

[Dogma-seq doc for reference](#)

Massively parallel profiling of accessible chromatin and proteins with ASAP-seq

Eleni Mimitou¹, Peter Smibert¹, Caleb Lareau²

1. Immunai
2. Departments of Genetics and Pathology, Stanford University, Stanford, CA 94305, USA

mimitoueleni@gmail.com, smibertp@gmail.com, clareau@stanford.edu

Abstract

While methods such as the Assay for Transposase Accessible Chromatin by sequencing (ATAC-seq) enable a comprehensive characterization of regulatory DNA, additional measurements are required to characterize the multi-faceted nature of eukaryotic cells. Here, we delineate the ATAC with Selected Antigen Profiling by sequencing (ASAP-seq) protocol, a scalable approach to quantifying proteins via oligo-tagged antibodies alongside accessible DNA in thousands of single-cells. Critically, our method utilizes a custom bridge oligo that enables the utilization of a variety of oligo-conjugated antibodies, enabling the utilization and repurposing of other commercial products. The ASAP-seq method can be completed with straightforward experimental and computational modifications existing single-cell ATAC-seq workflows but yields distinct modalities underlying complex cellular states, including estimation of protein abundance on the cell surface as well as intracellular and intranuclear factors.

Key Words Multimodal, single-cell, protein, accessible chromatin, ATAC, intracellular, gene regulation

1. Introduction

The massively parallel measurement of chromatin accessibility and transcriptomes within single cells has catalyzed a rapidly-increasing number of studies that characterize cellular heterogeneity in biological systems. Specifically, droplet-based single-cell ATAC-seq (scATAC-seq) and scRNA-seq enable the comprehensive characterization of genome-wide chromatin accessibility and cellular polyadenylated transcripts in thousands of individual cells within a single experiment. Though these methods have enabled many fundamental insights of the biology underlying complex tissues, both scATAC-seq and scRNA-seq suffer from data sparsity, wherein most accessible loci and most genes are not measured in most cells. This sparse sampling of the underlying cellular features can complicate downstream analyses and inferences, particularly in identifying cell state features that delineate closely related cell types. Hence, there has been increasing importance of the development and application of single-cell multi-omic technologies that measure multiple modalities from individual cells.

To this end, several recent methods have combined targeted detection of select protein markers with scRNA-seq [1–4]. Conceptually, these methodologies synthesize decades of knowledge of specific cell types and states obtained by cytometry-based approaches with the mostly unbiased readout from the transcriptome. Furthermore, the pairing of a sparse genome-wide RNA measurement with a sensitive protein-based quantification for a smaller number of targets simultaneously enables both systematic discovery of genes associated with cellular phenotypes while retaining high-confidence inference for a selected subset of proteins. Importantly, recent work has established sophisticated computational algorithms that demonstrate that combining multiplexed protein detection with scRNA-seq resolves cell types better than either modality alone [5], reinforcing the utility of this single-cell multi-omic approach.

Chromatin accessibility is an additional modality that is now routinely used to characterize single cells in high-throughput by a variety of different single cell Assay for Transposase-Accessible Chromatin (ATAC-seq) approaches [6–8]. In many circumstances, such as development, scATAC may provide a more sensitive measure of the continuum of cell states, particularly in differentiation settings where epigenetic reprogramming may be the first mover [9]. However, many complications arise from the accessible chromatin measurements derived from scATAC-seq. First, per-cell sparsity tends to be more extreme due in part to ~5-10x more features (accessible chromatin peaks) than scRNA-seq (genes). Additionally, inferences of gene activity scores rely on the (weighted) summation of accessibility fragments overlapping or near gene bodies. However, as many regulatory elements overlap gene bodies

but control the expression of other loci, this method provides an imperfect determination of the genes that are actively transcribed, much less translated, in any given cell. Thus, fine-grained cell type identification from scATAC-seq data alone is channeling, often requiring complementary scRNA-seq data for high-quality annotations.

To remedy these issues with scATAC-seq data, we recently introduced ATAC with Select Antigen Profiling by sequencing (ASAP-seq) to combine robust detection of proteins with chromatin accessibility[10]. In practice, the ASAP-seq workflow builds on the mitochondrial scATAC-seq (mtscATAC-seq; see other chapter) method that enables scATAC-seq to be performed on whole cells [11] (**Figure 1**). As a consequence of the modified protocol, the cell remains intact for high-quality estimation of protein abundances and accessible chromatin with minimal modifications to commercial products. Notably, ASAP-seq enables a number of tunable options, including 1) the use of “hashing” to multiplex samples; 2) the use of different types of antibody reagents for protein detection; 3) the detection of intracellular proteins through minor modifications; and 4) the ability to either enrich or deplete reads derived from mitochondrial DNA, which can be used for inferring clonal relationships between cells in a sample [11].

We note one key feature that is conceptually distinct for the proteo-genomic capture in ASAP. Specifically, in contrast to other methods that use exogenous oligonucleotides to either report on protein abundance or enable sample multiplexing [1–4, 12, 13], the oligonucleotide sequences that read out protein levels in ASAP-seq do not directly interact with the barcoding reagents from the parent ATAC-seq assay. Instead, ASAP-seq employs a bridging oligo to convert existing labeling oligonucleotides into a format that is compatible with the ATAC-seq kit, providing enhanced flexibility for reagent use in ASAP-seq (**Figure 2**). Here, the specification of this part of the protocol has important implications for the accessibility and usability of the assay as the bridging oligo enables the immediate use of a large catalog of existing and available reagents as well as combinations of different specifications of reagents.

In this method description, we outline the foundational steps requisite for enabling single-cell multi-omic profiling with ASAP-seq technology[10]. We outline a synthesized experimental and computational workflow that provides flexibility to quantify proteins for downstream integrative analyses and identifies critical steps associated with quality control of libraries. Taken together, ASAP-seq enables the high-confidence quantification of selected intracellular and surface antigens while retaining the comprehensive discovery of accessible chromatin loci and clonality underlying cells.

2. Materials

2.1 Cell processing, staining, fixation and lysis

1. Phosphate buffered saline (PBS) (any provider)
2. CITE-seq staining buffer: 2% BSA, 0.01% Tween in PBS
3. Human TruStain FcX™ (BioLegend 422301)
4. TotalSeq™-A or TotalSeq™-B oligo-labeled antibody reagents (individually or as panels) (BioLegend - see [note¹](#))
5. FACS buffer: PBS with 1% FBS. Filtered at 0.45µm, store at 4°C
6. DAPI (any provider, for example BioLegend 422801)
7. Formaldehyde, 16% (any provider, for example Thermo Fisher 28906)
8. Glycine solution, 2.5M (any provider, for example Ricca Chemical RMB19103-50C2)
9. Tris-HCl pH=7.5, 1M (any provider, for example Sigma-Aldrich T2194)
10. NaCl, 5M (any provider, for example Sigma-Aldrich 59222C)
11. MgCl₂, 1M (any provider, for example Sigma-Aldrich M1028)
12. NP40, 10% (Sigma-Aldrich, 74385)
13. Tween 20, 10% (Bio-Rad, 1662404)
14. Digitonin 5% (Thermo Fisher, BN2006)
15. BSA, 10% (any provider, for example Miltenyi Biotec 130-091-376)
16. Intracellular staining buffer (BioLegend, custom part number 900002577). Supplement with fresh DTT before use to 1mM final concentration
17. True-stain monocyte blocker (BioLegend, 426101)
18. DTT, 1M (any provider, for example Sigma-Aldrich 646563)
19. Flowmi Cell Strainer 40µm (Bel-Art, H13680-0040)
20. Bridge oligo A (BOA) or Bridge oligo B (BOB) (IDT, or other provider, see Table 1, [note²](#))

¹ While planning an experiment the first step is to choose a family of protein detection reagent. We recommend one of two types of TotalSeq™ reagents from BioLegend: TotalSeq™-A or TotalSeq™-B. These are commonly used for CITE-seq[2] or 10x Genomics “feature barcoding” with 3’ scRNA-seq, respectively. While it is theoretically possible to couple protein detection using TotalSeq™-C reagents to scATAC with a specifically designed bridge oligo, the sequence characteristics of the TotalSeq™-C species make bridge oligo annealing inefficient, and more importantly, the amplification handle on TotalSeq™-C reagent make it impossible to separately amplify the protein tag library from the ATAC library.

² Purchase the specific bridge oligo for the TotalSeq™ reagents you will use. Note that each family of totalseq reagent necessitates several dependent downstream reagents / processing steps (see tables 1 and 3). While we have not used antibody:oligo reagents other than TotalSeq™ products, the bridge oligo strategy, with appropriate design modifications, should be compatible with other products (for example, BD AbSeq reagents, 10x CellPlex reagents)

21. Indexing primers (IDT or other provided, see Table 1)

2.2 ASAP-seq library preparation

1. 10x Genomics Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit, 16 or 4 rxns
2. 10x Genomics Chromium Next GEM Chip H Single Cell Kit, 48 or 16 rxns
3. 10x Genomics Single Index Kit N, Set A, 96 rxns
4. 2x Kapa Hifi PCR mastermix
5. SPRI beads (AMPure XP beads or KAPA Pure beads)
6. Custom oligonucleotides for library prep (see Table 1)

2.3 Quality control and sequencing

1. Qubit dsDNA HS Assay Kit (Thermo Fisher Q32851 or Q33230)
2. Agilent Bioanalyzer High Sensitivity DNA Analysis Kit (or TapeStation or similar)
3. KAPA Library Quantification Kit for Illumina® Platforms (KAPA biosystems KK4835)
4. Illumina NovaSeq or NextSeq reagent kits

2.4 Software and references needed for computational analysis

1. Download cellranger-atac and relevant reference files.
(<https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac>). The most up-to-date reference files and versions of the software are available online. This software will be used to demultiplex sequencing libraries from an Illumina sequencing run and can be executed to process (See **Note³**)
2. Install an up-to-date version of the Python 3 library either for the system, the user, or through a conda environment (See **Note⁴**)

³ Recent versions of the software and commonly used reference genomes are available on the 10x support website <https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/installation>. Certain applications may require assembling a custom reference, particularly for cell inputs <https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/advanced/references>.

⁴ Some dependency packages are also required to run the workflow, depending on the exact use case, and are documented alongside the complementary tools.

3. Download the kite antibody tag preprocessing toolkit. The most up-to-date version of the software is available online at <https://github.com/pachterlab/kite>. This software is used to build a reference map of the oligonucleotide barcodes to the respective antibody clones.
4. Download the kallisto and bustools software binaries. Current versions of these software are available at <https://github.com/pachterlab/kallisto> and <https://github.com/BUStools/bustools>, respectively. These software are utilities used to efficiently count reads assigned to each antibody barcode for every cell while efficiently correcting for sequencing errors
5. Download the ASAP to kite script toolkit available here: https://github.com/caleblareau/asap_to_kite. This code is required to convert the ASAP-seq sequencing data into a format that are compatible with the existing kite | kallisto | bustools workflows. (See **Note⁵**)
6. mgatk package and dependencies (<https://github.com/caleblareau/mgatk>)
7. 10x scATAC barcode whitelist:
\$ wget https://teichlab.github.io/scg_lib_structs/data/737K-cratac-v1.txt This file is available in the distribution of CellRanger-ATAC but is more accessible from the indicated GitHub link.

3. Methods

3.1 Cell preparation, fixation and permeabilization

This section outlines the steps required to stain the cells with the conjugated antibodies, followed by fixation and permeabilization. The fixation steps are based on the mtscATAC-seq workflow (see separate chapter describing mtscATAC-seq in the same issue). Permeabilization can be performed using two alternative lysis buffers; LLL (Low Loss Lysis) and OMNI (based on OMNI-ATAC protocol; [14]), which is the default lysis buffer in the 10x Genomics scATAC kit. LLL is the lysis buffer described in mtscATAC kit, which due to lack of Tween 20 in its formulation, retains mtDNA fragments in the ATAC library that can be used for mtDNA variant tracing. In benchmarking experiments, either LLL or OMNI buffers yielded comparable ATAC

⁵ The current version of the asap_to_kite toolkit contains custom python scripts for performing this task of reformatting sequencing data. Depending on the library input (either TotalSeqA, TotalSeqB, a mix), these software will have to be run with custom parameters. See the github repository for more details.

and protein data and can be used interchangeably if mtDNA retention is not desired.

3.1.1 Cell staining

1. Obtain single cell suspensions (filter if needed) and measure viability and density. If viability is lower than 80% proceed with live cell enrichment and/or use best judgement depending on sample source / importance / cell numbers.
2. Resuspend 1-2 million cells in 100 μ l CITE staining buffer.
3. Add 10 μ l Fc Blocking reagent.
4. Incubate for 10 minutes at 4°C .
5. While cells are incubating in Fc Block, prepare the antibody pool (panel or titrated amounts).
6. Add antibody-oligo pool to cells.
7. Incubate for 30 minutes at 4°C .
8. Wash cells 3 times with 1 mL CITE staining buffer, spin at 300g for 5 min at 4°C for every wash to harvest cells.
9. Resuspend cells in 450 μ l **room temperature** PBS.

3.1.2 Cell fixation and permeabilization

1. Use about 0.5-1 million cells in 450 μ l PBS for the fixation reaction
2. Add 30 μ l 16% formaldehyde (1% final concentration), mix by pipetting and incubate at room temperature for 10 minute with occasional inversion.
3. Quench by adding glycine to final concentration 0.125 M
4. Wash with 1x ice-cold PBS by filling up the tube, invert 5 times
5. Spin at 400g, 5 minutes at 4°C
6. Discard supernatant and repeat wash with 1ml 1x ice-cold PBS
7. Spin 5 minutes 400g at 4°C , discard supernatant.
8. Resuspended cell pellet in 100 μ l chilled lysis buffer (LLL or OMNI buffer, Table 2, **see note⁶**), mix by pipetting.
9. Incubate on ice, 3 min for primary cells, 5 min for cell lines
10. Add 1 ml chilled wash buffer to the lysed cells, mix by pipetting
11. Spin at 500g, 5 min at 4°C. **If intracellular staining is desired go to section 3.1.3**
12. Remove supernatant, resuspend in 150 μ l 1x nuclei buffer (10x Genomics)

⁶ If mtDNA retention is desired, use LLL lysis buffer

13. Filter through 40 µm strainers. If the cell number is low, skip this step.
14. Count cells and adjust density according to 10x loading instructions.
15. Proceed to section 3.2

3.1.3 Intracellular staining

1. Resuspend cell pellet from step 3.1.2.11 in 40 µl intracellular wash buffer
2. Add 5 µl of FcX and 5 µl of monocyte block solution
3. Incubate on ice for 15 min
4. Add 50 µl of intracellular wash buffer, containing titrated amounts of conjugated intracellular markers (see **Note⁷**), incubate on ice for 30 min
5. Wash 3 times with the intracellular wash buffer, spin at 500g for 5 min at 4°C.
6. Remove supernatant, resuspend in 150 µl 1x nuclei buffer (10x Genomics)
7. Filter through 40 µm strainers. If the cell number is low, skip this step.
8. Count cells and adjust density according to 10x loading instructions.
9. Proceed to section 3.2

3.2 Transposition and barcoding

For this step proceed according to 10x Genomics Single Cell ATAC protocol (CG000168 Rev D for v1 and CG000209 Rev D for v1.1) with the below modifications:

1. During the barcoding reaction (step 2.1) spike in 0.5 µl of 1 µM bridge oligo. There is no dead volume in the reaction, so final volume will be 65.5 µl for v1 and 60.5 µl for v1.1
2. During GEM incubation (step 2.5), add a 5 min incubation at 40°C at the beginning of the protocol (see **Note⁸**). Incubation protocol: 40°C 5 min, 72°C 5 min, 98°C 30 sec, 98°C 10 sec, 59°C 30 sec, 72°C 1 min, cycle 12 times total, hold at 15°C.
3. During silane bead elution (step 3.1o) add 43.5 µl of Elution Solution I and subsequently recover 43 µl. Keep 3 µl aside to use as input (see **Note⁹**) in the tag library PCR and with the remaining 40 µl proceed to SPRI cleanup as per 10x protocol.

⁷ So far we have used about 0.5-1 µg of antibodies during the intracellular staining

⁸ this extra step is not essential when using TSA products, but increases efficiency in TSB capture

⁹ you can use either as input in the tag indexing reaction or combine when working with large antibody panels to increase input complexity.

4. During SPRI cleanup (step 3.2d), save the supernatant. For the bead bound fraction proceed as per 10x protocol. For the supernatant fraction, add 32 μ l SPRI, let bind for 5 min. Collect beads on magnet, wash twice with 80% EtOH, remove remaining ethanol and elute beads in 42 μ l EB (see **Note⁹**). This can be combined with the 3 μ l left aside after the silane purification, as input in the TSA/TSB indexing reaction:

50 μ l 2x KAPA mix
2.5 μ l primer P5 10 μ M
2.5 μ l indexing primer 10 μ M (RPxx or D7xx, see Table 2)
3-45 μ l input fragments (see note⁹)
100 μ l total

Incubation protocol: 95°C 3 min, 95°C 20 sec, 60°C 30 sec, 72°C 20 sec, 72°C 5 min, cycle 14-16 times total, hold at 4°C.

5. Proceed with indexing the ATAC library as described in section 4.2 of the 10x protocol. Usually 10 cycles provide sufficient material to perform library QC and sequencing. If native nuclei are run in parallel, a noticeable reduction in PCR yield can be observed with the fixed sample compared to native nuclei (presumably due to fixation).

3.3 Library QC, pooling and sequencing

3.3.1 Library QC:

We recommend to quantify all libraries in three sequential steps:

1. Qubit: use 1 μ l undiluted library for total nucleic acid mass
2. Fragment analyzer (see examples in **Figure 3**): preferably Agilent BioA (if not available TapeStation or PerkinElmer LabChip GX can be used). Run ~1-3ng of each library based on the Qubit read. The fragment analyzer will provide the size distribution of the library and a more accurate quantification of the expected-for-each-library size fragment/population.
3. KAPA qPCR: prepare 4nM dilutions of each library based on quantification by BioA and record dilution. Follow the KAPA manual instructions for quantification of the 'clusterable'

fragments (fragments containing P5/P7 sequences). This will be the most accurate concentration read for sequencing purposes (see [Note¹⁰](#)).

3.3.2 Pooling and sequencing:

Prepare equimolar concentrations of each library (concentration requirement will depend on instructions from the sequencing facility) and pool together in ratios that will satisfy sequencing depth requirements.

We recommend a minimum depth of:

- 30K reads/cell for the ATAC library
- 3K reads/cell for TSA/TSB hashtag libraries
- 150 (?) reads per antibody/cell for the surface panel. As a rule of thumb we allocate ~5-10K reads for panels up to 50 antibodies, 10-20K reads for panels up to 150 antibodies and >25K reads for panels >200 antibodies.

ATAC and TSA/TSB libraries can be sequenced on the same flow cell. We recommend that protein tag libraries should not occupy more than ~50% of the flow cell because beyond the first 10-15nt both Rd1 and Rd2 enter a low-diversity region (polyT/A or 10x capture sequence), resulting in a decreased data quality that can negatively impact ATAC fragment mapping. However, we note that we have not systematically evaluated relative loading abundances for the ATAC and protein tag library. We have used the Illumina NextSeq and NovaSeq reagents kits and respective sequencing platforms. A minimum of 75-cycle kit with recipe: [34, 8, 16, 34] is sufficient if you are not intending to retain mtDNA reads. For experiments that plan to retain mtDNA for genotyping, we recommend using longer reads to obtain high coverage of the mitochondrial genome for variant calling. In this setting, we typically utilize a 150-cycle kit with a [72, 8, 16, 72] recipe.

3.4 Demultiplex sequencing data

This section briefly summarizes the steps needed to demultiplex sequencing data to generate paired-end sequencing data associated with all libraries on the flow cell. In the ASAP-seq multimodal workflow, chromatin accessibility (and optionally mtDNA) are captured in the same

¹⁰ If KAPA qPCR is not an available option, use the molarity of the expected fragments as measured by BioA

library, whereas different libraries per protein modality (hash tags, protein abundance, different oligonucleotide backgrounds) will be present on distinct sequencing libraries.

1. Build the sample sheet csv that specifies the indices for both the ATAC library and the tag library (see [Note¹¹](#)).
2. Demultiplex sequencing data by running `cellranger-atac mkfastq`. For example,
`$ cellranger mkfastq --id=asap_seq_demux --run=/path/to/flow_cell --csv=sample_sheet.csv`
3. The resulting output folder (`asap_seq_demux`) will contain `.fastq.gz` files for both the scATAC and protein tag libraries.

3.5 Process sequencing data

This section outlines the steps to take raw sequencing data and generate counts matrices of features per cell. As a reference, we include **Box 1** that contains a summary of values from a real-world ASAP-seq library that was processed with the outlined workflow. In **Table 4**, we provide context for idea values associated with various steps in this pipeline, including ideas for debugging executions that don't meet quality control standards.

1. Build a mismatch aware antibody barcode map using kite. (see [Note¹²](#))
`$ python kite/featuremap/featuremap.py FeatureBarcodes.csv --header`
2. Build a kallisto index from the mismatch aware `.fasta` file produced by kite:
`$ kallisto index -i FeaturesMismatch.idx -k 15 FeaturesMismatch.fa`
3. For convenience in processing, define a bash variable related to the specific library/sample to run the subsequent steps: **`$ sample="ASAP_tag_Sample_ID"`**
4. Reformat sequencing reads for compatibility with kallisto using the `asap_to_kite` toolkit. A summary of the transformation of the data is shown in **Figure 4** and will depend on the antibody library used for the exact hyperparameters (see Table 3). **`$ python asap_to_kite_v2.py -f asap_seq_demux/flow_cell -s $sample -o "${sample}_pro"`**

¹¹ An online tool to facilitate building the sample sheet is available: <https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/using/bcl2fastq-direct>. We note that the index used for the tag libraries will not be available from the tool and must be entered manually.

¹² By default, the kite tool produces an off-by-one mismatch k-mer dictionary. While using the kallisto tool for read mapping, there is no

5. Pseudo-align the antibody barcode sequences to the established reference. **\$ kallisto bus -i FeaturesMismatch.idx -x 10xv2 -t 8 -o \$sample "\${sample}_pro_R1.fastq.gz" "\${sample}_pro_R2.fastq.gz"**
6. Correct for sequencing errors in the **\$ bustools correct -w 737K-cratac-v1.txt "\${sample}/output.bus" -o "\${sample}/output_corrected.bus"**
7. Sort the bus file. **\$ bustools sort -o "\${i}/output_sorted.bus" "\${i}/output_corrected.bus"**
8. Generate the protein x cell counts matrix. **\$ bustools count -o \${sample} -- genecounts -g FeaturesMismatch.t2g -e "\${sample}/matrix.ec" -t "\${sample}/transcripts.txt" "\${sample}/output_sorted.bus"**
9. Assess the number of unique UMIs captured **\$ bustools text "\${sample}/output_sorted.bus" -p | wc -l**
10. The resulting files, including **"\${sample}.mtx"**, **"\${sample}.barcodes.txt"**, and **"\${sample}.genes.txt"** provide a matrix markdown representation of the counts matrix.
11. In parallel, libraries for accessible chromatin can be processed with appropriate bioinformatics pipelines, such as cellranger-atac count. (see [Note¹³](#))
12. For ASAP-seq libraries where mtDNA was also captured, additional processing of the mitochondrial genome is facilitated by the mgatk package, which takes the outputs from the cellranger-atac count to generate per cell heteroplasmy estimates and a determination of high-confidence variants that may be useful for clonal lineage tracing. We refer to the mtscATAC-seq chapter for additional information on processing these mtDNA variants from the ASAP-seq libraries.

3.6 Perform multimodal analysis

For downstream analyses, we recommend utilizing the Seurat/Signac toolkit for multi-modal analyses of ASAP-seq data. While other tools for scATAC-seq analyses, such as ArchR, provide flexible metadata annotations that facilitate incorporating protein levels and mtDNA variants in broad cellColData slots, there is currently limited out-of-the-box functionality for the multi-modal capabilities derived from ASAP-seq. Conversely, the Seurat/Signac toolkit facilitates

¹³ The execution of this software can be performed modularly without information from the antibody tag libraries. Other single-cell ATAC preprocessing workflows can also be utilized at this point.

the normalization of protein abundances via the centered-log-ratio (CLR) transformation (used before annotation and association analyses with protein data) and complete suite of functions needed for analysis of the accessible chromatin arm of ASAP-seq. Notably, a comprehensive code repository of downstream analyses of ASAP-seq data, including tri-modal analyses of mtDNA, proteins, and chromatin accessibility and integrative analyses with CITE-seq data, is available online: https://github.com/caleblareau/asap_reproducibility. We refer to this code repository resource for more involved analysis vignettes but note the key functions required for downstream / multimodal analyses of ASAP-seq data.

1. Import chromatin accessibility counts via the customary import functions in the Signac package.

```
counts <- Read10X_h5(filename = "filtered_peak_bc_matrix.h5")
```

2. Create a Seurat object via the ChromatinAssay functionality

```
seurat_obj <- CreateSeuratObject(counts = CreateChromatinAssay(counts, ...), ...)
```

3. Import the output counts data from kite using the **read.table** or **fread** function into a data matrix object (**kite_counts**). Append this matrix to the Seurat object for just the cells that are present in the object.

```
seurat_obj[["ADT"]] <- CreateAssayObject(counts =  
  kite_counts[,colnames(seurat_obj)])
```

4. The resulting **seurat_obj** should have multiple assay slots that will enable both scATAC-based and protein-based analyses. We refer the reader to appropriate scATAC (https://satijalab.org/signac/articles/pbmc_vignette.html) and protein (https://satijalab.org/seurat/articles/multimodal_vignette.html) vignettes, respectively.
5. Generally speaking, we have elected to analyze our ASAP-seq data utilizing the chromatin accessibility counts for feature reduction, clustering, and two-dimensional projections while keeping the protein data as an independent annotation of cell features and cell states. However, we've also found that utilizing bioinformatics methods, like weighted nearest neighbor (WNN) can enhance dimensionality reduction [5]. This can be easily accomplished using the existing WNN tutorial but utilizing the latent semantic indexing (LSI) rather than principal components analysis (PCA) dimensions from here: https://satijalab.org/seurat/articles/weighted_nearest_neighbor_analysis.html

4. Notes

<in line>

Acknowledgements

C.A.L. is supported by a Stanford Science Fellowship and a Parker Institute of Cancer Immunotherapy Scholarship.

Display items

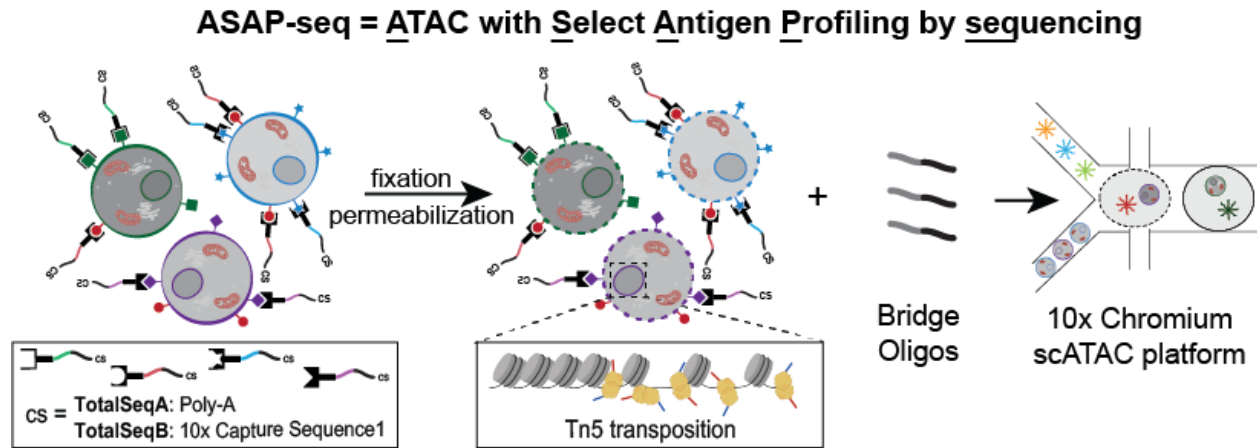


Figure 1. Schematic of the experimental assay. ASAP-seq allows whole cell input into the scATAC-seq workflow, maintaining the connection between nuclear content and cell surface marker information. Cells are stained with oligo-conjugated antibodies followed by fixation, permeabilization and Tn5 transposition. Bridge oligos are spiked in the barcoding mix prior to droplet formation to allow simultaneous barcoding of ATAC fragments and antibody-derived oligos.

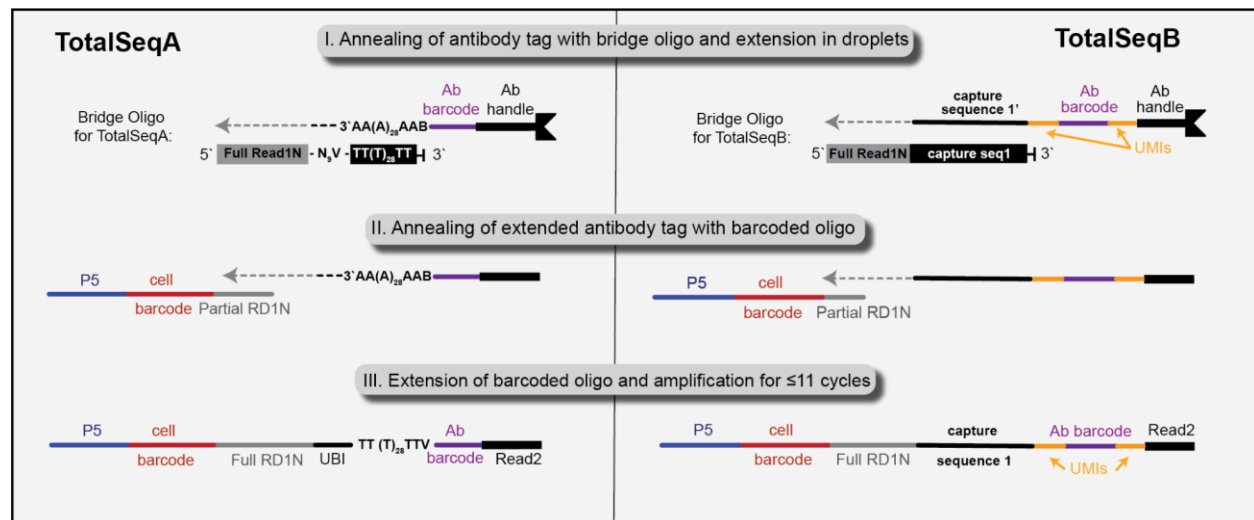


Figure 2. Barcoding scheme of the protein tags using the bridge oligo strategy. Bridge oligo A (BOA) and bridge oligo B (BOB) function as templates to extend the protein-derived oligos in droplets. While TSB tags (right) contain UMIs, UBIs (N₉V) are introduced to TSA tags via the bridge oligo (left) to allow molecule counting.

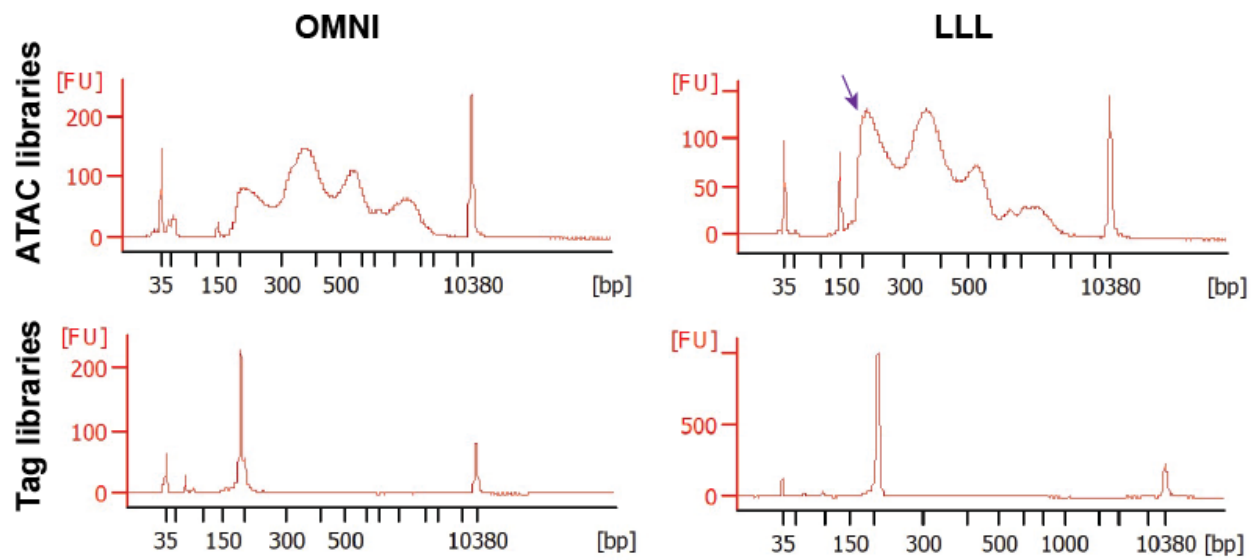


Figure 3. Representative fragment analyzer traces of the sequencing libraries. ATAC (top) and protein tag (bottom) libraries of fixed human PBMCs permeabilized with OMNI lysis buffer (A) or LLL lysis buffer (B). Note the increased abundance of the nucleosome free region (size <300 bp) in the LLL library that corresponds to the increased capture of mtDNA fragments (arrow).

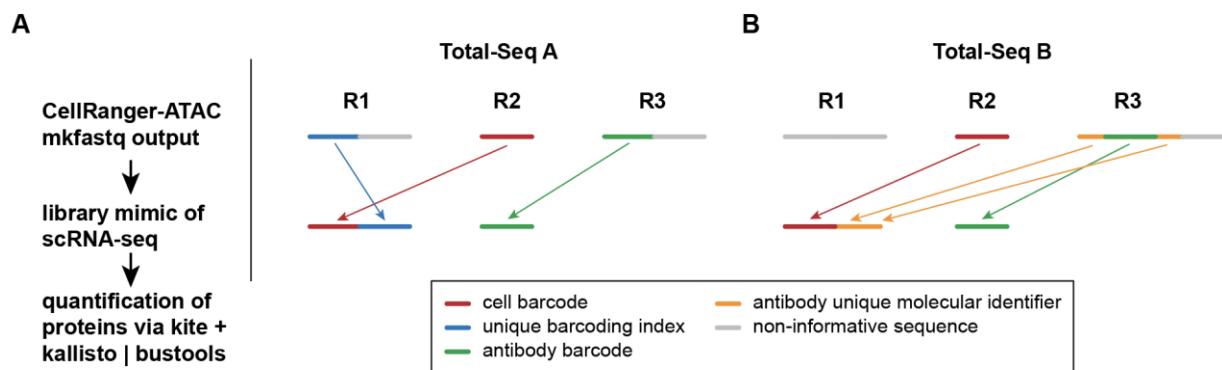


Figure 4. Schematic of library structure and computational preprocessing for ASAP-seq tag libraries. Colors represent specific technical attributes of the read library. Colored arrows represent the data transformations in the `asap_to_kite.py` tool. The resulting fastq files mimic scRNA-seq data and can be used in `kallisto | bustools` for single-cell protein abundance estimation. For Total-Seq B, both UMIs can be used in the mapping abundance but requires the execution of “`kallisto bus`” with custom parameters.

Step 3.5.5: Pseudoalignment

processed 9,809,110 reads, 8,498,079 reads pseudo aligned

Step 3.5.6: Correct

Found 737,280 barcodes in the whitelist

Number of hamming dist 1 barcodes = 20,309,952

Processed 8,498,079 bus records

In whitelist = 8,049,279

Corrected = 114,962

Uncorrected = 333,838

Step 3.5.7: Sort

Read in 8,164,241 BUS records

Step 3.5.9: Text

Read in 6,120,282 BUS records

Box 1. Example of quality metrics from running steps in section 3.5. Key metrics are indicated after the associated computational step.

Table 1: Oligo sequences		
Oligo	Sequence (shown 5'>3')	Notes
BOA	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGNNNNNNNNVTTTTTTTTTTTTTTTT TTTTTTTTTTTT/3InvdT	*Used to bridge TSA tags *3' modification to block extension *Brings a 10-nt UBI ending in V (non-T)
BOB	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTTGCTAGGACCGGCCTTAAAGC/3I nvdT/	*Used to bridge TSB tags *3' modification to block extension

P5	AATGATACGGCGACCACCGA	*Forward primer to amplify TSA and TSB tags
P7	CAAGCAGAAGACGGCATAACGAGAT	*Reverse primer to re-amplify already indexed tag libraries (optional)
RPxx	CAAGCAGAAGACGGCATAACGAGATxxxxxx xxGTGACTGGAGTTCCTTGGCACCCGAGA ATTCCA	*TruSeq Small RNA indexing primer, used to index TSA tags
D7xx	CAAGCAGAAGACGGCATAACGAGATxxxxxx xxGTGACTGGAGTTCAGACGTGTGC	*TruSeq DNA indexing primer, used to index TSB tags or TSA hashtags

Table 2: Permeabilization buffers			
Materials	Prepare fresh, keep on ice until use (see note³)		
	LLL Lysis buffer	OMNI Lysis Buffer	Wash buffer
1M Tris-HCl pH = 7.5	10 mM	10 mM	10 mM
5M NaCl	10 mM	10 mM	10 mM
1M MgCl ₂	3 mM	3 mM	3 mM
10% NP40	0.1%	0.1%	-
10% Tween 20	-	0.1%	-
5% Digitonin	-	0.01%	-

Table 3. Linking experimental reagents to downstream bioinformatics workflows				
Antibody family	Compatible Bridge oligo	Indexing PCR primer	Read configuration	asap_to_kite parameter

Totalseq-A (not hashtags)	BOA	RPxx	Fig. 4a	-j TotalSeqA (default)
Totalseq-A hashtags	BOA	D7xx	Fig. 4a	-j TotalSeqA (default)
Totalseq-B (including hashtags)	BOB	D7xx	Fig. 4b	-j TotalSeqB

Table 4. Interpretation of quality control metrics from ASAP-seq protein mapping			
Processing step	Metric	Target value	Debugging workflow
Step 3.5.5. Pseudoalignmen t ¹⁴	Antibody barcode pseudo- alignment rate	85%	1) Verify indicated reference matches the experimental input. 2) Verify the correct specification of the antibody library (e.g. TSA or TSB). 3) Run FastQC to look for overrepresented sequences that may correspond to known barcodes 4) If the FastQC quality-control report from the demultiplexed flow cell indicates poor base quality at bases associated with the antibody barcode (see Figure 4), consider using a two-mismatch barcode dictionary in step 3.5.2 .
<i>Calculation:</i> # pseudo-aligned / # reads * 100% <i>Example:</i> 8,498,079/9,809,110 * 100% = 86.6%			

¹⁴ By virtue of the kite reference, every kmer up to 1 mismatch will be accounted for and then collapsed during the quantification step. In this sense, though kallisto is a “pseudo-alignment” algorithm, the quantifications are absolute and effectively a fast dictionary-based quantification

Step 3.5.6. Correct	Bead barcode alignment rate	90%	<ol style="list-style-type: none"> 1) If very low (< 5%), check to see if the R2 was correctly handled for a reverse-complement 2) Check corresponding scATAC-seq library for barcode alignment rate 3) Examine top sequences for contamination or other repetitive sequences.
<p><i>Calculation:</i> $(\# \text{ whitelist} + \# \text{ corrected}) / \# \text{ pre-corrected records} * 100\%$</p> <p><i>Example:</i> $(8,049,279 + 114,962) / 8,498,079 * 100\% = \mathbf{96.1\%}$</p>			
Step 3.5.9. Text	UMI saturation rate	25%-50%	<ol style="list-style-type: none"> 1) This metric isn't interpretable if the top two values are not of reasonable quality 2) If >50%, sequencing is saturated and may represent a low-quality library (unless purposefully sequenced to saturation) 3) If <25%, additional sequencing is recommended
<p><i>Calculation:</i> $[1 - (\# \text{ Final UMIs} / \# \text{ pre-sorted records})] * 100\%$</p> <p><i>Example:</i> $[1 - (6,120,282 / 8,164,241)] * 100\% = \mathbf{25.0\%}$</p>			

References

1. Mimitou EP, Cheng A, Montalbano A, et al (2019) Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat Methods* 16:409–412
2. Stoeckius M, Hafemeister C, Stephenson W, et al (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 14:865–868
3. Peterson VM, Zhang KX, Kumar N, et al (2017) Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol* 35:936–939
4. Triana SH, Vonficht D, Jopp-Saile L, et al (2021) Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *bioRxiv*
5. Hao Y, Hao S, Andersen-Nissen E, et al (2021) Integrated analysis of multimodal single-cell data. *Cell* 184:3573–3587.e29
6. Satpathy AT, Granja JM, Yost KE, et al (2019) Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* 37:925–936
7. Lareau CA, Duarte FM, Chew JG, et al (2019) Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol* 37:916–924
8. Cusanovich DA, Daza R, Adey A, et al (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348:910–914
9. Ma S, Zhang B, LaFave LM, et al (2020) Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* 183:1103–1116.e20
10. Mimitou EP, Lareau CA, Chen KY, et al (2021) Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol*. <https://doi.org/10.1038/s41587-021-00927-2>
11. Lareau CA, Ludwig LS, Muus C, et al (2020) Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat Biotechnol*. <https://doi.org/10.1038/s41587-020-0645-6>
12. Stoeckius M, Zheng S, Houck-Loomis B, et al (2018) Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol* 19:224
13. McGinnis CS, Patterson DM, Winkler J, et al (2019) MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* 16:619–626
14. Corces MR, Trevino AE, Hamilton EG, et al (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 14:959–962