Comparative analysis of single-cell RNA Sequencing Platforms and Methods

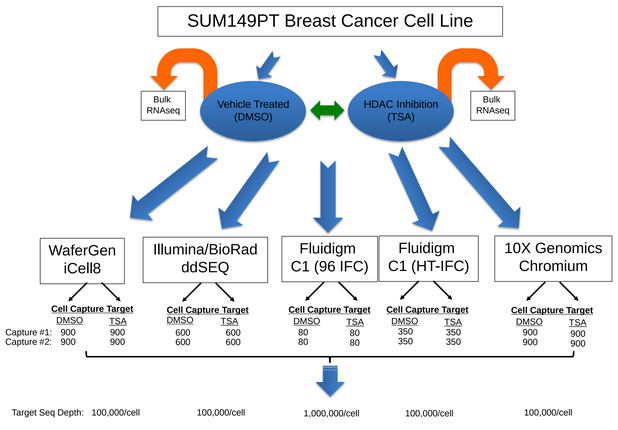
Sridar V. Chittur[1], Marcy Kuentzel[1], Michelle Zanche[9,10], Marlene Balys[9,12], Natalia Garcia-Reyero[2], Catharine Aquino Fournier[3], Hubert Rehrauer[3], Robert G. Steen[4], Sarah A. Boswell[4], Jin Jen[5], Christopher E. Mason[6], Jonathan Foox[6], Shahin Mohammadi[7,8], Jose Davila-Velderrain[7,8], Manolis Kellis[7,8], Jason R. Myers[9,10], Jacquelyn Lillis[9,10], Andrew McDavid[11], John M Ashton[9,10,12]

[1]Center for Functional Genomics, University at Albany-SUNY; [2]US Army ERDC Environmental Laboratory; [3]Functional Genomics Center Zurich, ETHZ; [4]Biopolymers Facility, Harvard Medical School; [5]Genome Analysis Core, Mayo Clinic; [5]Institute for Computational Biomedicine, Weill Cornell Medicine; [7]Computer Science and Artifical Intelligence Laboratory (CSAIL), MIT, [8]Broad Institute of MIT and Harvard; [9]Department of Microbiology & Immunology, [10]Genomics Research Center, University of Rochester, [11]Department of Biostatistics and Computational Biology, University of Rochester, [12]James P. Wilmot Cancer Institute, University of Rochester.

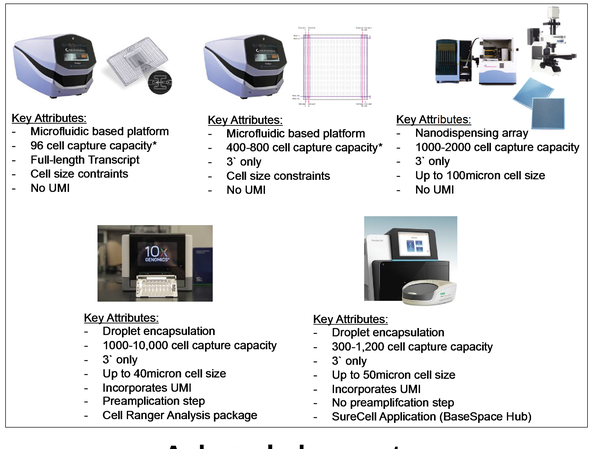
Introduction

The Genomics Research Group (GRG) generated data from hundreds of individual SUM149PT cells treated with the histone deacetylase inhibitor TSA vs. untreated controls across several scRNA-Seq platforms (Fluidigm C1, WaferGen iCell8, 10X Genomics Chromium Controller, and Illumina/BioRad ddSEQ). The goals of this project are to demonstrate RNA sequencing (RNA-Seq) methods for profiling the ultra-low amounts of RNA present in individual cells, and RNA amplification using the various currently available platforms. We will discuss the results of the study as well as technical challenges/lessons learned and present general guidelines for best practices in sample preparation and analysis.

Experimental Design



Platform Technologies



Key Attributes:
- Microfluidic based platform
- 96 cell capture capacity*
- Full-length Transcript
- Cell size contraints
- No UMI

Key Attributes:
- Microfluidic based platform
- 400-800 cell capture capacity*
- 3` only
- Cell size constraints
- No UMI

Key Attributes:
- Nanodispensing array
- 1000-2000 cell capture capacity
- 3` only
- Up to 100micron cell size
- No UMI

Key Attributes:
- Droplet encapsulation
- 1000-10,000 cell capture capacity
- 3` only
- Up to 40micron cell size
- Incorporates UMI
- Preamplication step
- Cell Ranger Analysis package

Key Attributes:
- Droplet encapsulation
- 300-1,200 cell capture capacity
- 3` only
- Up to 50micron cell size
- Incorporates UMI
- No preamplifcation step
- SureCell Application (BaseSpace Hub)

Acknowledgements

We sincerely appreciate the support of Fluidigm (96 and 800 cell IFCs, reagents), Wafergen (icell8 chips, reagents and capture support), 10x Genomics (Chromium Reagents and Illumina sequencing reagents) and Illumina (Sequencing reagents, SureCell WTA reagents, ddSEQ demo, and dedicated support for SureCell analysis). Special thanks to Marcy Kuentzel at the Center for Functional Genomics at SUNY Albany and Michelle Zanche and Marlene Balys at the University of Rochester Medical Center for generating the samples. Thanks to Steven Lotz at the Neural Stem Cell Institute for assistance wit the iCell8 system. Thanks also to Baldwin Dilone at the Harvard Biopolymers facility for the single cell capture and library prep and Flow Cytometry Facility at URMC for cell sorting support.

Results & Observations

Platform	Technology	Capacity	Capture Efficiency	Cell Size Dependence	Cell Monitoring	UMI	Ease of Use	Type of Profiling	Cell Capture to Library Cost (per sample)	Cost (per Cell)	Cell Capture + Library Generation Time
WaferGen iCell8	Ordered Array	1,000-2,000	~30%	No	Yes	No	Moderate	3' expression	$ 3,100.00	$3.00	7hrs
Fluidigm 96	Nanofluidics	96	~20%	Yes	Yes	No	Moderate	whole transcript	$ 2,000.00	$20.83	10hrs
Fluidigm HT	Nanofluidics	400 - 800	~18%	Yes	Yes	No	Hard	3' expression	$ 2,500.00	$6.25 - 3.13	12hrs
10x Genomics v2	Droplet Encapsulation	1,000-10,000	~65%	No	No	Yes	Easy	3' expression	$ 1,500.00	$1.50 - $0.15	10hrs
Illumina/BioRad ddSEQ	Droplet Encapsulation	300 - 1,200	~3%	No	No	Yes	Easy	3' expression	$ 1,200.00	$4.00 - $1.00	10hrs

Table 1. Cost analysis and key attributes of scRNA-seq platforms. Costs were determined using list prices for each technology available through respective vendor.

Basic Metrics & Preliminary Assessment

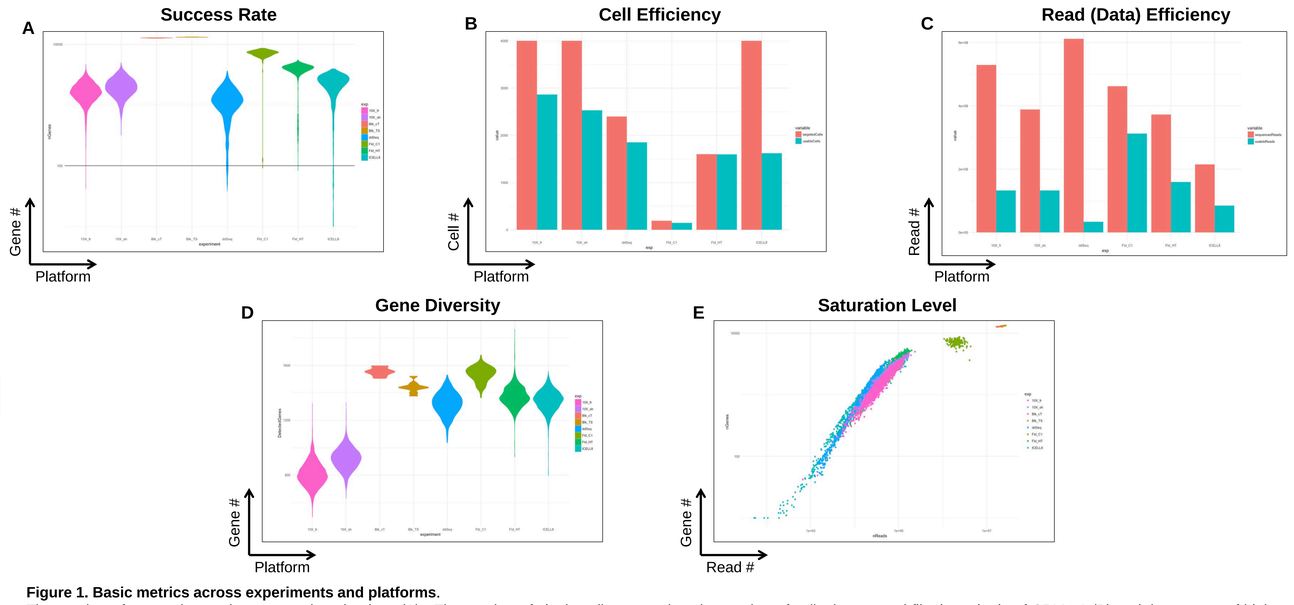


Figure 1. Basic metrics across experiments and platforms.
The number of genes detected across each technology (A). The number of single cells targeted vs the number of cells that passed filtering criteria of CPM >1 (B) and the amount of high quality "usable" data passing quality filter threshold (C). Gene diversity across platform technologies based on down sampling to 20,000 reads per cell (D). Unique gene detection for each platform based on sequencing depth (E).

Cell Outlier Detection & Removal

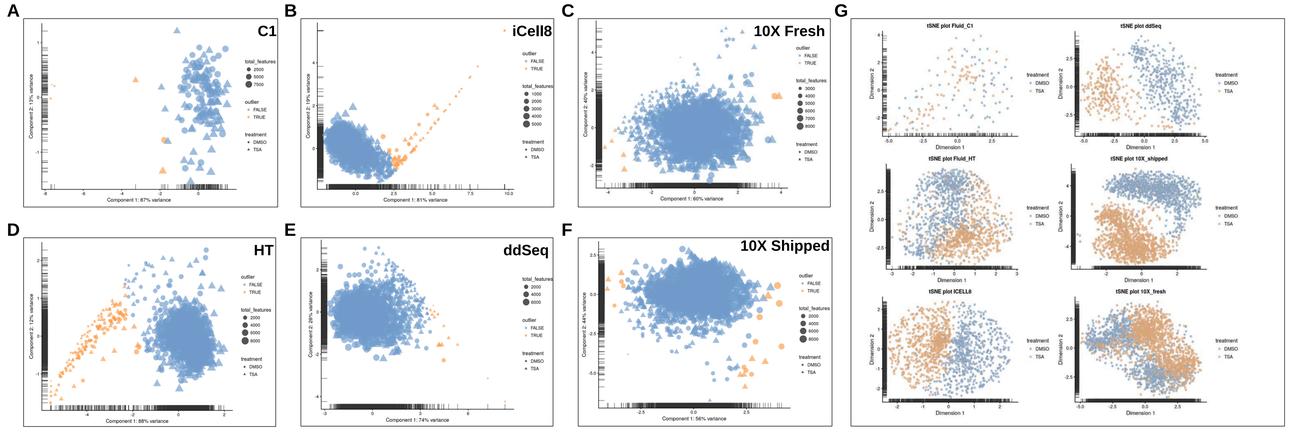


Figure 2. Cell outlier detection using Scater automated parameters.
Principal components for outlier detection results identify from Scater automated QC for data generated with Fluidigm C1 (A), WaferGen iCell8 (B), 10X Genomics with immediate processing (C), Fludigm HT (D), Illumina/BioRad ddSeq (E), or 10X Genomics with processing after overnight shipment (F). tSNE plots of single cell association based on treatment group (G).

Bulk RNA-Seq Differential Expression

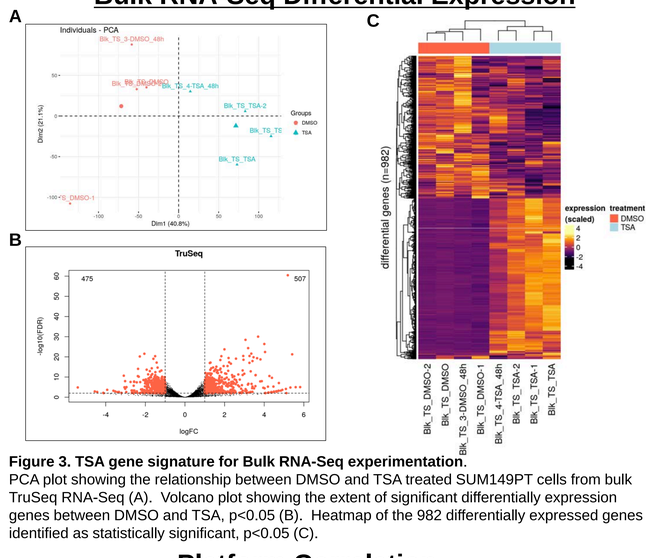
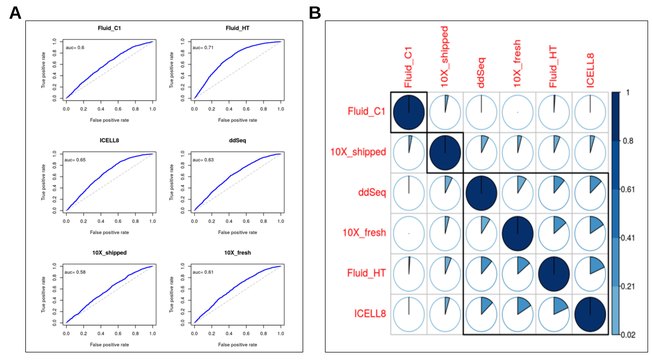


Figure 3. TSA gene signature for Bulk RNA-Seq experimentation.
PCA plot showing the relationship between DMSO and TSA treated SUM149PT cells from bulk TruSeq RNA-Seq (A). Volcano plot showing the extent of significant differentially expression genes between DMSO and TSA, p<0.05 (B). Heatmap of the 982 differentially expressed genes identified as statistically significant, p<0.05 (C).

Platform Correlation (Bulk RNA & Inter-platform)



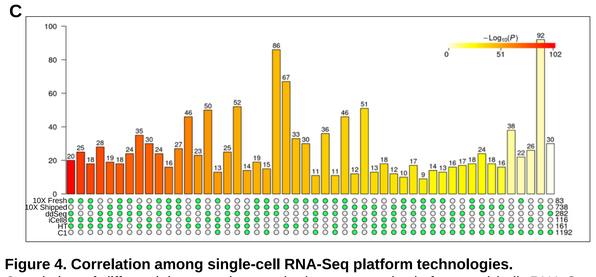


Figure 4. Correlation among single-cell RNA-Seq platform technologies.
Correlation of differential expression results between each platform and bulk RNA-Seq results (A). Correlation of differential expression results across platform technologies (B). Distribution of genes identified as differentially expression between DMSO vs TSA across all platforms tested (C).

Conclusions

- Each platform technology has some tradeoffs, such as in cell efficiency, data usability, or gene diversity.
- No platform reached gene detection saturation at sequencing depths tested, so all perform equally well in sensitivity under the tested conditions.
- Some platforms showed higher degree of outliers and differing ability to resolve treatment groups well.
- Using bulk RNA-Seq (TruSeq) as "ground truth", a low concordance was observed for differentially expressed genes across platforms.