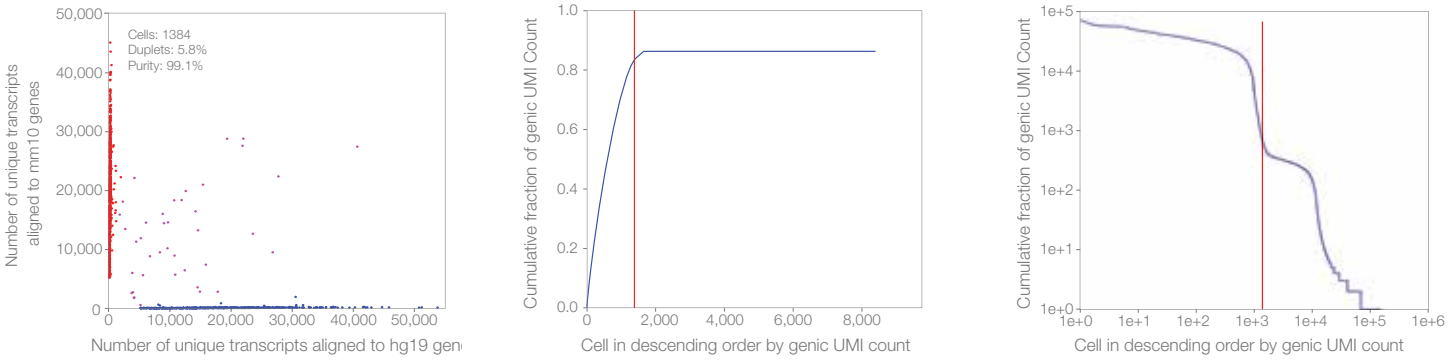

Fig. 3A. Two-species cell mixture demonstrates low crosstalk and high purity. Number of unique transcripts aligned to mouse (red) and human (blue) genome for each cell barcode. Unique transcripts mapping to both human and mouse are shown in purple and represent doublets (left panel). Cumulated fraction of unique transcripts assigned to cell barcodes in linear scale (middle panel) and log scale (right panel). The inflection point (knee) is used to determine the number of barcoded cells detected in the run.

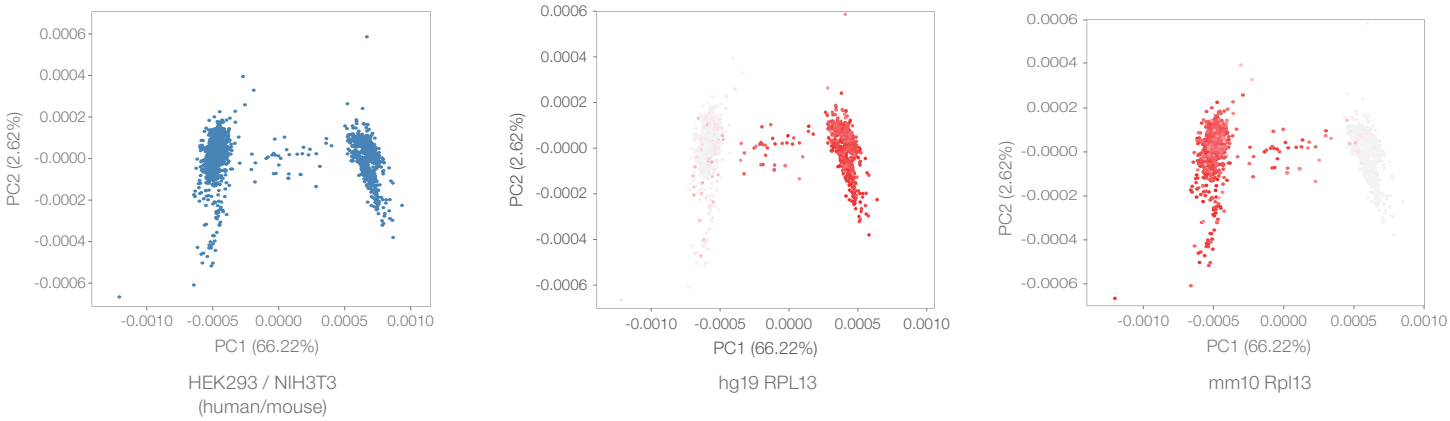

Fig. 3B. PCA clustering of 1:1 mixture of mouse and human cells detects distinct population. (Left) PCA analysis of 1,384 cells from a 1:1 ratio mixture of HEK293 and NIH3T3 cells using the Illumina BaseSpace single cell application. Cells color-coded by gene expression of human RPL13 gene (middle) or mouse Rpl13 (right).

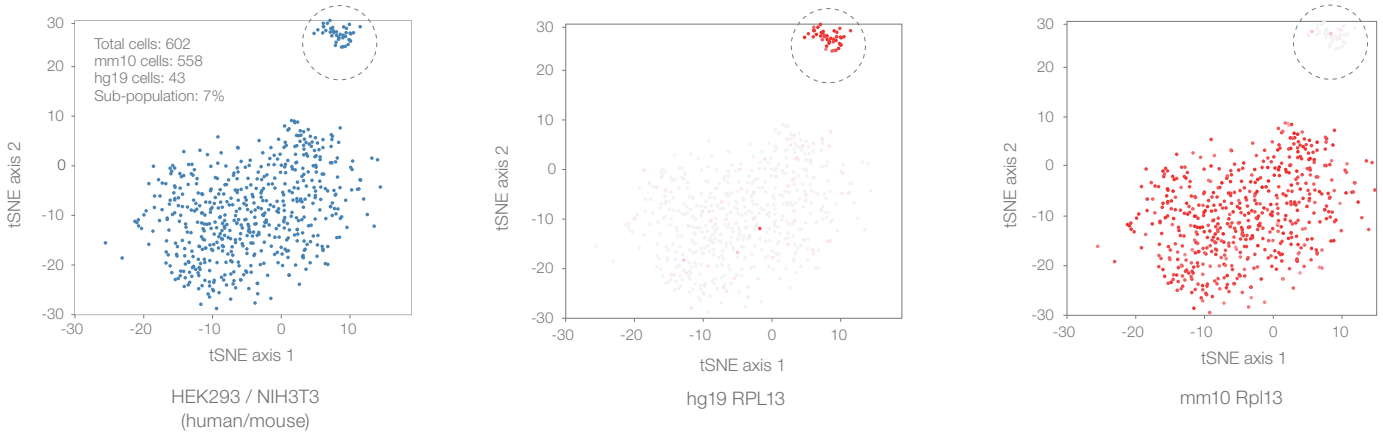

Fig. 3C. T-SNE analysis identifies a sub-population in a heterogeneous cell mixture. A mixture of mouse cells spiked with human cells. The human cells (representing 7% of the total cell population) are identified as a distinct cluster in t-SNE analysis based on gene expression profile (left). Cells color-coded by gene expression of human RPL13 gene (middle) or mouse Rpl13 (right) confirm the identity of the sub-population.

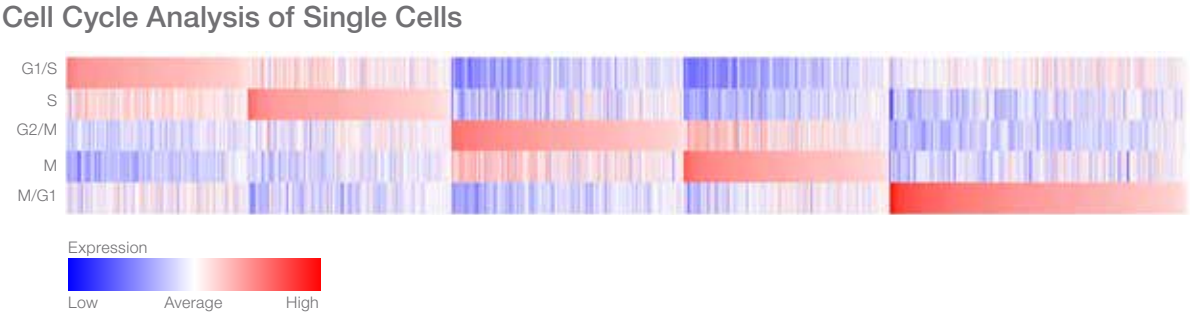

Fig. 4. Heat map based on the cell cycle by counting unique cell cycle transcripts. Cell cycle state based on unique transcript counts of genes in each cell cycle, normalized by total count for each cell for a mixture of HEK293 and NIH3T3 cells lines. Expression is centered by the median and scaled by the median absolute deviation for each cell cycle.

Sensitivity of Gene Detection Across Varied Cell Lines

Sensitivity of gene detection across varied cell lines and at varied read depths

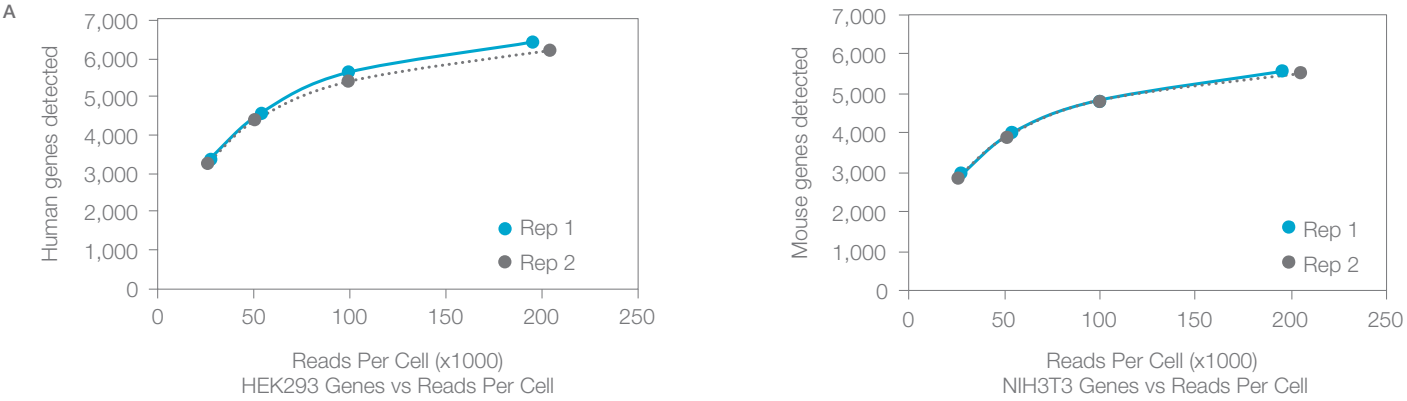

Fig. 5A. Replicate samples processed using a single cartridge were sequenced on NextSeq 550 and sequencing reads were sub-sampled to varied reads per cell ranging from 25,000 reads to 200,000 reads per cell. The median genes detected per cell are plotted at each sequencing depth.

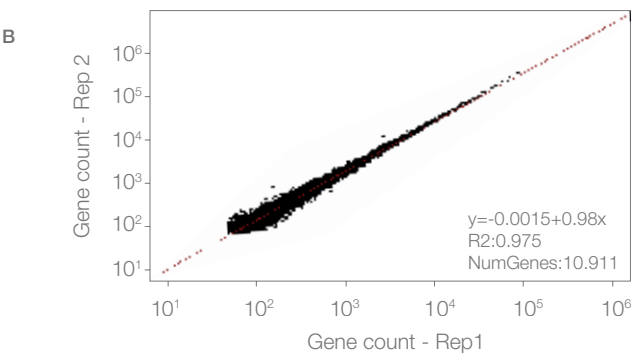

Fig. 5B. Reproducibility of gene expression for two replicates. Linear regression fit of gene counts (for genes with ≥ 50 counts) summed across all HEK293 cells, from two samples processed on a single chip shows high reproducibility.

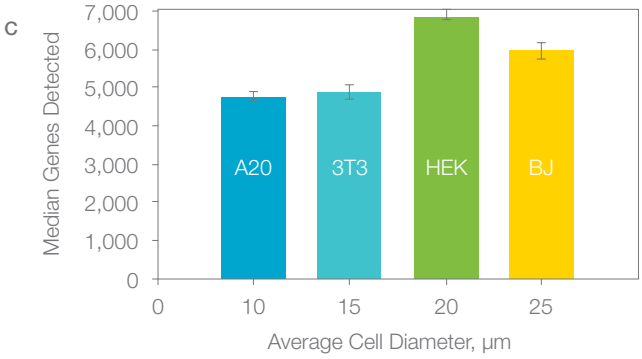

Fig. 5C. Sensitivity of gene detection across a panel of cells of varied diameter. The median genes detected per cell versus the cell diameter shows that recovery of transcripts is not limited by cell size.

Conclusions

- The Illumina Bio-Rad Single-Cell Sequencing Solution can reproducibly partition and analyze thousands of single cells in sub-nanoliter droplets from multiple cell lines in minutes with a simple protocol without pre-amplification.
- Analysis of human and mouse cell line mixing experiments demonstrates the ability of this platform to distinguish cells in a heterogeneous population by gene expression profiles.
- Robust chemistry allows for a high percentage assignment of transcripts to single cell barcodes in multiple cell lines.
- Transcriptional variation can be measured in single cells by analyzing changes in cell cycle gene expression.
- High sensitivity of gene expression is detected across a number of cell types and is not impacted by cell diameter

Visit www.bio-rad.com/ddSEQ for more information.

Pub No. 1070-2016-013
Bulletin 6883 Ver B 16-1124 1116