

Resource Summary Report

Generated by [dkNET](#) on Sep 11, 2026

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

DOI:10.17504/protocols.io.7hkhj4w

Type: Protocol

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

Protocol Information

URL: <https://dx.doi.org/10.17504/protocols.io.7hkhj4w>

Authors: Hadas Keren Shaul, Hadas Keren-Shaul, Ephraim Kenigsberg, Diego Adhemar Jaitin, Eyal David, Franziska Paul, Amos Tanay, Ido Amit

Group: Human Cell Atlas Method Development Community

Summary: 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 ($N \sim 10^3$)^{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-seq³⁸; Fig. 1b,c). Overview of the experimental protocol

MARS-seq2.0

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 cost-effective 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 development⁴² and in response to influenza virus infection⁴⁴, and detailed the heterogeneity of epithelial cells in the thymus⁴¹. 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 patients⁴⁷, 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 mice³⁸, 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. Optimization of MARS-seq2.0 background noise is a major problem in all massively parallel single-cell protocols. Our prior study¹⁰ 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 MARS-seq 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 μ l of lysis buffer containing the RT primer at 95 °C, followed by addition of 0.5 μ l 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

Lowering of RT volume: we reduced the volume of the RT reaction by eightfold from 4 μ l 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.

Optimization of lysis buffer: we optimized the lysis buffer to make it compatible with the aforementioned volume reduction.

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.

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.

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.

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).

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

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.

Generation of a complete analytical pipeline: sequenced reads are de-multiplexed and the molecular signal is separated from the background noise.

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)

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 (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 heatmap 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: 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. 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. 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. The amount of cell doublets in the MARS-seq2.0 data is extremely low and allows for identification of rare cell types and subpopulations. MARS-seq2.0 provides strand-specific information about the transcripts. However, the following limitations remain: 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-seq²⁴, should be used). The method is selective for polyadenylated RNA. The method requires access to FACS facilities, as well as liquid-handling robotics. 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

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.

Generation of a complete analytical pipeline: sequenced reads are de-multiplexed and the molecular signal is separated from the background noise.

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) < 0.05) 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 (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: 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. 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. 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. The amount of cell doublets in the MARS-seq2.0 data is extremely low and allows for identification of rare cell types and subpopulations. MARS-seq2.0 provides strand-specific information about the transcripts. However, the following limitations remain: 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-seq²⁴, should be used). The method is selective for polyadenylated RNA. The method requires access to FACS facilities, as well as liquid-handling robotics.

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) 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 (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: 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. 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. 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. The amount of cell doublets in the MARS-seq2.0 data is extremely low and allows for identification of rare cell types and subpopulations. MARS-seq2.0 provides strand-specific information about the transcripts. However, the following limitations remain: 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-seq²⁴, should be used). The method is selective for polyadenylated RNA. The method requires access to FACS facilities, as well as liquid-handling robotics. However, the following limitations remain: 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-seq²⁴, should be used). The method is selective for polyadenylated RNA. 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}, genetics⁵⁴, lineage⁵⁵ and signaling³⁷, 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 more⁵⁶.

Affiliations: Weizmann Institute of Science, Department of Immunology, Weizmann Institute of Science, Rehovot, Israel; Life Science Core Facility, Weizmann Institute of Science, Rehovot, Israel; These authors contributed equally: Hadas Keren-Shaul, Ephraim Kenigsberg, 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; These authors contributed equally: Hadas Keren-Shaul, Ephraim Kenigsberg, Department of Immunology, Weizmann Institute of Science, Rehovot, Israel, Department of Immunology, Weizmann Institute of Science, Rehovot, Israel, Department of Immunology, Weizmann Institute of Science, Rehovot, Israel, Department of Computer Science and Applied Mathematics and Department of Biological Regulation, Weizmann Institute, Rehovot, Israel; These authors jointly supervised this work: Amos Tanay, Ido Amit, Department of Immunology, Weizmann Institute of Science, Rehovot, Israel; These authors jointly supervised this work: Amos Tanay, Ido Amit

Version: 1

Publication Date: 2019

Proper 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 [dx.doi.org/10.17504/protocols.io.7hkhj4w](https://doi.org/10.17504/protocols.io.7hkhj4w)

Record Creation Time: 20210329T220534+0000

Record Last Update: 20210329T221016+0000

Data and Source Information

Source: