GRC: Evaluation of five single-cell RNAseq platforms using patient-derived AML samples

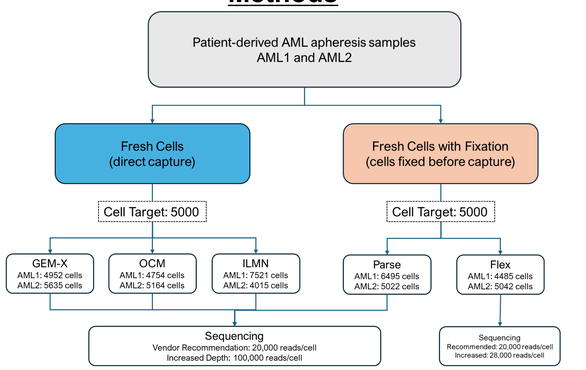
Elizabeth Pritchett[1], Jeffrey Malik[1], Marlene Rueby[1], Phil Spinelli[1], Cameron Baker[1], Hannah Aichelman[1], Jen Becker[1], Carter Long[1], John Ashton[1]

[1]Genomics Research Center, University of Rochester

Introduction

Single-cell RNA sequencing enables high-resolution profiling of complex tissues and has advanced our understanding of cellular heterogeneity. Many platforms are available, and selection of the appropriate methodology is essential for study design and outcomes. In this study, bone marrow samples were obtained from patients with MDS/AML after written informed consent in accordance with the Declaration of Helsinki and approval of University of Rochester institutional review board (IRB). We benchmarked five single-cell RNAseq platforms: 10X Genomics GEM-X, 10X Genomics Flex, 10X Genomics On-Chip Multiplexing (OCM), Parse Biosciences, and Illumina Single Cell (ILMN). In this study, we compared quality metrics between all approaches and investigated the impact of increased sequencing depth beyond vendor recommendations.

Methods



Samples (N=4, AML1-4) were collected via apheresis and processed for cryopreservation. Frozen aliquots were thawed and assessed for viability and debris. All samples exhibited >90% viability with minimal to no debris prior to fixation and/or cell capture. For the **Parse Biosciences Evercode WT v3** platform, ~750,000 cells per sample (N=2, AML1-2) were fixed prior to combinatorial barcoding and library preparation, targeting 5,000 cells per sample. For the **10X Genomics Flex** platform, ~750,000 cells per sample were fixed prior to gel bead-in-emulsion (GEM) generation using the Chromium X instrument. Cells designated for GEM-X and OCM workflows were counted and loaded to target 5,000 cells per sample into the microfluidic chip for GEM generation on the same instrument. For **Illumina's** platform, particle-templated instant partitions (PIPs) were generated using vendor-supplied equipment and protocols.

Library preparation for all platforms was performed according to the respective manufacturer's recommendations. Notably, both **Parse Biosciences** and **10X Genomics Flex** employ fixation-based approaches but differ in methodology: Parse uses a **combinatorial barcoding strategy** that enables full-transcript interrogation, whereas 10X Flex uses **probe-pairs** to hybridize directly to mRNA targets.

All resulting libraries were sequenced on the **NovaSeq X PLUS** platform at the **University of Rochester Genomics Research Center (GRC)**. All platforms recommend ~20,000 reads per cell. In this study, we sequenced at higher depth (100,000 reads per cell for GEM-X, OCM, Parse, and Illumina; 28,000 reads per cell for 10X Flex; hereafter 'increased'). All samples were downsampled to 20,000 UMIs (hereafter 'recommended'), and samples at both recommended and increased depths were analyzed using the standard Seurat v4 workflow to compare vendor recommendations to increased read depth per cell. scDblFinder and Harmony were used for doublet classification and cross-platform integration, respectively. All samples (AML1-4) are presented in initial quality metrics (Fig. 1a,b, 4a), but only samples present across all platforms (AML1-2) were used in cluster-based analysis (Fig. 2a,b,c, 3a,b, 4b).

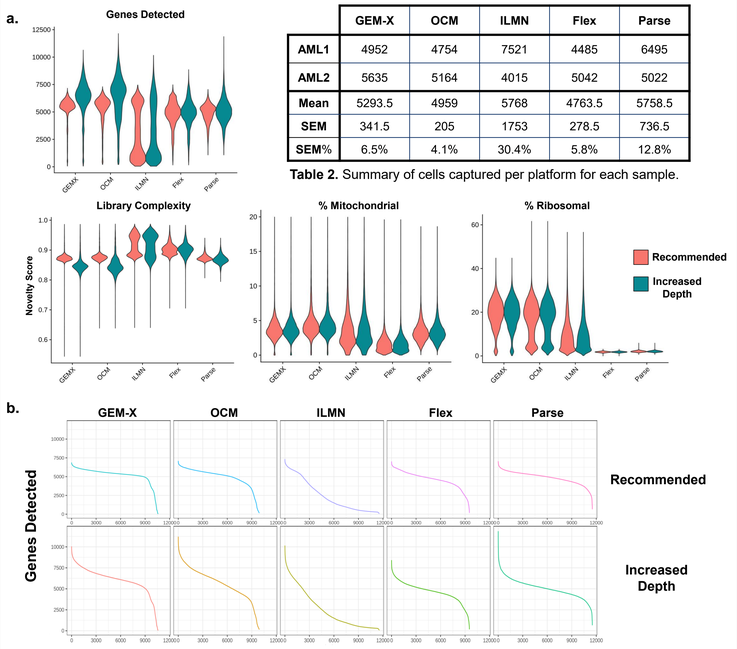
Scan the QR code for more detailed methods.

Results & Observations

Platform	Technology	Cell Fixation?	Type of Profiling	Capacity (cells)	UMI	Ease of Use (Lab Work)	Ease of Use (Analysis)	Automation Compatibility?	Cost (per sample)*
10X Genomics GEM-X	Droplet Encapsulation	No	3' + 5' Expression	20,000	Yes	Easy	Easy	No	$1,802
10X Genomics OCM	Droplet Encapsulation	No	3' + 5' Expression	5,000	Yes	Easy	Easy	No	$685
Illumina Single Cell T10	PIP Formation	No	3' Expression	10,000	Yes	Moderate	Hard	No	$600
Parse Biosciences Evercode WT v3	Cell Compartments + Combinatorial Indexing	Yes	Full Transcript	100,000 - 1 million	Yes	Moderate	Easy	Yes	$2,475
10X Genomics Flex	Droplet Encapsulation	Yes	Probe-Based Profiling	25,000 - 10 million	Yes	Easy	Easy	No	$600

Table 1. Summary of single-cell RNAseq platform key attributes and cost per sample. *Costs were determined using costs incurred to complete the study or list prices for each technology available through respective vendors. Contact vendor for most up-to-date pricing. Sequencing costs are not included in this estimate. "Capacity (cells)" for Fixation based methods indicate number of cells required for fixation.

Breakdown of metrics by platform



	GEM-X	OCM	ILMN	Flex	Parse
AML1	4952	4754	7521	4485	6495
AML2	5635	5164	4015	5042	5022
Mean	5293.5	4959	5768	4763.5	5758.5
SEM	341.5	205	1753	278.5	736.5
SEM%	6.5%	4.1%	30.4%	5.8%	12.8%

Table 2. Summary of cells captured per platform for each sample.

Figure 1a. Violin plots for Genes Detected (features), Library Complexity (novelty), % Mitochondrial, and % Ribosomal reads for all platforms at recommended and increased sequencing depth. Percentage ribosomal or mitochondrial content was generated using Seurat's *PercentageFeatureSet* with pattern matching corresponding to "RPS*|RPL*" and "MT-*" respectively. **Figure 1b.** Knee plots for all platforms at recommended and increased sequencing depth. Both **a** and **b** were generated using all samples.

Sample integration and clustering

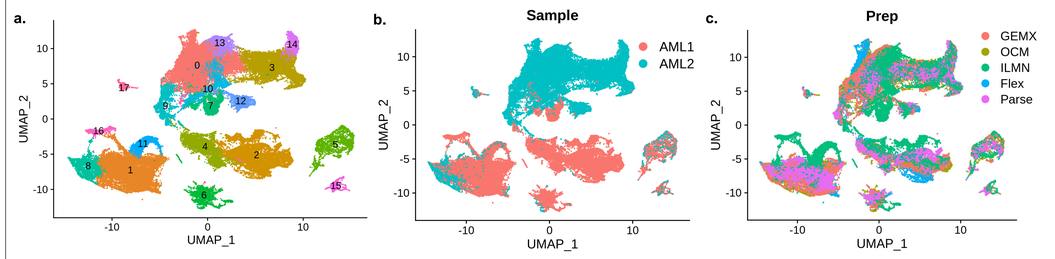


Figure 2a. UMAP of the integrated dataset used to determine impact of recommended versus increased sequencing depth. **Figure 2b.** Integrated dataset colored by sample. **Figure 2c**. Integrated dataset colored by platform. Only samples present across all platforms were used for clustering (AML1 and AML2, **a**,**b**,**c**).

Cluster breakdown by platform

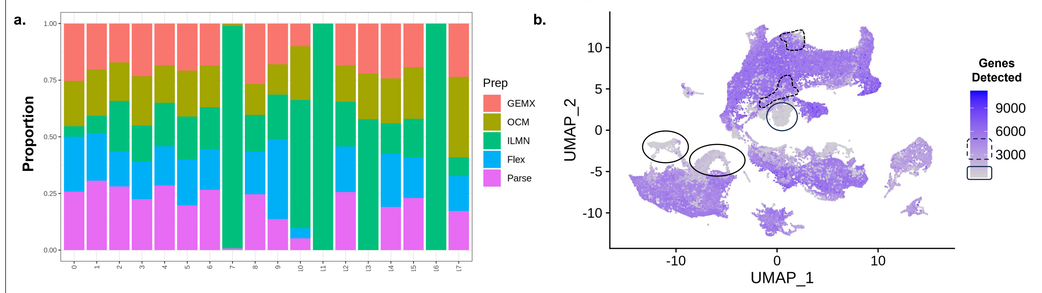


Figure 3a. Proportion of cells from each platform per cluster. Cells in clusters 7, 11, and 16 are predominantly from ILMN platform. **Figure 3b.** Cells within ILMN-dominated clusters are amongst those with fewest genes detected. By comparison, clusters with moderate amount of gene detection (10, 13) are represented by cells from multiple platforms. Only samples present across all platforms were used for clustering (AML1 and AML2, **a**,**b**).

Increased sequencing depth does not affect clustering

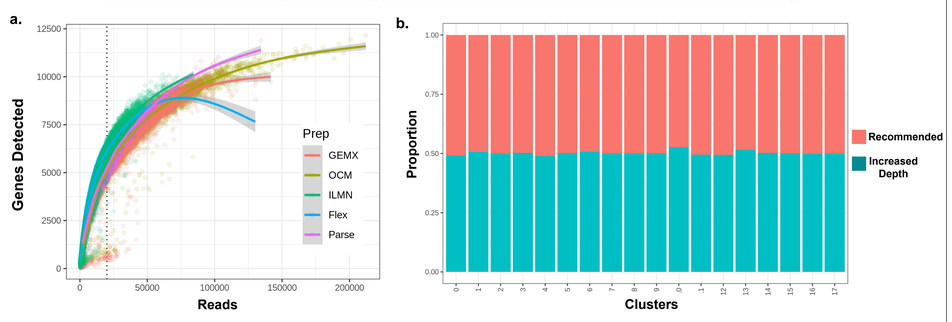


Figure 4a. Sequencing saturation curves for each prep type, generated using all samples. Dotted line represents number of reads recommended by vendors.
Figure 4b. Proportion of cells that come from recommended or increased sequencing depth per cluster, generated using the integrated dataset. Only samples present across all platforms were used for clustering (AML1 and AML2, **b**).

Conclusions

- Overall, platforms performed similarly with some exceptions for Illumina Single Cell T10.
- Increased read depth per cell led to increased gene detection but did not impact overall clustering for all platforms.
- All platforms tested are viable options for single cell RNAseq studies and platform selection will be based on many factors outside of standard metrics such as budget, ease of use, and experimental design or goals.

Future Directions

- Perform and compare cell-type identification across platforms.
- Compare recommended and increased read depths per cell within each platform.
- Further characterization of genes gained with increased sequencing depth.
- Investigate ILMN analysis workflow to determine if thresholds need to be adjusted for low-feature cells.
- Leverage Parse full-transcript coverage to investigate isoform expression.
- Integration of bulk RNAseq with scRNAseq.