Automating three-level combinatorial indexing for ultra-high-throughput single-cell sequencing on the SPT Labtech firefly®+

Sujay Raj[1], Sanja Nikolic[1], Rory Maizels[123], Daniel M. Snell[1], Federica Falconio[4]

[1] Francis Crick Institute, [2] Institute/Wellcome Sanger, [3] Institute/EMBL-EBI, [4] SPT Labtech

Introduction

The scalability bottleneck

Single-cell combinatorial indexing enables transcriptomic profiling at a massive scale without the requirement for specialised microfluidic hardware (Cusanovich *et al*., 2015, Cao *et al*., 2017, Rosenberg *et al*., 2018). By utilising fixed cells or nuclei as reaction vessels, this approach allows for exponential scaling through iterative rounds of pooling and barcoding. While traditional two-level indexing provides a baseline for discovery, expanding to three-level indexing (here 384 x 384 x 384) increases the potential indexing space to over 56 million combinations, significantly lowering collision rates for high-throughput experiments (Figure 1).

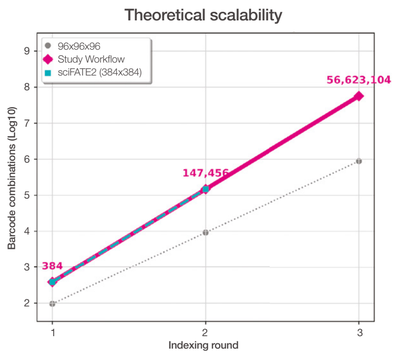

Figure 1. Theoretical scalability: Relationship between rounds of barcoding and theoretical barcoding space. As a reference, original combinatorial index approaches used 96 x 96 x 96 (grey line), and sciFATE2 used 96 x 384 (blue line).

Logistical complexity & automation

As protocols transition to three-level formats, the logistical complexity of managing thousands of discrete liquid handling steps across multiple 384-well plates presents a significant bottleneck. Manual execution is prone to human error, high cognitive load, and process variability. To address these challenges, we developed a high-throughput automated pipeline on the SPT Labtech firefly+. Building upon the protocol established by Maizels, Snell & Briscoe (2024a,b), this study focuses on the practical implementation of a three-level indexing framework to:

- Standardise the workflow and reduce technician touchpoints.
- Improve protocol robustness and spatial integrity across 384-well formats.
- Optimise recovery by mitigating cell loss associated with non-specific binding to plastic consumables.

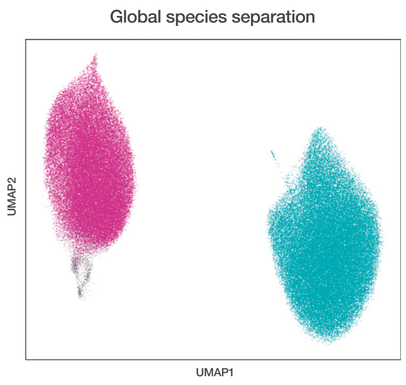

Figure 2. Global species separation: uniform manifold approximation and projection (UMAP) of transcript counts into multi-dimensional space, showing distinct species clusters even at relatively low sequencing depth.

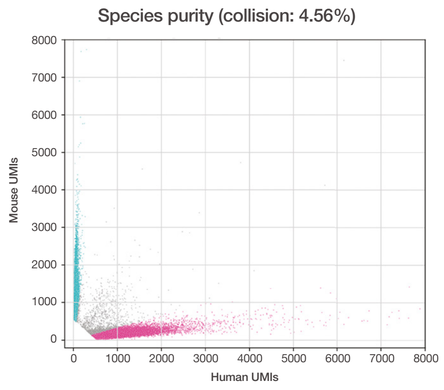

Figure 3. Species purity: the number of UMIs aligning to human and mouse genomes per sample (unique barcode combination), showing a doublet rate of 4.56%.

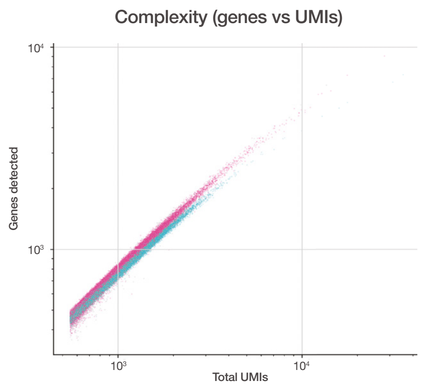

Figure 4. Complexity: correlation between UMIs and transcripts detected (log scale), separated by species.

The automated 4-method workflow

Workflow architecture

The library construction is partitioned into four discrete automated methods on the firefly+, interleaved with manual split-pool steps to facilitate high-throughput processing. This modular design allows for flexible run configurations while maintaining high precision throughout the indexing phases.

Instrument performance & efficiency

The automated workflow achieves a total of 106 minutes of instrument runtime, requiring only 6.5 minutes of active user interaction (Table). The remaining manual split-pool steps constitute the primary residual labour (~30 minutes each).

Phase	Human intervention	Deck Setup	firefly+	Thermal cycler (currently off-deck)	Split-Pool Steps	Approx. manual time
Method 1. RT set-up: Distribution of DNAse-toposiomerase (DT), cells, primers, and first strand reaction mix (FSRM)	2m 30s	10m	12m	RT incubation	1	60m
Method 2. Barcode ligation: Ligation mix, primers, and second strand synthesis mix (SSS)	2m 30s	10m	22min	Ligation incubation	1	90m
Method 3. Tagmentation & clean-up: TN55, DNA binding buffer (DBB), and SPRI-based purification	1m 30s	10m	62min	Tagmentation	0	120m
Method 4. PCR prep: Distribution of PCR mix and primers	0m 00s	10m	10min	Library amplification (PCR)	1	30m

Results

Pilot proof-of-principle: human-mouse species mix

To validate the automated indexing framework, a species mix pilot experiment was conducted utilizing 7.5M human (293T) and 7.5M mouse (mESC) cells. This run served to calibrate the firefly performance in handling high-volume cell suspensions and precise reagent distributions required for three-level indexing. Cells were fixed using methanol and stored at -80 for ~ 4 weeks prior to library preparation.

Data were analysed as previously described (Maizels, Snell & Briscoe, 2024a): briefly, raw fastq files were demultiplexed using the 96 x 384 x 384 combinations, and aligned to a composite mm10/GRCh38 reference using STAR-solo. QC thresholds were determined empirically, facilitating thresholding of UMIs and collision analysis.

Figure 5. Sequencing efficiency: histogram showing median sequencing depth per cell.

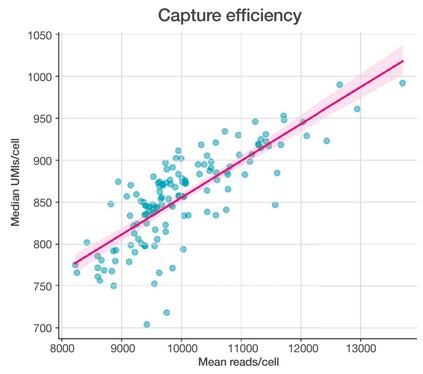

Figure 6. Capture efficiency: scatter plot showing correlation between mean reads per cell and median UMIs.

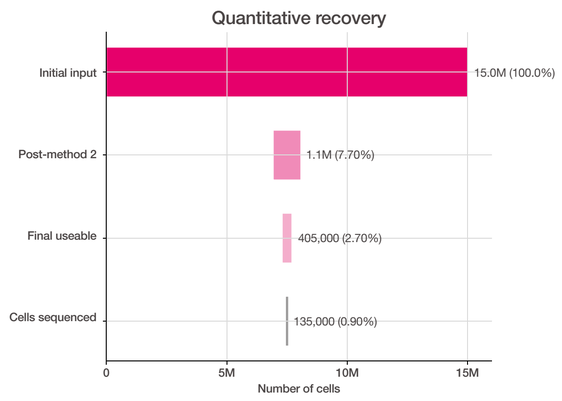

Figure 7. Quantitative recovery: funnel plot showing cell retention through different stages of the experiment.

1. Biological validation: purity & complexity

Barnyard analysis confirms the high spatial integrity and precision of the automated liquid handling:

- Species separation: clear global species separation is visualized via UMAP (Figure 2), with the barnyard plot demonstrating a collision rate of 4.56% (Figure 3). This indicates effective combinatorial indexing and limited barcode crosstalk during the automated “pool-and-split” sequence.
- Library complexity: a strong linear correlation between Genes vs. UMI counts confirms robust library construction, proving that the firefly air-displacement pipetting maintains the integrity of barcode transfers (Figure 4).
- Capture efficiency: median UMI capture shows a strong positive correlation with mean sequencing depth, suggesting that the libraries are high-complexity and non-saturated (Figures 5 & 6).

2. Quantitative recovery & yield

The recovery analysis identified a significant technical “funnelling” effect throughout the workflow (Figure 7):

- Post-Method 2 (RT & barcode ligation): cell recovery was measured at 1.1M cells (7.7% of initial input).
- Final library: The process yielded 405,000 useable cells (2.7% of input).
- Throughput proof: sequencing 1/3 of a 384-well plate to a depth of 1250M reads yielded 135,000 useable cells. This confirms the platform’s ability to recover high cell numbers even at pilot efficiency levels.

3. Process optimization: mitigating cell loss

The reduction from 15M initial input to 1.1M cells after the second indexing step highlights a critical technical bottleneck (Figure 7): non-specific binding to plastic consumables.

- The “sticking” challenge: cells and nuclei exhibit high affinity for the internal surfaces of syringes and pipette tips during the initial indexing rounds.
- Optimization strategy: current experiments on the firefly+ are evaluating specialized reagent formulations (e.g., BSA, PVP) and low-binding plastics (ULR syringes) to improve recovery efficiency and ensure maximum cell utilization across the three-level pipeline.

Conclusion

The implementation of three-level combinatorial indexing on the firefly+ provides a robust, scalable foundation for ultra-high-throughput single-cell transcriptomics. By automating the most labour-intensive indexing and clean-up phases, this platform eliminates the primary sources of manual error and provides a reproducible pipeline that significantly lowers the barrier to processing hundreds of thousands of cells per experiment.

Roadmap for firefly plus+

Development is ongoing to finalise a fully integrated, high-recovery pipeline:

1. Hardware upgrade: Integration of a 384-well ODTC to move all thermal cycling and incubation steps on-deck.
2. Full autonomy: utilising the added pooling capability of firefly+, automation of the remaining manual split-pool phases to achieve complete “walkaway” workflows.
3. Enhanced sensitivity: reduction of cell loss through low-bind reagents and plastics.
4. Cost reduction: implementation of targeted probe-based protocols to increase transcript resolution and further optimise sequencing expenditure.

References

Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).

Cusanovich, D. A. *et al.* Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).

Maizels, R. J., Snell, D. M. & Briscoe, J. Reconstructing developmental trajectories using latent dynamical systems and time-resolved transcriptomics. *Cell Syst.* **15**, 411-424.e9 (2024a).

Maizels, R. J., Snell, D. M. & Briscoe, J. A protocol for time-resolved transcriptomics through metabolic labeling and combinatorial indexing. *STAR Protoc.* **5**, 103356 (2024b).

Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).