

SHORT REPORT

Open Access



Comprehensive benchmarking of CITE-seq versus DOGMA-seq single cell multimodal omics

Zhongli Xu^{1,2}, Elisa Heidrich-O'Hare³, Wei Chen^{1,4*} and Richard H. Duerr^{3,5*} 

* Correspondence: wec47@pitt.edu; duerr@pitt.edu

¹Department of Pediatrics, University of Pittsburgh School of Medicine, UPMC Children's Hospital of Pittsburgh, 4401 Penn Ave, Pittsburgh, PA 15224, USA

³Department of Medicine, University of Pittsburgh School of Medicine, 3550 Terrace Street, Pittsburgh, PA 15261, USA

Full list of author information is available at the end of the article

Abstract

The recently developed method TEA-seq and similar DOGMA-seq single cell trimodal omics assays provide unprecedented opportunities for understanding cell biology, but independent evaluation is lacking. We explore the utility of DOGMA-seq compared to the bimodal CITE-seq assay in activated and stimulated human peripheral blood T cells. We find that single cell trimodal omics measurements after digitonin (DIG) permeabilization were generally better than after an alternative “low-loss lysis” (LLL) permeabilization condition. Next, we find that DOGMA-seq with optimized DIG permeabilization and its ATAC library provides more information, although its mRNA and cell surface protein libraries have slightly inferior quality, compared to CITE-seq.

Background

Massively parallel single cell omic measurements have advanced from single modal measurement of transcriptomes alone; to bimodal, simultaneous transcriptome and cell surface protein epitope measurements (CITE-seq [1] and REAP-seq [2]), transcriptome and chromatin accessibility measurements (sci-CAR [3], SNARE-seq [4], SHARE-seq [5]), and chromatin accessibility and cell surface protein epitope measurements (ASAP-seq [6] and ICICLE-seq [7]); to trimodal, simultaneous transcriptome, cell surface protein epitope, and chromatin accessibility measurements (DOGMA-seq [6] and TEA-seq [7]). DOGMA-seq and its sibling technique TEA-seq provide unprecedented opportunities to study complex cellular and molecular processes at single cell resolution, but a comprehensive independent evaluation is needed to compare these new trimodal assays to existing single modal and bimodal assays. In this study, we benchmarked and compared DOGMA-seq under two permeabilization conditions, as well as DOGMA-seq versus CITE-seq, using aliquots of the same ex vivo tissue culture activated and stimulated human peripheral blood T cell populations.



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Results and discussion

The DOGMA-seq and TEA-seq protocols require a cell permeabilization step following hashtag oligonucleotide (HTO) [8] and cell surface protein epitope antibody-derived tag (ADT) [1] labeling to enable preparation of its assay for transposase-accessible chromatin (ATAC) library. Alternative cell permeabilization conditions using either digitonin (DIG) or paraformaldehyde fixation with “low-loss lysis” (LLL) were introduced with the development of the DOGMA-seq [6] and TEA-seq [7] assays. Theoretically, DIG permeabilization is milder and will not lyse mitochondrial membranes, because it interacts with cholesterol, and the plasma membrane has a higher cholesterol content compared to the mitochondrial membrane [9]. We optimized the DIG condition used by TEA-seq developers [7], where cells were incubated with 0.01% DIG on ice for 5 min (min), and identified an optimal incubation time and concentration of DIG for permeabilization of peripheral blood mononuclear cells (PBMCs). We had previously determined that incubation with 0.01% DIG for 1 min was sufficient to permeabilize all cells based on acridine orange and propidium iodide (AOPI) staining. We then proceeded to test several dilutions of DIG (0.0025%, 0.005%, 0.0075%, and 0.01%) and found that incubation with 0.0075% DIG for 1 min was an optimal permeabilization condition, since 99.5% of the cells treated with 0.0075% DIG showed red fluorescence indicative of propidium iodide entry into cells with compromised membranes, but lesser concentrations of DIG resulted in lower proportions of permeabilized cells.

Next, we evaluated single cell trimodal omics measurements after our optimized DIG permeabilization condition compared to the alternative LLL condition (Fig. 1A). We collected PBMC from a healthy donor, enriched untouched T cells, activated and stimulated them under four different ex vivo tissue culture conditions, and then split each cell population into DIG and LLL permeabilization groups. Quality control metric comparisons between the two permeabilization groups are shown in the figures (Fig. 1B–I). The number of genes detected per cell and TSS enrichment scores are similar after DIG and LLL permeabilization, while DIG resulted in higher protein tag complexity and fraction of mtRNA but a lower fraction of mtDNA. These observations were consistent with those reported by DOGMA-seq developers [6]. However, we observed higher ATAC complexity but a lower fraction of unique molecular identifiers (UMIs) mapping to exons after DIG permeabilization, in contrast with the reported observations [6].

We calculated the fraction of cells with UMIs > 0 for all 163 protein tags in the Total-Seq™-A Human Universal Cocktail, V1.0 (BioLegend) to evaluate protein tag detection rate. As expected, we found higher detection rates for all 163 protein tags after DIG permeabilization (Fig. 1J), which is consistent with better preservation of the plasma membrane and cell surface proteins after DIG permeabilization. However, the differences in cell surface protein detection between the two permeabilization conditions did not attenuate differential protein tag signals upon differential stimulations at the pseudo-bulk level. For example, when we compared two groups of T cells that were both activated and cultured with IL-1 β and IL-23, and one of the two groups was also cultured with prostaglandin E2 (PGE2), we found high correlation of differential protein tag signals between the two permeabilization conditions (Fig. 1K).

To evaluate gene detection rates, we calculated the fraction of cells with UMIs > 0 for all 36,601 genes in the reference genome and found slightly higher detection rates for

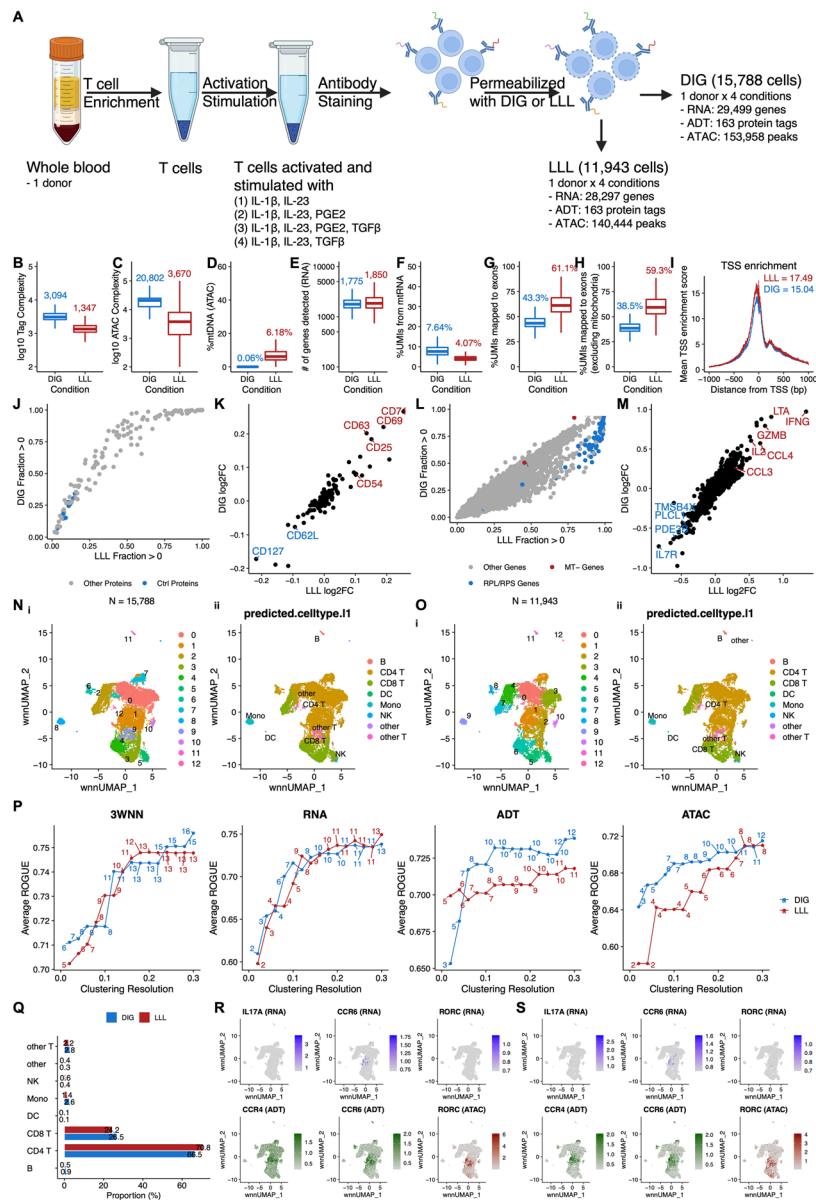


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Comparison between DIG and LLL conditions. **A** Overview of study design for comparison between DIG and LLL conditions. Each of four aliquots of T cells from a human donor were activated and stimulated under a different stimulation condition (total of four stimulation conditions in four tissue culture wells) in a 12-h tissue culture. Each of two aliquots of cells from each of the four tissue culture wells were then labeled with a unique hashtag (total of eight unique hashtags). Approximately equal numbers of the eight uniquely hashtagged cell populations were then pooled and labeled with antibody cocktail. An aliquot of the hashtag- and antibody cocktail-labeled pool of cells was permeabilized with DIG, and another aliquot of the hashtag- and antibody cocktail-labeled pool of cells was permeabilized with LLL prior to the subsequent DOGMA-seq library preparation steps. **B–H** Boxplot showing quality control metric comparisons between DIG and LLL conditions. Median values are indicated with corresponding colors. **B** Protein tag complexity per cell. Median ratio = 2.30. **C** ATAC fragment complexity per cell. Median ratio = 5.67. **D** Percentage of ATAC fragments mapped to mtDNA per cell. Median ratio = 0.01. **E** Number of genes per cell. Median ratio = 0.96. **F** Percentage of UMIs mapped to mtRNA per cell. Median ratio = 1.88. **G** Percentage of UMIs mapped to exons per cell. Median ratio = 0.71. **H** Percentage of UMIs mapped to exons per cell, excluding those mapped to mtRNA. Median ratio = 0.65. **I** TSS enrichment scores over distance from TSS. Maximum values are indicated with corresponding colors. **J** Pairwise comparison of protein tag detection frequencies under LLL (x-axis) and DIG (y-axis) conditions. Each point represents a single protein tag. Blue points highlight isotype control protein tags. Grey points are all other protein tags. **K** Correlation of protein tag fold change (\log_2) under DIG and LLL conditions in two groups of T cells that were both activated and cultured with IL-1 β and IL-23, and one of the two groups was also cultured with PGE2. Top upregulated protein tags are highlighted in red; downregulated protein tags are highlighted in blue. **L** Pairwise comparison of gene detection frequencies under LLL (x-axis) and DIG (y-axis) conditions. Each point represents a single gene. Blue points highlight ribosomal protein genes (RPL/S); red points highlight mitochondrial genes (MT-). Grey points are all other genes. **M** Correlation of gene fold change (\log_2) under DIG and LLL conditions in two groups of T cells that were both activated and cultured with IL-1 β and IL-23, and one of the two groups was also cultured with PGE2. Top upregulated genes are highlighted in red; downregulated genes are highlighted in blue. **N** (i) “Harmonized” 3WNN UMAP plot showing clusters identified in 3WNN clustering of DOGMA-seq data under DIG condition. (ii) “Harmonized” 3WNN UMAP plot showing cell types predicted after projection into the Azimuth PBMC reference. **O** (i) “Harmonized” 3WNN UMAP plot showing clusters identified in 3WNN clustering of DOGMA-seq data under LLL condition. (ii) “Harmonized” 3WNN UMAP plot showing cell types predicted after projection into the Azimuth PBMC reference. **P** Line plots showing average ROGUE (purity of identified clusters) for clustering based on 3WNN, RNA, ADT, and ATAC spaces, with resolution ranging from 0.02 to 0.3. Number of clusters identified in each clustering was labeled. **Q** Bar plot show that proportions of predicted cell types in DOGMA-seq data under DIG and LLL conditions are quite similar. **R, S** “Harmonized” 3WNN UMAP plots highlight canonical markers for Th17 cells in DOGMA-seq data. ATAC marker is motif activity (the deviations in chromatin accessibility across the set of regions related to the motif) calculated from ATAC-seq peaks. **R** Under DIG condition. Cluster 1 in Fig. 1Ni is enriched for Th17 cells based on CCR4 ADT, CCR6 ADT, and *RORC* ATAC signals. **S** Under LLL condition. Cluster 1 in Fig. 1Oi is enriched for Th17 cells based on CCR4 ADT, CCR6 ADT, and *RORC* ATAC signals.

some genes after LLL permeabilization (Fig. 1L). By default, the RNA sequencing reads from DOGMA-seq libraries were aligned to both exons and introns by cellranger-arc count. We also aligned RNA sequencing reads from both conditions to exons only or introns only and found that LLL outperformed DIG in gene detection rates when RNA sequencing reads were aligned to exons only (Additional file 1: Fig. S1D), but their performances were similar when the reads were aligned to introns only (Additional file 1: Fig. S1E). To further investigate the difference in gene detection rates, we calculated the proportion of exonic UMIs for each gene and dichotomized genes into exon-dominated genes (genes with proportion of exonic UMIs > 0.5) and intron-dominated genes (genes with proportion of exonic UMIs \leq 0.5), for DIG and LLL (Additional file 1: Fig. S1A–B). We found that genes defined as exon-dominated genes under both conditions had the largest increases in gene detection rates for LLL compared to DIG, whereas genes defined as intron-dominated genes under both conditions had the smallest differences (Additional file 1: Fig. S1C), suggesting that the differences in gene detection rates may be primarily due to differential detection of cytoplasmic genes. We

also found that gene fold changes with the addition of PGE2 to the cultures were highly correlated between the two permeabilization conditions at the pseudo-bulk level (Fig. 1M). Additionally, we estimated the ambient RNA contamination rate for each cell under DIG and LLL cell permeabilization conditions using DecontX [10], and we found that DIG had a lower contamination rate than LLL (Additional file 1: Fig. S1G), suggesting that the plasma membrane was more intact after DIG. We also used the decontaminated RNA count matrices (generated by DecontX) to calculate gene detection rates but found no difference (Additional file 1: Fig. S1H). We re-defined gene detection rates as the fraction of cells with UMIs > 1 to reduce the impact of genes with only one copy, which is not common for synthesized RNA. The differences in gene detection rates between DIG and LLL were smaller (Additional file 1: Fig. S1I) and decreased further when they were based on decontaminated RNA count matrices (Additional file 1: Fig. S1J), except for RPL/RPS genes. Based on these findings, we inferred that DIG and LLL have little difference in true gene detection rates for the majority of biologically meaningful genes, and that the higher observed gene detection rates in LLL were likely due to RPL/RPS genes and ambient RNA leaked from cytoplasm.

We next assessed DIG and LLL qualitatively by cluster analysis with a series of resolutions, ranging from 0.02 to 0.3. For both conditions, we performed clustering using RNA, ADT, and ATAC data separately. We also performed trimodal weighted nearest-neighbor [11] (3WNN) analysis to leverage all three types of data. The clusters identified in each cluster analysis with resolution = 0.2 are shown (Additional file 1: Fig. S1Ki-iv, Li-iv). To assess their performance, we calculated an entropy-based statistic, ROGUE [12], for each cluster analysis, which is a measurement of the purity of identified clusters. We found that DIG had higher average ROGUE (higher purity of identified clusters) than LLL, given similar resolution or number of clusters, when clustering based on ADT and ATAC spaces. However, DIG and LLL had comparable average ROGUE, given similar resolution or number of clusters, when clustering based on 3WNN and RNA spaces (Fig. 1P). These findings are consistent with our observation that DIG had higher quality ADT and ATAC libraries than LLL. To better visualize clusters identified in 3WNN analysis (resolution = 0.2) for DIG and LLL, we utilized all three types of data to perform embedding correction using Harmony (Fig. 1Ni, Oi) [13] and performed automatic cell annotation based on RNA data, using the Azimuth PBMC reference (Fig. 1Nii, Oii) [11]. In a head-to-head comparison, the proportions of predicted cell types were quite similar between DIG and LLL (Fig. 1Q, Additional file 1: Fig. S1F). Leveraging ATAC data, we can also locate clusters enriched for Th17 (Fig. 1R-S) and Th1 cells (Additional file 1: Fig. S1M-N) under both conditions, highlighting the value that the ATAC library of DOGMA-seq adds to transcriptome and ADT libraries for the study of lineage-specific T helper cells. Considering all of these observations, we conclude that DIG is the preferred DOGMA-seq cell permeabilization method.

To further investigate the information that DOGMA-seq provides, we performed a head-to-head comparison of DOGMA-seq with our optimized DIG permeabilization condition to CITE-seq (Fig. 2A). We collected PBMC from two healthy donors, enriched untouched T cells, activated and stimulated them under four different ex vivo tissue culture conditions. After the stimulation experiment, the cells were subjected to CITE-seq and DOGMA-seq, respectively. Using either assay, we observed comparable

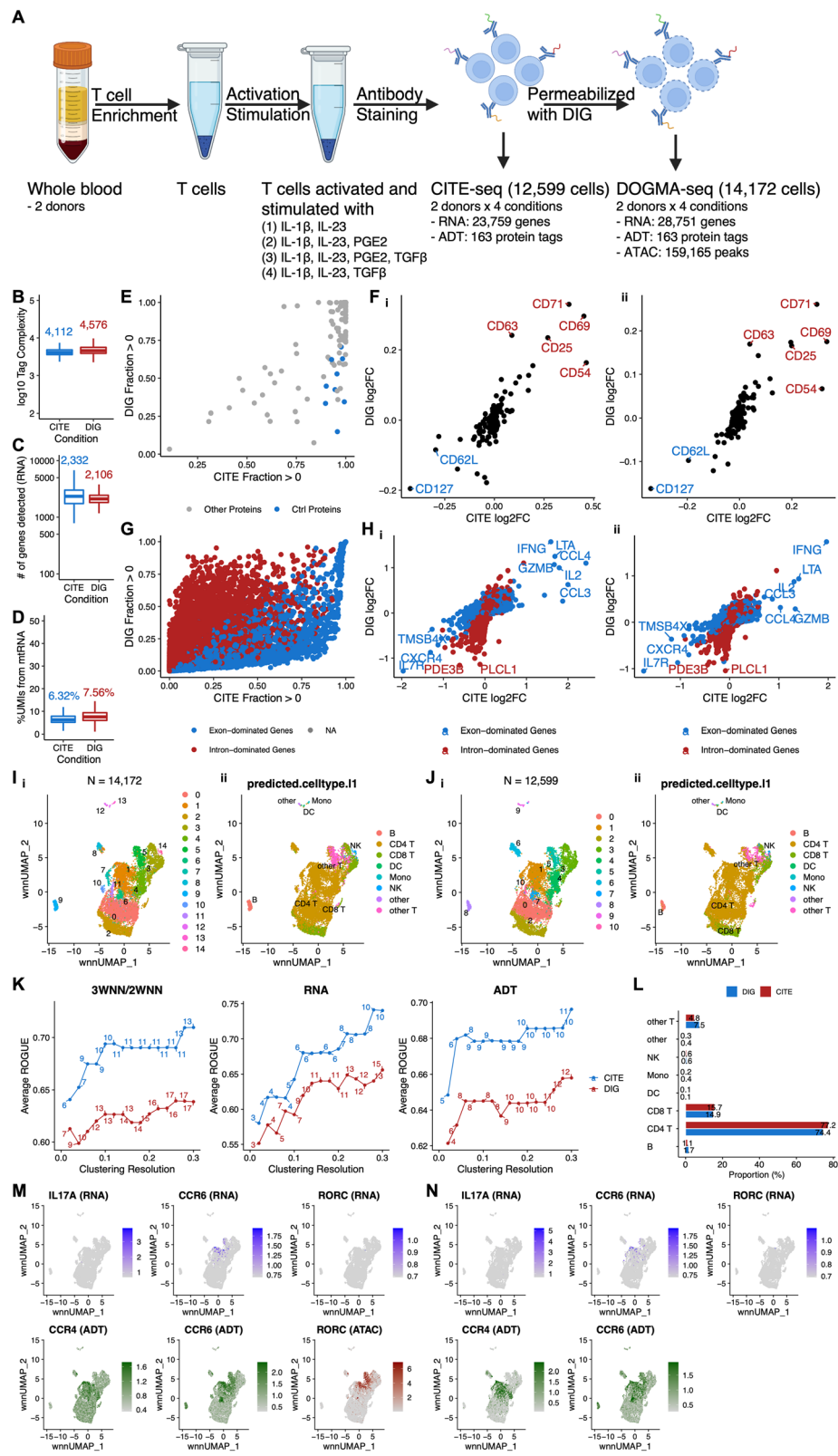


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Comparison between CITE-seq and DOGMA-seq. **A** Overview of study design for comparison between CITE-seq and DOGMA-seq. Each of four aliquots of T cells from two human donors were activated and stimulated under a different stimulation condition (total of four stimulation conditions in eight tissue culture wells) in a 12-h tissue culture. The cells from each of the eight tissue culture wells were then labeled with a unique hashtag (total of eight unique hashtags). Approximately equal numbers of the eight uniquely hashtagged cell populations were then pooled and labeled with antibody cocktail. Different aliquots of the hashtag- and antibody cocktail-labeled pool of cells then underwent either CITE-seq or DOGMA-seq with DIG library preparation. **B–D** Boxplot showing quality control metric comparisons between CITE-seq and DOGMA-seq. Median values are indicated with corresponding colors. **B** Protein tag complexity per cell. Median ratio = 0.90. **C** Number of genes per cell. Median ratio = 1.11. **D** Percentage of UMIs mapped to mtRNA per cell. Median ratio = 0.84. **E** Pairwise comparison of protein tag detection frequencies in CITE-seq (x-axis) and DOGMA-seq (y-axis). Each point represents a single protein tag. Blue points highlight isotype control protein tags. Grey points are all other protein tags. **F** (i) Correlation of protein tag fold change (\log_2) as detected by CITE-seq and DOGMA-seq in two groups of T cells from donor SB775372 that were both activated and cultured with IL-1 β and IL-23, and one of the two groups was also cultured with PGE2. Top upregulated protein tags are highlighted in red; downregulated protein tags are highlighted in blue. (ii) in samples from donor SB775393, as (i). **G** Pairwise comparison of gene detection frequencies in CITE-seq (x-axis) and DOGMA-seq (y-axis). Each point represents a single gene. Blue points highlight exon-dominated genes; red points highlight intron-dominated genes. An exon-dominated gene is defined as a gene with proportion of exonic UMIs (DOGMA-seq) > 0.5. An intron-dominated gene is defined as a gene with proportion of exonic UMIs (DOGMA-seq) \leq 0.5. NA means proportion of exonic UMIs is not available. **H** (i) Correlation of gene fold change (\log_2) as detected by CITE-seq and DOGMA-seq in two groups of T cells from donor SB775372 that were both activated and cultured with IL-1 β and IL-23, and one of the two groups was also cultured with PGE2. Selected intron-dominated genes are highlighted in red; selected exon-dominated genes are highlighted in blue. (ii) in samples from donor SB775393, as (i). **I** (i) “Harmonized” 2WNN UMAP plot showing clusters identified in 3WNN clustering of DOGMA-seq data. (ii) “Harmonized” 2WNN UMAP plot showing cell types predicted after projection into the Azimuth PBMC reference. **J** (i) “Harmonized” 2WNN UMAP plot showing clusters identified in 2WNN clustering of CITE-seq data. (ii) “Harmonized” 2WNN UMAP plot showing cell types predicted after projection into the Azimuth PBMC reference. **K** Line plots showing average ROGUE (purity of identified clusters) for clustering based on 3WNN/2WNN, RNA, and ADT spaces, with resolution ranging from 0.02 to 0.3. Number of clusters identified in each clustering was labeled. **L** Bar plots show that proportions of predicted cell types in DOGMA-seq and CITE-seq data are quite similar. **M** “Harmonized” 2WNN UMAP plots highlighting canonical markers for Th17 cells in DOGMA-seq data. ATAC marker is motif activity (the deviations in chromatin accessibility across the set of regions related to the motif) calculated from ATAC-seq peaks. Cluster 5 in Fig. 2li is enriched for Th17 cells based on CCR4 ADT, CCR6 ADT, and *RORC* ATAC signals. **N** “Harmonized” 2WNN UMAP plots highlight canonical markers for Th17 cells in CITE-seq data

quality control metrics for protein tag complexity, genes detected, and the fraction of mtRNA (Fig. 2B–D). The observations remained unchanged even when the donor difference was considered (Additional file 1: Fig. S2A–C).

To evaluate protein tag detection rates, we calculated the fraction of cells with UMIs > 0 for all 163 protein tags in the TotalSeq™-A Human Universal Cocktail, V1.0 and performed a pairwise comparison. As expected, we found that the protein tag detection rates of DOGMA-seq were lower in most cases (Fig. 2E), probably due to cell surface protein and/or ADT damage during permeabilization. It is also interesting that the detection rates of isotype control protein tags were higher in CITE-seq. Fortunately, the apparent cell surface protein and/or ADT damage did not attenuate differential protein tag signals upon differential stimulations at the pseudo-bulk level. When we compared two groups of T cells that were both activated and cultured with IL-1 β and IL-23, and one of the two groups was also cultured with PGE2, significant upregulation and downregulation of a few markers were detectable in both assays and could be replicated in samples from the other donor (Fig. 2Fi–ii).

We further compared transcript measurements by CITE-seq and DOGMA-seq. By default, the RNA sequencing reads from the CITE-seq library were aligned to exons

only by cellranger count, whereas those from the DOGMA-seq library were aligned to both exons and introns by cellranger-arc count. This is reasonable since the leakage of transcripts from permeabilized cells can be compensated in some way by additional alignment to introns. We calculated the fraction of cells with UMIs > 0 for all 36,601 genes in the reference genome and found that each assay had its advantage in gene detection rates (Additional file 1: Fig. S2G). We also confirmed that the detection rates of ribosomal genes were lower in DOGMA-seq [7]. We additionally calculated the proportion of exonic UMIs for each gene detected by DOGMA-seq (Additional file 1: Fig. S2H) and dichotomized genes into exon-dominated genes and intron-dominated genes. We observed that CITE-seq had higher detection rates for almost all exon-dominated genes and some intron-dominated genes, whereas DOGMA-seq had higher detection rates for most intron-dominated genes (Fig. 2G). Gene fold changes after the addition of PGE2 to the cultures were still correlated at the pseudo-bulk level, but exon-dominated genes and intron-dominated genes formed two lines with different slopes (Fig. 2Hi). This finding was replicable in samples from an independent donor using the same protocol (Fig. 2Hii). To perform fairer comparisons, we also aligned RNA sequencing reads from both assays to both exons and introns, exons only, and introns only. Generally, CITE-seq outperformed DOGMA-seq in gene detection rates when RNA sequencing reads were aligned to both exons and introns or exons only (Additional file 1: Fig. S2D-E), but their performances were similar when aligned to introns only (Additional file 1: Fig. S2F).

We next assessed CITE-seq and DOGMA-seq qualitatively by cluster analysis with a series of resolutions, ranging from 0.02 to 0.3. For DOGMA-seq data, we performed clustering using RNA, ADT, and ATAC data separately (Additional file 1: Fig. S3Aii–iv). We also performed 3WNN analysis to leverage all three types of data (Additional file 1: Fig. S3Ai). For CITE-seq data, we performed clustering for RNA and ADT data separately using the same parameters (including resolution) as those used for DOGMA-seq data (Additional file 1: Fig. S3Bii–iii). Bi-modal WNN (2WNN) analysis was also performed to leverage both RNA and ADT data (Additional file 1: Fig. S3Bi). We calculated ROGUE for each cluster analysis and found that CITE-seq had a higher average ROGUE than DOGMA-seq, given similar resolution or number of clusters, when clustering based on 3WNN (DOGMA-seq)/2WNN (CITE-seq), RNA, and ADT spaces (Fig. 2K). These findings are consistent with our observation that CITE-seq had higher quality RNA and ADT libraries than DOGMA-seq. We also realized that leveraging the ATAC library of DOGMA-seq in cluster analysis cannot fully compensate DOGMA-seq's inferior RNA and ADT library quality but provides additional information in cell type annotation. To better visualize clusters identified in 3WNN analysis of DOGMA-seq data and 2WNN analysis of CITE-seq data (resolution = 0.2), we utilized both RNA and ADT data to perform embedding correction using Harmony (Fig. 2Ii, Ji) [13]. Corroborated by Azimuth PBMC reference projection based on RNA data [11], the same cell types appeared on matched locations of “Harmonized” 2WNN Uniform Manifold Approximation and Projection (UMAP) [14] dimensionality reduction plots (Fig. 2Iii, Jii) and had comparable proportions in both assays (Fig. 2L, Additional file 1: Fig. S2I), suggesting that Harmony performed well without over-correction. The advantage of DOGMA-seq lies in its simultaneous measurement of

chromatin accessibility, enabling the identification of subtypes of T helper cells. Based on expression of canonical markers, cluster 5 in DOGMA-seq data was enriched for Th17 cells (Fig. 2M), and cluster 3 was enriched for Th1 cells (Additional file 1: Fig. S3C). These findings were further supported by peaks in genomic regions around canonical markers (Additional file 1: Fig. S3Ei-iv). However, Th17 and Th1 cells were harder to identify in CITE-seq data (Fig. 2N, Additional file 1: Fig. S3D). Again, these observations highlighted the value of DOGMA-seq to study lineage-specific T helper cells.

In addition to our comparisons of transcript measurements described above where CITE-seq RNA libraries were aligned to exons only and DOGMA-seq RNA libraries were aligned to both exons and introns because those are the default parameters and the common practice in the corresponding cellranger and cellranger-arc pipelines, we also performed alternative comparisons where CITE-seq and DOGMA-seq RNA libraries were aligned to both exons and introns. We found similar results regarding the correlation of log₂FC (Additional file 1: Fig. S4Di-ii). As expected, CITE-seq has more detected genes (Additional file 1: Fig. S4A-C) but a considerably lower average ROGUE (Additional file 1: Fig. S4E), implying more noisy RNA signals, possibly due to reduced mapping accuracy.

Conclusions

In summary, our study optimized the DIG cell permeabilization condition, performed a comprehensive comparison between the DIG and LLL cell permeabilization conditions, and found higher quality ADT and ATAC libraries after DIG permeabilization. We also noticed that LLL permeabilization causes higher contamination from ambient RNA, which should be considered in future analysis. We next performed a comprehensive transcriptome and ADT comparison between CITE-seq and DOGMA-seq. We found that DOGMA-seq with optimized DIG permeabilization and its ATAC library provides more information, even though its mRNA and cell surface protein libraries have slightly inferior quality, compared to CITE-seq. We also recognized the additional value of DOGMA-seq for studying lineage-specific T helper cells, with the potential to explore transcriptomic and epigenetic interactions between various subtypes of T cells. Our study provides a valuable and general guidance for technical and study design considerations for single cell multimodal omics experiments. Our study also shows the feasibility of using hashtags in DOGMA-seq.

Methods

Enrichment of and stimulation of peripheral blood T cells

T cells were enriched from 30 ml of whole blood from healthy human subjects, ages 18–35, following the MACSxpress[®] Whole Blood Pan T Cell Isolation Kit, human (Miltenyi Biotec) protocol. Four aliquots of one million cells each per study subject were stimulated overnight for 12 h in 1 ml X-VIVO[™] 15 Serum-free Hematopoietic Cell Medium (Lonza) with 10 µl ImmunoCult[™] Human CD3/CD28/CD2 T Cell Activator (STEMCELL Technologies), 50 ng/ml IL-1B (R&D Systems), and 50 ng/ml IL-23 (R&D Systems), and either 1 µM Prostaglandin E2 (PGE2) (Sigma) or 3 ng/ml TGFB (R&D Systems), or both, for a total of four stimulation conditions per subject. After the

overnight incubation, the cells were collected into 15 ml conical tubes through a 30- μ m strainer and washed with up to 14 ml of PBS/0.2% BSA, centrifuged at 400 g for 5 min at 4 °C, and resuspended in 50 μ l Pharmingen Stain Buffer (BSA) (BD Biosciences). An aliquot of each cell suspension was then stained with ViaStain™ AOPI Staining Solution (Nexcelom Bioscience), and viability counts were obtained using a Cellometer Auto 2000 Cell Viability Counter (Nexcelom Bioscience).

Cell staining with barcoded antibodies

After the addition of 5 μ l Human TruStain FcX™ (BioLegend), the cell suspension was incubated for 10 min on ice, and then up to 500,000 cells from each condition were stained with a unique TotalSeq™-A anti-human Hashtag antibody (BioLegend) [8] in 50 μ l for 30 min at 4 °C on a Laminar Wash™ 16-well strip (Curiox Biosystems). The cells were washed in the Laminar Wash Mini System (Curiox Biosystems) with 25 cycles at a flow rate of 10 μ l/s. Following collection of each well and a cell count, up to 1 million cells from the different conditions were pooled, centrifuged at 400g for 5 min at 4 °C and resuspended in 50 μ l PBS/0.2% BSA. Fifty microliters of TotalSeq™-A Human Universal Cocktail, V1.0 (BioLegend), at a 0.5 dilution, was added to the cell suspension, and then the total volume was split between 2 wells of a Laminar Wash™ 16-well strip for a 30 min incubation at 4 °C. The cells were washed as previously described, pooled in a 300 μ l volume of PBS/0.2% BSA, filtered through a 40 μ m Scienceware® Flowmi™ Cell Strainer (SP Bel-Art) and counted.

Single cell preparation for CITE-seq

A total of 30,000 cells were loaded into each of two wells of a Chromium Next GEM Chip G and run on the Chromium Controller (10x Genomics) for Gel Bead-in-emulsion (GEM) generation. The samples were processed according to the CG000204 Chromium Next GEM Single Cell 3' v3.1 Rev. D protocol (10x Genomics) for RNA library construction, while modifications for CITE-seq ADT and cell hashing HTO library preparation were performed according to the TotalSeq™-A Antibodies and Cell Hashing with 10x Single Cell 3' Reagent Kit v3 or v3.1 single index protocol (BioLegend). The modifications include the addition of ADT and HTO additive primers to a final concentration of 0.2 μ M in the cDNA amplification reaction (step 2.2) and using Illumina TruSeq™ Small RNA RPIx (with custom indexes) and Illumina TruSeq™ D70X_short primers, respectively, together with the SI-PCR primer, in ADT and HTO sample index PCRs. The ADT and HTO sample index PCRs were performed on the purified fraction of the supernatant recovered from cDNA amplification cleanup.

Single cell preparation for DOGMA-seq

The cells were processed according to the CG000338 Chromium Next GEM Multiome ATAC + Gene Expression Rev. D protocol (10x Genomics) for Gene Expression and ATAC library construction while modifications for DOGMA-seq were performed following the “Cell permeabilization with DIG” section of the DOGMA-seq protocol (NYGC Innovation Lab).

We first optimized the concentration of DIG that we would use for the DIG cell permeabilization step. We tested 0.0025%, 0.005%, 0.0075%, and 0.01% DIG with about 230,000 cells each, and a DIG concentration of 0.0075% was chosen (see Results).

Approximately 600,000 cells were centrifuged at 400g for 5 min at 4 °C and permeabilized in 100 µl chilled DIG lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl₂, 0.0075% DIG and 2 U/µl Protector RNase inhibitor (Roche)) for 1 min on ice, followed by the addition of 1 ml of chilled DIG wash buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl₂ and 1 U/µl RNase inhibitor) and pipette mixing, before centrifugation at 500g for 5 min at 4 °C. The supernatant was discarded, the pellet was resuspended in a very small volume of about 17 µl of diluted Nuclei Buffer (10x Genomics), and 5 µl was used for cell counting.

In a separate experiment, we compared DOGMA-seq's DIG and LLL protocols. Approximately 600,000 cells were permeabilized with DIG as described above and resuspended in 7.5 µl of Nuclei Buffer. Approximately 300,000 cells were fixed with 0.1% formaldehyde for 5 min at room temperature. Glycine was added as a quencher at 0.125 M. The cells were washed twice with PBS/1% BSA/RNase inhibitor, centrifuged at 400g for 5 min at 4 °C, and permeabilized in 100 µl chilled LLL lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP40, 1% BSA, 1 mM DTT and 2 U/µl RNase inhibitor) for 3 min on ice followed by the addition of 1 ml of chilled LLL wash buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 1% BSA, 1 mM DTT and 1 U/µl RNase inhibitor) and pipette mixing, before centrifugation at 500g for 5 min at 4 °C. The cells were resuspended in 7.5 µl of Nuclei Buffer. DIG and LLL-treated cells were also processed as described below. The cell recovery after DIG permeabilization was approximately 25–30 %, while the cell recovery after LLL permeabilization was approximately 50%.

Following the CG000338 Chromium Next GEM Multiome ATAC + Gene Expression Rev. D protocol, 30,000 cells were processed for each of two transposition reactions, which were then loaded into each of two wells of a Chromium Next GEM Chip J and run on the Chromium Controller (10x Genomics) for GEM generation.

During sample pre-amplification (step 4.2), ADT and HTO additive primers were added to a final concentration of 0.2 µM. After cleanup and elution in 100 µl, 30 µl was set aside for ADT and 10 µl was set aside for HTO sample index PCR reactions, which were performed using Illumina TruSeq™ Small RNA RPIx (with custom indexes) and Illumina TruSeq™ D70X_short primers, respectively, and the SI-PCR primer.

When generating the RNA library in the Chromium Next GEM Multiome ATAC + Gene Expression protocol, we chose to use the Single Index Plate T Set A instead of the Dual Index Plate TT Set A.

ADT and HTO primers

ADT additive: CCTTGGCACCCGAGAATT*C*C

HTO additive: GTGACTGGAGTTCAGACGTGTGC*T*C

SI-PCR: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC

RPI_SI-GA-A1-1: CAAGCAGAAGACGGCATAACGAGATAGTAAACCGTGACTGG
AGTTCCTTGGCACCCGAGAATTC*C*A

RPI_SI-GA-A1-2: CAAGCAGAAGACGGCATAACGAGATCCGTTTtaggtGACTGG
AGTTCCTTGGCACCCGAGAATTC*C*A

RPI_SI-GA-A1-3: CAAGCAGAAGACGGCATAACGAGATGACGCCGAGTGACTGG
AGTTCCTTGGCACCCGAGAATTC*C*A

RPI_SI-GA-A1-4: CAAGCAGAAGACGGCATAACGAGATTTACGGTTGTGACTGG
AGTTCCTTGGCACCCGAGAATTC*C*A

D701_S: CAAGCAGAAGACGGCATAACGAGATCGAGTAATGTGACTGGAGTT
CAGACGTGT*G*C

D702_S: CAAGCAGAAGACGGCATAACGAGATTCTCCGGAGTGACTGGAGTT
CAGACGTGT*G*C

D703_S: CAAGCAGAAGACGGCATAACGAGATAATGAGCGGTGACTGGAGTT
CAGACGTGT*G*C

D704_S: CAAGCAGAAGACGGCATAACGAGATGGAATCTCGTGACTGGAGTT
CAGACGTGT*G*C

Library quality control and quantitation

Initial cDNA and ATAC library sizes and estimated concentrations were determined in 2200 TapeStation System High Sensitivity D5000 assays (Agilent Technologies), and initial RNA, ADT, and HTO library sizes and estimated concentrations were determined in 2200 TapeStation System High Sensitivity D1000 assays (Agilent Technologies). Qubit™ dsDNA HS assays (Invitrogen), together with average library sizes from the TapeStation assays, was used to estimate molarity and to determine library dilutions in the KAPA Library Quantification qPCR assay (Roche Diagnostics), which was used to estimate molarity more accurately.

Library pooling for sequencing

Libraries from the DOGMA-seq experiment that compared DIG and LLL cell permeabilization were pooled and sequenced with libraries from other projects. The DOGMA-seq DIG and LLL RNA libraries were each represented at approximately 1.99 nM in a pool of libraries that was submitted to the UPMC Genome Center and sequenced using a NovaSeq 6000 S2 Reagent Kit v1.5 (100 cycles) (Illumina) on a NovaSeq 6000 sequencer (Illumina) with a configuration 28/8/0/102 (138 total cycles). DOGMA-seq DIG and LLL ADT libraries were each represented at approximately 2.20 nM and HTO libraries each at approximately 0.27 nM in a pool that was sequenced using a NovaSeq 6000 S1 Reagent Kit v1.5 (100 cycles) with a configuration 28/8/0/15 (51 total cycles). DOGMA-seq DIG and LLL ATAC libraries were each represented at approximately 2.5 nM in a pool that was sequenced using a NovaSeq 6000 S1 Reagent Kit v1.5 (100 cycles) with a configuration 53/8/24/53 and (138 total cycles).

The CITE-seq and other DOGMA-seq DIG RNA libraries were each pooled at approximately 2.15 nM and sequenced using a NovaSeq 6000 S4 Reagent Kit v1.5 (200 cycles) with a configuration 28/8/0/151 (187 total cycles). The CITE-seq and other DOGMA-seq DIG ADT libraries were each pooled at approximately 1.05 nM and

the CITE-seq and other DOGMA-seq DIG HTO libraries were each pooled at approximately 0.13 nM, and sequenced using a NovaSeq 6000 S4 Reagent Kit v1.5 (200 cycles) with a configuration 28/8/0/151 (187 total cycles). Finally, the other DOGMA-seq ATAC libraries were each pooled at approximately 5.0 nM and sequenced using a NovaSeq 6000 S1 Reagent Kit v1.5 (100 cycles) with a configuration 53/8/24/53 (138 cycles).

Raw sequencing data processing

For CITE-seq, the sequenced RNA library was processed and aligned to GRCh38 human reference genome using Cell Ranger software (version 6.1.2) from 10x Genomics, with UMI counts summarized for each barcode. To perform fairer comparisons, we additionally aligned the RNA library to exons only or both exons and introns, without or with the parameter `--include-introns`. To distinguish cells from the background, cell calling was performed on the full raw UMI count matrix, with the filtered UMI count matrix generated. ADT and HTO libraries were aligned using the kallisto | bustools workflow (kallisto version 0.46.0, bustools version 0.40.0) [15], with the full raw UMI count matrices generated, which were subsequently filtered according to the barcodes in the filtered UMI count matrix for RNA library. Cell demultiplexing was performed based on the filtered UMI count matrix for the HTO library [8] using the function `HTODemux()` from the R package Seurat (version 4.0.4) [11]. The sizes of matrices consisting of singlets are listed below:

RNA: 23,759 genes \times 12,599 cells; ADT: 163 protein tags \times 12,599 cells; merged from 8 samples (2 donors \times 4 conditions)

For DOGMA-seq, the sequenced RNA and ATAC libraries were processed and aligned to GRCh38 human reference genome using Cell Ranger ARC software (version 2.0.0), with UMI counts, peak counts, and fragments summarized for each barcode. To perform fairer comparisons, we additionally aligned the RNA library to both exons and introns or exons only, without or with the parameter `--gex-exclude-introns`. To distinguish cells from the background, joint cell calling was performed based on RNA UMIs per barcode and ATAC transposition events in peaks per barcode, with the filtered UMI count and peak count matrices generated. ADT and HTO libraries were aligned using the kallisto | bustools workflow (kallisto version 0.46.0, bustools version 0.40.0) [15], with the full raw UMI count matrices generated, which were subsequently filtered according to the barcodes in the filtered UMI count and peak count matrices for RNA and ATAC libraries. Cell demultiplexing was performed based on the filtered UMI count matrix for the HTO libraries [8] using the function `HTODemux()` from the R package Seurat (version 4.0.4) [11]. The sizes of matrices consisting of singlets are listed below:

DOGMA-seq data under DIG condition:

RNA: 29,499 genes \times 15,788 cells; ADT: 163 protein tags \times 15,788 cells; ATAC: 153,958 peaks \times 15,788 cells; merged from 4 samples (1 donors \times 4 conditions)

DOGMA-seq data under LLL condition:

RNA: 28,297 genes \times 11,943 cells; ADT: 163 protein tags \times 11,943 cells; ATAC: 140,444 peaks \times 11,943 cells; merged from 4 samples (1 donors \times 4 conditions)

DOGMA-seq data in comparison with CITE-seq:

RNA: 28,751 genes \times 14,172 cells; ADT: 163 protein tags \times 14,172 cells; ATAC: 159,165 peaks \times 14,172 cells; merged from 8 samples (2 donors \times 4 conditions)

Complexity analyses

Complexity analyses estimate number of distinct molecules for a cell, given number of distinct molecules observed for the cell and number of total molecules observed for the cell, according to Lander-Waterman equation. We adapted the custom code deposited by DOGMA-seq developers (https://github.com/caleblareau/asap_reproducibility/blob/master/global_functions/estimateLibraryComplexity.R) to perform complexity analysis. The code is a translation of the subcommand MarkDuplicates from Picard Tools [16] and takes the number of unique and duplicate fragments from the Cell Ranger ARC (chromatin) and bustools (protein tag) outputs as inputs. For protein tag, we defined unique and duplicate fragments as described by DOGMA-seq developers [6]. For chromatin, we used the sum of “atac_dup_reads” and “atac_fragments” columns and the “atac_fragments” column from the per_barcode_metrics.csv file.

Estimation of ambient RNA contamination rate

We used DecontX, which accurately predicted contamination levels in a mouse-human mixture dataset [10], to estimate ambient RNA contamination rate for each cell under DIG and LLL cell permeabilization conditions. We filtered out all cells in filtered matrices (including singlets used in the analysis and doublets detected based on HTO) from raw matrices and used them to empirically estimate the distribution of ambient RNA. Ambient RNA contamination rate was then separately estimated for each cell under DIG and LLL conditions. Decontaminated RNA count matrices were also separately generated for cells under DIG and LLL conditions.

Cluster analyses

After the matrices were imported into R, they consisted of singlets that were analyzed using the R package Seurat (version 4.0.4) [11] and Signac (version 1.3.0) [17]. For the RNA data of either assay, regularized negative binomial regression (SCTransform) [18] was used to normalize UMI count data, with glmGamPoi [19] invoked to improve the speed. Three thousand highly variable genes were identified and used in principal component analysis. The first 50 principal components were used in UMAP [14] dimensionality reduction and clustering using the original Louvain algorithm with a series of resolution, ranging from 0.02 to 0.3.

For ADT data of either assay, centered log ratio (CLR) transformation was used to normalize UMI count data. All 163 protein tags were set as variable features and used in principal component analysis. The first 30 principal components were used in UMAP dimensionality reduction and clustering using the original Louvain algorithm with a series of resolution, ranging from 0.02 to 0.3.

For ATAC data of DOGMA-seq, Latent Semantic Indexing (LSI) [20] was used to normalize and reduce the dimensionality of peak count data, with all the peaks selected as variable features. The first 50 (excluding the first component due to correlation with sequencing depth) LSI components were used in UMAP

dimensionality reduction and clustering using the SLM algorithm with a series of resolution, ranging from 0.02 to 0.3.

For CITE-seq and DOGMA-seq data, we also performed 2WNN (RNA and ADT) and 3WNN (RNA, ADT, and ATAC) dimensionality reduction [11] using the same numbers of components described above for each modality. Clustering was performed using the SLM algorithm with a series of resolution, ranging from 0.02 to 0.3.

To better visualize clusters identified in different assays or under different conditions, we performed embedding correction using Harmony [13]. For the comparison of CITE-seq and DOGMA-seq, we normalized the two data separately as described above and identified variable features for each modality (top 3000 for RNA and top 100 for ADT). We selected shared variable features between CITE-seq and DOGMA-seq as the variable features used in Harmony. We performed linear dimension reduction and ran Harmony on the linear components for each reduction with assay type as the covariate. UMAP dimensionality reduction was performed using “Harmonized” components for RNA and ADT data, with the “Harmonized” 2WNN UMAP plots generated.

For the comparison of DIG and LLL, we merged the two data sets and normalized them, due to smaller differences between the two data compared with the former comparison. Variable features were identified in the merged data for all three modalities and used in Harmony. We performed linear dimension reduction and ran Harmony on the linear components for each reduction with permeabilization method as the covariate. UMAP dimensionality reduction was performed using “Harmonized” components for RNA, ADT, and ATAC data, with the “Harmonized” 3WNN UMAP plots generated.

We also annotated cells in different assays or under different conditions using reference-mapping approach. We found anchors between our data and Azimuth PBMC reference [11], with `normalization.method = “SCT”`, `reference.reduction = “spca”`, `dims = 1:50`. Cell type labels were then transferred from the reference to our data.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02698-8>.

Additional file 1: Figures S1-S4. Figure S1. Additional comparison between DIG and LLL conditions. **Figure S2.** Additional comparison between CITE-seq and DOGMA-seq, part 1. **Figure S3.** Additional comparison between CITE-seq and DOGMA-seq, part 2. **Figure S4.** Alternative comparisons of transcript measurements where CITE-seq and DOGMA-seq RNA libraries were aligned to both exons and introns.

Additional file 2.

Acknowledgements

This research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided.

Review history

Review history is available as Additional File 2.

Peer review information

Stephanie McClelland was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team

Authors' contributions

Z.X., E.H., W.C., and R.D. conceived the project and designed the experiments. E.H. performed the experiments. Z.X. performed the data analysis. Z.X., E.H., W.C., and R.D. wrote the manuscript. W.C. and R.D. supervised the work. The authors read and approved the final manuscript.

Authors' information

Twitter handles: @zhongli_xu (Zhongli Xu)

Funding

This project was supported by grant U01DK062420 from the National Institutes of Health.

Availability of data and materials

The raw and processed data of CITE-seq and DOGMA-seq, including 8 samples for CITE-seq and 16 samples for DOGMA-seq, are deposited in Gene Expression Omnibus (GEO) (GSE200417) [21]. The analysis codes are available under MIT License at GitHub https://github.com/xzlandy/Benchmark_CITEseq_DOGMAseq [22] and at Zenodo [23].

Declarations**Ethics approval and consent to participate**

Study subjects were recruited for this study under a protocol approved by the University of Pittsburgh Human Research Protection Office.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Pediatrics, University of Pittsburgh School of Medicine, UPMC Children's Hospital of Pittsburgh, 4401 Penn Ave, Pittsburgh, PA 15224, USA. ²School of Medicine, Tsinghua University, Beijing, China. ³Department of Medicine, University of Pittsburgh School of Medicine, 3550 Terrace Street, Pittsburgh, PA 15261, USA. ⁴Department of Biostatistics, University of Pittsburgh School of Public Health, 130 De Soto Street, Pittsburgh, PA 15261, USA. ⁵Department of Human Genetics, University of Pittsburgh School of Public Health, 130 De Soto Street, Pittsburgh, PA 15261, USA.

Received: 10 January 2022 Accepted: 1 June 2022

Published online: 23 June 2022

References

1. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. 2017;14(9):865–8. <https://doi.org/10.1038/nmeth.4380>.
2. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol*. 2017;35(10):936–9. <https://doi.org/10.1038/nbt.3973>.
3. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*. 2018;361(6409):1380–5. <https://doi.org/10.1126/science.aau0730>.
4. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol*. 2019;37(12):1452–7. <https://doi.org/10.1038/s41587-019-0290-0>.
5. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell*. 2020;183:1103–1116.e20. <https://doi.org/10.1016/j.cell.2020.09.056>.
6. Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, Takeshima Y, et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol*. 2021;39(10):1246–58. <https://doi.org/10.1038/s41587-021-00927-2>.
7. Swanson E, Lord C, Reading J, Heubeck AT, Genge PC, Thomson Z, et al. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *Elife*. 2021;10:e63632. <https://doi.org/10.7554/eLife.63632>.
8. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol*. 2018;19(1):224. <https://doi.org/10.1186/s13059-018-1603-1>.
9. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016;48(10):1193–203. <https://doi.org/10.1038/ng.3646>.
10. Yang S, Corbett SE, Koga Y, Wang Z, Johnson WE, Yajima M, et al. Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol*. 2020;21(1):57. <https://doi.org/10.1186/s13059-020-1950-6>.
11. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
12. Liu B, Li C, Li Z, Wang D, Ren X, Zhang Z. An entropy-based metric for assessing the purity of single cell populations. *Nat Commun*. 2020;11(1):3155. <https://doi.org/10.1038/s41467-020-16904-3>.
13. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16(12):1289–96. <https://doi.org/10.1038/s41592-019-0619-0>.
14. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv. 2018.

15. Melsted P, Boeshaghi AS, Liu L, Gao F, Lu L, Min KHJ, et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nat Biotechnol.* 2021;39(7):813–8. <https://doi.org/10.1038/s41587-021-00870-2>.
16. Broad Institute. Picard toolkit. <http://broadinstitute.github.io/picard/>. Accessed 15 Jun 2021.
17. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. *Nat Methods.* 2021; 18(11):1333–41. <https://doi.org/10.1038/s41592-021-01282-5>.
18. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 2019;20(1):296. <https://doi.org/10.1186/s13059-019-1874-1>.
19. Ahlmann-Eltze C, Huber W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics.* 2021;36:5701–2. <https://doi.org/10.1093/bioinformatics/btaa1009>.
20. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science.* 2015;348(6237):910–4. <https://doi.org/10.1126/science.aab1601>.
21. Xu Z, Heidrich-OHare E, Chen W, Duerr RH. Comprehensive benchmarking of CITE-seq versus DOGMA-seq single cell multimodal omics. *Datasets. Gene Expression Omnibus.* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE200417>. Accessed 21 May 2022.
22. Xu Z, Heidrich-OHare E, Chen W, Duerr RH. Comprehensive benchmarking of CITE-seq versus DOGMA-seq single cell multimodal omics. *GitHub.* https://github.com/xzlandy/Benchmark_CITEseq_DOGMAseq. Accessed 21 May 2022.
23. Xu Z, Heidrich-OHare E, Chen W, Duerr RH. Comprehensive benchmarking of CITE-seq versus DOGMA-seq single cell multimodal omics. *Zenodo.* 2022. <https://doi.org/10.5281/zenodo.6598749>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

