

# Unlocking the potential for single-cell sequencing at scale with combinatorial indexing

Frank Steemers, Patrick Boyd, Maggie Nakamoto, Ashley Woodfin, Jerushah Thomas, Dmitry Pokholok, Dominic Skinner, Lin Lin, Utsab Ray, Nathan Pacheco, Felix Schlesinger, Toumy Guettouche, Jason Koth.

## Scale Biosciences, 3210 Merryfield Row San Diego CA 92121

### Background

#### Problem

Throughput of single-cell experiments has not scaled, primarily due to high costs, making single-cell screens and profiling of rare cell populations or large patient cohorts prohibitively expensive.

#### Solution

Here we increase the throughput of on-market systems by adding upstream combinatorial indexing technology using the ScaleBio™ scATAC pre-indexing kit, and with the stand-alone ScaleBio™ Single Cell RNA Sequencing Kit, which using three levels of indexing.

#### Workflow

- ATAC: Nuclei from up to 24 samples are distributed across a plate for barcoded tagmentation. Nuclei can then be pooled and superloaded onto existing single-cell capture systems with up to 100,000 nuclei per channel while using the additional barcode to recover data from multiplets, maintaining a theoretical effective doublet rate of 3-5%.
- RNA: A 50:50 mix of human K562 and mouse 3T3 cells were processed through the 3-level scRNAseq workflow. Sequencing of these libraries demonstrates strong RNAseq metrics and low background and doublet rates. Similar observations could be made for PBMCs processed through this workflow, with major cell types easily identified.

Figure 2: ScaleBio scATAC and scRNA-seq workflows produce high quality data with low background and multiplet rate



#### **Results**

Combinatorial indexing can be effectively applied to single cell applications to increase throughput while decreasing cost.

Figure 1: Combinatorial indexing technology can be used to increase single cell sequencing throughput, while decreasing cost.



- A. Combinatorial indexing uses repeated rounds of distributing cells across plates and subsequent well-based barcoding. Although multiple cells receive the same barcode in each round, repeated rounds creates a unique combinatorial barcode for the molecules in each individual cell.
- B. Schematic shows how 3 rounds of indexing produce 3.5 million unique index combinations. After plate-based workflow libraries are then sequenced and indices demultiplexed to assign reads to individual cells or nuclei

			Mouse cells	57.0%	
Mito. Reads	2.1%		Estimated	7 70/	-   р
Cells above threshold	42,831		doublets	5.5%	
% Reads in cells	95.0%		Background*	0.25%	
Mean Reads per cell	20,756			-	-
Median unique reads per cell	10,446				С
Saturation in cells	0.23				
Cells			3T3 Cells		
		ABCDEEGH			
. A 50:50 mix of hum through the Scale B	an K562 an io scRNA-se	d r eq	nouse 3T3 cells workflow. 10 w	were mixed p ells of the fina	rior to I distr

- low background.
- Barnyard plot shows number of mouse vs human unique reads in each cell. This plot shows distinct human and mouse cell clusters with low doublet rates (~3.3%) and low background.
- . (Left) Plot showing the number of nuclei found as singlets, doublets, triplets, etc in 10X droplets. (Right) The unique reads per cell remain constant even as the droplets are loaded with up to 5 nuclei per droplet.



Metric	Value
Mean reads/cell	63,327
Median molecules/cell	15,195
Saturation	63%

- being processed ribution plate were carried forward for sequencing, targeting approximately 13,000 cells.
- E. Plotting human and mouse unique molecules shows a low multiplet rate as expected (<5%) and low background rate, demonstrating that this workflow can generate clean data from large numbers of cells while maintaining a low doublet rate. High median unique molecule counts can also be observed demonstrating that fixed and permeabilized cells retain RNA adequately throughout the plate-based workflow.

Figure 3: ScaleBio ATACseq pre-indexing workflow produces good cell recovery, strong ATAC metrics, and successful cell-type identification with zebrafish brain and retina.



sequencing, the ScaleBio bioinformatics pipeline demultiplexes the cell barcodes to create a summary output, as well as files compatible with downstream secondary analysis applications.

B. The brain and retinas were dissected from 3 zebrafish; nuclei isolation via Dounce homogenization was then performed on each tissue. Up to 50,000 nuclei from each tissue were loaded into each well of a Scale plate, with samples being evenly spread across the tagmentation plate. 90,000 nuclei were then loaded onto a single channel of the 10X Chromium. C. Sequencing metrics are shown for one of each sample type, demonstrating high percent read alignment and reads in cells, and strong sensitivity. Clean knee plots can be observed enabling confident cell calling.

Figure 4: ScaleBio scRNA-seq kit produces high quality data with high cell recovery, UMI, and gene counts

cDC2
NK\_CD56bright

CD4 Proliferating
CD4 TEM





- A. Knee plot shows uniform signal, confident cell calling, and low background across the PBMC population.
- B. Metrics show good cell recovery and strong unique molecule recovery (>4k unique molecules/cell) with good gene diversity (>2k genes/cell).
- C. Plot depicting extrapolated molecules detected vs read depth
- D. UMAP showing predicted cell clusters from standard Seurat processing. Clusters show distinct localization with well-defined cell populations.
- E. UMAP identifying PBMC subtypes from 4D using Azimuth.
- F. Box plot showing abundances of each cell type as assigned by Azimuth.
- G. Dot plot showing expression PBMC lineage markers across cell types identified in figure 4D.



3

Lens

Retinal

Neurons

- D. UMAP of all 6 samples shows clear clustering of sample types with distinct localization of brain and retina samples.
- E. UMAP showing clustering of cells and identities resolves the expected cell types found in each tissue.

#### Conclusions

- The ScaleBio<sup>™</sup> scATAC pre-indexing kit can be used to increase throughput of existing scATACseq methods by enabling superloading of a 10X lane while maintaining low effective doublet rates and background and high sensitivity, nuclei recovery, and cell-type resolution.
- The ScaleBio<sup>™</sup> Single Cell RNA Sequencing Kit shows strong cell recovery with high sensitivity and low background. Data from PBMCs shows that the kit can recover the expected major cell populations from PBMCs.

© SCALE BIOSCIENCES, INC. 2023

