Dec 26, 2019

MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing

MARS-seq2.0

DOI
dx.doi.org/10.17504/protocols.io.7hkhj4w

Hadas Keren Shaul[1], Hadas Keren-Shaul[2,3,4], Ephraim Kenigsberg[5,4], Diego Adhemar Jaitin[2], Eyal David[2], Franziska Paul[2], Amos Tanay[6,7], Ido Amit[2,7]

[1]Weizmann Institute of Science; [2]Department of Immunology, Weizmann Institute of Science, Rehovot, Israel;
[3]Life Science Core Facility, Weizmann Institute of Science, Rehovot, Israel;
[4]These authors contributed equally: Hadas Keren-Shaul, Ephraim Kenigsberg;
[5]Precision Immunology Institute, Icahn Institute for Data Science and Genomic Technology, and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA;
[6]Department of Computer Science and Applied Mathematics and Department of Biological Regulation, Weizmann Institute, Rehovot, Israel;
[7]These authors jointly supervised this work: Amos Tanay, Ido Amit

Human Cell Atlas Method …

Hadas Keren Shaul

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.7hkhj4w

Protocol Citation: Hadas Keren Shaul, Hadas Keren-Shaul, Ephraim Kenigsberg, Diego Adhemar Jaitin, Eyal David, Franziska Paul, Amos Tanay, Ido Amit 2019. MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. **protocols.io https://dx.doi.org/10.17504/protocols.io.7hkhj4w**

Protocol status: Working
We use this protocol and it's working

Created: September 19, 2019

Last Modified: December 26, 2019

Protocol Integer ID: 27916

Abstract

Human tissues comprise trillions of cells that populate a complex space of molecular phenotypes and functions and that vary in abundance by 4–9 orders of magnitude. Relying solely on unbiased sampling to characterize cellular niches becomes infeasible, as the marginal utility of collecting more cells diminishes quickly. Furthermore, in many clinical samples, the relevant cell types are scarce and efficient processing is critical. We developed an integrated pipeline for index sorting and massively parallel single-cell RNA sequencing (MARS-seq2.0) that builds on our previously published MARS-seq approach. MARS-seq2.0 is based on >1 million cells sequenced with this pipeline and allows identification of unique cell types across different tissues and diseases, as well as unique model systems and organisms. Here, we present a detailed step-by-step procedure for applying the method. In the improved procedure, we combine sub-microliter reaction volumes, optimization of enzymatic mixtures and an enhanced analytical pipeline to substantially lower the cost, improve reproducibility and reduce well-to-well contamination. Data analysis combines multiple layers of quality assessment and error detection and correction, graphically presenting key statistics for library complexity, noise distribution and sequencing saturation. Importantly, our combined FACS and single-cell RNA sequencing (scRNA-seq) workflow enables intuitive approaches for depletion or enrichment of cell populations in a data-driven manner that is essential to efficient sampling of complex tissues. The experimental protocol, from cell sorting to a ready-to-sequence library, takes 2–3 d. Sequencing and processing the data through the analytical pipeline take another 1–2 d.

Introduction

The remarkably rich repertoire of transcriptional states of cells within tissues has been studied for many decades. However, only recently have experimental and computational advances in the field of single-cell genomics opened the way for unbiased dissection of tissues into single cells and the de novo characterization of cell types, subtypes, and transcriptional states[1–13]. Single-cell RNA sequencing (scRNA-seq) is emerging as a key molecular tool for elucidating biological complexity, promising to contribute to a variety of fields in both basic research and medicine[14–17]. Methods for single-cell genome-wide expression analysis are continuously being developed, offering increasing coverage, precision and throughput[1,10,18–31]. Here, we describe a robust method for massively parallel scRNA-seq that combines indexed FACS sorting (recording of surface marker levels for each sorted single cell) and robotics with multiple layers of molecular barcoding. Although drop-based and microwell-based methods increased the throughput of scRNA-seq methods[17,29,30], the complexity of the cellular hierarchy remains a major challenge for unbiased sampling of mammalian tissues. Human tissues comprise trillions of cells that populate a complex space of molecular phenotypes and functions such that states within this space may vary in their abundance by 4, 5 and up to 9 orders of magnitude[32,33]. A classic example of this remarkable property of the human cell state space is the difference in abundance between highly abundant red blood cells ($N \sim 10^{13}$) and rare hematopoietic stem cells ($N \sim 10^{3-4}$)[34,35]. A well-designed single-cell technology must consider the orders-of-magnitude variation in abundance of human cell lineages and develop experimental, computational and statistical approaches to overcome it.

Because cells, and especially combinations of stem cells, progenitors and various types of mature cells, span many orders of magnitude in abundance, relying solely on unbiased sampling to accurately characterize rare populations becomes infeasible, as the marginal utility of collecting more cells diminishes quickly (Fig. 1a). Combining unbiased transcriptional maps with protein markers enables iterative refinement of cell sampling using FACS-based enrichment and/or depletion strategies for more efficient and deeper sampling of key rare subpopulations (Fig. 1a). Combining index sorting and scRNA-seq has many other potential utilities, including refinement and generation of new sorting panels for

both diagnostics and basic research and for mapping fluorescent reporter markers[36,37] for lineage tracing and CRISPR pooled screens (CRISP-seq[38]; Fig. 1b,c).

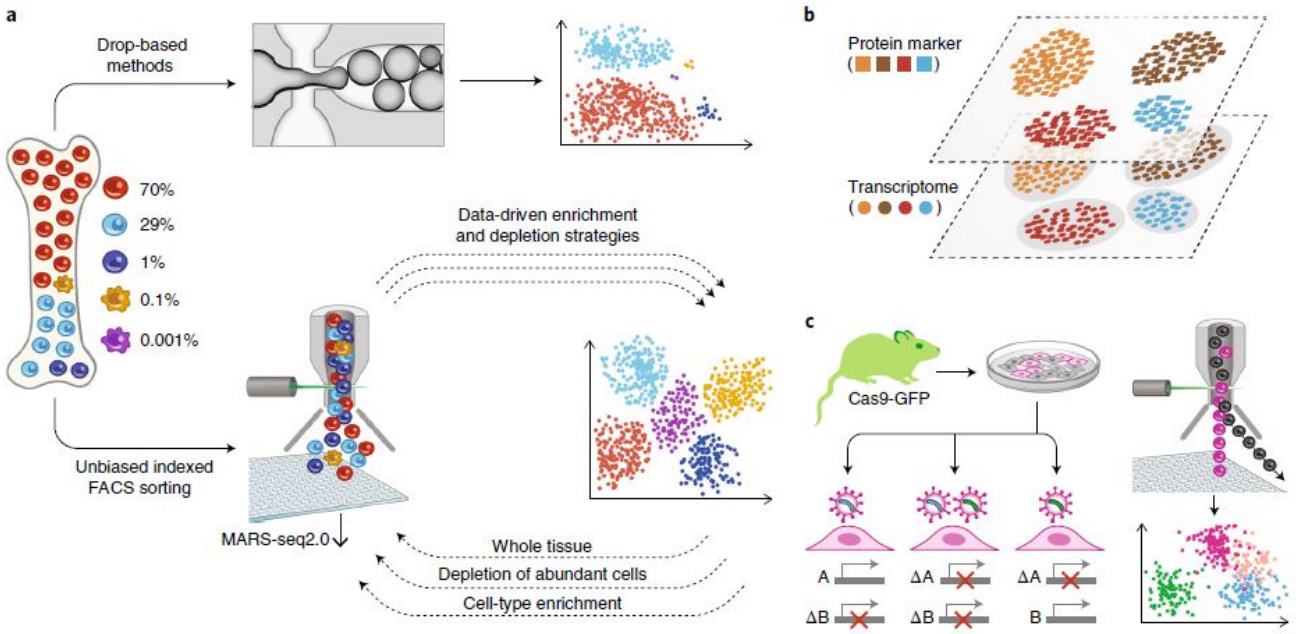

Fig. 1 | MARS-seq2.0 enables efficient sampling of rare subpopulations. a, Schematic presenting the ability of MARS-seq2.0, using unbiased indexed FACS sorting, to perform data-driven enrichment and depletion strategies (represented by forward and back arrows) to efficiently identify rare subpopulations, in comparison to drop-based methods. Bone is shown as an example of a heterogeneous organ characterized by scRNA-seq. **b**, Indexed FACS sorting (recording of surface marker levels for each sorted single cell), which is an optional element of MARS-seq2.0, enables correlation of unbiased transcriptional mapping with protein markers. **c**, MARS-seq2.0 can be integrated with CRISPR-pooled screens (CRISP-seq[38]). A perturbation of several targets can be made in vitro or in vivo, and an affected cell is detected by the cell transcriptome. Using a gene editing system in which cells derived from mice expressing Cas9-GFP are transduced with a pooled lentiviral library harboring guide RNAs, it is possible to perturb gene A (right side), gene B (left) or both (middle) in the same cell.

Overview of the experimental protocol

MARS-seq2.0 provides a complete experimental and computational framework for massively parallel scRNA-seq. It integrates index sorting into the scRNA-seq pipeline, minimizes doublets (two independent cells that are captured and processed together) and technical cell-to-cell contamination rates (background noise), and supports processing of tens of thousands of cells a day in a simple and costeffective manner. MARS-seq2.0 is based on our MARS-seq protocol, which we presented in February 2014 (ref. [10]). So far, we have used MARS-seq2.0 to process > 1 million cells from mice, non-model organisms and human studies[10,35,36,38–42], generating between ~100 and ~10,000 unique molecular identifiers (UMIs; molecular tags that serve to reduce quantitative bias) for each, with an average of ~1,700 UMIs and ~700 genes per cell (Supplementary Fig. 1). Using MARS-seq2.0, we identified a unique microglia type associated with restricting development of Alzheimer's disease[40,43], characterized hematopoietic progenitors[9,35], mapped the cellular composition of the lung during development[42] and in response to influenza virus infection[44], and detailed the heterogeneity of epithelial cells in the thymus[41]. We also used MARS-seq2.0 for whole-organism mapping of early metazoan cell type–specific transcription[45,46]. Further, we applied MARS-seq2.0 to large clinical studies characterizing the heterogeneity of plasma cells in healthy donors and multiple myeloma patients[47], as well as infiltrating immune cells in lung and

melanoma patients[48,49]. MARS-seq2.0 was also combined with other modalities, including single-cell CRISPR-pooled screens to dissect immune circuits in vitro and in mice[38], and photoactivatable reporters (NICHE-seq), single-molecule fluorescence in situ hybridization and paired-cell sequencing (gene expression profiling of pairs of joined cells) to spatially reconstruct immune niches[36,50,51].

In the MARS-seq2.0 framework, single cells are index-sorted by FACS into 384-well plates, lysed and automatically barcoded (with a cellular and a molecular barcode) by reverse transcription (RT) using a poly-T primer in a low-volume (500 nl) reaction. Then the single cells are pooled together, using a liquid-handling robot for subsequent molecular reactions to be performed on the pooled and labeled material. Each group of single cells amplified together is referred to as an 'amplification batch'. The pooled cDNA is converted into double-stranded DNA (dsDNA) and is linearly amplified by T7 RNA polymerase in vitro transcription (IVT). The amplified RNA (aRNA) is fragmented and a pool barcode is added by RNA–DNA ligation, allowing efficient sequencing of 8,000–10,000 cells in a single run. Illumina adaptors are then added by RT and PCR, which enables quantification of the mRNA (Fig. 2). During library preparation, we routinely evaluate the complexity of the libraries, yielding a quality control (QC) score that is used to monitor MARS-seq2.0 performance before sequencing. On the basis of these scores, amplified libraries suitable for sequencing of 20–40 384-well plates are pooled and sequenced together on a single flow cell. Sequenced reads are automatically processed in the MARS-seq2.0 computational pipeline (http://compgenomics.weizmann.ac.il/tanay/? page_id=672), in which the molecular signal is separated from the background noise.

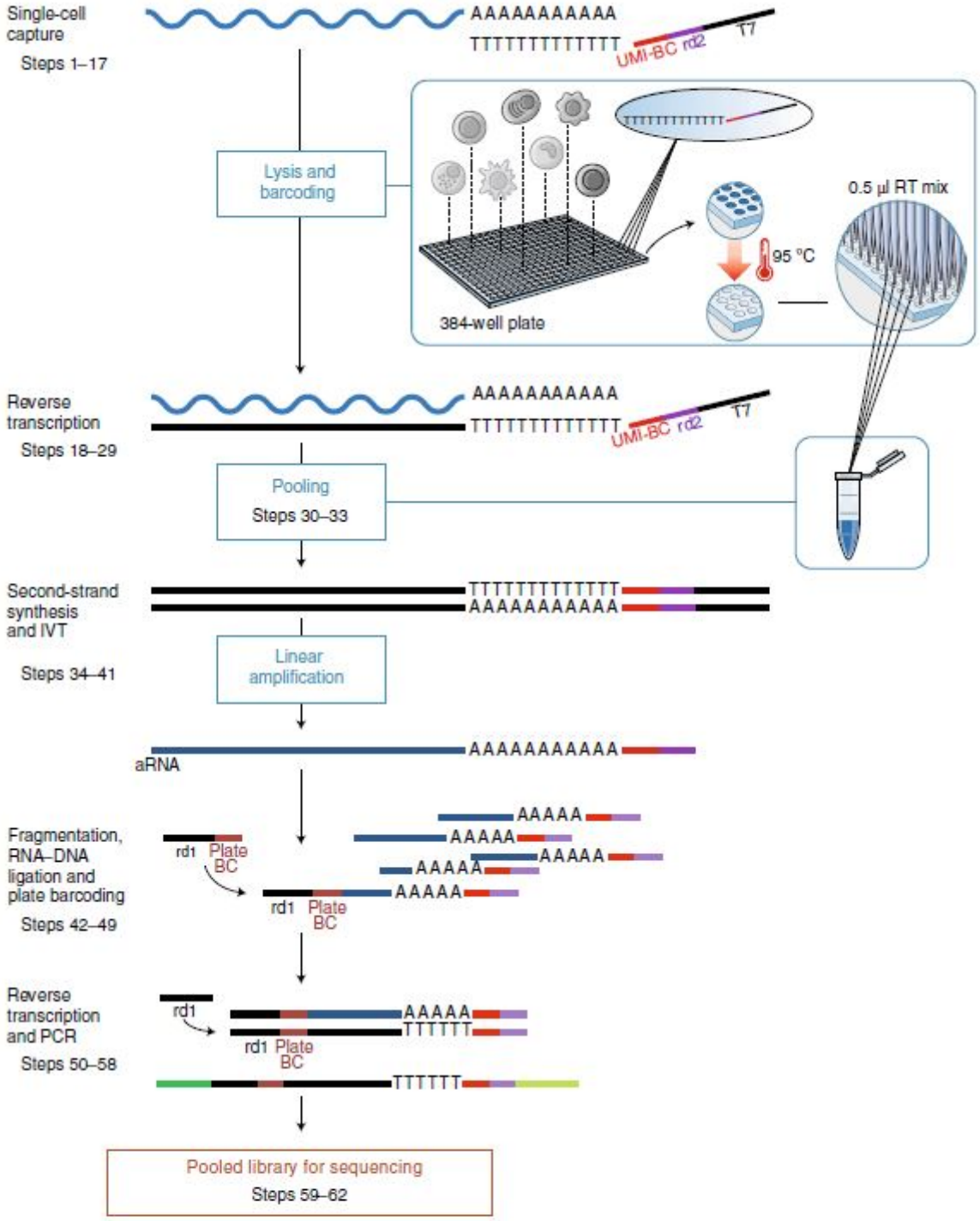

Fig. 2 | Schematic highlighting key experimental steps in MARS-seq2.0. The pipeline is based on FACS sorting of single cells, cellular barcoding and random molecular tagging of poly-A-tailed RNA molecules. Barcoded samples are pooled for library generation and next-generation sequencing. Plate BC, a sequence containing a plate (or amplification batch) barcode; rd1, rd2, Illumina sequences necessary for sequencing; T7, T7 RNA polymerase promoter sequence; UMI-BC, a sequence containing the unique molecular identifier and cell barcode.

Optimization of MARS-seq

Background noise is a major problem in all massively parallel single-cell protocols. Our prior study[10] suggested that this cell-to-cell contamination noise could partially be associated with a molecular process that randomly links a gene, a UMI and a cell barcode during library preparation. To enhance the experimental stability and detection rate while controlling for potential sources of systematic noise in MARS-seq2.0, we set out to optimize RNA retention and amplification. Starting from the MARSseq protocol, we optimized all steps of the single-cell processing, including lysis conditions, reaction volume, primer design and the composition of enzymatic reactions. We evaluated the possibility of reducing the RT reaction volume by analyzing 768 single mouse embryonic stem (ES) cells: after sorting, we performed complete evaporation of the 2 µl of lysis buffer containing the RT primer at 95 °C, followed by addition of 0.5 µl of RT mix, using a liquid-handling robot. For these conditions, RT primer quantities were considerably reduced. In low-volume reactions, we did not observe a substantial increase in the number of molecules measured per cell as compared with those of high-volume reactions, but rather saw an improved signal-to-noise ratio as evaluated by the number of molecules in the wells containing cells as compared with our negative-control empty wells (Supplementary Fig. 2a). This is also a consequence of the 200-fold reduction in the RT primer final concentration that prevents well-to-well contamination during the pooling and downstream enzymatic reactions. In addition, such low-volume reactions are cost effective (reducing enzyme consumption) and scalable. Because second-strand-synthesis enzymes are potentially a major sensitivity-limiting factor but can also generate unwanted background noise post pooling, we further evaluated several different second-strand-synthesis mixtures. We observed a small but considerable inverse correlation between the number of molecules obtained in the different conditions and the signal-to-noise ratio, suggesting that although some mixtures may be slightly more efficient in converting the hybrid RNA–cDNA template into a double-stranded DNA, this is accompanied by unwanted hybrid molecular products reaching a noise level of up to 40 % of the true signal in several of the mixtures (Supplementary Fig. 2b). These results demonstrate that it is critical to optimize and control for noise level in scRNA-seq analysis and that this factor can be affected by various steps during single-cell library production.

Key procedural advancements in MARS-seq2.0

MARS-seq2.0 introduces important modifications of the MARS-seq method in almost every part of the protocol. These improvements are related to throughput, robustness, noise reduction and costs. Specifically, we have performed the optimizations discussed in the following sections.

Experimental improvements

1. Lowering of RT volume: we reduced the volume of the RT reaction by eightfold from 4 µl to 500 nl. Lowering of the RT volume was possible due to evaporation of the sample in the 384-well plate before addition of the RT mix. This is a direct sixfold reduction in the cost of the single-cell library preparation because the RT step is the most costly part of the protocol.
2. Optimization of lysis buffer: we optimized the lysis buffer to make it compatible with the aforementioned volume reduction.
3. Reduction of RT primer concentration: we reduced the concentration of the RT primer from 200 to 1 nM (enabled by the lower volume), which further reduced the contamination and background noise.
4. Optimization of RT primer composition: we modified the cell barcode (7 bp) and the UMI (8 bp) to optimize yield with longer oligos that enable efficient error correction.
5. Primer removal by exonuclease I is performed in each well before pooling in order to maximize the exclusion of any RT primer leftovers that were not used in the RT and could be a source of potential noise.

6. Optimization of second-strand-synthesis enzymes: we tested all major commercially available second-strand-synthesis mixtures, in different dilutions, and identified a specific mixture and concentration that maximize yield and reduce PCR contaminants (major problem in the original MARS-seq and all IVT-based protocols).
7. Optimization of barcoded ligation adaptor: we modified the barcoded ligation adaptor sequence (barcode of 4 bp, 5 Ns) to improve barcoding of amplification batches and reduce the noise level between them.

Optimization of the conditions indicated above resulted in a sixfold reduction in the cost of library production (from $0.65 to $0.10 per cell) and reduced the background level (from 10–15 % to 2 %) as evaluated by the number of molecules in the wells containing cells as compared with our negative-control empty wells.

Analytical and QC improvements

1. QC scheme to monitor MARS-seq2.0 performance during library preparation: we generated a QC pipeline based on quantitative real-time polymerase chain reaction (qPCR) of housekeeping genes in order to evaluate the complexity of the libraries before sequencing.
2. Generation of a complete analytical pipeline: sequenced reads are de-multiplexed and the molecularsignal is separated from the background noise.
3. QC measurements following sequencing: the analytical pipeline automatically generates a set of > 20 QC measurements (e.g., number of reads, percentage mapping to exons and number of UMIs) that can assist the users in evaluating and optimizing many important features in their single-cell data.

MARS-seq2.0 evaluation

To evaluate the efficiency of the method and data quality, we applied MARS-seq2.0 to 2,256 mouse ES cells and mouse embryonic fibroblasts (MEFs). We detected 11.3 million mRNA molecules (medians of ~6,000 and ~4,000 molecules per ES and MEF cell, respectively, Supplementary Fig. 3a) while sequencing 80,000–100,000 reads per cell (Supplementary Fig. 3b). As expected, the number of detected transcripts per cell is correlated with cell size (Supplementary Fig. 3c). Because MARS-seq2.0 is designed to allow deep population sampling and characterization of known and novel subpopulations, we quantify our detection rate by estimating the distribution of mean gene expression on independent pools of 100 single cells, revealing a dynamic range spreading over three orders of magnitude (Supplementary Fig. 3d). By comparing a pool of 1,128 ES cells with a pool of 752 MEF cells, we detected 2,350 genes with a significant (false-discovery rate (FDR) < 10−5) fourfold difference in expression (Supplementary Fig. 3e). In many cases, these expression differences are also pronounced at the single-cell level (Supplementary Fig. 3f–i).

Advances in microfluidics technologies allow processing of several thousands of cells in parallel.
However, current microfluidic approaches display a relatively high rate (1–15 %) of sequencing of two cells (or more) in a single droplet (doublets)[30,31,52,53], which may complicate downstream analysis.
To evaluate the doublet rate in MARS-seq2.0, we mixed human and mouse ES cells and analyzed the percentage of doublets. Our analysis demonstrates that MARS-seq2.0 has a negligible amount of doublet cells (< 0.2 %; 2 out of 1,041 cells) and provides high confidence in cell identity (Supplementary Fig. 4).

The computational framework

To implement an effective error-correction algorithm and to ensure the quality of sequenced libraries, we designed a computational framework that closely follows the different experimental steps in MARS-seq2.0. Our framework models RNA-seq data generation as a sequential process of multiple samplings and amplifications steps (see Supplementary Methods, Supplementary Fig. 5), in which single mRNA molecules in a single cell are distributed and sampled in different pools. The computational pipeline quantifies the number of tagged mRNA molecules per gene that went through this complex sampling process, while eliminating several types of experimental artifacts. In addition, it generates a graphical and user-friendly library diagnostics report for each amplification batch, highlighting important technical parameters that can be variable between batches (see Supplementary Figs. 6–8 for representative diagnostic reports). The report provides a detailed analysis of the reads' fraction spent on sequencing of primer and other sequences used during library preparation, which can highlight biases such as those caused by an excess of primers. Our pipeline automatically filters sequencing and other polymerization errors that occur in earlier experimental steps (i.e., IVT) and generate spurious UMIs or alter cellular barcodes.

Different sources of barcode contamination may substantially bias biological interpretation of the data by mixing genes between different single-cell subpopulations in the same amplification batch.

We therefore developed an experimental design that allows us to routinely estimate the total cell-to-cell contamination levels by measuring the number of molecules associated with four negative-control wells (empty wells that do not contain cells) in each amplification batch. To visually relate the number of molecules per cell (Supplementary Fig. 3a) to the original single-cell position on the 384-well plate matrix, a plate-view heat-map is generated that shows how the number of recovered genes and spike-in molecules are distributed across the plate position. This can highlight potential sorting, robotic or other problems. Although technical QC is critical, as demonstrated above, we stress that the ultimate way to evaluate the effectiveness of a specific scRNA-seq method is to examine the novelty and richness of the biological insights emerging from the data and that these may depend on multiple factors, including the quality of biological materials.

Strengths and limitations of the protocol

The key strengths of MARS-seq2.0 are as follows:

1. The combination of indexed FACS sorting and scRNA-seq enables refinement and enrichment of the cells of interest, which is especially critical for analysis of rare subpopulations and processing of rare cells in human clinical samples.
2. The improvements made in MARS-seq2.0 provide a cost-effective library preparation method that allows analysis of any number of single cells without compromising the cost per cell value. This is in contrast to other high-throughput methods, such as droplets[30,31], that become cost effective only when processing a very large number of cells at once.
3. The detailed experimental and analytical QC pipelines allow the user to carefully monitor the quality of the libraries and data and to troubleshoot specific problems in the scRNA-seq process.
4. The amount of cell doublets in the MARS-seq2.0 data is extremely low and allows for identification of rare cell types and subpopulations.
5. MARS-seq2.0 provides strand-specific information about the transcripts.

However, the following limitations remain:

1. MARS-seq2.0 is a 3'-based scRNA-seq method and will yield information only about the 3' end of the transcript. Therefore, it is not suitable for identifying alternative splicing isoforms or specific sequences at the 5' end of the gene

(for these purposes, single-cell full-length RNA-sequencing methodologies, such as SMART-seq[24], should be used).
2. The method is selective for polyadenylated RNA.
3. The method requires access to FACS facilities, as well as liquid-handling robotics.

Future applications

The reduction in cost per sample in MARS-seq2.0, and especially the ability to refine the population of interest by enriching or depleting cell populations in a data-driven manner, allows the researcher to focus on specific populations, including very rare cell types or states in animal models and clinical samples[47,49].

Furthermore, techniques for simultaneous acquisition of RNA and other molecular signatures of single cells, for example, spatial location[36,51], genetics[54], lineage[55] and signaling[37], are critical for deep molecular understanding of physiological processes and diseases. MARS-seq2.0 presents a flexible platform for combination of unbiased transcriptional mapping with a large number of fluorescent markers that enables collection of multiple information tiers on the same single cell, such as time, spatial location, signaling, genetics, epigenetics and many more[56].

Guidelines

Experimental design

Tissue dissociation into single cells

The input required for MARS-seq2.0 is a single-cell suspension. Each tissue or cell line should be optimized and calibrated to determine the appropriate method for dissociation into single cells. In general, we recommend minimizing the time and lowering the reaction temperature of any enzymatic digestion, without compromising the type of cells extracted. Failure to do so can result in expression of stress-related genes in the data as a result of the dissociation protocol[57].

FACS

In MARS-seq2.0, single cells are introduced into a 384-well plate via FACS sorting. Gating of the appropriate cell population in the FACS is critical for analysis of high-quality cells. It is important to exclude dead cells, debris and doublets (see detailed recommendations in Supplementary Manual 1).

QC of libraries before sequencing

At several key steps throughout the MARS-seq2.0 library preparation protocol (after pooling, Step 33; after linear amplification, Step 53, and after the final library has been generated, Step 56), QC is performed by qPCR of a housekeeping gene. Primers for the 3' end (no more than 200 bp upstream of the polyA site) of a housekeeping gene should be designed and calibrated according to standard qPCR primer requirements. We have provided primer sequences that we used for mouse (ActB) and human (GAPDH) genes. In the case that the experiment is performed on a species for which it is difficult to design primers, it is possible to proceed with library preparation without the qPCR QCs and evaluate library quality on the basis of only DNA quantity (using a Qubit assay, Step 57) and quality (using a TapeStation assay, Step 58) results.

Combination of scRNA-seq with other platforms

Following RNA amplification by IVT, it is possible to continue with half of the aRNA for scRNA-seq preparation by MARS-seq2.0 and to use the other half for specific analysis of targeted sequences, such as B-cell receptor and T-cell receptor sequencing, at the single-cell level[47,49], or for examining guide RNA expression by unique guide index (UGI-seq), as we performed in CRISP-seq[38].

Sequencing depth

All single-cell protocols involve considerable amplification, which introduces non-uniform representation of distinct molecules (identified as UMIs) into the read space[25,58,59]. Because MARS-seq2.0 also allows massive multiplexing of many thousands of cells, the number of reads per cell (or cells in one sequencing flow cell) should be designed to balance depth and cost effectiveness. We sought to test how the number of recovered molecules is dependent on sequencing depth by downsampling our sequenced reads in a logarithmic manner, randomly selecting, for example, 50 %, 25 % or 12.5 % of the total read pool and processing each of these downsampled sequenced libraries. Interestingly, the number of recovered molecules per cell rapidly converges at 8,250 molecules per cell at 200,000 reads/ cell, indicating that further sequencing in this case will provide little new information and will probably have an unimportant benefit on the analysis (Supplementary Fig. 9a). Repeating the saturation analysis for groups of single cells with different ranges of total molecule count (i.e., smaller and larger cells; Supplementary Fig. 9b) provides MARS-seq2.0 users with a tool to determine

sequencing depth, given the distribution of cell sizes in specific studies. For example, small cells (such as those in some hematopoietic niches) with potentially only 1,000 RNA molecules per cell can be sequenced at a low depth of ~ 10,000 reads per cell, whereas larger cells (e.g., tissue culture, ES cells) are saturated only when sequencing ≥ 200,000 reads. In experiments in which the distribution of cell types or their mRNA content is unknown, an efficient sequencing approach could be to sequence a single sample to saturation and determine the optimal sequencing depth for the following samples according to the empirical saturation curve.

Troubleshooting

Troubleshooting advice can be found in Table 1.

	Step	Problem	Possible reason	Solution
	6	Low cell viability	Inadequate single-cell preparation / sensitive cells	Optimize tissue dissociation
	21	No signal in qPCR in test sorting efficiency	Inadequate sorting conditions	Carefully follow the 'FACS calibration for 384-well plate protocol' (Supplementary Manual 1)
	40	No signal or low signal (high cycle threshold (CT) >30) in QC1 (this	Low cell viability	Use DAPI to assess cell viability in the FACS

		is also dependent on cell type)		
			Inadequate sorting conditions	Carefully follow the 'FACS calibration for 384-well plate protocol'
			Unsuccessful sorting session	Check the test sort from the same session
			A mistake in the RT process	Check protocol and reagents used
			Use of uncalibrated qPCR primers, or primers of poorly expressed genes	Check qPCR primer design. If necessary, order new primers and repeat QC1
	67	No enrichment between QC1 and QC2	Material was lost in clean ups; a mistake was made in the molec	Repeat the experiment

			ular reactions	
			Use of a qPCR primer that is not at the 3′ end of the transcript	Check qPCR primer design. If necessary, order new primers and reevaluate QC2
	74	Low concentration of final library (<1 ng/µl)	Not enough amplification in the final enrichment	Repeat final enrichment (Steps 69/70) with more PCR cycles
	75	Additional small fragments in library profile	Insufficient AMPure XP clean up in the final step (Step 72)	Repeat clean up of final library (Step 72)
		Large fragments in library profile, usually together with a very high Qubit value	Too much amplification in the final enrichment	Repeat final enrichment (Steps 69/70) with fewer PCR cycles

Table 1: Troubleshooting table

Timing

- For Steps 1–41, the preparation time depends on the number of 384-well plates to be processed together. The following is an example of an estimated time course for single-cell sorting of 24 plates and preparing libraries from eight of them at once (keeping the rest at −20 °C), assuming there is no limitation on the number of PCR cyclers available for use.
- From Step 42 onward, the number of plates is no longer relevant for the timing, and as many plates as desired can be processed in parallel.
- Steps 1–4, cell capture plate preparation: ~50 min for 24 plates
- Steps 5–12, single-cell sorting: variable, ~3 h for 24 plates
- Steps 13–21, testing single-cell sorting efficiency: ~2 h (testing of 1 plate)
- Steps 22–30, barcoding and RT reaction: ~3 h for 8 plates (robotic addition of RT is done separately for each plate; thermal cycler incubation can be done in parallel, in several thermal cyclers, to save time)
- Steps 31–35, primer removal by exonuclease I: ~3 h for 8 plates (robotic addition of exonuclease I is done separately for each plate; thermal cycler incubation can be done in parallel, in several thermal cyclers, to save time)
- Steps 36–41, pooling and sample cleanup: ~3 h for 8 plates (robotic pooling is done separately for each plate; sample cleanup can be done in parallel to save time)
- Steps 42–50, second-strand-synthesis and IVT amplification: ~2.5 h + overnight incubation
- Steps 51–55, RNA fragmentation: ~30 min
- Steps 56–61, ligation of plate barcode: ~2.5 h
- Steps 62–68, second reverse transcription: ~1.5 h
- Steps 69–72, enrichment of final library: ~1 h
- Steps 73–75, QC assessment before sequencing: ~1 h
- Steps 76–80, sequencing: ~12 h
- Steps 81–86, running the analytical pipeline: ~1 d

Anticipated results

Step 21
A signal in the qPCR in the single cell test sort indicates that the sorting session was successful. The cycle threshold (CT) value obtained depends on the cell type used and should be monitored for consistency between experiments. For the ES and MEF cells sorted in this protocol, a CT of ~32–33 was expected. Only plates that were sorted successfully should be used for further processing of MARS-seq2.0 libraries.

Step 40
QC1 provides a good estimation of the amount of starting material and an indication of a successful cell sorting and high-quality cells. Cells with higher RNA content (e.g., ES cells) are expected to yield a lower CT value than cells with a lower RNA content (e.g., dendritic cells). Normally, QC1 values for mouse cells (ActB) range between 24 and 29. QC1 values for human cells (GAPDH) range between 27 and 32). In this protocol, the QC1 value for mouse ES cells was ~26 and that for MEFs was ~24.

Step 67
QC2 provides information about the amount of material after the linear amplification. The difference between QC2 and QC1 values reflects the enrichment obtained by the IVT. Enrichment of ~5 cycles is expected in a successful library and it can range from 3 to 8 cycles. In this protocol, the QC2 value for mouse ES cells was ~19 and that for MEFs was ~17.

Step 72
QC3 provides an indication of the amount of successfully amplified library within the final product. It is expected to yield ~9 cycles difference between QC2 and QC3 if 15 cycles were used for the final PCR enrichment at Step 55. QC3 should also be inversely proportional to the Qubit values, i.e., a sample with a higher Qubit value should have a lower CT as compared with other samples. In this protocol, we used 14 or 12 cycles for mouse ES and MEF cells, respectively, to yield a QC3 of ~11 for each.

Step 74
The amount of library (as measured by Qubit assay in ng/µl) indicates a successful library amplification process and is also dependent on the number of PCR cycles that were used for enrichment. A typical library should yield between 1 and 10 ng/µl dsDNA. Higher Qubit values are also acceptable, providing that the TapeStation profile is good (see discussion of Step 58 below). In this protocol, we used 14 or 12 cycles for mouse ES cells and MEF cells, respectively, to yield a library concentration of ~9 ng/µl.

Step 75
A good library profile is critical to evaluating the quality of the library and to determining its average peak. A typical library peak is between 300 and 500 bp with no residual primer-dimers or adaptors that appear as short fragments at ~100 bp. See an example of a good-quality library profile in Supplementary Fig. 10.

Reporting Summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The sequenced data analyzed in this study have been deposited in GEO: GSE123392. Published data analyzed in this paper are available at https://doi.org/10.1016/j.cell.2015.11.013. Access to all published sequenced data generated using the method described here can be found within the relevant publications.

Code availability
Users can access the code freely on our website: http://compgenomics.weizmann.ac.il/tanay/?page_ id=672.

References

1. Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods 6, 377–382 (2009).
2. Tang, F. et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat. Protoc. 5, 516–535 (2010).
3. Shalek, A. K. et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498, 236–240 (2013).

4. Moignard, V. et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat. Cell Biol. 15, 363–372 (2013).
5. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343, 193–196 (2014).
6. Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 509, 371–375 (2014).
7. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344, 1396–1401 (2014).
8. Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNAseq. Science 347, 1138–1142 (2015).
9. Paul, F. et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell 163, 1663–1677 (2015).
10. Jaitin, D. A. et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343, 776–779 (2014).
11. Pollen, A. A. et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat. Biotechnol. 32, 1053–1058 (2014).
12. Grun, D. et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525, 251–255 (2015).
13. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell 167, 285 (2016).
14. Grun, D. & van Oudenaarden, A. Design and analysis of single-cell sequencing experiments. Cell 163, 799–810 (2015).
15. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. 16, 133–145 (2015).
16. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize wholeorganism science. Nat. Rev. Genet. 14, 618–630 (2013).
17. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. Mol. Cell 58, 610–620 (2015).
18. Islam, S. et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 21, 1160–1167 (2011).
19. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2, 666–673 (2012).
20. Islam, S. et al. Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. Nat. Protoc. 7, 813–828 (2012).
21. Ramskold, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. 30, 777–782 (2012).
22. Sasagawa, Y. et al. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol. 14, R31 (2013).
23. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods 10, 1096–1098 (2013).
24. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181 (2014).
25. Islam, S. et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods 11, 163–166 (2014).
26. Wu, A. R. et al. Quantitative assessment of single-cell RNA-sequencing methods. Nat. Methods 11, 41–46 (2014).
27. Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell 162, 184–197 (2015).

28. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 14, 955–958 (2017).
29. Gierahn, T. M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods 14, 395–398 (2017).
30. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202–1214 (2015).
31. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187–1201 (2015).
32. Regev, A. et al. The Human Cell Atlas. Elife 6, e27041 (2017).
33. Trapnell, C. Defining cell types and states with single-cell genomics. Genome Res. 25, 1491–1498 (2015).
34. Ballas, S. K. Erythrocyte concentration and volume are inversely related. Clin. Chim. Acta 164, 243–244 (1987).
35. Giladi, A. et al. Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. Nat. Cell Biol. 20, 836–846 (2018).
36. Medaglia, C. et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. Science 358, 1622–1626 (2017).
37. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods 14, 865–868 (2017).
38. Jaitin, D. A. et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. Cell 167, 1883–1896.e15 (2016).
39. Paul, F. & Amit, I. Plasticity in the transcriptional and epigenetic circuits regulating dendritic cell lineage specification and function. Curr. Opin. Immunol. 30, 1–8 (2014).
40. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's Disease. Cell 169, 1276–1290.e17 (2017).
41. Bornstein, C. et al. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. Nature 559, 622–626 (2018).
42. Cohen, M. et al. Lung single-cell signaling interaction map reveals basophil role in macrophage imprinting. Cell 175, 1031–1044.e18 (2018).
43. Matcovitch-Natan, O. et al. Microglia development follows a stepwise program to regulate brain homeostasis. Science 353, aad8670 (2016).
44. Steuerman, Y. et al. Dissection of influenza infection in vivo by single-cell RNA sequencing. Cell Syst. 6, 679–691.e4 (2018).
45. Sebe-Pedros, A. et al. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. Cell 173, 1520–1534.e20 (2018).
46. Sebe-Pedros, A. et al. Early metazoan cell type diversity and the evolution of multicellular gene regulation. Nat. Ecol. Evol. 2, 1176–1188 (2018).
47. Ledergor, G. et al. Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. Nat. Med. 24, 1867–1876 (2018).
48. Lavin, Y. et al. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. Cell 169, 750–765.e17 (2017).
49. Li, H. et al. Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. Cell 176, 775–789.e18 (2019).
50. Halpern, K. B. et al. Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. Nat. Biotechnol. 36, 962–970 (2018).

51. Halpern, K. B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature 542, 352–356 (2017).
52. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. 36, 89–94 (2018).
53. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049 (2017).
54. Han, K. Y. et al. SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells. Genome Res. 28, 75–87 (2018).
55. Biase, F. H. et al. Rainbow-Seq: combining cell lineage tracing with single-cell RNA sequencing in preimplantation embryos. iScience 7, 16–29 (2018).
56. Giladi, A. & Amit, I. Single-cell genomics: a stepping stone for future immunology discoveries. Cell 172, 14–21 (2018).
57. Wu, Y. E., Pan, L., Zuo, Y., Li, X. & Hong, W. Detecting activated cell populations using single-cell RNA-seq. Neuron 96, 313–329.e6 (2017).
58. Grun, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. Nat. Methods 11, 637–640 (2014).
59. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat. Methods 9, 72–74 (2011).
60. Rais, Y. et al. Deterministic direct reprogramming of somatic cells to pluripotency. Nature 502, 65–70 (2013).

Acknowledgements

We thank the J. H. Hanna lab for the donation of v6.5 mouse embryonic stem (ES) cells, DR4 mouse embryonic fibroblast (MEF) cells and WIBR3 human ES cells. We thank G. Brodsky for help with the artwork. This work was funded by the Chan Zuckerberg Initiative (CZI) (A.T. and I.A.), European Research Council consolidator grants (A.T. and I.A.), ERC-COG (724824-scAssembly; A.T.), ERC-COG (724471-HemTree2.0; I.A.) and the I-Core program (A.T. and I.A.).

Author contributions

H.K.-S. developed the experimental protocols; designed, performed, and analyzed experiments; and wrote the manuscript. E.K. designed experiments, developed the computational methods, and performed bioinformatics analysis. D.A.J. developed experimental protocols. E. D. performed bioinformatic analysis. F.P. contributed to the experimental design. A.T. directed the project and developed computational methods. I.A. directed the project and designed the experiments.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-019-0164-4.

Materials

Biological materials

A fresh single-cell suspension.
We received v6.5 mouse ES cells, DR4 MEF cells and WIBR3 human ES cells from the J. H. Hanna lab60 (https://hpscreg.eu/cell-line/WIBRe001-A, Weizmann IRB-ESCRO approval no. 40-1 Hanna) and grew them in DMEM with 15 % (vol/vol) FBS. In this study, ~10 × 10^6 cells were trypsinized to form a single-cell suspension and were filtered with a 40-µm strainer. For the human–mouse doublet analysis experiment, human ES cells (WIBR3) and mouse ES cells (v6.5) were dissociated similarly, counted and mixed in 1:1 ratio. Single common myeloid progenitor cells were derived from the bone marrow of C57BL/6 mice (Envigo, cat. no. C57BL/6JOlaHsd) as previously described[9]. The single-cell suspension should be **stored on ice (4 °C)** until sorting.

C57BL/6 inbred mice **Envigo Catalog #**C57BL/6JOlaHsd

Reagents

- EB8 (10 mM Trizma hydrochloride solution (dilution of 1.0 M, pH 8, (Sigma-Aldrich, cat. no. 648314) in nuclease-free water (Sigma-Aldrich, cat. no. W4502))). Store at room temperature (20–25 °C) for up to 6 months.

Tris Buffer, 1.0 M, pH 8.0 **Millipore Sigma Catalog #**648314

Molecular grade H20 **Sigma Aldrich Catalog #**W4502

- EB7.5 (10 mM Trizma hydrochloride solution (dilution of 1.0 M, pH 7.5, (Calbiochem, cat. no. 9285) in nuclease-free water)). Store at room temperature for up to 6 months.

OmniPur® TRIS Solution, 1.0M pH 7.5 **Millipore Sigma Catalog #**9285-OP

- Triton X-100 (Sigma-Aldrich, cat. no. T8787). Store at room temperature per manufacturer's instructions.

Triton X-100 **Sigma Aldrich Catalog #**T8787

Safety information

CAUTION Triton X-100 is harmful if swallowed. It can cause skin corrosion and serious eye damage. It is hazardous to the aquatic environment. Handle it using appropriate safety equipment.

- RNasin Plus ribonuclease inhibitor (Promega, cat. no. N2111)

RNasin(R) RNase Inhibitor, 2,500u **Promega Catalog #**N2111

- Superscript III reverse transcriptase (Thermo Fisher Scientific, cat. no. 18080093), supplied with superscript buffer (5×) and DTT (100 mM)

SuperScript™ III Reverse Transcriptase **Thermo Fisher Scientific Catalog #**18080085

- Fast SYBR Green master mix (Thermo Fisher Scientific, cat. no. 4385614)

Fast SYBR™ Green Master Mix **Thermo Fisher Scientific Catalog #**4385614

Note

CRITICAL Store the stock at −20 °C. Aliquots of 1 ml can be stored at 4 °C for up to 1 year.

- Exonuclease I (NEB, cat. no. M0293S), supplied with exonuclease I buffer (10×)

Exonuclease I (E.coli) - 3,000 units **New England Biolabs Catalog #**M0293S

- AffinityScript reverse transcriptase (Agilent, cat. no. 600109), supplied with AffinityScript buffer (10×) and DTT (100 mM)

AffinityScript Multiple Temperature Reverse Transcriptase **Agilent Technologies Catalog #**600109

- dNTP mix (Larova, cat. no. DMIX25)

dNTP Mix 25 **LAROVA GmbH Catalog #**DMIX25

Note

CRITICAL Make aliquots of the dNTP mix and store them at −20 °C for up to 1 year. After aliquot preparation, avoid more than one freeze–thaw cycle.

- External RNA Controls Consortium (ERCC) RNA spike-in mix (Thermo Fisher Scientific, cat. no. 4456740)

ERCC RNA Spike-In Mix **Thermo Fisher Catalog #**4456740

Note

CRITICAL Store at −80 °C.
Prepare 1-µl aliquots of a 1:10 dilution and store at −80 °C for up to 1 year.
Prepare1-µl aliquots of a 1:100 dilution in nuclease-free water and store at −80 °C for up to 1 year.

- Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63881)

Agencourt AMPure XP SPRI beads **Beckman Coulter Catalog #**A63881

Note

CRITICAL Store at 4 °C per manufacturer's instructions. Bring AMPure XP beads to room temperature before use.

- Absolute ethanol (Sigma, CAS number 64-17-5)

Note

CAUTION Ethanol is highly flammable. Handle it using appropriate precautions.

- NEBNext mRNA Second Strand Synthesis Module (NEB, cat. no. E6111), supplied with secondstrand-synthesis buffer (10×)

NEBNext® Ultra II Non-Directional RNA Second Strand Synthesis Module **New England Biolabs Catalog #**E6111

Note

CRITICAL Make aliquots of the second-strand-synthesis buffer and store them at −20 °C for up to 1 year.

- HiScribe T7 High Yield RNA Synthesis Kit (NEB, cat. no. E2040), supplied with T7 RNA polymerase buffer (10×) and ribonucleotide triphosphates (rNTPs): A, C, U, G, 100 mM each

HiScribe T7 High Yield RNA Synthesis Kit - 50 rxns **New England Biolabs Catalog #**E2040S

Note

CRITICAL Mix the four rNTPs supplied in equal ratios, divide into aliquots and store at −20 °C for up to 1 year.

- Turbo DNase (Ambion; cat. no. AB-AM2238), supplied with Turbo DNase buffer (10×)

TURBO™ DNase (2 U/µL) **Thermo Fisher Scientific Catalog #**AM2238

- RNA fragmentation reagent (10×; Ambion, cat. no. AB-AM8740), supplied with stop solution

RNA Fragmentation Reagents **Thermo Fisher Catalog #**AM8740

- T4 RNA ligase 1 (NEB, cat. no. M0204S), supplied with T4 ligation buffer (10×) and PEG (50% (wt/vol))

T4 RNA Ligase 1 (ssRNA Ligase) - 1,000 units **New England Biolabs Catalog #**M0204S

- Dimethyl sulfoxide (DMSO; Sigma-Aldrich, cat. no. D8418)

DMSO (dimethyl sulfoxide) **Sigma Aldrich Catalog #**D8418

- Adenosine triphosphate (ATP) solution (100 mM; Thermo Fisher Scientific, cat. no. R0441)

ATP Solution (100 mM) **Thermo Fisher Catalog #**R0441

Note

CRITICAL Prepare aliquots of small volume (5 µl) and store at −80 °C for up to 1 year. Use ATP aliquots only once (discard any remaining solution).

- KAPA HiFi HotStart ReadyMix (KAPA Biosystems, cat. no. KK2601)

HotStart ReadyMix (KAPA HiFi PCR kit) **Kapa Biosystems Catalog #**KK2601

- Qubit dsDNA HS Assay Kit (Invitrogen, cat. no. Q32854)

Qubit dsDNA HS Assay kit **Thermo Fisher Scientific Catalog #**Q32854

- TapeStation High Sensitivity D1000 Reagents (Agilent, part no. 5067-5585)

High Sensitivity D1000 Reagents **Agilent Technologies Catalog #**5067-5585

Equipment

- FACS sorter with plate-sorting capability. (BD Biosciences, FACSAria SORP model).
- Automated liquid-handling platform (Agilent, Bravo model with 96ST head). The MARSseq2.0 scripts for the Bravo instrument are available as Supplementary Software. The four scripts are named ‘Cell capture plate preparation’, ‘RT mix addition’, ‘Exonuclease I addition’ and ‘Pooling’. They are loaded into the Bravo robot as specified at relevant steps in the Procedure
- CPAC heating/cooling unit for a 96-well plate (INHECO), integrated in the Bravo
- CPAC heating/cooling unit for a 384-well plate (INHECO), integrated in the Bravo

Note

CRITICAL It is possible to use other liquid-handling platforms that are calibrated for performing the same protocols, for example, the Nanodrop II or the Nanodrop Express (BioNex).

- Filtered robotic pipette tips (Axygen, cat. no. VTF-384-50UL-R-S)

Equipment

Axygen® Automation Tips for Agilent®	NAME
Automation Pipet Tips	TYPE
Axygen® Scientific	BRAND
VTF-384-50UL-R-S	SKU
https://ecatalog.corning.com/life-sciences/b2c/US/en/Genomics-&-Molecular-Biology/Automation-Consumables/Automation-Pipet-Tips/Axygen%C2%AE-Automation-Tips-for-Agilent%C2%AE-Velocity11-%7C-VPrep%C2%AE-and-Bravo/p/VTF-384-50UL-R-S	LINK

- Thermal cycler for a 384-well plate (Eppendorf, Mastercycler pro 384)
- Thermal cycler for a 96-well plate (Eppendorf, Mastercycler Nexus eco)
- 1.5-ml LoBind tubes (Eppendorf, cat. no. 022431021)
- Microplate Centrifuge (500g; QSR Technologies)
- Benchtop mini-centrifuge (2000g; Genereach, GE Cubee)

Note

CRITICAL This centrifuge is small and can be placed next to the FACS machine to ensure immediate spindown of the cell capture plate after the single-cell sorting.

- 384-well plates (Eppendorf, cat. no. 951020539)
- 96-well plates (Eppendorf, cat. no. 951020401)

Equipment

Eppendorf twin.tec® PCR Plate 96	NAME
Plate	TYPE
Eppendorf	BRAND
951020401	SKU
https://online-shop.eppendorf.us/US-en/Laboratory-Consumables-44512/Plates-44516/Eppendorf-twin.tec-PCR-Plates-PF-8180.html	LINK

- Aluminum sealers (Sigma-Aldrich, cat. no. A2350-100EA)

Equipment

AlumaSeal® II film	NAME
aluminum foil film	TYPE
AlumaSeal	BRAND
A2350-100EA	SKU
https://www.sigmaaldrich.com/catalog/product/sigma/a2350?lang=en&region=US&cm_sp=Insite-_-prodRecCold_xviews-_-prodRecCold10-1	LINK

- Qubit fluorometer (Invitrogen, cat. no. Q32857)
- TapeStation 2200 (Agilent, cat. no. G2965A)
- A compatible Illumina DNA sequencing instrument: NextSeq 500 sequencing system with a High Output v2 kit (75 cycles; Illumina, cat. no. FC-404-2005).
- Computer running a Linux system (≥8 GB RAM; CentOS-7: https://www.centos.org/download/)
- TapeStation High Sensitivity D1000 ScreenTape (Agilent, part no. 5067-5584)

Equipment	
High Sensitivity D1000 ScreenTape	NAME
High Sensitivity DNA Assay	TYPE
Agilent	BRAND
5067-5584	SKU
https://www.agilent.com/store/en_US/Prod-5067-5584/5067-5584	LINK

Software

- MARS-seq2.0 computational pipeline: available at http://compgenomics.weizmann.ac.il/tanay/?page_id=672 and as Supplementary Manual 2

Data

- An example dataset can be found in the Gene Expression Omnibus (GEO): GSE123392, AB345, showing data obtained from mouse ES and MEF cells

Reagent setup

RT primer (RT1)
The RT primer includes a polyT tail, a 7-bp well barcode (X), an 8-bp UMI (NNNNNNNN; for controlling PCR bias), an Illumina rd2 sequence and a T7 promoter: 5′-CGATTGAGGCCGGTAA TACGACTCACTATAGGGGCGACGTGTGCTCTTCCGATCT XXXXXXXNNNNNNNNTTTTTTTTTTTTTTTTTTTTN-3′. Barcodes were designed to have an edit distance of at least two from any other barcode and almost constant GC content (three or four Gs or Cs). The full primer list is available at http://compgenomics.weizmann.ac.il/tanay/?page_id=672. See also Supplementary Table 1. RT1 primers are shipped dry and dissolved in nuclease-free water to a stock concentration of 25 µM. Upon suspension of the RT primer plate, prepare aliquots of further dilutions: 1 µM, 50 nM and 12.5 nM (working concentration). All RT primer plates are kept at −20 °C and are stable for several years. We did not observe any decrease in yield due to freeze–thaw cycles.

Note

CRITICAL Always spin down (500g, room temperature, 30 s) the RT1 primer plate before opening to avoid contamination of RT primers between wells.

Mouse QC primers
ActB forward: 5′-GGAGGGGGTTGAGGT GTT-3′; reverse: 5′-TGTGCACTTTTATTG GTCTCAAG-3′. Primers are shipped dry and dissolved in nuclease-free water to a stock concentration of 100 µM. Prepare working aliquots of 12.5 µM. The primers can be stored at −20 °C for several years.

Human QC primers
GAPDH forward: 5′-TCCCCCACCACACTGAATCT-3′; reverse: 5′-ACAAGGTGCGGCTCCCTA-3′. Primers are shipped dry and dissolved in nuclease-free water to a stock concentration of 100 µM. Prepeare working aliquots of 12.5 µM. The primers can be stored at −20 °C for several years.

Barcoded ligation adaptor (plate adaptor)
The ligation adaptor contains a pool barcode of 4 bp (X), 5 Ns and a partial Illumina Read1 sequencing adaptor: 5′ 5Phos/XXXXNNNNNAGATCGGAAGAGCGTCGTGTAG/3SpC3 3′ (5Phos, phosphorylation; 3SpC3, C3 spacer). The pool barcode enables pooling of several plates in the Illumina sequencing. The full list of ligation adaptor sequences is available at http://compgenomics. weizmann.ac.il/tanay/?page_id=672. See also Supplementary Table 2.

Note

CRITICAL This oligo has a phosphate on the 5′ end (5′Phos) in order to ligate this end to the RNA, and has a blocker at the 3′ end (3SpC3) to prevent ligation of adaptors to each other. Ligation adaptors are shipped dry and suspended in nuclease-free water to a stock concentration of 100 µM. Prepare working aliquots of 50 µl of 100 µM ligation adaptors in 8-well strips for ease of use. Store the ligation adaptors at −20 °C for up to 1 year.

Second RT primer (RT2)
The primer is complementary to the ligation adaptor: 5′-CTACACGACGCTCTTCCGATCT-3′. Store the primer at −20 °C.

Enrichment primers
The forward primer contains the Illumina P5-Read1 sequences: 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′; the reverse primer contains the P7-Read2 sequences: 5′-CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′. Primers are shipped dry and dissolved in nuclease-free water to a stock concentration of 100 µM. Prepeare working aliquots of enrichment primer mix (forward and reverse together) at 25 µM. The primers can be stored at −20 °C for several years.

Safety warnings

Please see SDS (Safety Data Sheet) for hazards and safety warnings.

Before start

- **CRITICAL** The MARS-seq2.0 protocol described here is designed for use with 384 different RT1 primers, with each primer being uniquely present in the 384-well plate. If a user is interested in reducing costs, it is possible to use only 192 different RT1 primers, with each primer being represented twice in the 384-well plate. In the latter case, following RT reaction and barcoding, single cells are pooled from each 384-well plate into two amplification batches for the downstream process. For more information and relevant robotic scripts, contact the authors.
- **CRITICA**L Unless specified otherwise, all plate centrifugation (spin down) should be done at 500 g for 30 s at room temperature.
- **CRITICAL** Unless specified otherwise, all tube centrifugation (spin down) should be done at 2000 g for 5 s at room temperature.
- **CRITICAL** Vortex and spin down all reagents before use.

Making MARS-seq2.0 384-well cell capture plates (Timing ~50 min for 24 plates)

1 Turn on the Bravo liquid-handling platform and wait for initialization to be completed. Load the 'cell capture plate preparation' protocol file for the 96ST head. The appropriate protocol file can be found in the Supplementary Software.

2 Prepare cell capture plate master mixes in four 96-well plates: lysis buffer ([M] 0.1 % volume Triton X-100 in nuclease-free water) supplemented with [M] 0.1 U/µl RNase inhibitor (RNasin Plus) and [M] 1 millimolar (mM) RT1 primer from group 1 (barcodes 1–96), group 2 (barcodes 97–192), group 3 (barcodes 193–288) or group 4 (barcodes 289–384).
Barcode groups are listed in Supplementary Table 1.
To prepare about twenty-four 384-well plates, use a multichannel pipette to mix 57.5 µL of lysis buffer with 5 µL of [M] 12.5 nanomolar (nM) RT1 primer stock per well of a 96-well plate.

3 Place the 96-well cell capture plate master mixes, a box containing filtered tips and the destination 384-well plates in the appropriate locations in the Bravo robot. The cell capture plate preparation script mixes the group 1 master mix plate, aspirates 2 µL from it and dispenses it into destination 384-well plate 1.

Then 2 µL is again aspirated from group 1 master mix plate to be dispensed into the other destination 384-well plates. If more than three cell capture plates are prepared, the robotic script will pause and prompt the user to replace the filled destination plates with new plates. Once group 1 master mix has been added to all desired plates, the tips are replaced and the process is repeated for groups 2, 3 and 4. The entire process takes ~ 00:50:00 for 24 plates.
See Fig. 3a for a schematic of the plates' locations in this protocol.

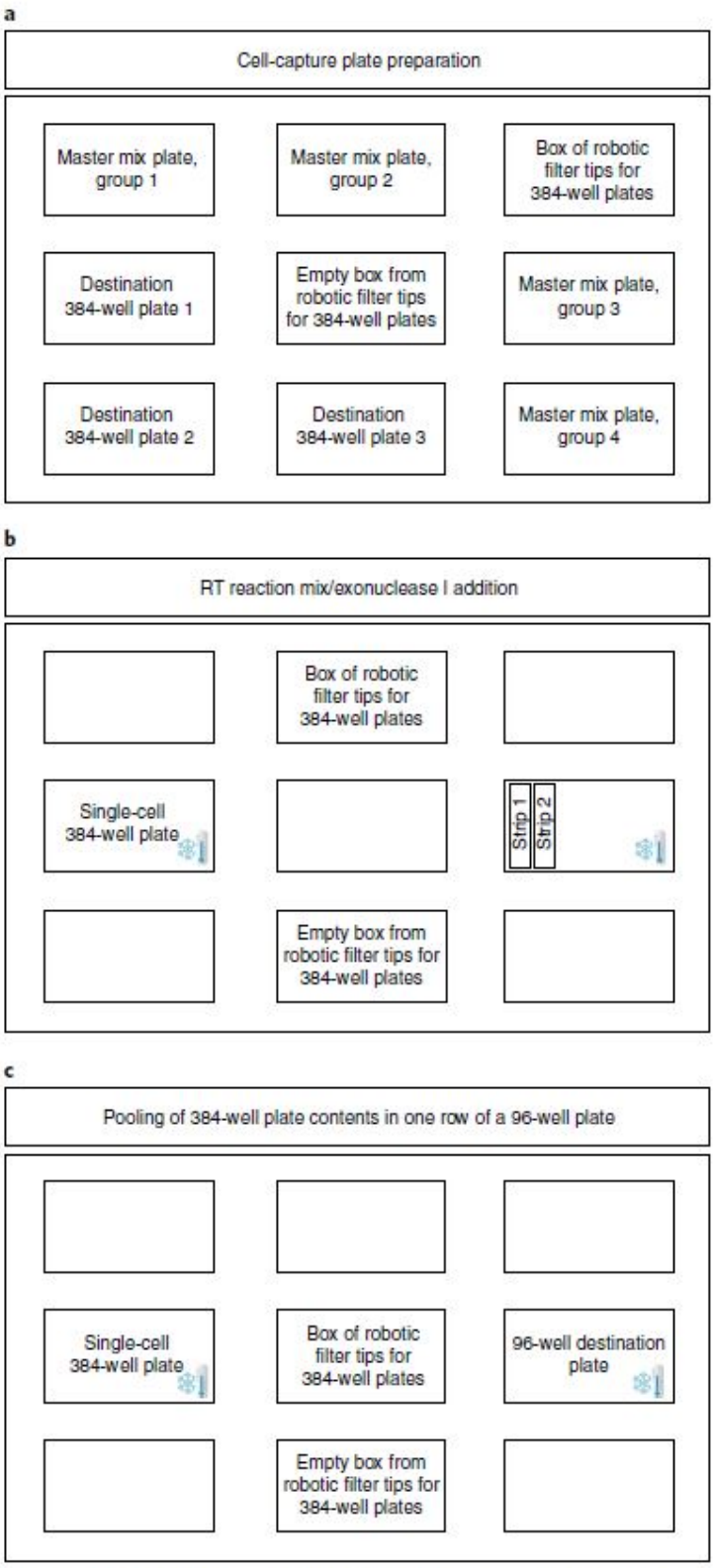

Fig. 3 | Schematic of plate location in the Bravo liquid-handling robot automated protocols. a, Cell capture plate preparation. **b**, RT mix addition/exonuclease I addition. **c**, Pooling of 384-well plate contents in one row of a 96-well plate. Snowflake/thermometer symbols indicate samples placed at 4 °C.

4 Seal the prepared cell capture plates and freeze at -20 °C until use.

Cell capture plates can be kept at -20 °C for 1–2 years.

Sorting of single cells into MARS-seq2.0 capture plates (Timing ~7 min per capture plate, if the cell population sorted is not a limiting factor, i.e., not a rare cell population)

5 Thaw the cell capture plates from Step 4 and spin them down. Keep the plates on ice until sorting.

6 Prepare a single-cell suspension for FACS (see 'Biological materials' section).

Note

We recommend gating on side scatter area (SSC-A) versus forward scatter area (FSC-A) to collect live cells, and then on forward scatter width (FSC-W) versus FSC-A to sort only singlets.

? TROUBLESHOOTING

7 Calibrate the FACS machine for single-cell sorting into plates.

Note

See Supplementary Manual 1 for detailed information on single-cell sorting optimization.

8 Place a cell capture plate in the appropriate location in the FACS instrument and sort the desired population.

Note

We recommend leaving between two and four empty wells in each 384-well plate as a no-cell control during data analysis.

CRITICAL STEP Keep the single-cell suspension on ice until FACS sorting.

CRITICAL STEP The time from single-cell suspension to FACS sorting into the capture plates should be minimal. We recommend not exceeding 4–5 h.

9 Immediately after sorting of each plate, seal the plate with an aluminum sealer, spin it down to ensure cell immersion into the lysis solution, snap-freeze on dry ice and store at -80 °C until further processing.

CRITICAL STEP The sealer used must have the ability to sustain low temperatures; otherwise the seal can break off during freezing at −80 °C.

CRITICAL STEP It is very important to spin down the cell capture plate immediately after singlecell sorting and keep it on dry ice until storage at −80 °C.

10 Sorted cell capture plates can be kept at −80 °C for several years.

Indexed single-cell sorting (Timing ~14 min per capture plate, if cell population sorted is not a limiting factor)

11 Indexed FACS sorting is an optional element of the MARS-seq2.0 pipeline that allows correlating transcription and protein level by recording the level of surface markers for each sorted single cell.

12 If you desire to record fluorescent marker levels for each single cell, activate the FACS Diva 'index sorting' function during single-cell sorting.
See Supplementary Manual 3 for a detailed protocol on how to extract index sorting data.

Testing of single-cell sorting efficiency (Timing ~2 h for 1 or 2 plates)

13 We recommend testing the single-cell sorting efficiency in each new project or as a routine procedure following each sorting session. In Steps 13–19, one or two plates from the sorting session are used for testing of the single-cell sorting efficiency and not for library preparation.

14 Thaw one of the sorted 384-well plates from Step 9 for 00:05:00 at Room temperature and spin it down.

15 Pre-heat a 384-well thermal cycler to 72 °C.

16 Incubate the sorted 384-well plate for 00:03:00 at 72 °C to denature secondary structures and

immediately transfer the plate to an ice block.

17 Prepare a RT mix to add to 2–4 columns per plate according to the table below, and add 2 µL of RT mix per reaction well:

	Component	Amount (µl) per 1 reaction (1×)	Amount (µl) per 2 columns (36×)
	AffinityScript buffer (10×)	0.4	14.4
	DTT 100 mM	0.4	14.4
	dNTPs 100 mM	0.16	5.76
	AffinityScript enzyme	0.2	7.2
	Nuclease-free water	0.84	30.24
	Sorted capture plate (Step 14/15)	2	
	Total reaction volume	4	

18 Seal the plate, spin it down and incubate for 00:02:00 at 42 °C, for 00:50:00 at 45 °C, and for 00:05:00 at 85 °C.

19 After RT, spin down the plate and dilute the cDNA 1:20 as follows:
(i) bring the volume in the columns in which RT was performed to 10 µl by adding 6 µL of EB8 to each well;
(ii) add 1 µL of this diluted cDNA to 19 µL of EB8 in a separate dilution plate.

20 Perform quantitative qPCR for mouse ActB or human GAPDH. Prepare the qPCR master mix according to the table below (for duplicates):

	Component	Amount (µl) per 1 reaction (1×)	Amount (µl) per 2 columns (75×)
	Primer mix (12.5 µM	0.2	15

	each forward and reverse primers; Reagent setup)		
	Fast SYBR Green master mix	5.8	435
	1:20 diluted cDNA	6	

21 Run qPCR for 45 cycles as shown in the table below:

	Cycle number	Denature	Anneal
	1	95 °C, 20 s	
	2–46	98 °C, 3 s	60 °C, 30 s

? TROUBLESHOOTING

Barcoding and RT reaction (Timing ~1.5 h for 1 plate)

22 Turn on the Bravo liquid handler, initialize and load the 'RT mix addition' script.

23 Thaw the 384-well plates (from Step 9) to be processed at Room temperature for 00:05:00 and spin down. Keep on ice until the next step. Pre-heat a 384-well thermal cycler to 95 °C.

24 Prepare an RT reaction mix sufficient for the number of plates to be processed:

	Component	Amount (µl) per 1 well (1x)	Amount (µl) per 1 plate (520x)
	Superscript III buffer (5×)	0.1	52
	DTT 100 mM	0.05	26
	dNTPs 100 mM	0.02	10.4
	ERCC spike-in mix	0.01	5.2

	(1:80,000)		
	RNasin (40 U/µl)	0.0125	6.5
	Superscript III enzyme	0.01	5.2
	Nuclease-free water	0.2975	154.7
	Single cell + RT1 oligo (384-well plates; Step 22, after evaporation)	-	-
	Total	0.5	260

25 Prepare an ERCC spike-in mix (1:80,000) in three steps to use in the RT reaction mix above.
First, prepare a 1:1,000 dilution by adding 9 µL of EB7.5 to a 1:100 ERCC aliquot.
Then, prepare a 1:10,000 dilution by mixing 2 µL of 1:1,000 ERCC with 18 µL of EB7.5.
Last, prepare a 1:80,000 dilution by mixing 5 µL of 1:10,000 ERCC with 35 µL of EB7.5.

26 Divide the RT mix among two 8-well strips, 16 µL per well, and place the strips on a 4 °C, 96-well
CPAC heating/cooling unit in the Bravo robot.

27 *Evaporation*. Remove the aluminum seal of the plate (Step 23) and place it at 95 °C for 00:03:00 for
complete evaporation of the lysis buffer. Immediately transfer the plate to a cooling 384-well block
for RT reaction mix addition using the robot.

CRITICAL STEP Evaporation time should be adjusted depending on the thermal cycler used (typically between 3 and 4 min) in order to ensure complete evaporation.

28 Activate the RT reaction mix addition script. This automated protocol adds 0.5 µL from the RT
reaction mix (Step 24) to each well in the 384-well plate (placed on a 4 °C, 384-well CPAC heating/
cooling unit). The tips are replaced and the process is repeated until all wells in the entire 384-well
plate are supplemented with RT reaction mix. The entire process takes 17 min per 384-well plate.
See Fig. 3b for a schematic of the plates' locations in this protocol.

In the RT reaction, the final Triton X-100 concentration is [M] 0.4 % volume and the RT primer concentration is [M] 4 nanomolar (nM) .

29 Following RT mix addition, seal the plate, spin down and transfer it to a 384-well thermal cycler for
the RT program (00:02:00 at 00:02:00 , 00:50:00 at 50 °C , and 00:05:00 at 85 °C).

30 Samples can be frozen at -20 °C until further processing (we have stored samples at this stage for up to 1 year).

Primer removal by exonuclease I (Timing ~1 h for 1 plate)

31 Spin down the 384-well plate after RT, and keep it on ice until the next step. Load the 'Exonuclease
I addition' Bravo script.

32 Prepare an exonuclease I reaction mix sufficient for the number of plates to be processed:

	Component	Amount (µl) per 1 well (1×)	Amount (µl) per 1 plate (520×)
	Exonuclease buffer (10×)	0.15	78
	Exonuclease I enzyme	0.015	7.8
	EB8	0.835	434.2
	cDNA (Step 29/30)	0.5	-
	Total reaction volume	1.5	-

33 Divide the exonuclease I mix among two 8-well strips, 32 µL per well, and place the strips in a 4 °C ,
96-well CPAC heating/cooling unit in the Bravo robot.

34 Activate the 'Exonuclease I addition' script. This automated protocol adds 1 µL from the exonuclease reaction mix to each well in the 384-well plate (placed in a 4 °C 384-well CPAC heating/cooling unit). The entire process takes 17 min per 384-well plate.
See Fig. 3b for a schematic of the plates' locations in this protocol.

35 Following Exonuclease addition, seal the plate, spin down and transfer to a 384-well thermal cycler
for the exonculease program (00:30:00 at 37 °C and 00:10:00 at 80 °C).

Pooling of barcoded 384 single-cell samples and cleanup (Timing ~1 h for 1 plate)

36 Load the 'Pooling' Bravo script. The pooling protocol pools together 384 single cells that are to be amplified together as an amplification batch. Prepare a 96-well pooling plate: dispense 55 µL of [M] 0.1 % volume Triton X-100 into each well in row H of a clean 96-well plate for tip pre-wash and place it in the appropriate location in the Bravo robot (destination 96-well plate).

37 Spin down the 384-well plate after the exonculease program and place it in the appropriate location in the Bravo robot. Activate the 'pooling' script: the process begins by pre-washing the tips and then aspirating the contents of the 384-well plate and dispensing them into row A of the 96-well destination plate. This is the first step of the pooling, which results in 12 wells in a 96-well plate (1 row), each well containing single cells from 2 columns of the 384-well plate. If more than one 384-well plate is pooled, tips will be replaced and the second 384-well plate will be pooled into row B of
the 96-well destination plate, and so on. The process takes 15 min per plate.
See Fig. 3c for a schematic of the plates' locations in this protocol.

38 Spin down the 96-well plate containing the pooled cDNA.

39 Using a multi-channel pipette with six barrels, transfer the contents of the last six wells, 48 µL in each well, to the first six wells in each row of the 96-well plate. This is the second step of the pooling, which results in 96 µl in each of the six wells.

40 Purify the product with 0.9 volumes of AMPure XP beads: add 86.4 µL to each well and pipette to mix. Incubate for 00:05:00 at Room temperature and transfer the contents of each row to a LoBind Eppendorf tube. Discard the 96-well plate used for pooling. Continue with AMPure XP cleanup in the LoBind Eppendorf tube and elute in 19 µL of EB8. Transfer 0.9 µL of the pooled sample to 19.1 µL of EB8 for QC1 (Box 1), and 17 µL to the first column in a new 96-well plate for the next step. At the end of this process, all single cells in each 384-well plate are transferred into one well in a 96-well plate. If processing several 384-well plates in parallel, each plate is represented in one well in a 96-well plate, one after the other in the same column.

CRITICAL STEP All AMPure XP cleanups are performed according to the manufacturing protocol. The amount of beads added in each step and the volume used for elution are indicated in the

specific step.

CRITICAL STEP From this step onward, the protocol is done in a 96-well plate, moving from one column to the next in each step. Each 384-well plate is represented in one well (one pool).

? TROUBLESHOOTING

Note

Box 1: Experimental QC by qPCR of a housekeeping gene
See instructions on how to perform qPCR in Steps 20 and 21. This is an important measurement for estimating library quality:
QC1: an estimation of the amount of starting material (Step 40). QC1 is performed after cells in one plate (or amplification batch) are barcoded, have undergone RT and are pooled together before the second-strand-synthesis step. Cells with high RNA content are expected to yield a lower CT value than cells with a lower RNA content (Anticipated results).
QC2: a measurement of library linear amplification (see Step 53). QC2 is performed after the pooled single cells undergo in vitro transcription and reverse transcription (see Anticipated results).
QC3: a measurement for final library quantity (see Step 56). QC3 is performed on the final amplified library (see Anticipated results).

CRITICAL Primers for qPCR should be designed for a housekeeping gene on the 3′ end of the transcript (no more than 200 bp upstream of the polyA site). Recommended primer sequences for mouse and human are provided (Reagent setup).

CRITICAL Although we recommend performing the above QC steps in order to maintain a foolproof process in the long run, these QC measurements can be skipped or sampled sporadically.

41 Samples can be frozen at -20 °C until further processing (we have stored samples at this stage for up to 1 year).

Second-strand synthesis and IVT amplification (Timing ~2.5 h + overnight incubation)

42 Convert the pooled cDNA to double-stranded DNA using a second-strand-synthesis kit in which the enzyme is diluted 1:8 in nuclease-free water:

	Component	Volume (µl) per well
	Second-strand-synthesis buffer (10×)	2
	Second-strand-synthesis enzyme	1

	(diluted 1:8 in nuclease-free water)	
	Purified cDNA pool (Step 40/41)	17
	Total reaction volume	20

CRITICAL STEP Mix the second-strand-synthesis reaction well with a 20-µl pipette.

43 Spin down the plate and incubate at 16 °C for 02:00:00 .

44 Purify the product with 1.4 volumes of AMPure XP beads: add 10 µL of EB8 and 42 µL of AMPure XP beads. Continue with AMPure XP cleanup and elute in 8 µL of EB8. Do not transfer the supernatant to a new well; IVT is performed with the beads.

45 Linearly amplify the cDNA by IVT. Prepare the following reaction mix and add 12 µL of IVT mix
per well:

	Component	Volume (µl) per well
	rNTPs (mixture of A, C, U, G)	8
	T7 RNA polymerase buffer (10×)	2
	T7 RNA polymerase enzyme	2
	Purified dsDNA (Step 44)	8
	Total reaction volume	20

46 Incubate at 37 °C overnight.

CRITICAL STEP Following IVT incubation overnight, the AMPure XP beads sink to the bottom. This is normal.

47 Following IVT, add 10 µL of Turbo DNase I mix to digest the DNA template:

	Component	Volume (µl) per well
	Turbo DNase buffer (10×)	3
	Turbo DNase I	2
	Nuclease-free water	5
	IVT reaction (Step 46)	20
	Total reaction volume	30

48 Incubate at 37 °C for 00:15:00 .

49 Purify the aRNA product with 1.2 volumes of AMPure XP beads: add 20 µL of EB7.5 and 60 µL of AMPure XP beads. Continue with AMPure XP cleanup. Elute with 19 µL of EB7.5. Transfer 18 µL to a well in a new column in the same 96-well plate.

CRITICAL STEP Make sure to avoid transferring the beads, as this will interfere with the next step.

50 Samples can be frozen at -80 °C until further processing (we have stored samples at this stage for up to 1 year).

RNA fragmentation (Timing ~30 min)

51 **CRITICAL** The aRNA (Step 49) is fragmented by chemical fragmentation into a median size of ~300 nucleotides by 00:03:00 of incubation at 70 °C in Zn2+ RNA fragmentation solution.

CRITICAL If aRNA (Step 49) was frozen, thaw the 96-well plate on ice and spin down before the next step.

52 Add 2 µL of 10× Zn2+ fragmentation reagent to the aRNA from Step 49.

53 Incubate for exactly 00:03:00 in a thermal cycler pre-heated to 70 °C. Transfer immediately to a cooling block and add 2 µL of stop solution.

54 Cool for 00:01:00 – 00:02:00 and spin down.

55 Purify the product with 2 volumes of AMPure XP beads: add 44 µL of AMPure XP beads. Continue with AMPure XP cleanup. Elute with 5.5 µL of EB7.5. Transfer 5 µL to a well in new column and continue to ligation.

CRITICAL STEP Drying time of the beads is considerably shorter than that used in the AMPure XP cleanup in other steps.

CRITICAL STEP Make sure not to transfer beads in the elution step, as this will interfere with the next step.

Ligation of plate barcode (Timing ~2.5 h)

56 Ligate a plate adaptor to the fragmented RNA (RNA:ssDNA ligation): select a unique plate adaptor
(Supplementary Table 2) for each library to be sequenced together and record it.

57 Add 1 µL of ligation plate adaptor ([M] 100 micromolar (µM)) to the purified fragmented RNA (from Step 55) and incubate for 00:03:00 in a thermal cycler pre-heated to 70 °C . Transfer immediately to a cooling block.

58 Prepare a ligation master mix in a LoBind Eppendorf tube by adding the ingredients below according to the order specified. The master mix should contain all the reagents, except for the fragmented RNA + ligation adaptor from Step 57.

	Component	Volume (µl) per well
	T4 ligation buffer (10×)	2
	100% (vol/vol) DMSO	1.9

	100 mM fresh ATP	0.2
	50% (wt/vol) PEG	8
	T4 RNA ligase 1	2
	Fragmented RNA + ligation adaptor (from Step 57)	6
	Total reaction volume	20

CRITICAL STEP When preparing a mix for ligation, follow these guidelines:
(i) vortex and spin down after addition of each reagent,
(ii) place DMSO and PEG at 37 °C until use,
(iii) work very slowly with PEG to ensure the correct volume is added and
(iv) store ATP in small aliquots at −80 °C.

59 Add 14.1 µL from the ligation mix to the aRNA and ligation adaptor (from Step 57), mix very slowly
until homogeneous and spin down the plate.

60 Incubate at 22 °C for 02:00:00 .

61 Purify the ligation product with 1.5 volumes of AMPure XP beads: add 80 µL of EB7.5 + 100 µL of
AMPure XP beads. Continue with AMPure XP cleanup. Elute with 13.7 µL of EB7.5 and transfer
13.2 µL to a well in a new column of the plate.

CRITICAL STEP Allow a longer magnetization time for the AMPure XP beads (7–8 min).

CRITICAL STEP Drying time of the AMPure XP beads may be longer (7–8 min).

Second reverse transcription (Timing ~1.5 h)

62 **CRITICAL** The ligated product is reverse-transcribed with a primer complementary to the ligated adaptor.

63 Add 1 µL of [M] 25 micromolar (µM) RT2 primer to the purified ligated aRNA (from Step 61), mix by pipetting, seal and spin down the plate.
You should have 14.2 µL per well.

64 Incubate the plate for 00:03:00 in a thermal cycler pre-heated to 72 °C. Transfer immediately to a cooling block. Spin down the plate.

65 Prepare the RT mix according to the table below (omitting the ligated product + RT primer) and add 5.8 µL of the mix per reaction well:

	Component	Volume (µl) per well
	AffinityScript buffer (10×)	2
	100 mM DTT	2
	100 mM dNTP mix	0.8
	AffinityScript RT enzyme	1
	Ligated product + RT primer (from Step 50)	14.2
	Total reaction volume	20

66 Seal the plate, spin down and incubate it for 00:02:00 at 42 °C, 00:50:00 at 45 °C, and 00:05:00 at 85 °C.

67 Purify the cDNA with 1.5 volumes of AMPure XP beads: add 10 µL of EB8 and 45 µL of AMPure XP
beads. Continue with AMPure XP cleanup and elute in 25 µL of EB8. Transfer 24 µL to a new
column. Take 1.2 µL of the cDNA and add it to 18.8 µL of EB8 for QC2 (Box 1).

? TROUBLESHOOTING

68 **PAUSE POINT** Samples can be frozen at -20 °C until further processing (we have stored samples
at this stage for up to 1 year).

Enrichment of final library (Timing ~1 h)

69 The library is completed by amplifying the cDNA using PCR: transfer 11.5 µL of cDNA (Step 67) to a well in a new column and add 12.5 µL of KAPA HiFi HotStart ReadyMix + 1 µL of [M] 25 micromolar (µM) enrichment primer mix (Reagent setup).

70 Vortex and spin down. The remaining half of the cDNA can be kept as a backup at -20 °C.

71 Run a PCR program as follows:

	Cycle no.	Denature	Anneal	Extend	Final
	1	98 °C, 2 min			
	2–3	98 °C, 20 s	57 °C, 30 s	72 °C, 30 s	
	4–15	98 °C, 20 s	65 °C, 30 s	72 °C, 30 s	
	16			72 °C, 5 min	
	17				10 °C, ∞

CRITICAL STEP The optimal number of PCR cycles is a tradeoff between insufficient amplification and too much amplification (which can cause PCR bias). The number of cycles is dependent on the cell type used and should be calibrated for each project. Typically, the number of amplification cycles required for enrichment is between 11 and 17. In the table above, we have used 14 cycles (2 cycles at 57 °C and 12 cycles at 65 °C annealing temperature) as an example. If needed, the PCR can be repeated with an adjusted cycle number (altering the number of cycles performed at 65 °C annealing temperature) using the cDNA backup.

CRITICAL STEP Tracking QC values in several experiments from the same cell type can help in estimating, empirically, the necessary enrichment. As QC2 (Step 67) indicates the amount of material in the sample before the final amplification, it can link the number of PCR cycles done to the final library concentration obtained at the end of the process.

72 Purify the amplified library with 0.7 volumes of AMPure XP beads: add 35 µL of EB8 and 42 µL of AMPure XP beads. Continue with AMPure XP cleanup. Elute with 20 µL of

EB8. Transfer 19 µL of the final library to a new well. Take 2 µL of the final library and add it to 38 µL of EB8 for QC3 (Box 1).

QC of libraries before sequencing (Timing ~1 h)

73 **CRITICAL** All libraries should be carefully examined before sequencing.

74 *Library concentration*. Measure the final library concentration (ng/µl) using the dsDNA HS Assay Kit and a Qubit fluorometer. A typical library should yield between 1 and 10 ng/µl of dsDNA.

CRITICAL STEP For better accuracy, it is recommended to use 2 µL of the library for Qubit measurement.

? TROUBLESHOOTING

75 *Library profile*. Evaluate the library profile and the mean molecule size using a 2200 TapeStation instrument or equivalent.

? TROUBLESHOOTING

Sequencing (Timing ~12 h)

76 **CRITICAL** Successful libraries (passing QC checks) can be sequenced.

77 Calculate the molarity of each pool, using the following formula:

$$\mathrm{nM} = 1,000 \times [\mathrm{ng}/\mu\mathrm{l}] / (0.649 \times [\mathrm{bp}]),$$

where [ng/µl] is the concentration obtained with the Qubit assay (Step 74) and [bp] is the library profile peak obtained with TapeStation (Step 75).

78 Dilute each library pool to [M] 4 nanomolar (nM) (or another equivalent molar concentration of the pools) with EB8
and mix the different pools in equal ratios.

79 Libraries of 20–40 384-well plates can be mixed and sequenced together on a NextSeq 500 High Output v2 kit flow cell to yield ~0.5 × 10^9 reads.

80 Sequence libraries with the following parameters: rd1 = 75 bp, rd2 = 15 bp, no indices.

Analytical pipeline (Timing ~1 d)

81 **CRITICAL** Detailed instructions on how to use the MARS-seq2.0 analytical pipeline and create a single-cell database (SCDB) can be found in the MARS-seq2.0 pipeline user manual on our website http://compgenomics.weizmann.ac.il/tanay/?page_id=672. See Supplementary Manual 2.

82 Make sure you have the correct SCDB directory structure in your Linux storage.

83 Make sure that the parameters defined in scdb/config/config.txt are correct.

84 Add the metadata on the sequenced library according to the SCDB as described in detail in the user
manual.

85 Add the .fastq files (raw reads) obtained at Step 80 to the SCDB.

86 Run the pipeline locally or on an SGE (sun grid engine) cluster.