nature protocols

https://doi.org/10.1038/s41596-025-01308-8

Protocol

Check for updates

A scalable, low-cost, sample hashing workflow for multiomic single-cell analysis using the Seq-Well S^3 platform

Daniela D. Russo[1,2,3,4], Sarah L. Quinn[2,3,4], Olha Kholod[5], Michal A. Elovitz[6], Douglas A. Lauffenburger[7], Pardis Sabeti[4,8,9,10], Boris Julg[2], Andrea G. Edlow[11,12,13], Brittany A. Goods[5], Alex K. Shalek[2,3,4,14] & Sergio Triana[2,3,4,10,14]

Abstract

In-depth analyses of clinical samples have the potential to provide unparalleled insights into the cellular mechanisms that underlie both health and disease, as well as therapeutic and prophylactic responses. However, these specimens are often paucicellular, necessitating the use of workflows that maximize the amount of information that can be learned. Here we provide a detailed protocol for generating and analyzing single-cell multiomic data from low-input samples with the Seq-Well S^3 platform. We further describe a matched pipeline for sample hashing that reduces costs and sources of technical variation in the resulting data while also enhancing throughput. In brief, our streamlined and efficient methodology involves: (1) optionally staining single-cell suspensions with antibody–oligonucleotide conjugates for cell surface protein quantification and/or sample multiplexing; (2) generating Seq-Well S^3 sequencing libraries; (3) optionally producing bulk-RNA sequencing libraries via SMART-seq2 to support genetic demultiplexing; and (4) computationally analyzing the resulting data. Each step herein has been designed to leverage readily available reagents and standard laboratory equipment, substantially lowering barriers to entry for researchers. The overall Protocol can yield high-quality multiomic insights from samples in under a week.

Key points

- This Protocol provides detailed guidelines for performing single-cell multiomic analyses on the Seq-Well S^3 platform. The workflow integrates single-cell RNA sequencing with surface protein quantification using oligonucleotide-tagged antibody staining and antibody-based cell hashing. SMART-seq2 bulk RNA sequencing can optionally be employed for genetic demultiplexing.
- This Protocol is well suited to precious clinical samples, cost-sensitive applications and resource-limited environments due to its low input sample requirements, reliance on widely commercially available reagents and limited instrumentation needs.

Key references

Mead, B.E. et al. *Nat. Biomed. Eng.* **6**, 476–494 (2022): https://doi.org/10.1038/s41551-022-00863-9

Kotliar, D. et al. *Cell* **183**, 1383–1401.e19 (2020): https://doi.org/10.1016/j.cell.2020.10.002

Hughes, T. K. et al. *Immunity* **53**, 878–894.e7 (2020): https://doi.org/10.1016/j.immuni.2020.09.015

[1]Department of Immunology, Harvard Medical School, Boston, MA, USA. [2]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [3]Institute for Medical Engineering and Science, Department of Chemistry, and Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Thayer School of Engineering, Program in Quantitative Biomedical Sciences, and Department of Molecular and Systems, Dartmouth College, Hanover, NH, USA. [6]Nuttall Women's Health, New York, NY, USA. [7]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [9]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [10]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [11]Department of Obstetrics and Gynecology, Massachusetts General Hospital, Boston, MA, USA. [12]Vincent Center for Reproductive Biology, Massachusetts General Hospital, Boston, MA, USA. [13]Harvard Medical School, Boston, MA, USA. [14]These authors contributed equally: Alex K. Shalek, Sergio Triana. e-mail: strianas@broadinstitute.org

Introduction

The exploration of cellular heterogeneity with single-cell genomics methods, such as single-cell RNA sequencing (RNA-seq), has transformed basic and clinical biology. Collectively, these approaches have revealed new cell types and states[1,2] and helped delineate molecular processes driving healthy development and aging[3–5], as well as autoimmunities[6–8], infections[9–11] and cancers[12–14]. Recent experimental advances have further enhanced discovery potential by enabling examination of multiple complementary molecular layers—such as surface protein expression and the transcriptome—at once in a single cell[15,16]. These concurrent measurements support detailed analyses of intricate cellular activity, such as post-transcriptional[17] or translational[18] regulation. These same strategies can also be used to facilitate sample multiplexing, improving throughput while reducing costs[15]. However, the full potential of these technologies is often not realized due to constraints related to sample availability, labor or costs. The development of cost-effective, multiplexable single-cell multiomic workflows amenable for use with paucicellular inputs would empower the characterization of rare or clinical samples[19,20] and increase the applicability of single-cell genomic techniques in resource constrained environments[21,22]; parallelized sample processing would additionally permit increased sample throughput without increased labor or cost to complete the experiment successfully. Achieving these goals is essential to power basic and applied studies nearly anywhere[23].

Development and overview of the protocol

Previously, we demonstrated that Seq-Well S^3 is a portable, efficient and simple single-cell RNA-seq workflow[24,25]. We have since described how to combine the method with oligo-tagged antibodies[20,26]. This enables both quantification of cell surface proteins, enhancing the assay's information content, and antibody-based cell hashing, lowering cost per sample, increasing throughput and mitigating batch effects[20,24,26,27]. Herein, we provide a refined and complete protocol for combining Seq-Well S^3 single-cell RNA-seq, cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) surface protein quantification[15] and oligo-labeled antibody-based cell hashing (Fig. 1a). We also include details on how to perform SMART-seq2 (ref. 28) bulk RNA-seq on excess cells to support, where desired, higher fidelity sample demultiplexing through incorporation of genetic hashing. Finally, we include detailed computational workflows to enable analyses of the generated data (Fig. 1b). Collectively, this Protocol substantially increases the amount of information that can be learned from a sample even with low input numbers of cells, while reducing cost and labor requirements.

Our Protocol is composed of several experimental (Fig. 1a) and computational steps (Fig. 1b). CITE-seq leverages oligonucleotide-labeled antibodies to tag cell surface proteins, permitting simultaneous detection of surface protein and RNA expression within individual cells using a single molecular capture reagent. It can also be used for multiplexing samples before processing, if each sample is prelabeled with a unique oligonucleotide-labeled antibody. However, successful application of CITE-seq requires effective antibody staining, which, in turn, depends on the availability of a sufficient number of input cells. Recognizing this limitation, we also introduce an optimized pipeline that leverages SMART-seq2 bulk RNA-seq to enable demultiplexing based on single nucleotide polymorphisms that differ between sample donors ('genetic hashing')[29,30]. Coupling both genetic and oligo-labeled antibody hashing permits computational rescue of cells unassignable via either method alone, and increases the confidence of cell assignments to the corresponding sample of origin[31].

To further enhance the utility of these approaches, we also provide a streamlined computational workflow tailored for demultiplexing and integrated analysis of the generated transcriptomic and proteomic data (Fig. 1b). As one embodiment, we profile the transcriptomes and surface proteomes of peripheral blood mononuclear cells. We note that this Protocol can be applied, with minor optimization, to a wide range of sample types, highlighting its versatility and applicability for revealing previously unappreciated cellular phenotypes, dissecting mechanisms of disease pathogenesis and identifying potential therapeutic targets[20–22,26,27].

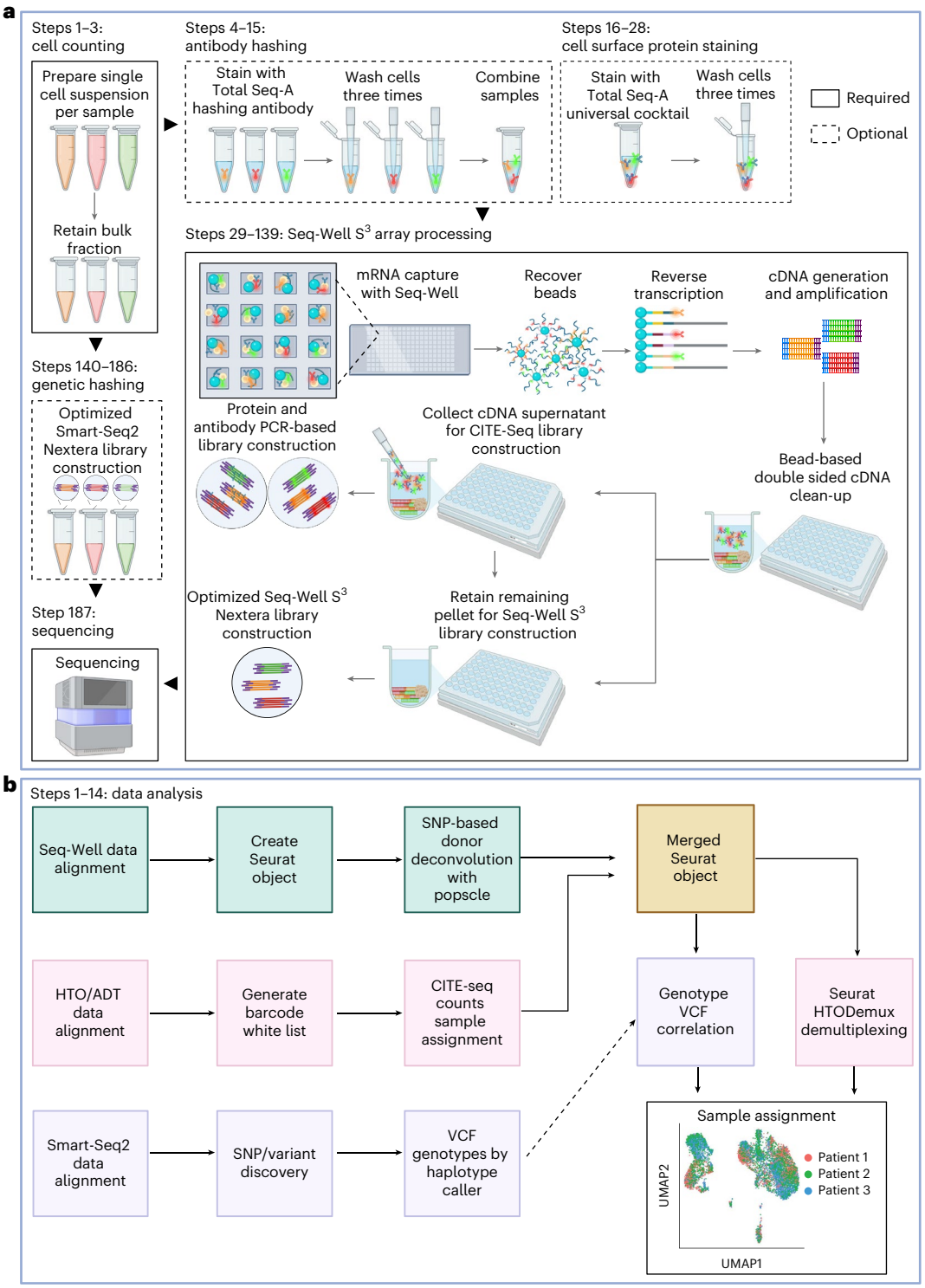

Fig. 1 | Schematic workflows for single-cell multiomic sequencing library preparation and data analysis processing. The dashed line boxes indicate optional steps of the Protocol. **a**, Sample preparation and generation of sequencing libraries. Four library types (scRNA, HTO, ADT and bulk RNA) will undergo Illumina sequencing to generate raw reads for data analysis. **b**, Each library type will undergo its respective preprocessing indicated by each colored box. Subsequent data integration occurs resulting in a final Seurat object composed of various data levels. The figure was created with BioRender.com (Shalek, A. (2025)) https://BioRender.com/v1czkcc.

Applications of the Protocol

Seq-Well S^3 is an open-source picowell platform with transparent processing chemistry, facilitating optimizations to enable work across a variety of systems including cancer[2], infectious disease[19,21,22,27,32,33] and inflammation[34,35] to name a few. By integrating CITE-seq, the presented protocol enables improved clustering and discovery of transcriptomic features via incorporation of canonical surface markers. The possibility of multiplexing samples using CITE-seq additionally enables batch processing to reduce potential sources of technical variation and increases the number of donors that can be profiled within a specific budget, enhancing power for certain analyses and experiments. As has been shown previously, this combination provides a potent discovery platform in the realm of oncology, immunology and fundamental human biology[36], facilitating dissection of disease mechanisms and identification of therapeutic targets[2,37].

Comparison with other methods

Current methods for interrogating dissociated single cells fall into four broad categories, including plate, droplet, split-pool-based and picowell methods[38]. Plate-based approaches enable the most comprehensive views of individual cell transcriptomes as sequencing is not limited to the 3′ or 5′ end of genes. This capability is crucial for accurately reconstructing T cell receptors, B cell receptors and oncogenes, among others, offering critical insights into cellular function and disease pathology[39–41]. However, plate-based workflows can be time-consuming and costly, limiting throughput and thus power. By leveraging early 3′ or 5′ barcoding, droplet-based techniques overcome these issues, enabling thousands of cells to be processed in parallel. As a result, these strategies have become the de facto standard for many applications[42,43], especially given recent molecular advances that have further enhanced their efficiency and reliability. However, droplet technologies require high cell concentrations—typically between 200 and 500 cells per microliter—to achieve optimal cell capture, which can pose challenges in studies with low cell numbers, such as rare cell populations or clinical biopsies, resulting in low yields and/or increased costs per cell. Split-pool workflows combine high throughput with minimal equipment requirements[44] but often yield reduced data quality and limited gene capture due to difficulties manipulating nucleic acids in fixed cells. These methods are also not well-suited for applications involving low numbers of cells. Picowell techniques integrate aspects of both plate-based and droplet-based methods to achieve intermediate throughput without extensive equipment requirements, making them versatile and applicable to various settings[24,25,42,45]. Moreover, the gravity-based cell loading typically used in picowells platforms permits gentle handling of fragile cells, further enhancing usability in diverse research contexts. Multiple picowell-based workflows have been described, including Cyto-seq[46], Microwell-seq[47,48] and the Seq-Well system[24,45].

Among these, Seq-Well S^3 offers certain advantages while performing comparably to both droplet and other picowell methods. One key strength is its low cell numbers/concentration requirement, which sets it apart from droplet-based approaches like those of 10x Genomics. Specifically, Seq-Well S^3 operates effectively at concentrations as low as 15,000 cells per milliliter[49], whereas droplet-based methods generally require higher cell concentrations, typically ~200–500 cells per microliter (200,000–500,000 cells per microliter). As a result, if desired, Seq-Well S^3 can be performed without concentrating samples, reducing possible material losses. This difference in cell concentration requirements makes Seq-Well S^3 particularly beneficial for applications where low-input samples are the norm, such as rare cell population studies[50], small biopsy analyses[19] or clinical sample applications[27,51]. Seq-Well S^3 also maintains high gene counts and data quality across a wide range of cell types, including fragile and low-abundance cells. This makes it a suitable platform for studies in clinical or other sensitive research contexts, where sample quality and quantity are often limiting factors.

The harmonized Seq-Well S^3 protocol we present here is designed to enable the processing of precious samples with minimal equipment, lowering barriers to access. In addition, the inclusion of sample hashing using both CITE-seq and genetics further enhances throughput, minimizes batch effects and reduces costs and labor. Unlike antibody-based hashing methods, the genetic approach we outline does not require substantially higher cell counts or substantially alter overall costs, making it an attractive alternative for cost-sensitive applications.

While microarray, whole-genome sequencing and bulk RNA-seq can be used to acquire patient-specific genetic data to enhance genetic deconvolution, the SMART-seq2 protocol we detail here offers a cost-effective, low-coverage genotyping alternative. This alternative enables researchers to balance cost, resolution and sample requirements in a way that suits their specific project objectives.

Limitations of the protocol

Despite the aforementioned versatility and robustness, the presented protocol does have limitations. First, it requires fresh cells and is not compatible with fixed samples; while Seq-Well S^3 and genetic hashing are compatible with nuclei, CITE-seq antibody staining is not. Second, the incorporation of CITE-seq necessitates more cells, as some are lost during washes to remove free floating antibodies. CITE-seq cell loss is dependent upon cell recovery during the washing steps; thus, it is most useful when a large number of input cells is available to ensure sufficient recovery post staining. CITE-seq is implemented in this protocol for both antibody based hashing and staining specific cell surface proteins of interest. These are complementary approaches, with antibody hashing enabling sample demultiplexing independent of donor genotype; however, its inclusion increases the number of washes and potential cell loss. When sample input is limited, cell surface protein staining can be performed with genetic hashing alone to enhance overall cell recovery as fewer washes are required. Third, the cost per array is fixed, making the per-cell cost variable and dependent on yield, which underscores the importance of optimal loading[45,51]. Fourth, while hashing increases sample throughput, it reduces the cells recovered from each sample, so we recommend conducting a pilot study to estimate effect sizes accurately, enabling better cohort design (that is, how to best distribute the ~6,000 cells recovered per array across samples). Finally, while the protocol requires limited peripherals, it still calls for a few basic tools. Given these considerations, users should evaluate the potential limitations of the protocol when adapting it to their specific research needs and whether a different approach might be more enabling.

Experimental design

Sample preparation

This Protocol relies on single-cell suspensions, which can be prepared via the user's preferred tissue dissociation method. After cellularization, cells are processed through the Seq-Well S^3 protocol and, if cell counts and budget allow, via SMART-seq2 to create bulk RNA-seq data for improved genetic demultiplexing. For successful genetic hashing, each participant's genetic makeup needs to be unique. To maintain RNA quality, we recommend quickly lysing and then freezing leftover cells, either from the same sample or other specimens from the same patient. This permits bulk RNA processing at a later time. Perform this step before any antibody tagging to ensure that genotype sequences are derived from the samples themselves, not from the hashing oligonucleotides.

We tag individual samples using unique Total Seq-A anti-human Hashtag Antibodies from BioLegend, which specifically target CD298 and β2 microglobulin (B2M). These markers are ubiquitously present on most human cells, making these antibodies nearly universally applicable. We have also achieved success with the BD Single-Cell Multiplexing Kit but show data here using the BioLegend Total Seq-A antibody system for simultaneous multiplexing and cell surface protein identification.

In our protocol, we adopt a modular approach for sample tagging and surface protein detection. While it is feasible to perform the antibody-based multiplexing and surface protein staining simultaneously to reduce cell losses, CITE-seq panels for protein detection are expensive, particularly when applied to individual samples. Performing these reactions on a per-sample basis can thus be cost prohibitive, especially in projects with large sample cohorts or conducted in low-resource settings.

By multiplexing samples first using unique hashtag oligos (HTOs), we substantially reduce the number of CITE-seq reactions needed for surface protein detection. This approach limits costs while maintaining the flexibility to add surface protein readouts at later stages if required. This strategy also ensures that the protocol is adaptable to different experimental goals and

budgetary constraints while still providing high-quality data. By modularizing the process, we have enhanced the protocol's usability across a wider range of research contexts without compromising data integrity or efficiency.

Single-cell mRNA capture

The capture process adheres to the Seq-Well S^3 methodology, as we have detailed in prior publications[24,25,43]. In brief, this method involves loading a suspension of single cells onto a picowell array coloaded with mRNA capture beads and performing a series of downstream processing steps on those beads. This overall process is comparable to other approaches that use barcoded beads to tag mRNA from cells with both a cell barcode and a unique molecular identifier (UMI), including Drop-seq[52] and 10x Genomics[53] workflows.

After lysis and reverse transcription, the beads are retrieved from the array and undergo second strand DNA synthesis (S^3). To mitigate potential biases associated with complementary DNA amplification, material is evenly distributed across 24 wells of a standard 96-well plate. This distribution minimizes variations in amplification efficiency and maximizes the uniformity of the sequencing library, thereby enhancing the accuracy and reliability of the resulting data.

Isolation of short fragments

Post complementary DNA amplification, the protocol deviates from standard Seq-Well S^3 to selectively isolate the shorter DNA fragments from the cell labeling procedures. This isolation step is essential for ensuring purity and specificity downstream. AMPure XP can be used to bind DNA fragments, with the efficiency of binding directly correlated to the beads' concentration within the solution and the size of the DNA. We leverage this length dependence to effectively sequester longer DNA fragments on AMPure XP beads, leaving the shorter, tagged fragments in the supernatant.

We then transfer the supernatant, enriched with the short DNA fragments of interest, to a fresh 96-well plate. Given that subsequent stages include additional rounds of PCR, we distribute the supernatant across two wells of the 96-well plate to mitigate the potential for bias in amplification, minimizing any disparities that could impact the integrity of the data.

Generation and sequencing of multiple library types

Three separate sequencing libraries are generated per experiment: (1) one dedicated to gene expression analysis (GEX); (2) one for HTOs for downstream sample demultiplexing; and (3) one for antibody-derived tags (ADTs) for assessing surface protein levels. Amplified cDNA from each library will undergo the appropriate fragmentation to yield optimal size fragments for sequencing with Illumina's short-read technology. Each library is created separately from the others, and the library is assigned a unique barcode, facilitating the multiplexing of numerous arrays within a single sequencing run, further enhancing throughput and efficiency.

For optimal data quality and depth, we recommend a minimum sequencing target of 25,000 reads per cell for gene expression libraries. A total of 25,000–75,000 reads per cell reduces the chance of dropout, especially for lowly expressed genes[24,25] and achieves sequencing saturation. We recommend referring to the most recent sequencing recommendation from BioLegend for guidance regarding HTO and ADT library sequencing requirements[54]. The library structure outlined in previous Seq-Well S^3 protocols may be used, although we recommend increasing the read 2 length from 50 to 90 bp to permit better genotyping and genetic demultiplexing.

Leveraging genetic demultiplexing for limited samples

Genetic demultiplexing can be performed alone or with antibody-based cell hashing to assign cells to their samples of origin. While the combination results in the most effective cell assignment, genetic demultiplexing can be used by itself to avoid the cell loses associated with antibody washing steps (for example, when working with scarce cellular material). Genetic demultiplexing requires the generation of a genotyping reference, for example, via low input bulk RNA-seq. This reference can be made with either the same cells that are run on Seq-Well S^3, or a less precious sample from the same individual, such as peripheral blood mononuclear cells.

This flexibility ensures that even when the primary single-cell sample is limited, researchers can still perform robust genetic hashing and downstream analyses without compromising data quality or wasting valuable material.

To support a reference-informed implementation, we present the use of SMART-seq2 as a low-cost, efficient method for bulk RNA-seq that can be employed to genotype donor samples. SMART-seq offers flexibility and scalability, making it suitable for a range of input sizes, including very low cell numbers. As detailed below, we recommend preserving cells from the same or a different tissue at the time of sample collection. These cells should be lysed and frozen promptly to maintain RNA integrity for later processing.

For bulk RNA-sequencing, we leverage SMART-seq2 in this Protocol to enable genotype profiling, optimized for inputs ranging from a single cell to millions[55]. The process starts with the extraction of total RNA from a bulk cell population. When working with low cell inputs (for example, <20,000 cells), this can be done using lysis buffer followed by RNAClean XP SPRI beads. These beads can be used to capture nucleic acids from the sample while effectively removing residual lysis agents and non-nucleic acid cellular materials that could interfere with downstream RNA processing. When working with larger cell numbers (for example, >20,000 cells), other common bulk RNA extraction methods should be used instead (that is, Qiagen RNeasy Kits for RNA Purification). We recommend aiming for at least 1,000 cells as input for bulk RNA-seq, if possible, to assure a better single-nucleotide polymorphism (SNP)/exon coverage

Following extraction, the isolated RNA is subjected to reverse transcription in the presence of a template-switching oligo. The template-switching oligo enables addition of a PCR amplification handle to the 5′ end of a cDNA; use of the same PCR handle on the oligo-dT primer and the template-switching oligo enables subsequent PCR with a single oligonucleotide and preferential amplification of longer transcripts (due to hairpin suppression of shorter ones). After synthesizing cDNA, the samples undergo limited-cycle PCR amplification to increase the quantity of cDNA, followed by the incorporation of sequencing adapters and unique indices during library preparation. These indices are crucial for enabling the multiplexing of multiple libraries in a single sequencing run. The libraries undergo a final size selection step to isolate fragments ideally suited for clustering on Illumina platforms, permitting a more robust quantification and thereby enhancing the quality and utility of the data generated. Successful sequencing has been performed with read lengths as short as 35 bp (although longer read length will improve the accuracy of the genotyping) and attributing 3 million reads per sample.

Downstream analysis

We employ a systematic approach to ensure the integration of diverse data types—specifically, scRNA, HTO, ADT and bulk RNA-seq libraries. Each of these data streams is processed individually to maintain the integrity and specificity of the information it contains, before being unified into a cohesive dataset for in-depth analysis. The process begins with the demultiplexing of raw sequencing files, followed by alignment to a species- or analyte-appropriate reference. From here, computational pipelines are applied individually to each library type. For the single-cell RNA-seq data, processing involves the generation of count matrices. Once the single-cell RNA-seq component is structured into a Seurat object, the data from the HTO and ADT libraries are incorporated using the 'HTODemux()' function in Seurat[56]. This function is designed to demultiplex cell hashing data and annotate cells based on their barcode identities, thereby enriching the Seurat object with cell-specific protein expression data derived from the ADT library. The bulk RNA-seq data, while not integrated at the level of count matrices, is leveraged to give patient level labels to the output data obtained from genotype-free demultiplexing software[57].

It is important to note that while we describe this process using specific tools and languages (for example, Seurat in R), the principles of data processing and integration can be applied using different software packages and programming environments, including Python and frameworks such as Scanpy[58]. Alternative tools may be employed based on compatibility, specific features or personal preference, which can also effectively accomplish the task of merging and analyzing these complex datasets. Ultimately, the goal is to create a layered,

multidimensional representation of the data that permits nuanced biological interpretations and robust statistical analysis.

Expertise needed to implement the protocol

The protocol—encompassing single-cell suspension preparation, Seq-Well S^3, CITE-seq and SMART-seq2, and all downstream analysis steps—is designed with simplicity and accessibility in mind. It can be proficiently followed by scientists with foundational training in cell culture and general molecular biology techniques. While the initial stages of the protocol, including cell isolation and basic sequencing preparation, are straightforward, more advanced steps such as CITE-seq and SMART-seq2 may require a deeper understanding of single-cell and RNA-seq methodologies. Sequencing, in particular, might necessitate collaboration with specialized core facilities, given the need for advanced equipment such as an Illumina (or similar) sequencer. Due to the large amount of output data, preprocessing, including sequence alignment to the genome, should occur on a cluster platform that provides a greater amount of RAM than standard laptops. Cloud-based solutions are also available for use and are recommended for some computationally demanding tasks. For the computational protocol, data analysis has been streamlined for execution on a standard consumer-grade personal computer. This approach makes the computational aspect of the protocol accessible to a broad range of users, although a basic understanding of bioinformatics will be beneficial. The overall design of this protocol ensures that while it is approachable for those at the early stages of their scientific career, it also provides the depth and comprehensiveness required for advanced research applications.

Materials

Reagents

Common reagents

- Cell Staining Buffer (BioLegend, cat. no. 420201)
- Human TruStain FcX Fc Blocking reagent (BioLegend, cat. no. 422302)
- AMPure XP (SPRI Beads for DNA) (Beckman Coulter, cat. no. A63881)
- Qubit dsDNA HS Assay Kit (Thermo Fisher, cat. no. Q32854)
- UltraPure Water (Thermo Fisher, cat. no. 10977023)
- KAPA HiFi Hotstart Readymix (Roche, cat. no. KK-2602)
- High Sensitivity D5000 Reagents (Agilent, cat. no. 5067-5593)
- 200 proof ethanol (Decon Labs, cat. no. 2716)
- Trypan blue (Thermo Fisher, cat. no. 15250061)
- 2-Mercaptoethanol, Molecular Biology Grade (Sigma, cat. no. 444203-250ML)
  ▲ **CAUTION** 2-Mercaptoethanol is highly corrosive and toxic, and poses a health and environmental hazard. Avoid skin contact or inhalation. Wear full personal protective equipment when handling and always use within a chemical hood. Apply other safety measures if necessary and adhere to local guidelines when disposing.
- MgCl2 (Sigma, cat. no. 63069-100ML)
- dNTP (10mM) (New England Biolabs, cat. no. N0447L)
- SuperaseIN RNAse Inhibitor (Thermo Fisher, cat. no. AM2694)
- Maxima H-RT and 5× buffer (Thermo Fisher, cat. no. EPO0753)
- Nextera XT library prep kit (Illumina, cat. no. FC-131-1096)
- Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples) (Illumina, cat. no. 20091654)
  ▲ **CRITICAL** Any of the Illumina DNA/RNA UD Indexes set are compatible.
- RPMI-1640 (Gibco, cat. no. 11875093)
- L-glutamine (Corning, cat. no. 25-005-CI)
- Fetal bovine serum (Thermo Fisher, cat. no. A3840002)
- Penicillin–streptomycin (Corning, cat. no. 30-001-CI)
- HEPES buffer solution (Corning, cat. no. 25-060-CI)

Seq-Well S^3-specific reagents

- Sodium carbonate (Sigma, cat. no. 223530-500G)
- Bovine serum albumin (Sigma, cat. no. A9418-100G)
- 10% Sarkosyl (Sigma, cat. no. L7414)
- 5 M NaCl (Thermo Fisher Scientific, cat. no. 24740011)
- 1× PBS (Thermo Fisher Scientific, cat. no. 10010023)
- 1 M Tris–HCl pH 8.0 (Thermo Fisher Scientific, cat. no. 15568025)
- 8% (wt/vol) PEG8000 (Sigma, cat. no. 83271-500ML-F)
- 10 mM Tris pH 8.0 + 1 mM EDTA (Sigma, cat. no. 93283)
- Tween-20 (Sigma, cat. no. P9416-50ML)
- 0.5 M EDTA (Thermo Fisher Scientific, cat. no. 15575020)
- 0.5% (wt/vol) SDS (Thermo Fisher Scientific, cat. no. 24730020)
- Guanidine thiocyanate (Sigma, cat. no. AM9422)
- Klenow Exo- (New England BioLabs, cat. no. M0212S)
- Exonuclease I (*Escherichia coli*) with buffer (New England BioLabs, cat. no. M0293)

Surface protein staining (ADT)-specific reagents

- TotalSeq-A Human Universal Cocktail, V1.0 (BioLegend, cat. no. 399907)

Antibody cell hashing (HTO) specific reagents

- TotalSeq-A0251 Antibody anti-human Hashtag 1 (BioLegend, cat. no. 394601)
- TotalSeq-A0252 Antibody anti-human Hashtag 2 (BioLegend, cat. no. 394603)
- TotalSeq-A0253 Antibody anti-human Hashtag 3 (BioLegend, cat. no. 394605)
- TotalSeq-A0254 Antibody anti-human Hashtag 4 (BioLegend, cat. no. 394607)
- TotalSeq-A0255 Antibody anti-human Hashtag 5 (BioLegend, cat. no. 394609)
- TotalSeq-A0256 Antibody anti-human Hashtag 6 (BioLegend, cat. no. 394611)
- TotalSeq-A0257 Antibody anti-human Hashtag 7 (BioLegend, cat. no. 394613)
- TotalSeq-A0258 Antibody anti-human Hashtag 8 (BioLegend, cat. no. 394615)
- TotalSeq-A0259 Antibody anti-human Hashtag 9 (BioLegend, cat. no. 394617)
- TotalSeq-A0260 Antibody anti-human Hashtag 10 (BioLegend, cat. no. 394619)
- TotalSeq-A0296 Antibody anti-human Hashtag 11 (BioLegend, cat. no. 394651)
- TotalSeq-A0262 Antibody anti-human Hashtag 12 (BioLegend cat. no. 394623)
- TotalSeq-A0263 Antibody anti-human Hashtag 13 (BioLegend, cat. no. 394625)
- TotalSeq-A0264 Antibody anti-human Hashtag 14 (BioLegend, cat. no. 394627)
- TotalSeq-A0265 Antibody anti-human Hashtag 15 (BioLegend, cat. no. 394629)
- TotalSeq-A0276 Antibody anti-human Hashtag 16 (BioLegend, cat. no. 394681)
- TotalSeq-A0277 Antibody anti-human Hashtag 17 (BioLegend, cat. no. 394653)
- TotalSeq-A0278 Antibody anti-human Hashtag 18 (BioLegend, cat. no. 394655)
- TotalSeq-A0279 Antibody anti-human Hashtag 19 (BioLegend cat. no. 328909)
- TotalSeq-A0280 Antibody anti-human Hashtag 20 (BioLegend, cat. no. 328911)
- TotalSeq-A0281 Antibody anti-human Hashtag 21 (BioLegend, cat. no. 328901)
- TotalSeq-A0282 Antibody anti-human Hashtag 22 (BioLegend, cat. no. 328903)
- TotalSeq-A0297 Antibody anti-human Hashtag 23 (BioLegend, cat. no. 328905)
- TotalSeq-A0284 Antibody anti-human Hashtag 24 (BioLegend, cat. no. 328907)

Genetic cell hashing specific reagents

- RNaseZap RNase Decontamination Solution (Thermo Fisher, cat. no. AM9780)
- RNAClean XP (Beckman Coulter, cat. no. A63987)
- Buffer RLT Plus (Qiagen, cat. no. 1053393)
- Betaine (Sigma, cat. no. B0300-5VL)

Common oligonucleotides

- dN-SMRT, 34-nt DNA oligonucleotide
  5′-AAG CAG TGG TAT CAA CGC AGA GTG ANN NGG NNN B-3′,
  where N can be any base (A, T, C, G) and B can be C, G or T

- SMART PCR Primer for library amplification, 30-nt custom DNA oligonucleotide with 3 riboguanines at the 3′ end (IDT, custom DNA oligonucleotide):
  5′-AAG CAG TGG TAT CAA CGC AGA GTG AATrGrGrG-3′
- P5-SMART Hybrid Oligo, 66-nt custom DNA oligonucleotide for single primer cDNA amplification with the asterisk indicating a phosphodiester bond (IDT, custom DNA oligonucleotide):
  5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG CCT GTC CGC G-G AAG CAG TGG TAT CAA CGC AGA GT*A *C-3′
- Custom read 1, 37-nt custom DNA oligonucleotide that allows the sequencing of the template, 12-bp cell barcode and 8bp UMI as the first read on Illumina sequencers (IDT, custom DNA oligonucleotide):
  5′-GCC TGT CCG CGG AAG CAG TGG TAT CAA CGC AGA GTA C-3′

Surface protein staining (ADT)-specific oligonucleotides

- ADT additive primer, 20-nt custom DNA oligonucleotide that binds to the Universal Cocktail antibody flanking sequences and Small RPIx index primer to permit downstream computational identification of cell surface proteins (IDT, custom DNA oligonucleotide with PAGE purification):
  5′-CCT TGG CAC CCG AGA ATT *C*C-3′
- TruSeq Small RNA RPIx, 66-nt custom DNA oligonucleotide that binds to the ADT additive and Illumina Nextera N7xx primers. This provides each surface protein library with a unique identifier for computational demultiplexing following sequencing on an Illumina sequencer (IDT, custom DNA oligonucleotide with PAGE purification):
  5′-CAA GCA GAA GAC GGC ATA CGA GAT [unique 6-nt index] GTG ACT GGA GTT CCT TGG CAC CCG AGA ATT C*C*A-3′

Antibody cell hashing (HTO) specific oligonucleotides

- HTO additive primer v2, 34-nt custom DNA oligonucleotide that binds to the hashing antibody flanking sequence and D7xx index primer to permit downstream computational demultiplexing of unique hashes (IDT, custom DNA oligonucleotide with PAGE purification):
  5′-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT*C* T-3′
- TruSeq D7xx_Long, 66-nt custom DNA oligonucleotide that binds to the HTO additive and Illumina Nextera N7xx primers. This provides each hashing library with a unique identifier for computational demultiplexing following sequencing on an Illumina sequencer (IDT, custom DNA oligonucleotide with PAGE purification):
  5′-CAA GCA GAA GAC GGC ATA CGA GAT [unique 8-nt index] G TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA T*C*T-3′

SMART-seq2 specific oligonucleotides

- Oligo dT primer, 30-nt poly-dT stretch and a 25-nt universal 5′ anchor sequence DNA oligonucleotide with a 5′ biotin modification. This primer is responsible for the priming of the reverse transcription reaction on polyadenylated RNA sequences (IDT, custom DNA oligonucleotide with RNase-free HPLC purification and 5′ biotin modification):
  5′-AAG CAG TGG TAT CAA CGC AGA GTA CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TVN-3′, where N can be any base (A, T, C, G) and V can be A, C or G
- Genetic template switching oligo, 30-nt DNA oligonucleotide with three riboguanines at the 3′ end (rGrGrG). Begins the process of template switching in the newly synthesized DNA sequences (IDT, custom DNA oligonucleotide with RNase-free HPLC purification):
  5′-AAG CAG TGG TAT CAA CGC AGA GTA CAT rGrGrG-3′
- ISPCR, 23-nt DNA oligonucleotide that will bind the 5′ end of the newly synthesized single strand template DNA to create full cDNA amplification using PCR (IDT, custom DNA oligonucleotide):
  5′-AAG CAG TGG TAT CAA CGC AGA GT-3′

Equipment

Common equipment

- 2100 Bioanalyzer instrument (Agilent, cat. no. G2939BA)
  Optional: if a Bioanalyzer is not available to you, you may substitute this equipment with any gel electrophoresis set-up with the appropriate ladder in order to measure the fragment size of DNA
- 96-well plates, clear (Thermo Fisher, cat. no. AB2396)
- PCR clear adhesive seals (Thermo Fisher, cat. no. AB0558)
- Aluminum foil seals (Biorad, cat. no. MSF1001)
- Individually wrapped 25 mL reservoirs (VWR, cat. no. 89094-664)
- Loading tips (Agilent, cat. no. 5067-5599)
- High Sensitivity D5000 ScreenTape (Agilent, cat. no. 5067-5592)
- Qubit fluorometer (Thermo Fisher, cat. no. Q32854)
- Qubit tubes (Thermo Fisher, cat. no. Q32856)
- Twelve-channel, 200 µL pipette, L12-200XLS+ (Rainin, cat. no. 17013810)
- Twelve-channel, 20 µL pipette, L12-20XLS+ (Rainin, cat. no. 17013808)
- Pipette Tips RT LTS 200µL FL 960A/10 (Rainin, cat. no. 30389240)
- Pipette Tips RT LTS 20µL 960A/10 (Rainin, cat. no. 30389200)
- 1.5 mL DNA LoBind tubes, PCR clean microcentrifuge tubes (Eppendorf, cat. no. 22431021)
  ▲ **CRITICAL** Use of Lo-Bind microcentrifuge tubes is necessary as non-LoBind tubes can result in a substantial loss of product throughout the protocol because of retention of DNA in the non-LoBind plastic.
- 2 mL DNA LoBind tubes, PCR clean microcentrifuge tubes (Eppendorf, cat. no. 22431048)
- 15 mL conical tubes (VWR, cat. no. 76176-950)
- 50 mL conical tubes (VWR, cat. no. 21008-951)
- Vortex (Fisher Scientific, cat. no. 02-215-414)
- Mini tabletop centrifuge (Genesee Scientific, cat. no. 31-500)
- Centrifuge (Thermo Scientific, cat. no. 75009710)
- 4 °C fridge
- −40 °C freezer
- −80 °C freezer
- Refrigerator
- Neubauer improved hemacytometers (VWR, cat. no. 82030-468)
- Light microscope (Leica DM1000)
- 500 mL Vacuum Filter Unit with 0.2µm Nylon Membrane (Corning, cat. no. 430773)
- Bead bath that is able to reach 37 °C
- DynaMag-96 Side Skirted Magnet (Invitrogen, cat. no. 12027)
- DynaMag-2 magnet (Invitrogen, cat. no. 12321D)
- Thermocycler, 96-well plate compatible
- HulaMixer sample mixer (Thermo Fisher, cat. no. 15920D)
- Serological pipette controller (INTEGRA, cat. no. 155 000)
- Tube racks

Seq-Well-specific equipment

- Bead loaded Seq-Well arrays
- Functionalized membranes for Seq-Well array sealing
- Glass slides, precleaned (Thermo Fisher, cat. no. 10010049)
- Plasma cleaner (Harrick Plasma, cat. no. PDC-001-HP)
- Pump for plasma cleaner (Harrick Plasma, cat. no. PDC-VPE)
- Agilent Clamps (Agilent Technologies, cat. no. G2534A)
- Sterile disposable serological pipets, plugged, polystyrene, 10 mL (VWR, cat. no.89130-898)
- Sterile disposable serological pipets, plugged, polystyrene, 25 mL (VWR, cat. no.89130-900)
- Fine point high-precision forceps (Fisher Scientific, cat. no. 22-327379)
- Incubator with a range of 37–52 °C
- Vacuum aspirator system

Software

- STAR v2.5.4b (https://github.com/alexdobin/STAR)
- GATK 4 or later (https://gatk.broadinstitute.org)
- BWA 0.7.15-r1140 (http://www.htslib.org)
- Picard 2.16.0-SNAPSHOT (https://broadinstitute.github.io/picard/)
- Samtools 1.3.1 (using htslib 1.3.1) (https://sourceforge.net/projects/samtools/files/samtools/1.3.1/)
- Python 2.7

Data

▲ **CRITICAL** The following data can be obtained from the Broad Institute public references (https://console.cloud.google.com/storage/browser/gcp-public-data--broad-references/hg38/v0) before data analysis.

- Primary assembly of human reference genome (GRCh38) in FASTA format (Homo_sapiens_assembly38.fasta)
- Reference FASTA index for human reference genome (GRCh38) index (Homo_sapiens_assembly38.fasta.fai)
- BWA human reference dictionary (Homo_sapiens_assembly38.dict)
- Known indel sites in VCF format (Mills_and_1000G_gold_standard.indels.hg38.vcf.gz and Homo_sapiens_assembly38.known_indels.vcf.gz)
- Index of known indel sites VCF (Mills_and_1000G_gold_standard.indels.hg38.vcf.gz.tbi and Homo_sapiens_assembly38.known_indels.vcf.gz.tbi)
- BWA reference files:

```
ref_dict: Homo_sapiens_assembly38.fasta.64.dict
ref_amb: Homo_sapiens_assembly38.fasta.64.amb
ref_bwt: Homo_sapiens_assembly38.fasta.64.bwt
ref_pac: Homo_sapiens_assembly38.fasta.64.pac
red_sa: Homo_sapiens_assembly38.fasta.64.sa
```

- Gene annotation file for the primary assembly of the human reference genome in GTF format (https://www.gencodegenes.org/human)
- dbSNP data of known polymorphic sites in the GRCh38 build of the human reference genome, in VCF format (Homo_sapiens_assembly38.dbsnp138.vcf)
  Index file for dbSNP data (Homo_sapiens_assembly38.dbsnp138.vcf.idx)

Reagent setup

Seq-Well S^3 preparation

This Protocol assumes you have obtained functionalized Seq-Well arrays and membranes for sealing. Before the start of the procedure, Seq-Well arrays must have undergone bead loading as described in the Seq-Well S^3 protocol[24,25]. It is recommended to do these 1 d to 1 week before beginning the experiment.

▲ **CRITICAL** If you bead load arrays just before cell loading, cover arrays with bead loading buffer. If prepared ahead of time, arrays should be covered with 5 mL of array quenching buffer. Arrays can sit in array quenching buffer for up to 10 d.

▲ **CRITICAL** It is highly recommended to map out how samples will be hashed before beginning. A unique hashtag must be used for each sample that will be loaded onto a shared Seq-Well S^3 array.

Complete RP10 medium

- 10% (vol/vol) fetal bovine serum
- 1% (vol/vol) L-glutamine
- 1% (vol/vol) penicillin–streptomycin
- 1% (vol/vol) HEPES buffer

Filter the complete RP10 medium through the 500 mL vacuum filter unit. Place the filtered medium in a 37 °C-bead bath before proceeding to the protocol. After use, the medium can be stored at 2–8 °C for up to 3 weeks.

Complete RLT buffer

▲ **CRITICAL** Must be made fresh each time

- 1% (vol/vol) 2-mercaptoethanol
- 99% vol/vol buffer RLT Plus

Prelysis buffer

- 5 M guanidine thiocyanate (Sigma, cat. no. AM9422)
- 1 mM EDTA

  ▲ **CRITICAL** Guanidine thiocyanate will take ~2 h to go into solution, it is recommended to prepare this solution at least 1 d before use.

  ▲ **CRITICAL** Prelysis buffer is photosensitive. It is recommended to store the prelysis buffer in a clear glass bottle, wrapped in aluminum foil to protect it from light.
  In these conditions, the prelysis buffer can be stored at room temperature (15–25 °C) for ~6 months.

Bead loading buffer (50 mL)

- 2.5 mL 2 M sodium carbonate
- 42.5 mL UltraPure water
- 5 mL bovine serum albumin (100 mg/mL)
- Titrate with glacial acetic acid to reach pH 10

  ▲ **CRITICAL** Prepare this fresh, just before use.

Array quenching buffer (50 mL)

- 2.5 mL 2 M sodium carbonate
- 500 µL 1 M Tris–HCl
- 47 mL UltraPure Water

  Array quenching buffer can be made in bulk and stored at room temperature (15–25 °C) for ~6 months.

Complete lysis buffer (50 mL)

- 47.25 mL prelysis buffer
- 250 µL 10% (vol/vol) Sarkosyl
- 500 µL 2-mercaptoethanol

  ▲ **CRITICAL** Prepare this fresh, just before use.

Hybridization buffer (50 mL)

- 20 mL 5 M NaCl
- 26 mL of PBS
- 4 mL 8% (wt/vol) PEG8000

  Hybridization buffer can be made in bulk and stored at room temperature (15–25 °C) for up to 1 year.

Wash buffer (50 mL)

- 20 mL 5 M NaCl
- 150 µL 1 M MgCl2
- 1 mL 1 M Tris–HCl pH 8.0
- 24.85 mL UltraPure water
- 4 mL 8% vol/vol PEG8000

  You will need a minimum of 50 mL of wash buffer per array. Wash buffer can be made in bulk and stored at room temperature (15–25 °C) for up to 1 year.

TE–TW buffer

- 49.5 mL UltraPure Water
- 500 µL 1 M Tris–HCl pH 8.0
- 100 µL 0.5 M EDTA
- 50 µL Tween-20

  TE with Tween (TE–TW) can be made in bulk and stored at room temperature (15–25 °C) for up to 1 year.

TE–SDS

- 49.5 mL UltraPure water
- 500 µL 1 M Tris–HCl pH 8.0
- 100 µL 0.5 M EDTA
- 250 µL 10% (wt/vol) SDS

  TE–SDS can be made in bulk and stored at room temperature (15–25 °C) for up to 1 year.

Procedure 1: sample preparation and sequencing

Sample preparation

▲ **CRITICAL** If you plan to perform both surface protein staining and antibody hashing, best results will be obtained with a starting input of at least 200,000 cells for each unique sample.

▲ **CRITICAL** This procedure can be done in 5 mL polystyrene Falcon tubes or in a 96 well, round (U)-bottom, tissue cultured surface treated plate (such as Thermo Fisher, cat no. 163320). If using tubes, add 2 mL of stain buffer instead of 200 µL in Step 11.

Sample preparation: cell counting

● **TIMING 30 min**

▲ **CRITICAL** Ensure single-cell suspension media does not contain DNase I. This will interfere with the effectiveness of the antibody activity.

1. Prepare a single-cell suspension. The specifics of this process vary greatly depending on sample type and, therefore, are outside the scope of this Protocol. If your sample contains red blood cell contamination, red cell lysis is required.
2. Place 10 µL of cells in 100 µL of complete RLT lysis buffer for each sample. Add 10 µL per well of this mixture to 10 wells of a new 96-well plate and immediately put the plate into a −80 °C freezer. This lysate will be used to perform genetic hashing later. It is recommended to add samples into two separate 96-well dishes to be frozen separately.
3. Manually count the number of cells for each sample prepared in Step 1 by combining 10 µL of cells and 10 µL of Trypan blue in the well of a 96-well plate. For pipette mix, add 10 µL to a Neubauer improved hemacytometer. Count the cells under a light microscope.

(Optional) Sample preparation: stain with hashing antibodies

● **TIMING 55 min**

4. Dilute 1 million cells in 45 µL of Cell Staining Buffer in the well of a 96-well plate. If you have less than 1 million cells, centrifuge at 350–500*g* for 5 min, 4 °C. Remove the lysate, then resuspend all cells in 45 µL of Cell Staining Buffer. If you have less than 200,000 cells, we recommend multiplexing of samples via genetic multiplexing exclusively.
5. Add 5 µL of Human TruStain FcX Fc Blocking reagent to the cells from Step 4. The final blocking volume should now be 50 µL.
6. Incubate for 10 min at 4 °C.
7. While the cells incubate, to avoid potential antibody aggregates centrifuge the antibody hashes in a tabletop centrifuge at max speed (≥800*g*) for 1 min, room temperature.
8. Mix 49 µL Cell Staining Buffer + 1 µL of a unique antibody hash to create a 50 µL antibody pool.

   ▲ **CRITICAL STEP** A unique antibody hash is needed to create the antibody pool for each sample.

9. Carefully add 50 µL of antibody pool to the 50 µL blocked cell suspension. Then incubate for 30 min at 4 °C.
10. After incubation, add 100 µL of Cell Staining Buffer to the cell suspension, then centrifuge at 350–500*g* for 5 min, 4°C.
11. Remove the supernatant, resuspend the cell suspension in 200 µL of Cell Staining Buffer then centrifuge at 350–500*g* for 5 min, 4 °C.
12. Repeat Step 11 two more times two additional times for a total of three washes.
13. Remove the supernatant, then resuspend in 200 µL of complete RP10 media.
14. Manually count the cells for each sample as outlined in Step 3.
    ◆ **TROUBLESHOOTING**
15. Pool an equal number of cells from each sample that will share a Seq-Well array. If the volume of the pool is less than 50 µL, add Cell Staining Buffer to achieve a final volume of 50 µL. If the pool volume exceeds 50 µL, centrifuge the pool at 350–500*g* for 5 min, 4 °C. Remove the lysate and resuspend in 50 µL of Cell Staining Buffer. Proceed directly to cell surface protein staining.

(Optional) Prepare samples: stain cell surface protein antibodies

● **TIMING 55 min**

16. Equilibrate the lyophilized vials of TotalSeq-A Human Universal Cocktails for 5 min, allowing them to reach room temperature.
17. Centrifuge the vials at 10,000*g* for 30 s, room temperature.
18. Resuspend each vial in 50 µL of Cell Staining Buffer. Replace the cap and vortex for 10 s.
19. Incubate for 5 min at room temperature.
20. Vortex each vial, then centrifuge 14,000*g* at 2–8 °C for 10 min.
    ▲ **CRITICAL STEP** This step ensures removal of protein aggregates
21. Carefully transfer the entire contents of the 50 µL vial, avoiding the bottom of the tube, to the 50 µL cell pool from Step 15, then incubate for 30 min at 4 °C.
    ▲ **CRITICAL STEP** Ensure the bottom of the tube is not touched. If unsure, centrifuge the vial again before adding.
22. After incubation, add 100 µL of Cell Staining Buffer to the cell suspension, then centrifuge at 350–500*g* for 5 min, 4 °C.
23. Remove the supernatant, resuspend the cell suspension in 200 µL of Cell Staining Buffer then centrifuge at 350–500*g* for 5 min, 4 °C.
24. Repeat Step 23 two additional times for a total of three washes.
25. Remove the supernatant, then resuspend in 220 µL complete RP10 media.
26. Manually count the cells for each sample as outlined in Step 3.
    ◆ **TROUBLESHOOTING**
27. Place the cells on ice until you are ready to load the cells onto the Seq-Well array.
28. Adjust to your desired cell concentration for Seq-Well, then proceed with Cell Loading as described in the Seq-Well S^3 protocol.
    ▲ **CRITICAL STEP** Typically, 15,000 cells in 200 µL of complete RP10 are loaded onto each Seq-Well array. When cell hashing, you may opt to 'overload' the array by adding 30,000 cells in 200 µL of complete RP10 onto a Seq-Well array.

Seq-Well S^3 processing

▲ **CRITICAL** This part uses the stained cell suspension retrieved in Step 28. The following steps are described in more detail in the Seq-Well S^3 protocol (refs. 24,25), specifically beginning with the sections labeled 'Cell Loading' ending just before the whole transcriptome amplification (WTA) PCR step. It is crucial to familiarize oneself with the Seq-Well protocol and techniques before starting this part. Here, we provide a brief recapitulation of the standard Seq-Well protocol steps that are unchanged by the addition of our staining protocols.

Seq-Well S^3 processing: cell loading

● **TIMING 30 min**

29. Obtain your bead loaded Seq-Well array and functionalized membrane. These should be previously prepared and stored in four-well dishes ('Reagent setup' section in Materials).
30. Aspirate the buffer surrounding your Seq-Well array.
31. Add 5 mL of complete RP10 media to your array. Allow to rest for 1 min.
32. Aspirate, then add 5 mL of fresh complete RP10 media.
33. Incubate for 5–10 min at room-temperature.
34. Aspirate the media, ensuring the array will not move.
35. Use a 1,000 µL pipette tip to center the array, ensuring no side of the array is touching the four-well dish.
36. Add 200 µL of single cell suspension from Step 28 to the array in a drop-wise fashion. Then, incubate for 10 min at room temperature.
    - Place the array on a rocker at 50 rpm to facilitate movement of the cell suspension
    - Intermittently, gently rock in a four-well dish in the *x* and *y* direction to facilitate movement of the cell suspension

    ▲ **CRITICAL STEP** If the array touches the edges of the four-well dish, capillary action will force the cells away from the array surface, causing cell loss. We recommend keeping the array in the center of the dish.
37. Wash the array with 5 mL of 1× PBS by adding buffer gently to the corner of each well, gently rock in the *x* and *y* direction, then aspirate the 1× PBS. Repeat this four times. After the final aspiration, add 5 mL of media that does not contain serum (RPMI or PBS have been successfully used).

    ▲ **CRITICAL STEP** Any media that is compatible with your cells can be used here. Ensure final media does not have serum. Excess serum on the array surface can impede successful membrane attachment.

Seq-Well S^3 processing: membrane sealing

● **TIMING 5 min**

▲ **CRITICAL** This process is made easier if performed over a dark colored surface.

38. Remove and dab dry one array and one membrane with a Kim wipe by first removing a single array from the four-well dish and placing it on a clean surface (such as the lid of the four-well dish). Carefully remove the excess moisture from the slide portion of the array (avoiding the PDMS surface). Then remove a functionalized membrane from the four-well dish and hold the sides of the slide in hand with membrane facing up. Remove excess moisture surrounding the membrane with a Kim wipe without touching the membrane itself.
39. Carefully position the membrane in the center of a microscope slide and create a small membrane overhang (2–3 mm) on one of the short ends of the microscope slide.
40. Obtain a clean microscope slide in the hand not holding the membrane. Invert the membrane slide so the treated surface of the membrane faces the top of your array. Position the overhang just slightly above the start of the array and apply firm pressure on the overhang using the clean microscope slide.
41. While maintaining pressure with the clean microscope slide, gently apply the membrane using only the weight of the slide.
42. Transfer the sealed array to an open clamp and place your clean microscope side on top.
43. Close the clamp and tighten to the point of resistance, then place in a 37 °C incubator for 40 min.

Seq-Well S^3 processing: lysis and hybridization

● **TIMING 90 min**

44. Remove the clamp from the incubator, open the clamp, remove the array and submerge it in 5 mL of complete lysis buffer with the top slide still attached ('Reagent setup' section in Materials)

45. Once top slide has detached, remove top slide from dish and place arrays, still held in the four-well dish, on a horizontal rotator for 20 min at 50–60 rpm at room temperature (15–25 °C).
46. Aspirate the lysis buffer and wash with 5 mL of hybridization buffer.
    ▲ **CAUTION** Use a separate waste container for lysis buffer because guanidine thiocyanate can react with bleach in tissue culture traps creating cyanide gas. Hybridization buffer washes may also contain trace lysis buffer and should also use a separate waste container.
47. Aspirate the hybridization buffer, add 5 mL of fresh hybridization buffer and rotate for 40 min at 50–60 rpm.

Seq-Well S^3 processing: bead recovery

● **TIMING 30 min**

48. Remove array from hybridization buffer, carefully pull away the membrane using fine-tipped tweezers and place array in a 50 mL conical tube which contains 30–40 mL of wash buffer ('Reagent setup' section in Materials).
49. Rest the array on the edge of the 50 mL conical and vigorously dispense 1 mL wash buffer over the array surface to dislodge the beads.
50. Use a clean microscope slide to gently scrape the array surface to remove remaining beads.
51. Place the empty array back in the four-well dish, cap the 50 mL conical and pellet beads for 5 min at 1,000*g*, room temperature.
52. Observe the array under a microscope to ensure adequate removal of beads. Then dispose of the array.
53. Aspirate all but ~1 mL of excess wash buffer. Collect the beads using a 1,000 µL pipette and transfer to a new 1.5 mL microcentrifuge tube.

Seq-Well S^3 processing: reverse transcription

● **TIMING 2.5 h, option to run overnight**

54. Prepare the reverse transcription mix:

Reagent name	Volume (1 reaction)
UltraPure Water	40 µL
Maxima 5× RT Buffer	40 µL
30% PEG8K	80 µL
10 mM dNTPs	20 µL
RNase inhibitor	5 µL
Template switch oligo (100 µM)	5 µL
Maxima RNase H-minus RT	10 µL

55. Centrifuge microcentrifuge tube from Step 53 for 1 min at 800*g*, room temperature.
56. Remove supernatant and resuspend in 250 µL of 1× Maxima RT Buffer.
57. Centrifuge microcentrifuge tube for 1 min at 800*g*, room temperature.
58. Aspirate supernatant, then resuspend beads in 200 µL of reverse transcription mix.
59. Incubate for 30 min at room temperature followed by 90 min at 52 °C (or overnight, preferred), both with end-over-end rotation.
    ■ **PAUSE POINT** Can be stored at 4 °C for up to 2 weeks in TE–TW ('Reagent setup' section in Materials)

Seq-Well S^3 processing: exonuclease I treatment

● **TIMING 90 min**

60. Wash beads once with 500 µL of TE–TW ('Reagent setup' section in Materials), once with 500 µL of TE–SDS ('Reagent setup' section in Materials) and twice with 500 µL of TE–TW.
61. Aspirate TE–TW, then wash with 500 µL of 10 mM Tris–HCl.

62. Aspirate and resuspend beads with 200 µL of exonuclease I mix.

Reagent name	Volume (1 reaction)
10× ExoI Buffer	20 µL
UltraPure Water (Thermo Fisher 10977023)	170 µL
ExoI	10 µL

63. Incubate for 1 h at 37 °C with end-over-end rotation.

Seq-Well S^3 processing: second strand synthesis (S^3)

● **TIMING 90 min**

64. Centrifuge microcentrifuge tube for 1 min at 800*g*, room temperature and discard supernatant.
65. Prepare 550 µL of 0.1 N NaOH from a >1 N stock solution for each array.
66. Add 500 µL 0.1 N NaOH to the tube from Step 64 and incubate at room temperature for 5 min with end-over-end rotation.
67. Centrifuge microcentrifuge tube for 1 min at 800*g*, room temperature and discard supernatant.
68. Wash samples once with 500 µL of TE–TW, once with 500 µL of TE, then resuspend with 200 µL of second strand synthesis mix.

Reagent name	Volume (1 reaction)
Maxima 5× RT Buffer	40 µL
30% PEG8K	80 µL
10 mM dNTPs	20 µL
dN-SMRT oligo (1mM)	2 µL
Klenow enzyme	5 µL
UltraPure water	53 µL

69. Incubate for 1 h at 37 °C with end-over-end rotation.
70. Wash beads twice with 500 µL of TE–TW, then once with 500 µL of TE.

Seq-Well S^3 processing: WTA

● **TIMING 2 h**

71. Prepare the WTA amplification reaction mix as outlined below (the columns outline different compositions depending on whether only HTO, ADT or HTO and ADT oligos were used). For each array you will need to prepare 24 reactions.

Reagent name	HTO (1 reaction)	ADT (1 reaction)	HTO + ADT (1 reaction)
KAPA HiFi Hotstart Readymix (Roche KK-2602)	25 µL	25 µL	25 µL
Water	14.4 µL	14.4 µL	14.2 µL
ISPCR (100 µM)	0.4 µL	0.4 µL	0.4 µL
HTO additive primer (10 µM)	0.2 µL	0 µL	0.2 µL
ADT additive primer (10 µM)	0 µL	0.2 µL	0.2 µL

72. Wash beads once with 500 µL of water, pellet beads, remove supernatant and resuspend in 240 µL of water.
73. Add 40 µL of WTA amplification reaction mix per reaction to a 96-well plate.
74. Pipette mix the bead solution, then add 10 µL to the 40 µL of WTA amplification reaction mix.
    ▲ **CRITICAL STEP** You must pipette mix the bead mixture before adding to PCR mastermix each time. This means for each array; you will pipette mix the bead mixture 24 times.
75. Use the WTA thermocycler program to perform whole-transcriptome amplification.

Temperature	Time	Number of cycles
95 °C	5 min	1 cycle
98 °C	20 s	4 cycles
65 °C	45 s	
72 °C	3 min	
98 °C	20 s	9–12 cycles (9 cycles optimal for cell line or larger cells like macrophages; 12 cycles optimal for primary cells)
67 °C	20 s	
72 °C	3 min	
72 °C	5 min	1 cycle
4 °C	Hold	Infinite hold

■ **PAUSE POINT** Proceed to purification step or store the WTA product at 4 °C for 1–5 d or −20 °C for 5 d up to 1 year.

Seq-Well S^3 processing: separation of short and long WTA products and PCR clean-up

● **TIMING 1 h**

▲ **CRITICAL** Allow 20–30 min for the AMPure XP SPRI beads to equilibrate to room-temperature before use. Ensure you make fresh 80% (v/v) ethanol for this procedure. Please refer to the AMPure XP SPRI bead official protocol[59] for more details.

▲ **CRITICAL** To quickly gauge if the AMPure SPRI beads are ready to use, quickly invert the bottle. If you observe no beads stuck to the bottom, the beads are ready to use. If you see a substantial number of beads stuck to the bottom, incubate another 10 min at room temperature.

76. Separation of short and long WTA products (Step 76 – 80). Remove plate from thermocycler and centrifuge at 800*g* for 1 min, room temperature.
77. Add 30 µL of room-temperature AMPure SPRI to each well containing PCR product. This creates a 0.6× volumetric ratio with the WTA product. Thoroughly pipette mix each well such that all product in the well is a uniform color.
78. Incubate for 3 min at room-temperature, and then place on the DynaMag-96 Side Skirted Magnet for 2 min.
79. Collect the supernatant and place it in a new 96-well plate. These supernatants will contain your short WTA products. At this step, you may condense the supernatant corresponding to a single Seq-Well array into 12 wells. This is best done by transferring the first row of supernatants to a new 96-well plate, following the same positioning from the original plate (one supernatant in each well). Then adding the second row of supernatant to the first row of the new 96-well plate, still following the same positioning. In the new 96-well plate, you should have 12 wells corresponding to a single array.
80. Place the new 96-well plate containing your supernatant on ice, or store at 4 °C.
81. Long WTA product PCR clean-up (Steps 81–92). Add 100 µL per well of fresh 80% (vol/vol) ethanol to the plate that still contains the SPRI beads.
82. Move the 96-well plate to the right, back to the center, then to the left on the DynaMag-96 Side Skirted Magnet. The plate should be moved a minimum of 5 times each to the right and left to ensure adequate bead movement to increase the beads exposure to the ethanol wash.
83. Incubate on the magnet for 30 s at room-temperature then remove the 100 µL of ethanol.
84. Repeat Steps 81 to 83, yielding three ethanol washes in total.
85. After removing the final ethanol wash, wait 1 min. Then use a 20 µL multichannel pipette to remove residual ethanol and allow beads to dry for 5–10 min.
    ▲ **CRITICAL STEP** Beads are dry once they appear matte in color, but before they develop visible cracks.
86. Once the beads are dry, remove the plate from the magnet and add 26 µL of UltraPure water per well to elute dried beads.
87. Incubate for 1 min off the magnet at room-temperature. Then place samples back onto the magnet for 1 min or until supernatant appears clear.

88. Transfer the 25 µL of water to a new 96-well plate. This supernatant will contain your long WTA product. At this step, you may want to condense the supernatants to 12 wells per array as explained in Step 79.
89. Add a volume of SPRI beads to the new 96-well plate that is 0.8× the volume of the eluent in the wells.
90. Repeat Steps 78–85, then elute the dried beads in 16 µL of UltraPure water.
91. Transfer the 15 µL of UltraPure water to a new 96-well plate. At this step, you may condense the eluent corresponding to a single Seq-Well array into 3 wells of the new 96-well plate.
92. Place this new 96-well plate on ice if continuing immediately or store if pausing.
    ■ **PAUSE POINT** Proceed to next step or store the purified WTA product at 4 °C for 1–5 d or −20 °C for 5 d up to 1 year.
93. Short WTA product PCR clean-up (Step 93–109). When ready to proceed, retrieve the 96-well plate containing your supernatants (Step 92).
94. Pool an equal volume of supernatant from each well corresponding to one Seq-Well array, following either option A or B.
    (A) **If performing cell hashing and 'overloading' the Seq-Well array:**
        (i) Pool the WTA reactions into two separate pools, thereby limiting the total number of bead barcodes per pool.
        (ii) If you condensed the supernatants into 12 wells, pool 30 µL from 6 of the 12 wells into a 1.5 mL microcentrifuge tube.
        (iii) If you did not condense, pool 20 µL from 12 of the 24 wells into a 1.5 mL microcentrifuge tube.
    (B) **In the absence of cell hashing and/or overloading:**
        (i) Pool the WTA reactions into a single pool.
        (ii) If you condensed the supernatants into 12 wells, pool 20 µL from each of the 12 wells into a 1.5 mL microcentrifuge tube.
        (iii) If you did not condense, pool 10 µL from each of the 24 wells into a 1.5 mL microcentrifuge tube.
95. Transfer 160 µL of pooled supernatant to a new 1.5 mL microcentrifuge tube.
96. To 160 µL of pooled 0.6× SPRI supernatant, add 140 µL SPRI beads. The total concentration of PEG is now equivalent to a 2× SPRI, enabling the short amplicons to bind to the SPRI beads.
97. Incubate for 3 min at room temperature, and then place the DynaMag-2 Magnet for 2 min.
98. Remove and discard the supernatant.
99. Add 800 µL of fresh 80% (v/v) ethanol to each microcentrifuge tube. Rotate each tube 180° six times to allow beads to pass through the ethanol solution to the opposite side of the tube.
100. Incubate on the magnet for 30 s at room-temperature then remove the 800 µL of ethanol.
101. Repeat Steps 99 and 100 two more times, yielding three bead washes in total.
102. After removing the final ethanol, wait 1 min. Then use a 200 µL multichannel pipette to remove residual ethanol and allow beads to dry for 5–10 min.
    ▲ **CRITICAL STEP** Beads are dry once they appear matte in color, but before they develop visible cracks.
103. Once the beads are dry, remove the plate from the magnet and add 51 µL of UltraPure water to elute dried beads.
104. Incubate for 1 min off the magnet at room-temperature. Then place samples back onto the magnet for 1 min or until supernatant appears clear.
105. Transfer 50 µL of water to a new 96-well plate.
106. Add a volume of SPRI beads to the new 96-well plate that is 2.0× the volume of the eluent in the wells.
107. Repeat Steps 78–85, then elute the dried beads in 16 µL of UltraPure water.
108. Transfer the 15 µL of sample to a new 96-well plate.
109. Place this new 96-well plate on ice or store if pausing.
    ■ **PAUSE POINT** Proceed to quantification of product or store the short WTA product at 4 °C for 1–5 d or −20 °C for 5 d up to 1 year.

Seq-Well S^3 processing: quantification of product

● **TIMING 30 min**

110. Use Qubit to quantify the concentration of both short and long WTA products according to the manufacturer's instructions. Use 1 µL of the cDNA product as input. Expect 1–5 ng/µL DNA for short WTA product. Keep in mind that the target amplicon is a small portion of this mixture.
◆ **TROUBLESHOOTING**

111. Run an Agilent D5000 High Sensitivity Screentape or BioAnalyzer High Sensitivity Chip according to the manufacturer's instructions. Use 1 µL of the WTA product as input.
◆ **TROUBLESHOOTING**
- For long WTA products, expect a distribution between 400 and 1,500 bp with a peak greater than ~900 bp (Fig. 2a,b).
- For the short WTA products, expect a hump centered ~300 bp. This is a complex mixture of short products, including the amplified ADT. Step-out PCR will more selectively amplify the ADT (Fig. 2c,d).

■ **PAUSE POINT** Proceed to single-cell RNA-seq library preparation, or store the cDNA product at 4 °C for 1–5 d or −20 °C for 5 d up to 1 year.

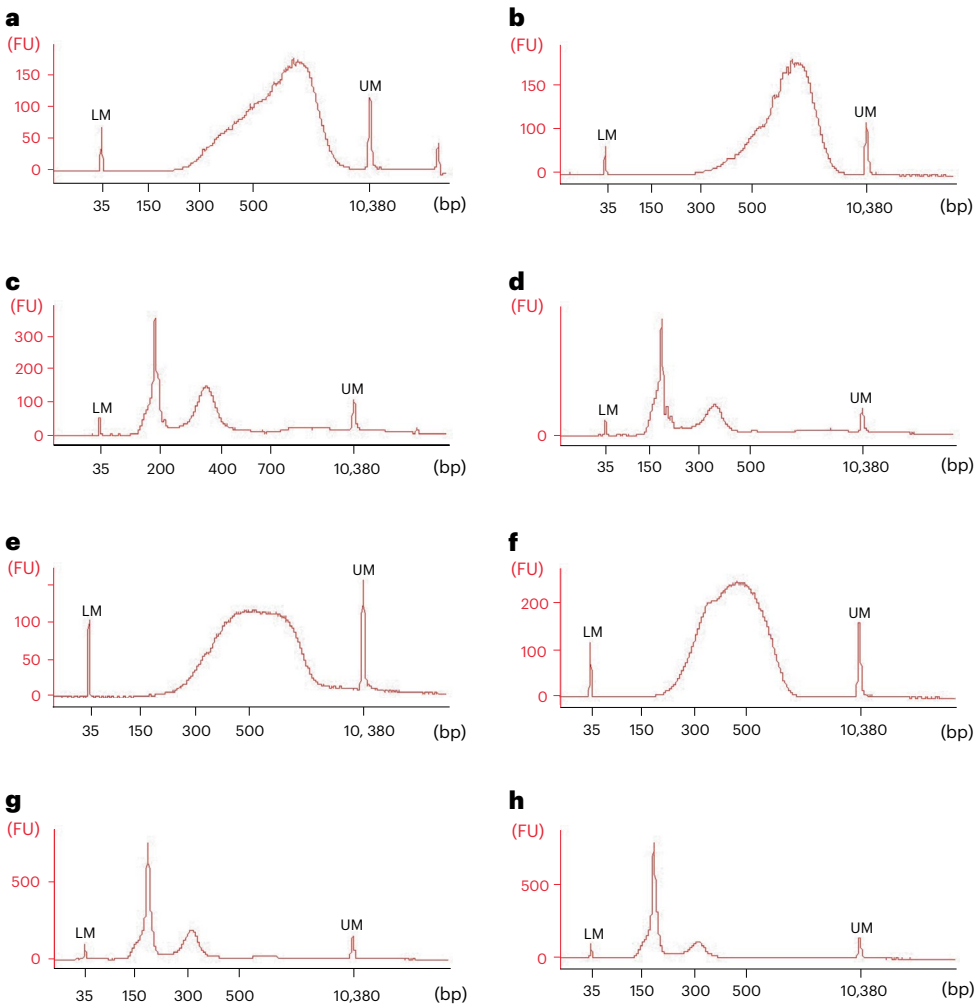

Fig. 2 | Example BioAnalyzer electropherograms of amplified cDNA and final sequencing libraries. a–**d**, cDNA concentration and average fragment size post-WTA PCR clean-up are respectively 746 bp and 5.45 nM for Seq-Well S3 (**a**), 668 bp and 7.24 nM for Smart-seq2 (**b**), 380 bp and 13.21 nM for HTO (**c**) and 358 bp 9.14 nM for ADT (**d**). **e**–**h**, Final library concentration and average fragment size are respectively 678 bp and 8.3 nM for Seq-Well S3 (**e**), 457 bp and 18.49 nM for Smart-seq2 (**f**), 356 bp and 11.87 nM for HTO (**g**) and 342 bp 10.80 nM for ADT (**h**). Traces were obtained post-WTA PCR clean-up for **a**–**d** and post-sequencing library PCR clean-up for **e**–**h**. Lower marker (LM) is visible at 35 bp and upper marker (UM) at 10,380 bp.

Seq-Well S^3 processing: single-cell RNA-seq library preparation

● **TIMING 1.5 h**

▲ **CRITICAL** Purified long fractions of the cDNA should be processed following the remaining steps outlined in the Seq-Well S^3 protocol. In brief, this includes preparing the libraries, performing a final PCR clean up and then preparing the samples for sequencing following manufacturer's instructions. A more detailed version of library preparation protocols can be found in the original Seq-Well S^3 protocol.

112. In a new 96-well plate, dilute your long WTA product from Step 92 such that you have 800 pg of cDNA in 5 µL of water.
113. To each well add 10 µL of Tagment DNA buffer and 5 µL of Amplicon Tagment Mix, pipette mixing up and down 5× after each addition.
114. Seal with a PCR plate seal and centrifuge 1 min at 800*g*, room temperature.
115. Incubate the samples for 5 min at 55 °C, then cool to 10 °C.
▲ **CRITICAL STEP** Ensure your incubator has already reached 55 °C before beginning the incubation step.
116. Place on ice for 30 s, then centrifuge for 30 s at 800*g*, room temperature.
117. Immediately add 5 µL Neutralize Tagment Buffer to each well and pipette mix up and down five times. Incubate for 5 min at room temperature.
Add to each well, in the order listed:

Reagent	Amount
RNAse-free H20	8 µL
10 µM New-P5-SMART PCR hybrid oligo	1 µL
10 µM Nextera N700 oligo	1 µL
Nextera PCR Mix	15 µL

Seal with a PCR plate seal, centrifuge 1 min at 800*g*, room temperature.

118. Place the 96-well plate in the thermocycler and carry out the Seq-Well S^3 Library Amplification PCR program:

Temperature	Time	Number of Cycles
72 °C	5 min	1 cycle
95 °C	30 s	1 cycle
95 °C	10 s	12 cycles
55 °C	30 s	
72 °C	30 s	
72 °C	2 min	1
4 °C	Hold	Hold

119. Perform two 0.6× volumetric SPRI bead purifications.
▲ **CRITICAL STEP** For the remainder of this procedure, we will only focus on processing the short fractions which contain our HTO and ADT products.

Seq-Well S^3 processing: step-out PCR for short fraction of WTA products

● **TIMING 45 min**

120. Dilute short cDNA product from Step 109 to 175 pg/µL using UltraPure water as the diluent. Ensure a minimum of 44 µL of product.
121. In microcentrifuge tubes, create separate HTO and ADT Step-out PCR Mixes following the Step-out PCR Mix table below.

Reagent name	HTO reaction (1 reaction)	ADT reaction (1 reaction)
KAPA HiFi Hotstart Readymix (Roche KK-2602)	25 µL	25 µL
Diluted short WTA product (175 pg/µL)	20 µL	20 µL
P5-SMART Hybrid Oligo (10 µM)	2.5 µL	2.5 µL

Reagent name	HTO reaction (1 reaction)	ADT reaction (1 reaction)
TrueSeq Small RNA RPIx (10 µM)	0 µL	2.5 µL
TrueSeq D70x_Long (10 µM)	2.5 µL	0 µL

▲ **CRITICAL STEP** Each sample that will be pooled into a common sequencing run must have a unique index.

122. Place samples in the thermocycler and run the PCR step-out thermocycler program:

Temperature	Time	Number of cycles
95 °C	5 min	1 cycle
98 °C	20 s	13–15 cycles (14–15 cycles for ADT) (13–15 cycles for HTO)
60 °C	45 s	
70 °C	1 min	
72 °C	2 min	1
4 °C	Hold	Hold

Seq-Well S^3 processing: sequencing library PCR clean-up

● **TIMING** 1 h

▲ **CRITICAL** Allow 20–30 min for the AMPure SPRI beads to equilibrate to room temperature before use. Ensure you make fresh 80% (vol/vol) ethanol for this procedure. Please refer to the AMPure SPRI bead official protocol for more details.

▲ **CRITICAL** To quickly gauge if the AMPure SPRI beads are ready to use, quickly invert the bottle. If you observe no beads stuck to the bottom, the beads are ready to use. If you see a lot of beads stuck to the bottom, incubate another 10 min at room temperature.

123. Place 50 µL of each library into a new 1.5 mL microcentrifuge tube.
124. Add 60 µL of room-temperature AMPure SPRI to each microcentrifuge tube containing PCR product. This creates a 1.2× volumetric ratio with the library. Thoroughly pipette mix each well such that all product in the well is a uniform color.
125. Incubate for 3 min at room temperature, and then place on the DynaMag-Tube Magnet for 2 min.
126. Remove the supernatant, then add 600 µL of fresh 80% (vol/vol) ethanol to each tube.
127. Rotate the tube, allowing the beads to shift. Rotate the tube a minimum of five additional times to ensure adequate bead movement to increase the beads' exposure to the ethanol wash.
128. Incubate on the magnet for 30 s at room-temperature then remove the 600 µL of ethanol.
129. Repeat Steps 126–128, yielding three ethanol washes in total.
130. After removing the final ethanol wash, wait 1 min. Then use a 200 µL pipette to remove residual ethanol and allow beads to dry for 5–10 min.
    ▲ **CRITICAL STEP** Beads are dry once they appear matte in color but before they develop visible cracks.
131. Once the beads are dry, remove the tube from the magnet, and add 31 µL of UltraPure water to elute dried beads.
132. Incubate for 1 min off the magnet at room temperature. Then place samples back onto the magnet for 1 min or until supernatant appears clear.
133. Transfer the 30 µL of sample to a new 96-well plate.
134. Perform a second SPRI cleanup by adding 1× volumetric ratio of AMPure XP SPRI beads to each well. Thoroughly pipette mix each well such that all product in the well is a uniform color.
135. Incubate for 3 min at room temperature and then place on the DynaMag-96 Side Skirted Magnet for 2 min.
136. To complete this second SPRI clean-up, follow Steps 81–85.
137. Once the beads are dry, remove the plate from the magnet, and add 31 µL of UltraPure water to elute dried beads.
138. Incubate for 1 min off the magnet at room-temperature. Then place samples back onto the magnet for 1 min or until supernatant appears clear.
139. Transfer the 30 µL of water to a new 96-well plate.

SMART-seq2 for genetic hashing

▲ **CRITICAL** Create an RNA-clean workspace. This includes wiping all surfaces, micro-pipettes, serological pipette controller, tube racks and tube magnets with RNaseZap and ensuring its adequate removal using 70% (vol/vol) ethanol. Failure to remove all detergents and surfactants in the RNaseZap can inhibit enzymes downstream.

Genetic hashing: RNA isolation using SPRI beads

● **TIMING 40 min**

▲ **CRITICAL** Allow 20–30 min for the RNAClean XP SPRI beads to equilibrate to room temperature before use. Ensure you make fresh 80% (vol/vol) ethanol for this procedure. Please refer to the RNAClean XP SPRI bead official protocol for more details. It is recommended to perform RNA isolation with SPRI and reverse transcription on a maximum of two plates at a time.

▲ **CRITICAL** To quickly gauge if the RNAClean XP SPRI beads are ready to use, quickly invert the bottle. If you observe no beads stuck to the bottom, the beads are ready to use. If you see a lot of beads stuck to the bottom, incubate another 10 min at room temperature.

140. Obtain the bulk-lysate that was saved in Step 2 and thaw it on ice. Add a 2.2× volumetric ratio of RNAClean XP SPRI beads to the bulk-lysate. Thoroughly pipette mix each well such that all product in the well is a uniform color.
141. Incubate for 5 min at room temperature and then place the DynaMag-96 Side Skirted Magnet for 2 min.
142. Remove and discard the supernatant.
143. Add 100 µL of fresh 80% (v/v) ethanol to the plate that still contains the SPRI beads.
144. Move the the 96-well plate to the right, back to the center, then to the left on the DynaMag™-96 Side Skirted Magnet. The plate should be moved a minimum of five times each to the right and left to ensure adequate bead movement to increase the beads' exposure to the ethanol wash.
145. Incubate on the magnet for 30 s at room temperature then remove the 100 µL of ethanol and discard.
146. Repeat Steps 141 to 143, yielding three ethanol washes in total.
147. After removing the final ethanol wash, wait 1 min. Then use a 20 µL multichannel pipette to remove residual ethanol and allow beads to dry for 5–10 min.
    ▲ **CRITICAL STEP** Beads are dry once they appear matte in color, but before they develop visible cracks.
148. Once bead pellet has dried, remove plate from magnet and elute RNA from beads by resuspending dried beads in 4 µL of genetic hybridization mix.

Reagent name	Volume (1 reaction)
Oligo dT primer	1 µL
10 mM dNTPs	1 µL
RNase Inhibitor	0.1 µL
UltraPure water (Thermo Fisher 10977023)	1.9 µL

    ▲ **CRITICAL STEP** Resuspending a bead pellet in 4 µL of genetic hybridization mix is not trivial. It works best if you position your pipette toward the well wall when you dispense the mix, then pipette up and down.
149. Seal the 96-well plate tightly with a PCR plate seal, bring to centrifuge, tilt the plate on a 45° angle and hit the side of the plate against your gloved hand. Turn the plate 180° and hit the edge of the plate against the side of your hand again. This ensures the beads are resuspended and well mixed in the wells.
150. Centrifuge plate 1 min at 800*g*, room temperature.
151. Incubate for 3 min at 72 °C in the thermocycler and then immediately place on ice for 30 s.

Genetic hashing: reverse transcription

● **TIMING 2.5 h**

152. Centrifuge plate for 1 min at 800*g*, room temperature.
153. Unseal the plate, add 7 µL of genetic reverse transcription mix per well prepared as outlined below, pipette mix well by pipetting up and down ten times.

Reagent name	Volume (1 reaction)
UltraPure water (Thermo Fisher 10977023)	0.7 µL
Maxima 5× RT Buffer	2 µL
Betaine (5 M)	2 µL
$MgCl_2$ (100 mM)	0.9 µL
Genetic Template Switching Oligo (10 µM)	1 µL
RNase inhibitor	0.25 µL
Maxima RNase H-minus RT	0.15 µL

154. Seal plate with PCR plate seal, then centrifuge for 1 min at 800*g*, room temperature.
155. Place samples in the thermocycler and run the genetic reverse transcription thermocycler program.

Temperature	Time	Number of cycles
42 °C	90 min	1 cycle
50 °C	2 min	10 cycles
42 °C	2 min	
70 °C	15 min	1 cycle
4 °C	Hold	Hold

■ **PAUSE POINT** Proceed to WTA or store at −20 °C for up to 1 week.

Genetic Hashing: WTA

● **TIMING 2.25 h**

156. Remove the plate from the thermocycler and centrifuge for 1 min at 800*g*, room temperature.
157. Add 14 µL of genetic amplification mix to each well.

Reagent name	Volume (1 reaction)
UltraPure water (Thermo Fisher 10977023)	1 µL
ISPCR (10 µM)	0.5 µL
KAPA HiFi Hotstart Readymix (Roche KK-2602)	12.5 µL

158. Place samples in the thermocycler and run the genetic amplification thermocycler program.

Temperature	Time	Number of cycles
98 °C	3 min	1 cycle
98 °C	15 s	18 cycles
67 °C	20 s	
72 °C	6 min	
72 °C	5 min	1
4 °C	Hold	Hold

■ **PAUSE POINT** Proceed to the WTA PCR clean-up or store the WTA product at 4 °C for 1–5 d or −20 °C for up to 1 year.

Genetic hashing: WTA PCR clean-up

● **TIMING 1 h**

▲ **CRITICAL** Allow 20–30 min for the AMPure XP SPRI beads to equilibrate to room temperature before use. Ensure you make fresh 80% (vol/vol) ethanol for this procedure. Please refer to the AMPure XP SPRI bead official protocol for more details.

▲ **CRITICAL** To quickly gauge if the AMPure XP SPRI beads are ready to use, quickly invert the bottle. If you observe no beads stuck to the bottom, the beads are ready to use. If you see a lot of beads stuck to the bottom, incubate another 10 min at room temperature.

159. Add a 0.8× volumetric ratio of room-temperature AMPure XP SPRI beads to each well containing WTA PCR product. Thoroughly mix each well such that all product in the well is a uniform color.
160. Incubate for 3 min at room temperature, and then place the DynaMag-96 Side Skirted Magnet for 2 min.
161. Remove and discard the supernatant.
162. Add 100 µL of fresh 80% (vol/vol) ethanol to the plate that still contains the SPRI beads.
163. Move the 96-well plate to the right, back to the center, then to the left on the DynaMag-96 Side Skirted Magnet. The plate should be moved a minimum of five times each to the right and left to ensure adequate bead movement to increase the beads exposure to the ethanol wash.
164. Incubate on the magnet for 30 s at room-temperature then remove the 100 µL of ethanol.
165. Repeat Steps 162–164, yielding three ethanol washes in total.
166. After removing the final ethanol wash, wait 1 min. Then use a 20 µL multichannel pipette to remove residual ethanol and allow beads to dry for 5–10 min.
    ▲ **CRITICAL STEP** Beads are dry once they appear matte in color but before they develop visible cracks.
167. Once the beads are dry, remove the plate from the magnet and 51 µL of UltraPure water to elute dried beads.
168. Incubate for 1 min off the magnet at room temperature. Then place samples back onto the magnet for 1 min or until supernatant appears clear.
169. Transfer the 50 µL of UltraPure water to a new 96-well plate.
170. Add a volume of SPRI beads to the new 96-well plate that is 0.8× the volume of the eluent in the wells.
171. Repeat Steps 160–168, then elute the dried beads in 16 µL of UltraPure water.
172. Transfer 15 µL of each sample to a new 96-well plate.
173. Place this new 96-well plate on ice or store if pausing.
    ■ **PAUSE POINT** Proceed to quantification and normalization of cDNA or store at 4 °C for 1–5 d or −20 °C for 5 d up to 1 year.

Genetic hashing: quantification and normalization of cDNA

● **TIMING 20 min**

174. Quantify the amount of cDNA in each of your samples utilizing a Qubit fluorometer according to manufacturer's instructions. We recommend combining 199 µL of the Qubit working solution and 1 µL of your sample in a fresh Qubit tube.
    ◆ **TROUBLESHOOTING**
    ▲ **CRITICAL STEP** You may use alternative quantification methods such as Nanodrop or other fluorescence-based kits at this step. We simply provide our recommended method optimizing cost and accuracy.
175. Use 1 µL of the purified PCR product to measure the fragment size distribution using the Agilent HS DNA BioAnalyzer or similar instrument according to the manufacturer's instructions.
    ◆ **TROUBLESHOOTING**
    ▲ **CRITICAL STEP** you want your distribution to be ~1,600–10,000 bp.
    ▲ **CRITICAL STEP** You may use alternative methods such as gel electrophoresis and Agilent Tapestation systems for identifying fragments peaks at this step. We simply provide our recommended method optimizing cost and accuracy.
    ◆ **TROUBLESHOOTING**

176. In a new 96-well plate, dilute 5 µL of each sample to 0.2 ng/µL.
■ **PAUSE POINT** Proceed to library preparation or store at 4 °C for 1–5 d or −20 °C for 5 d up to 1 year.

Genetic hashing: library preparation

● **TIMING 1 h**

177. In a new 96-well plate, add the following reagents in the order in which they are listed, pipette mixing up and down 5× between each addition:

Reagent	Amount
Tagment DNA Buffer	2.5 µL
Diluted PCR product	1.25 µL
Amplicon Tagment Mix	1.25 µL

178. Seal with a PCR plate seal, centrifuge 1 min at 800*g*, room temperature.
179. Incubate the samples for 10 min at 55 °C, then cool to 10 °C.
180. Place on ice for 30 s, then centrifuge for 30 s at 800*g*, room temperature.
181. Immediately add 1.25 µL Neutralize Tagment Buffer to each well and pipette mix up and down five times. Incubate for 5 min at room temperature.
182. Add 3.75 µL Nextera PCR Master Mix (Included in the Nextera XP kit) and 2.5 µL of combined Index primer solution to each well. Mix well by pipetting up and down five times.
183. Seal with a PCR plate seal, 1 min at 800*g*, room temperature.
184. Place the 96-well plate in the thermocycler and carry out the genetic library amplification PCR.

Temperature	Time	Number of cycles
72 °C	3 min	1 cycle
95 °C	30 s	1 cycle
95 °C	10 s	12 cycles
50 °C	30 s	
72 °C	60 s	
72 °C	5 min	1
4 °C	Hold	Hold

Genetic hashing: sequencing library PCR clean-up

● **TIMING 1 h**

185. Pool 5 µL from each sample.
186. Perform two 0.6× volumetric SPRI cleanups following Steps 159–173.

Sequencing

● **TIMING varies**

187. Follow manufacturer's instructions for sequencing.
◆ **TROUBLESHOOTING**

Procedure 2: data analysis

Sequence alignment

● **TIMING varies**

▲ **CRITICAL** Ensure all necessary software is installed before beginning the analysis.

1. Perform demultiplexing and generate FASTQ files for single-cell and bulk gene expression, HTO and ADT libraries using Illumina bcl2fastq (v2.20.0.422, https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) or bcl convert

(https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/bcl-convert.html). Here, we assume the output files are located in a directory named fastqs (./fastqs/). It is recommended to decide upon a sample name and have each library type follow the same naming scheme from here on.

◆ **TROUBLESHOOTING**

Example: sample name: Sample1

Gene expression library name: GEX_Sample1
HTO library name: HTO_Sample1
ADT library name: ADT_Sample1

2. Align single-cell gene expression libraries to the reference genome using the STARSolo[60] pipeline (https://github.com/alexdobin/STAR/blob/master/docs/STARsolo.md) with the following parameters:
`--genomeDir`: this parameter specifies the directory where the genome index files are located. Replace path `/path/to/genome_index_files` with path on your local machine or high performance cloud server.
`--readFilesIn`: this parameter specifies the path to the input read files. Typically, these would be the FASTQ files containing your RNA-seq data. You can provide one or two files depending on whether you have single-end or paired-end data. Replace path `/path/to/input_files` with path on your local machine or hpc server.
`--soloType CB_UMI_Simple`: this parameter sets the type of single-cell processing to be done. `CB_UMI_Simple` indicates that STAR should expect a simple cellular barcode (CB) and UMI structure without any advanced features like whitelist checking or correction.
`--soloCBstart`: this specifies the starting position (1-based) of the CB within the read sequence. `1` means the barcode starts at the first base of the read.
`--soloCBlen`: this defines the length of the cell barcode. `12` means the barcode is 12 bases long.
`--soloUMIstart`: this indicates the starting position of the UMI right after the cell barcode. `13` means the UMI starts at the thirteenth base of the read.
`--soloUMIlen`: this sets the length of the UMI. An `8` here means the UMI is 8 bases long.
`--outSAMtype BAM SortedByCoordinate`: this tells STAR to output the aligned reads in BAM format, sorted by coordinate.
`--outSAMattributes CR UR CY UY CB UB`: these options specify which extra attributes to include in the SAM/BAM output:
   - `CR`: cell barcode sequence as originally read
   - `UR`: UMI sequence as originally read
   - `CY`: phred quality of the cell barcode sequence
   - `UY`: phred quality of the UMI sequence
   - `CB`: corrected cell barcode sequence (if correction is applied, otherwise the same as CR)
   - `UB`: processed UMI sequence, typically the same as UR unless processing or correction is done

`--outFileNamePrefix ./sample_name/Solo.out/`: file path for output files. Before alignment, each sample should have a separate folder, with a secondary folder named Solo.out to indicate the files are output of the STARsolo pipeline.

Example of script:

```
STAR --genomeDir /path/to/genome_index_files --readFilesIn /path/to/
input_files --soloType CB_UMI_Simple --soloCBstart 1 --soloCBlen 12
--soloUMIstart 13 --soloUMIlen 8 --outSAMtype BAM SortedByCoordinate
--outSAMattributes CR UR CY UY CB UB --outFileNamePrefix ./sample_name/
Solo.out/
```

The output will be in the form of count, barcode and feature matrices for downstream processing.

Quality control and alignment of antibody hashing libraries and cell surface protein markers

● **TIMING** varies

3. Using your command line, run the python package CITE-seq-Count[61] with each HTO and/or ADT library and the following parameters, using either the standard CITE-seq-Count (option A) or the custom CITESeq_counts.sh wrapper which we have created to facilitate computational processing (option B).

(A) **Standard CITE-seq-Count (example for single HTO library):**

(i) Run the following script

```
pip install CITE-seq-Count
celln=$(cat "/directory/to/starsolo/barcodes/barcodes.tsv" |
wc -l)
CITE-seq-Count -R1 "HTO_fastq_read1.fastq" \
-R2 "HTO_fastq_read1.fastq" \
-t HTO.csv \
-cbf 1
-cbl 12 \
-umif 13 -umil 20 \
-cells $celln \
-wl "/directory/to/starsolo/barcodes/barcodes.tsv" \
-o "HTO_samplename" > "HTO_samplename.out"
```

-R1: specifies the path to the FASTQ file containing read 1 (paired-end sequencing)
-R2: specifies the path to the FASTQ file containing read 2 (paired-end sequencing)
-t: points to a CSV file containing the list of tags (that is, antibodies) and their corresponding DNA sequences that are used to detect specific proteins on the cells.
-cbf: indicates the starting position (first position) of the CB in the read.
-cbl: indicates the ending position (last position) of the CB in the read.
-umif: Indicates the ending position (last position) of the UMI in the read.
-cells: the number of expected cells to be specified, which can help in filtering out barcodes that are unlikely to correspond to actual cells.
-wl: specifies the path to a whitelist file containing a list of valid barcodes.
-o: sets the output directory where the results will be written.

(B) **Custom run_CITESeq_counts.sh wrapper** (https://github.com/ShalekLab/MultiOmics_SeqWellS3/blob/main/run_CITESeq_Counts.sh). This wrapper can be run depending on your data types:

(i) If only antibody hashing was performed (HTO only), run the wrapper as follows:

```
chmod +x ./run_CITESeq_Counts.sh
./run_CITESeq_Counts.sh HTO "sample1 sample2" "idx1 idx2"
```

where `"sample1 sample2"` are each of the sample names, which should be consistent between the fastq file and STARsolo output folder name, and `"idx1 idx2"` are each of the sample number values assigned to each fastq file.

(ii) If only cell surface protein staining was performed (ADT only), run the wrapper as follows:

```
chmod +x ./run_CITESeq_Counts.sh
./run_CITESeq_Counts.sh ADT "sample1 sample2" "idx1 idx2"
```

where `"sample1 sample2"` are each of the sample names which should be consistent between the fastq file and STARsolo output folder name, and `"idx1 idx2"` are each of the sample number values assigned to each fastq file.

(iii) If both antibody hashing and cell surface protein staining were performed (HTO and ADT):

```
chmod +x ./run_CITESeq_Counts.sh
./run_CITESeq_Counts.sh Both "sample1_HTO sample2_HTO" "idx1_HTO
idx2_HTO" "sample1_ADT sample2_ADT" "idx1_ADT idx2_ADT"
```

where `"sample1 sample2"` are each of the sample names which should be consistent between the fastq file and STARsolo output folder name, `"idx1_HTO idx2_HTO"` are each of the sample number values assigned to each fastq file for each HTO specific files, and `"idx1_ADT idx2_ADT"` are each of the sample number values assigned to each fastq file for each HTO specific files

Example for samples GEX_1 and GEX_2

Fastq file names:
Gene expression data: GEX_Sample1_S1_R1_001.fastq.gz and
GEX_Sample1_S1_R2_001.fastq.gz
GEX_Sample2_S2_R1_001.fastq.gz and
GEX_Sample2_S2_R2_001.fastq.gz
HTO data: HTO_Sample1_S3_R1_001.fastq.gz and
HTO_Sample1_S3_R2_001.fastq.gz
HTO_Sample2_S4_R1_001.fastq.gz and
HTO_Sample2_S4_R2_001.fastq.gz
ADT data: ADT_Sample1_S5_R1_001.fastq.gz and
ADT_Sample1_S5_R2_001.fastq.gz
ADT_Sample2_S6_R1_001.fastq.gz and
ADT_Sample2_S6_R2_001.fastq.gz

-HTO only:

```
chmod +x ./run_CITESeq_Counts.sh
./run_CITESeq_Counts.sh HTO "Sample1 Sample2" "3 4"
```

-ADT only:

```
chmod +x ./run_CITESeq_Counts.sh
./run_CITESeq_Counts.sh ADT "Sample1 Sample2" "5 6"
```

-HTO and ADT:

```
chmod +x ./run_CITESeq_Counts.sh
./run_CITESeq_Counts.sh Both "Sample1 Sample2" "3 4" "5 6"
```

4. Load gene expression, HTO and ADT matrices into your R environment and perform downstream analyses utilizing the Seurat v5 package[56]. All codes used in this protocol for this analysis can be found in via GitHub at https://github.com/ShalekLab/MultiOmics_SeqWellS3/tree/main.

Cell assignment using sample specific genotype profiles

● **TIMING** varies

5. Align your bulk gene expression libraries using a sample-appropriate aligner. We have had greatest success using STAR for alignment of bulk RNA-seq samples.

STAR --genomeDir hg38_index \
--readFilesIn sample_1.R1.fastq.gz sample_1.R2.fastq.gz \

```
--readFilesCommand zcat \
--sjdbGTFfile Homo_sapiens.GRCh38.97.gtf \
--runThreadN 16 \
--outSAMtype BAM SortedByCoordinate \
--outFilterType BySJout \
--outFileNamePrefix sample_1
```

6. Generate an unmapped BAM file (uBAM) for each of your bulk RNA-seq libraries utilizing Picard's[62] FastqToSam

```
java -Xmx8G -jar picard.jar FastqToSam \
FASTQ=6484_snippet_1.fastq \ #first read file of pair
FASTQ2=6484_snippet_2.fastq \ #second read file of pair
OUTPUT=6484_snippet_fastqtosam.bam \
READ_GROUP_NAME=H0164.2 \ #required; changed from default of A
SAMPLE_NAME=NA12878 \ #required
LIBRARY_NAME=Solexa-272222 \ #required
PLATFORM_UNIT=H0164ALXX140820.2 \
PLATFORM=illumina \ #recommended
SEQUENCING_CENTER=BI \
RUN_DATE=2014-08-20T00:00:00-0400
```

7. Perform data preprocessing for variant discovery by following the publicly available Workflow Description Language file and accompanying JSON file of inputs indicated by GATK's Best Practices[63–65] (https://github.com/gatk-workflows/gatk4-data-processing/tree/master). It is recommended to run these pipelines via a GATK container image on Docker as outlined on the GATK website (https://gatk.broadinstitute.org/hc/en-us/articles/360035889991--How-to-Run-GATK-in-a-Docker-container). The main steps of the pre-processing pipeline can be also be run through a wrapper script found in via GitHub at https://github.com/ShalekLab/MultiOmics_SeqWellS3/blob/main/gatk_preprocessing with the following code and parameters:

```
./gatk_preprocessing.sh \
-s 'sample1 sample2' \
-r reference_name \
-f ref_fasta.fa \
-i ref_fasta_index.fa.fai \
-d reference_dictionary.dict \
-v dbSNP_vcf.vcf \
-x dbSNP_vcf_index.vcf.gz.tbi \
-b 'sample1.unmapped.bam sample2.unmapped.bam'\
-u .bam \
-k known_indel_sites.vcf
```

where `"sample1 sample2"` are each of the sample names and should be consistent with the name of your unmapped bam files, 'reference_name' is the name of the reference genome build utilized, 'ref_fasta.fa' is the fasta file for the reference genome utilized as a .fa or .fasta file, 'ref_fasta_index.fa.fai' is the index for the reference fasta file utilized as a .fai file, 'reference_dictionary.dict' is the sequence dictionary for the reference being utilized as .dict file, `'sample1.unmapped.bam sample2.unmapped.bam'` refers to the name of each unmapped bam file per sample in a list form, '.bam' refers to the suffix of your bam files which should be consistent across all bam files, and 'known_indel_sites.vcf' refers to a list of known indel sites for the utilized genome as a .vcf file/.

8. Identify unique patient variants utilizing HaplotypeCaller[65], which leverages local de novo assembly of haplotypes in an active region to call SNPs and indels simultaneously.

```
gatk --java-options "-Xmx4g" HaplotypeCaller \
-I 'sample1.unmapped.bam sample2.unmapped.bam'\
-R ref_fasta.fa \
-O 'Sample1.vcf.gz Sample2.vcf.gz'
```

where `"sample1.unmapped.bam sample2.unmapped.bam"` are each of the sample bam files generated in Step 7, 'ref_fasta.fa' is the fasta file for the reference genome utilized as a .fa or .fasta file (same file as used in Step 7), and `'Sample1.vcf.gz Sample2.vcf.gz'` refers to the name of each output file vcf file that will be output from HaplotypeCaller.

9. Aggregate the vcfs from all samples that were hashed together using the merge function from bcftools with the following code and parameters:

```
git clone --recurse-submodules https://github.com/samtools/htslib.git
git clone https://github.com/samtools/bcftools.git
cd bcftools
make
cd.
./bcftools merge -m none --force-samples --threads 18 Sample1.vcf.gz
Sample2.vcf.gz -o "aggregated.vcf"
```

`-m none`: this option specifies the merging method. The `none` method means that the input files are simply concatenated without any merging of records from different files that refer to the same position.
`--force-samples`: forces the use of sample names as given in the VCF files. This option is useful when merging files with overlapping sample names; it prevents bcftools from automatically renaming samples to avoid collisions.
`--threads`: specifies the number of threads to use for processing.
`Sample1.vcf.gz Sample2.vcf.gz`: these are the input files to be merged.
`-o "aggregated.vcf"`: specifies the output file name for the merged VCF file.

10. Sort vcf to match your bam file by running the sort_vcf_same_as_bam.sh script which is part of the popscle_help_tools respository on GitHub (https://github.com/aertslab/popscle_helper_tools/blob/master/sort_vcf_same_as_bam.sh). The inputs for running the script will be the aggregated vcf file from Step 9 (aggregated.vcf) and one bam file from any of the samples contained within the vcf file (sample.bam).
chmod +x ./sort_vcf_same_as_bam.sh

```
./sort_vcf_same_as_bam.sh sample.bam aggregated.vcf > sorted.vcf
```

11. For each sample, use the sorted, genome specific vcf to generate pileups around known variants using the 'dsc-pileup' tool. This tool is part of the popscle package available through bioconda or GitHub (https://github.com/statgen/popscle)[54]. This identifies the number of reads from each allele at each of the common SNP locations when run with the following parameters:

```
conda install bioconda::popscle
popscle dsc-pileup --sam sample.bam --vcf sorted.vcf --out sample_name/
pileup_out--tag-group CB --tag-UMI UB --skip-umi --group-list starsolo_
output/barcodes.tsv
```

`--sam sample.bam`: specifies the input SAM or BAM file which contains the aligned sequencing reads.
`--vcf sorted.vcf`: specifies the input VCF file that contains the variants to be analyzed against the BAM file. This VCF should be sorted and indexed.
`--out sample_name/pileup_out`: specifies the base name of the output files.
`--tag-group CB`: this option specifies the BAM tag that contains the CB information.
`--tag-UMI UB`: indicates the BAM tag that contains the UMI information, specified here as UB.

`--skip-umi`: this option is to ignore UMI information when processing the data.
`--group-list starsolo_output/barcodes.tsv`: specifies a file containing a list of valid CBs.

12. For each sample use the freemuxlet tool, which is also part of the popscle package, without genotypes, to deconvolute the identities of barcoded cells with the following code and parameters[54]:

```
popscle freemuxlet --nsample 3 --out sample_name/freemux_merged --plp
sample_name/pileup_out
```

`--nsample 3`: this parameter specifies the number of different samples (or individuals) that are assumed to be mixed in the dataset.
`--out sample_name/freemux_merged`: sets the base name for the output files produced by freemuxlet. Here, the output files will be saved in a directory called sample_name and the files will start with the prefix freemux_merged. The tool generates multiple output files containing various information about the de-multiplexing results, such as probabilities of each cell belonging to each sample.
`--plp sample_name/pileup_out`: specifies the input pileup file generated by popscle dsc-pileup or another compatible tool. This pileup file contains the allele counts at variant positions for each cell. freemuxlet will use this file to assess genetic variation across cells and assign them to the correct sample.

13. Create an association between the genotype clusters and patient data using the 'vcf-match-sample-ids' tool (https://github.com/hyunminkang/apigenome/blob/master/scripts/vcf-match-sample-ids) with the following code:

```
perl vcf-match-sample-ids.txt \
-vcf1 \
sample_name/freemux_merged/demuxlet_merged.clust1.vcf.gz \
-vcf2 sorted.vcf > "Sample_Assignment.csv"
```

where -vcf1 is the sample output from running freemuxlet and -vcf2 is the aggregated vcf generated in Step 10.

14. Output will be the correlation between each donor genotype and freemuxlet genotype. Match donor and freemuxlet genotype by their highest correlation and append results to your Seurat metadata. All analysis codes can be found via GitHub at https://github.com/ShalekLab/MultiOmics_SeqWellS3/tree/main.
◆ **TROUBLESHOOTING**

Troubleshooting

Problem 1

After cell staining, the remaining number of cells is too low to continue with Seq-Well S^3. (Procedure 1: Step 14 and 26)

We recommend a minimum of 120,000 cells to perform cell surface protein staining and antibody hashing. If only performing cell surface protein staining or antibody hashing, a minimum of 60,000 cells are recommended. Ensure not to disturb the pellet when aspirating. You may decide to leave a small amount of liquid behind between washes to ensure you are not aspirating cells. Moreover, you may choose to use V bottom plates to help with cell retention.

Problem 2

Low cDNA yield at post-WTA quantification (Procedure 1: Step 110 and 174).

Low cDNA yield can be due to several issues. It can be due to lack of cDNA resulting from low quality samples as mentioned later in Problem 4. However, if the shape of your cDNA size

distribution looks ideal, the low yield can be indicative of too few PCR cycles during WTA. You can perform the WTA step again and add additional cycles to increase the amount of cDNA. In addition, post PCR clean-up, elute the product in a smaller volume of water such as 10 µL to concentrate your WTA product.

Problem 3
BioAnalyzer trace following post-WTA PCR clean-up is abnormal (large peak between 20–80 bp) (Procedure 1: Step 111 and 175).

This is a phenomenon we refer to as primer dimers. When samples are lower quality, highly degraded or mRNA content is low, there are more primers in solution than mRNA to bind to. This results in primers dimerizing with themselves and appear as a large peak. An additional PCR clean-up using SPRI beads may help remove additional primer dimers. Caution is advised if DNA concentration is already low, as additional clean-up will further decrease the concentration.

Problem 4
SMART-seq2 fails or results in low quality libraries (Procedure 1: Step 174 and 187).

It is recommended to retain two separate bulk RNA samples per sample in case of sample failure. If this is not possible, you can demultiplex samples based on genetic differences but you will not be able to assign the relevant clinical metadata to the original samples.

Problem 5
Unable to demultiplex due to overlapping barcodes (Procedure 2: Step 1).

Ensure unique barcodes are placed on a given lane of your sequencing run. All pooled barcodes should have a Hamming distance of at least 2 from all other barcodes (input of bcl2fastq).

Problem 6
You notice a high proportion of doublets in your data (Procedure 2: Step 14).

This is often a result of cell clumps caused by improper dissociation into single-cell suspension or overloading an array with a higher number of cells than recommended. Ensure you perform a manual cell count immediately before loading your cells. If you have a high proportion of dead cells, you may consider dead cell removal via magnetic or fluorescent-activated cell sorting.

Timing

Procedure 1: sample preparation and sequencing

Sample preparation

Steps 1–3, counting cells: 30 min
Steps 4–15, stain with hashing antibodies: 55 min
Steps 16–28, stain with cell surface protein antibodies: 55 min

Seq-Well S^3 processing

Steps 29–53, cell loading, membrane sealing, lysis, hybridization and bead recovery: 3.5 h
Steps 54–59, reverse transcription: 2.5 h or option to run overnight
Steps 60–75, exonuclease I treatment, second strand synthesis, WTA: 4.5 h
Steps 76–109, separation of short and long WTA products and PCR clean-up: 1 h
Steps 110–111, quantification of product: 30 min
Steps 112–119, scRNA library preparation: 1.5 h
Steps 120–122, step-out PCR for short fractions of WTA products: 45 min
Steps 123–139, sequencing library PCR clean-up: 1 h

Smart-seq2 for genetic hashing (7.75 h)
Steps 140–151: RNA isolation using SPRI: 40 min
Steps 152–155: reverse transcription: 2.5 h
Steps 156–158: WTA: 2.25 h
Steps 159–173: WTA PCR clean-up: 1 h
Steps 174–176: quantification and normalization of cDNA: 20 min
Steps 177–184: library preparation: 1 h
Steps 185–186: sequencing library PCR clean-up: 1 h

Sequencing
Step 187: varies

Procedure 2: data analysis
Steps 1–2: sequence alignment: varies
Steps 3–4: quality control and alignment of antibody hashing libraries and cell surface protein markers: varies
Steps 5–14: cell assignment using sample specific genotype profiles: varies

Anticipated results

The integration of Seq-Well S^3, CITE-seq and SMART-seq2 in our comprehensive multiomic Protocol offers a convenient approach to decipher the complexities of cellular biology at a single-cell resolution. Completion of this Protocol results in the generation of all necessary libraries and the preliminary processing required for effective multiomic integration (Fig. 1).

One of the cornerstones of our protocol's robustness is the implementation of quality control measures at various stages of sample and library preparation. Bioanalyzer traces serve as a crucial tool for assessing the quality of cDNA at multiple points throughout the workflow—post-cDNA purification and after final library preparation. These traces provide

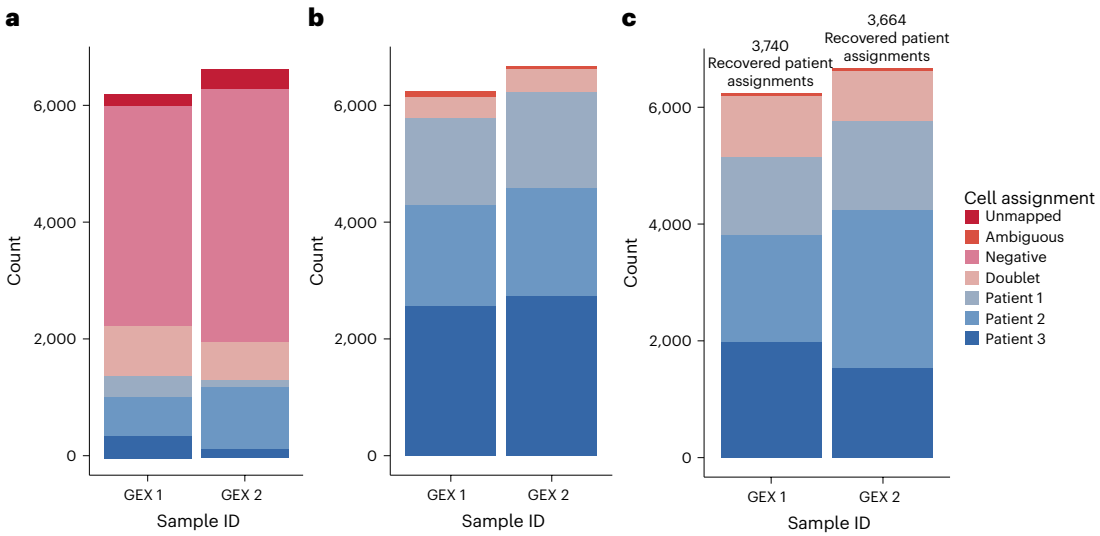

Fig. 3 | Comparative analysis of cell assignments across multiplexing methods. a, Cell assignments based on antibody hashing with categories for negatives, doublets, three patient samples, and an unmapped category. **b**, Genetic assignments with categories for ambiguous, doublets, and the same patient samples. **c**, Final assignments following the combined multiplexing approach with the same assignment nomenclature as the genetic assignments. Ambiguous assignments are substantially reduced in the final combined approach, demonstrating its efficacy in cell-type discrimination. Doublet rates are consistent across methods, ensuring reliable data interpretation. The bar plots represent the distribution of cell assignments using antibody hashing for **a**, genetic hashing for **b** and a combined approach for **c** for two separate Seq-Well arrays, labeled gene expression 1 (GEX 1) and gene expression 2 (GEX 2).

Table 1 | Patient label concordance between antibody and genetic hashing methods

	Concordant cells (%)	Cohen's kappa	Adjusted Rand index
Array 1 (GEX_1)	89	0.84	0.71
Array 2 (GEX_2)	94	0.89	0.83

Percentage of cells bearing the same label assignment across hashing modalities, Cohen's kappa for rater reliability of class observation and adjusted Rand index of similarity were calculated to determine the consistency in label assignment. Each measurement of concordance yielded high association of patient labels on a per array basis.

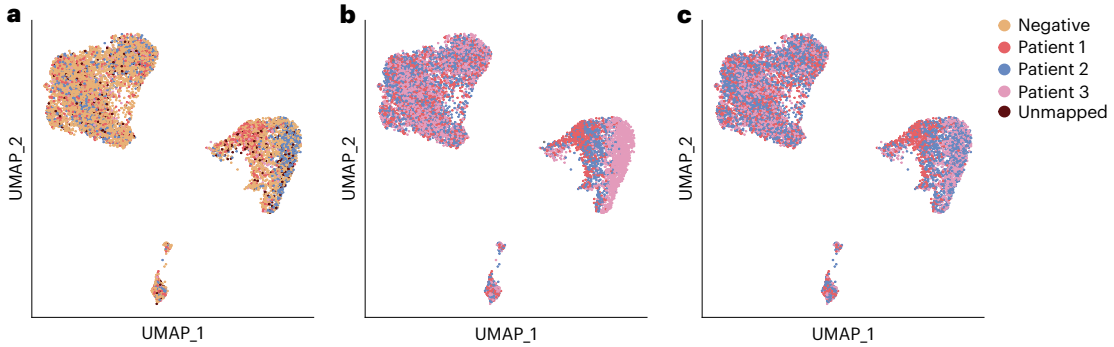

Fig. 4 | UMAP visualization demonstrating clustering independent of cell assignment method based on RNA data alone. UMAP plots representing cell clusters derived solely from RNA input data across different cell assignment methods. **a**–**c**, Clustering based on antibody assignment (**a**), genetic assignment (**b**) and final assignment using a combined approach (**c**). Each dot represents a single cell, color coded by assignment: ambiguous (brown), negative—referring to a cell with no detectable hashtag information (blue) and specific patient samples (green for patient 1, red for patient 2, purple for patient 3 and orange for unmapped). The clustering patterns are consistent across the various multiplexing methods, demonstrating that the clusters are based solely on RNA input and not influenced by the cell assignment method. This uniformity underscores the robustness of the RNA-based data in revealing true biological variation, independent of multiplexing strategy.

essential insights into the size distribution and concentration of cDNA samples, helping to confirm that the preparations are within optimal ranges for high-quality sequencing results (Fig. 2)

We also compared the performance and robustness of the antibody-based, genetic and combined multiplexing strategies described in this Protocol. All strategies resulted in robust cell identification with minimal batch-to-batch variation, as evidenced by the consistent results across independent arrays. Each approach reliably identified patient-specific cells, but their respective strengths and limitations inform the optimal choice for different experimental contexts (Fig. 3). Antibody multiplexing offers a streamlined workflow and yields more stringent label assignments and less overall ambiguity. However, its reliance on staining efficiency results in a substantial fraction of negative cells—cells that present with no antibody information (Fig. 3a). Genetic multiplexing, by contrast, is independent of staining and leverages endogenous SNP variation. This method achieves higher overall label assignment rates, as virtually all cells containing sufficient genetic information can be assigned, but at the cost of increased ambiguity in some cases (Fig. 3b). The integration of modalities yields a synergistic effect: the combined approach leverages the stringency of antibody assignments and the completeness of genetic assignments, resulting in both a higher number of recovered patient assignments and fewer cells with ambiguous and negative labels (Fig. 3c). Notably, both antibody and genetic multiplexing strategies independently assign patient labels with high concordance, as evidenced by quantitative measures including the percentage of cells with matching label assignments, Cohen's kappa statistic for rater reliability and the adjusted Rand index for clustering similarity. Each of these measures indicates a high degree of agreement between the two approaches on a per-array basis, and discrepancies are largely confined to cells that are ambiguous or negative in the individual modalities (Table 1). This underscores the reliability of both modalities for accurate sample assignment. Taken together, these data suggest that any of the approaches may be utilized and the decision should be guided by

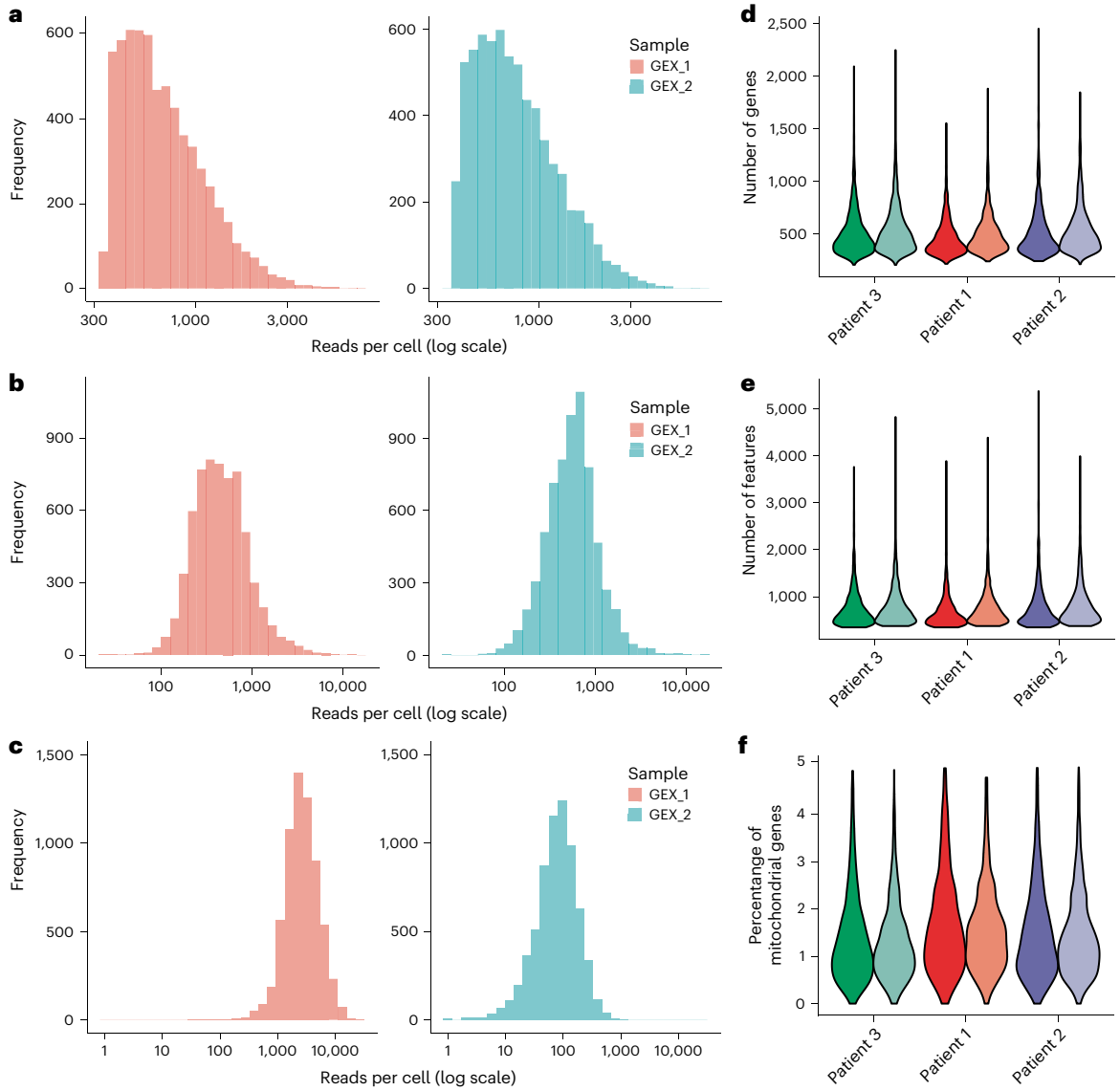

Fig. 5 | Quality control metrics following filtering of single-cell data. **a**–**c**, Histograms showing read count distributions per cell for each array (**a**), antibody hashtag (**b**) and cell-surface protein antibody (**c**). **d**–**f**, Violin plots depicting the distribution of the number of genes detected per cell (**d**), total number of molecules detected per cell (**e**) and the percentage of mitochondrial genes per cell (**f**), respectively. These metrics were used to assess and filter low-quality cells before downstream analyses.

project-specific considerations such as throughput, desired stringency and available resources. Our results demonstrate that each approach is reliable, and a combined strategy can further enhance assignment confidence and completeness.

The integration of CITE-seq with Seq-Well S^3 permits increased sample throughput via sample multiplexing as well as the simultaneous detection of protein markers and RNA expression. The outcomes are expected to provide insights across various fields such as immunology, oncology, and developmental biology. The design of our protocol effectively supports sample multiplexing by employing UMIs and hashing techniques, enabling the concurrent processing of multiple samples without sacrificing data integrity. The protocol's capability to handle overloading of Seq-Well arrays with computational doublet detection ensures scalability of experiments while maintaining high data quality. To validate the consistency of the outputs, we conducted tests on two technical replicate Seq-Well arrays.

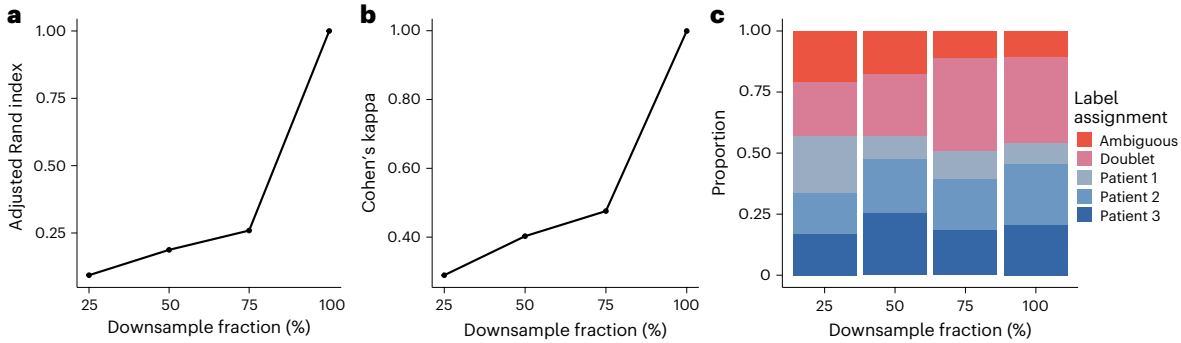

Fig. 6 | Effects of data downsampling on the reproducibility and composition of patient label assignments. a, Adjusted Rand index (ARI) comparing patient label assignments at downsample fractions 25%, 50%, 75% and 100% relative to the full dataset (100%). **b**, Cohen's kappa values for patient label assignments at the same downsample fractions, similarly benchmarked against the 100% dataset. **c**, A stacked bar plot illustrating the proportion of cells assigned to each patient or labeled as ambiguous/doublet at each downsample fraction.

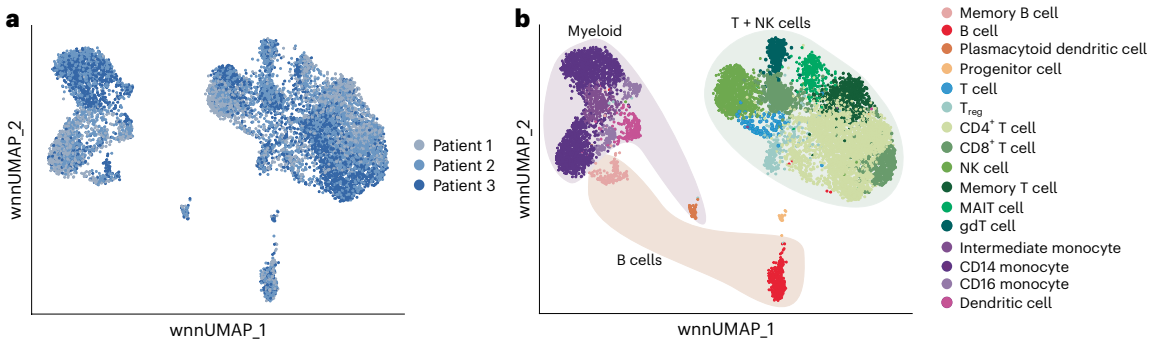

Fig. 7 | Integrated analysis of RNA and protein data in single-cell populations using weighted nearest neighbor analysis. UMAP visualizations generated by integrating RNA and protein data through weighted nearest neighbor analysis to reveal detailed cellular heterogeneity across patients and cell types. **a**, Distribution of cells from three different patients (shown in different shades of blue) showing absence of patient-specific clustering patterns. **b**, Major immune cell types identified based on combined RNA and protein expression profiles, illustrating an ability to more easily identify a diverse array of immune cell types including $CD4^+$ T cells, $CD8^+$ T cells, monocytes, neutrophils, natural killer (NK) cells, dendritic cells, mucosal-associated invariant T (MAIT) cells, gamma delta (gdT) T cells and B cells, among others, each color coded uniquely for clarity. This integrated approach highlights the nuanced differences in cell type distribution across patient samples, providing a comprehensive overview of the cellular landscape. T_{reg}, T regulatory cells.

After appropriate data filtering, we retained 6,244 and 6,672 cells from each array, respectively. Our findings reveal a robust capacity to distinguish between multiplexed samples using a hybrid approach that combines antibody and genetic markers (Fig. 4, Supplementary Fig. 1 and Supplementary Table 1) while still maintaining high quality data as evidenced by our post sequencing quality control metrics (Fig. 5). High quality data can be determined in various ways. Here, we consider high-quality GEX, HTO and ADT libraries to have the majority of cells with more than 500 transcripts per cell (Fig. 5a–c). Moreover, good data quality is indicated by cells with more than 250 but fewer than 2,500 genes (Fig. 5d) and less than 5% of mitochondrial genes in a cell. Moreover, as patient label assignment concordance was substantial between antibody and genetic methods (Table 1), it is possible to utilize only one form of multiplexing. This trend was less pronounced with decreased sequencing depth of the bulk genetic data, highlighting the need for adequate sequencing to improve the identification of a more robust SNP profile for each patient (Fig. 6). Importantly, Seq-Well S^3 processing shows no observable batch effects between arrays, as evidenced in uniform manifold approximation and projection (UMAP) visualizations and composition of clusters (Fig. 7). This consistency is maintained across various methods of multiplexing, ensuring reliable and scalable experiments.

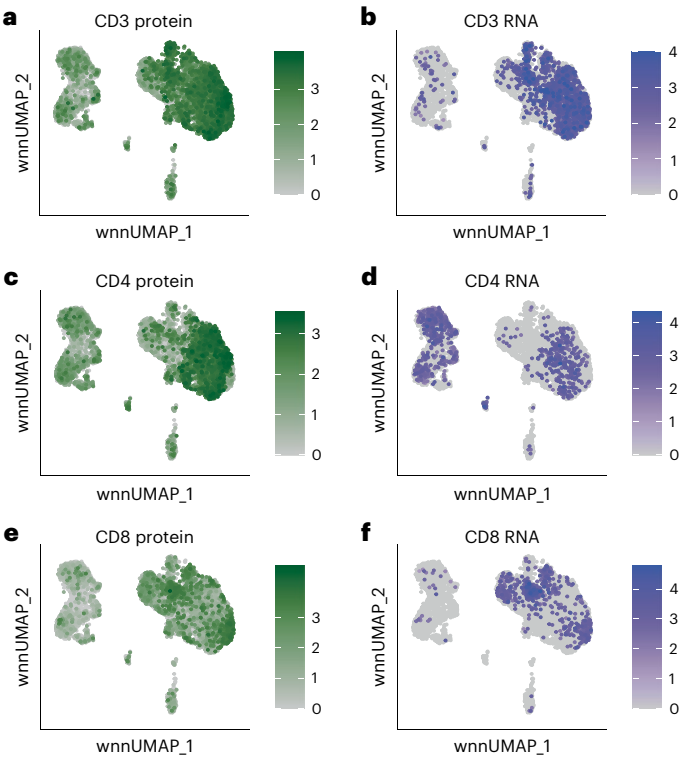

Fig. 8 | Comparative visualization of feature plots illustrating CD4 and CD8 expression at protein and RNA levels. Feature plots demonstrating the expression of T cell markers CD4, CD8 and CD3 in low-dimensional space, highlighting the distinct advantages of integrating RNA and protein data for accurate cell type identification. **a**,**b**, Expression of CD4 at the protein (**a**) and RNA (**b**) levels, respectively. **c**,**d**, Expression of CD8 at the protein (**c**) and RNA (**d**) levels, respectively. **e**,**f**, Expression of CD3 at the protein (**e**) and RNA (**f**) levels, respectively. The color gradients indicate expression levels, with green representing protein and blue representing RNA. These feature plots reveal the clearer delineation and higher expression visibility of T cell subsets when protein data are incorporated alongside RNA, especially evident in the more diffuse RNA signal compared with the concentrated protein signal. This integrated analysis approach, using weighted nearest neighbor methods, enhances the resolution of T cell type identification, crucial for studies requiring precise cellular characterization.

The addition of proteomic information complements transcriptomic data and enables the detailed characterization of cell types, states and functional pathways, providing a dual readout of cellular activity at the single-cell level. Specifically, multiomic analysis provides increased granularity in cell type identification, facilitating the distinction between closely related cell types (Fig. 8). This is particularly valuable in cases such as differentiating between $CD4^+$ and $CD8^+$ T cells, cell subsets canonically defined by sets of surface markers and with substantial gene expression overlap, where conventional singleomic methods might fall short (Fig. 8 and Supplementary Fig. 2). Furthermore, as mRNA and protein levels often do not correlate well—given stochasticity and differences in the timing and regulation of transcription and translation, as well as differences in RNA and protein degradation rates[66,67]—they provide complementary information. In sum, this comprehensive protocol enables a more precise understanding of cellular diversity and function in a streamlined system.

Reporting summary

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

Data availability

Raw sequencing data, as well as processed data for single-cell RNA-seq, CITE-seq and bulk RNA-seq data are deposited at GEO under GSE266385.

Code availability

Command line and R code created by the authors for this Protocol are available via GitHub at https://github.com/ShalekLab/MultiOmics_SeqWellS3.

Received: 21 July 2024; Accepted: 14 November 2025;
Published online: 06 March 2026

References

1. Andreatta, M. et al. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat. Commun.* **12**, 2965 (2021).
2. Raghavan, S. et al. Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119–6137.e26 (2021).
3. Kotliar, D. et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **8**, e43803 (2019).
4. Zhu, H. et al. Human PBMC scRNA-seq-based aging clocks reveal ribosome to inflammation balance as a single-cell aging hallmark and super longevity. *Sci. Adv.* **9**, eabq7599 (2023).
5. Jeffries, A. M. et al. Single-cell transcriptomic and genomic changes in the ageing human brain. *Nature* **646**, 657–666 (2025).
6. Yang, X., Hou, X., Zhang, J., Liu, Z. & Wang, G. Research progress on the application of single-cell sequencing in autoimmune diseases. *Genes Immun.* **24**, 220–235 (2023).
7. Scheinecker, C., Göschl, L. & Bonelli, M. T_{reg} cells in health and autoimmune diseases: new insights from single cell analysis. *J. Autoimmun.* **110**, 102376 (2020).
8. Zhao, M. et al. The application of single-cell RNA sequencing in studies of autoimmune diseases: a comprehensive review. *Clin. Rev. Allergy Immunol.* **60**, 68–86 (2021).
9. Huang, W., Wang, D. & Yao, Y.-F. Understanding the pathogenesis of infectious diseases by single-cell RNA sequencing. *Microb. Cell* **8**, 208–222 (2021).
10. Wang, B. et al. Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection. *Nat. Microbiol.* **8**, 1846–1862 (2023).
11. Swaminath, S. & Russell, A. B. The use of single-cell RNA-seq to study heterogeneity at varying levels of virus–host interactions. *PLOS Pathog.* **20**, e1011898 (2024).
12. Nofech-Mozes, I., Soave, D., Awadalla, P. & Abelson, S. Pan-cancer classification of single cells in the tumour microenvironment. *Nat. Commun.* **14**, 1615 (2023).
13. Liu, Z. L. et al. Single cell deciphering of progression trajectories of the tumor ecosystem in head and neck cancer. *Nat. Commun.* **15**, 2595 (2024).
14. Kang, J. et al. Systematic dissection of tumor-normal single-cell ecosystems across a thousand tumors of 30 cancer types. *Nat. Commun.* **15**, 4067 (2024).
15. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
16. Heumos, L. et al. Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550–572 (2023).
17. Sidhaye, J. et al. Integrated transcriptome and proteome analysis reveals posttranscriptional regulation of ribosomal genes in human brain organoids. *eLife* **12**, e85135 (2023).
18. Bathke, J., Konzer, A., Remes, B., McIntosh, M. & Klug, G. Comparative analyses of the variation of the transcriptome and proteome of Rhodobacter sphaeroides throughout growth. *BMC Genomics* **20**, 358 (2019).
19. Genshaft, A. S. et al. Single-cell RNA sequencing of liver fine-needle aspirates captures immune diversity in the blood and liver in chronic hepatitis B patients. *Hepatology* **78**, 1525 (2023).
20. Liu, N. et al. Scalable, compressed phenotypic screening using pooled perturbations. *Nat. Biotechnol* **43**, 1324–1336 (2025).
21. Morang'a, C. M. et al. scRNA-seq reveals elevated interferon responses and TNF-α signaling via NFkB in monocytes in children with uncomplicated malaria. *Exp. Biol. Med.* **249**, 10233 (2025).
22. Kotliar, D. et al. Single-Cell Profiling of ebola virus disease in vivo reveals viral and host dynamics. *Cell* **183**, 1383–1401.e19 (2020).
23. Tarazona, S. et al. Harmonization of quality metrics and power calculation in multi-omic studies. *Nat. Commun.* **11**, 3092 (2020).
24. Hughes, T. K. et al. Second-strand synthesis-based massively parallel scRNA-seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity* **53**, 878–894.e7 (2020).
25. Drake, R. S. et al. in *Single Cell Transcriptomics: Methods and Protocols* (eds. Calogero, R. A. & Benes, V.) 57–104 (Springer US, 2023); https://doi.org/10.1007/978-1-0716-2756-3_3
26. Mead, B. E. et al. Screening for modulators of the cellular composition of gut epithelia via organoid models of intestinal stem cell differentiation. *Nat. Biomed. Eng.* **6**, 476–494 (2022).
27. Ziegler, C. G. K. et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* **181**, 1016–1035.e19 (2020).
28. Trombetta, J. J. et al. Preparation of single-cell RNA-seq libraries for next generation sequencing. *Curr. Protoc. Mol. Biol.* **107**, 4.22.1–4.22.17 (2014).
29. Cardiello, J. F. et al. Evaluation of genetic demultiplexing of single-cell sequencing data from model species. *Life Sci. Alliance* **6**, e202301979 (2023).
30. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
31. Brown, D. V. et al. A risk-reward examination of sample multiplexing reagents for single cell RNA-seq. *Genomics* **116**, 110793 (2024).
32. Gideon, H. P. et al. Multimodal profiling of lung granulomas in macaques reveals cellular correlates of tuberculosis control. *Immunity* **55**, 827–846.e10 (2022).
33. Wilk, A. J. et al. Pro-inflammatory feedback loops define immune responses to pathogenic Lentivirus infection. *Genome Med.* **16**, 24 (2024).
34. Ordovas-Montanes, J. et al. Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649–654 (2018).
35. Tzouanas, C. N. et al. Hepatic adaptation to chronic metabolic stress primes tumorigenesis. *Cell* **189**, 35–460 (2026).
36. Nyquist, S. K. et al. Cellular and transcriptional diversity over the course of human lactation. *Proc. Natl Acad. Sci. USA* **119**, e2121720119 (2022).
37. Cheng, J., Liao, J., Shao, X., Lu, X. & Fan, X. Multiplexing methods for simultaneous large-scale transcriptomic profiling of samples at single-cell resolution. *Adv. Sci.* **8**, 2101229 (2021).
38. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
39. Pai, J. A. & Satpathy, A. T. High-throughput and single-cell T cell receptor sequencing technologies. *Nat. Methods* **18**, 881–892 (2021).
40. Mhanna, V. et al. Adaptive immune receptor repertoire analysis. *Nat. Rev. Methods Primer* **4**, 1–25 (2024).
41. Irac, S. E., Soon, M. S. F., Borcherding, N. & Tuong, Z. K. Single-cell immune repertoire analysis. *Nat. Methods* **21**, 777–792 (2024).
42. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
43. Jovic, D. et al. Single-cell RNA sequencing technologies and applications: a brief overview. *Clin. Transl. Med.* **12**, e694 (2022).
44. Rosenberg, A. B. et al. SPLiT-seq reveals cell types and lineages in the developing brain and spinal cord. *Science* **360**, 176–182 (2018).
45. Gierahn, T. M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
46. Fan, H. C., Fu, G. K. & Fodor, S. P. A. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**, 1258367 (2015).
47. Chen, H. et al. High-throughput Microwell-seq 2.0 profiles massively multiplexed chemical perturbation. *Cell Discov.* **7**, 1–4 (2021).
48. Han, X. et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e17 (2018).
49. Ding, Y. et al. A deep learning model to predict a diagnosis of Alzheimer disease by using ^{18}F-FDG PET of the brain. *Radiology* **290**, 456–464 (2019).
50. Jaiswal, S. et al. Identification and characterization of the T cell receptor (TCR) repertoire of the cynomolgus macaque (*Macaca Fascicularis*). *BMC Genomics* **23**, 647 (2022).
51. Prakadan, S. M. et al. Genomic and transcriptomic correlates of immunotherapy response within the tumor microenvironment of leptomeningeal metastases. *Nat. Commun.* **12**, 5955 (2021).
52. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
53. Mastering biology to advance human health. *10x Genomics* https://www.10xgenomics.com/ (2025).
54. TotalSeq™–a cell surface protein and hashing single-cell staining guide. *BioLegend* https://www.biolegend.com/nl-be/protocols/totalseq-a-cell-surface-protein-and-hashing-single-cell-staining-guide (2025).
55. Gonye, A. L. K. et al. Protocol for bulk RNA sequencing of enriched human neutrophils from whole blood and estimation of sample purity. *STAR Protoc.* **4**, 102125 (2023).
56. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
57. Neavin, D. et al. Demuxafy: improvement in droplet assignment by integrating multiple single-cell demultiplexing and doublet detection methods. *Genome Biol.* **25**, 94 (2024).
58. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
59. AMPure XP Beads Protocol | PCR & NGS Library Cleanup. *Beckman Coulter Life Sciences* https://www.beckman.com/reagents/genomic/cleanup-and-size-selection/pcr/ampure-xp-protocol (2025).
60. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. Preprint at *bioRxiv* https://doi.org/10.1101/2021.05.05.442755 (2021).
61. Roelli, P., bbimber, Flynn, B., santiagorevale & Gui, G. Hoohm/CITE-seq-Count: 1.4.2. *Zenodo* https://doi.org/10.5281/zenodo.2590196 (2019).
62. Picard Tools—by Broad Institute. *Github* https://broadinstitute.github.io/picard/ (2025).
63. Van der Auwera, G. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media, Incorporated, 2020).
64. Van der Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1–11.10.33 (2013).
65. Poplin, R. et al. Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at *bioRxiv* https://doi.org/10.1101/201178 (2018).
66. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
67. Triana, S. et al. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* **22**, 1577–1589 (2021).

Acknowledgements
A.K.S. was supported, in part, by the NIH NIDA DP1 Avant-Garde Pioneer Award (1DP1DA053731) and the Bill and Melinda Gates Foundation. A.K.S., B.A.G., A.G.E., D.A.L. M.A.E. and D.D.R. were supported, in part, by NIH 5U19AI167899. S.T. was supported in part by a Pew Latin American Fellowship Program in Biomedical Sciences. B.A.G is supported in part through the Geisel School of Medicine at Dartmouth's Center for Quantitative Biology through a grant from the National Institute of General Medical Sciences (NIGMS, grant no. P20GM130454) of the NIH.

Author contributions
D.D.R. designed the wet-laboratory and computational protocols, performed wet-laboratory experiments, analyzed the data, and wrote the manuscript. S.L.Q. performed wet-laboratory experiments and wrote the manuscript. O.K analyzed the data. M.A.E., D.A.L., P.S., B.J. and A.G.E. provided intellectual input, and provided insights into data outputs. S.H.T. conceived the study, designed the wet-laboratory and computational protocols, analyzed the data, supervised the study, provided insights into data outputs and wrote the manuscript. A.K.S. and B.A.G. conceived the study, provided intellectual input, supervised the study and provided insights into data outputs. All authors revised and approved the final version of the protocol and manuscript.

Competing interests
P.C.S. is a cofounder of and shareholder in Delve Bio. P.C.S. is a co-founder and equity holder in Delve Biosciences and Lyra Labs, and a board member and equity holder in Polaris Genomics. P.C.S was formerly a co-founder of Sherlock Biosciences and board member of Danaher Corporation, until December 2024. A.G.E. reports compensation for consulting for Mirvie, Inc. and Merck Pharmaceuticals, and research funding was from Merck Pharmaceuticals. A.K.S. reports compensation for consulting and/or SAB membership from Honeycomb Biotechnologies, Cellarity, J&J, Ochre Bio, Bio-Rad Laboratories, Relation Therapeutics, IntrECate Biotherapeutics, Parabilis Medicines, Danaher, Quotient Therapeutics, Passkey Therapeutics and Dahlia Biosciences unrelated to this work.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-025-01308-8.

Correspondence and requests for materials should be addressed to Sergio Triana.

Peer review information *Nature Protocols* thanks Doron Betel, Ping Zhu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2026

nature portfolio

Corresponding author(s): Sergio H. Triana, Alex K. Shalek, Britt A. Goods

Last updated by author(s): Jul 16, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	Command line and R code created by the authors for this protocol are available at https://github.com/ShalekLab/MultiOmics_SeqWellS3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw sequencing data, as well as processed data for scRNA-seq, CITE-seq, and bulk RNA-seq data are deposited at GEO under GSE266385.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	D-eidentified human peripheral blood mononuclear cells were obtained from collaborators at the Ragon Institute by the Shalek Lab. IRB oversight and approval was provided by MIT (Protocol 1603517498).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size calculation was not performed. A sample size of three donors was utilized as we found that this was a good number for multiplexing on a single array to generate a complete dataset for the purposes of this paper.
Data exclusions	N/A
Replication	We included two technical replicates to show consistency in our findings for the method.
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	TotalSeq™-A Human Universal Cocktail, V1.0 (BioLegend, cat. no. 399907), TotalSeq™-A0251 Antibody anti-human Hashtag 1, (BioLegend, cat. no. 394601), TotalSeq™-A0252 Antibody anti-human Hashtag 2 (BioLegend, cat. no. 394603), TotalSeq™-A0253 Antibody anti-human Hashtag 3 (BioLegend, cat. no. 394605)
Validation	Antibodies used in this study were provided from BioLegend as a cocktail and all recommended procedures from the manufacturer were followed for their handling and use.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A