

METHODOLOGY

Open Access



SCITO-seq2: ultra-high-throughput single-cell transcriptome and epitope sequencing

Su-Hyeon Lee^{1†}, Bo-Yeong Jin^{2,3†}, Cho-Rong Lee⁴, Doo Ri Kim⁵, Areum Shin⁵, Sung-Gyoo Park⁴, Yae-Jean Kim⁵, Seong Heon Kim⁶, Murim Choi^{2*} and Byungjin Hwang^{1,7*}

[†]Su-Hyeon Lee and Bo-Yeong Jin contributed equally to this work.

*Correspondence:
murimchoi@snu.ac.kr;
bjhwang113@yuhs.ac

¹ Department of Biomedical Sciences, Yonsei University College of Medicine, Seoul 03722, Republic of Korea

² Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

³ BK21 FOUR Biomedical Science Program, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

⁴ College of Pharmacy and Research Institute of Pharmaceutical Science, Seoul National University, Seoul, Republic of Korea

⁵ Division of Infectious Diseases and Immunodeficiency, Department of Pediatrics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

⁶ Department of Pediatrics, Seoul National University Children's Hospital & College of Medicine, Seoul, Republic of Korea

⁷ Brain Korea 21 Project, Yonsei University College of Medicine, Seoul, Republic of Korea

Abstract

We introduce SCITO-seq2, an enhanced successor to SCITO-seq that integrates probe-based RNA detection with the established ultra-high-throughput protein profiling. SCITO-seq2 achieves robust quantification of transcripts and surface proteins across more than 100,000 cells, with a shared pool barcoding strategy ensuring precise matching of molecular profiles within multiplexed droplets. SCITO-seq2 is compatible with cell hashing technology, allowing efficient sample multiplexing. We demonstrate its utility in autoimmune diseases, including childhood systemic lupus erythematosus and CTLA4 haploinsufficiency with autoimmune infiltration, enabling the detection of minor immune clusters and disease-specific protein signatures. This platform establishes a scalable, streamlined, and cost-effective next-generation single-cell multi-omics workflow.

Keywords: Single-cell multi-omics, Ultra-high-throughput sequencing, Childhood systemic lupus erythematosus, CTLA4 haploinsufficiency with autoimmune infiltration

Background

Dissecting cellular heterogeneity is fundamental to elucidating biological variation in healthy and pathological states. Whereas any single type of omics approach cannot fully explain the mechanism underlying disease pathophysiology, recent advances in single-cell multi-omic technologies (transcriptomics, proteomics [1–5], epigenomics [6–13], etc.) have revolutionized our ability to uncover novel disease-related biomarkers. For example, CITE-seq [5] enables concurrent analysis of transcriptome and epitope profiles within individual cells, facilitating the discovery of novel cell types and states [14, 15]. However, as interest in large-scale single-cell studies, such as tissue atlas construction [16–18], developmental trajectory mapping [19, 20], and perturbation experiments [21], has increased, the throughput limitations of current systems have become increasingly apparent.



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

Combinatorial indexing strategies, which involve multiple rounds of indexing, have emerged as a scalable solution for single-cell technology [22–24]. These methods employ sequential DNA barcode incorporation via PCR or ligation, with successive pooling and splitting steps ensuring generation of unique single-cell barcodes, and have enabled million-scale single-cell experiments. Building upon this foundation, droplet-based two-step indexing methods were subsequently developed, offering simplified options for massive-scale experiments [25–27]. These methods pre-index the genome or transcriptome before super-Poisson loading of pooled cells into microdroplets containing hundreds of thousands of barcoded beads. Then, the multiple cells within a droplet are deconvoluted using a combination of pre-indexed and droplet barcodes, achieving markedly increased yield with reduced manual intervention. This technological framework has recently spawned several ultra-high-throughput multi-omic methods capable of jointly profiling RNA with ATAC [28, 29] or the VDJ repertoire [30]. However, there remains a need for an ultra-high-throughput technology for simultaneous RNA and protein profiling that can overcome the limitations of RNA-only measurements in capturing certain biological signals [5]. While our previous SCITO-seq method established ultra-high-throughput protein profiling and provided proof-of-concept RNA detection, efficient RNA capture remains challenging for comprehensive multi-omic analysis [27].

Here, we present SCITO-seq2, a scalable method for simultaneous profiling of the transcriptome and surface epitopes at an unprecedented scale ($> 10^5$ cells), which represents a substantial improvement in throughput over conventional CITE-seq [5]. Our approach combines probe-based hybridization capture with splint-oligo-hybridized antibodies, where a shared pool barcoding strategy enables precise matching of RNA and protein profiles even within multiplexed droplets. This design is further compatible with cell hashing technology [31], enabling efficient multiplexing of samples in a single high-throughput experiment. The application of SCITO-seq2 to peripheral blood mononuclear cells (PBMCs) demonstrated a strong correlation between surface protein and gene expression. We subsequently applied SCITO-seq2 to investigate patients with autoimmune conditions, including childhood systemic lupus erythematosus (SLE) and CTLA4 haploinsufficiency with autoimmune infiltration (CHAI), demonstrating the utility of the high-resolution multi-modal analysis.

Results

SCITO-seq2 enables ultra-high-throughput multi-omics sequencing

The implementation of ultra-high-throughput analysis in microfluidics-based single-cell partitioning presents two primary challenges: the generation of multiplets (droplets containing multiple cells) and their subsequent demultiplexing for data analysis. Resolving multiplets necessitates that each molecular modality within a droplet carries a pre-indexed pool barcode (PBC) encoding sample identity. We achieved this by integrating commercially available probe-based whole-transcriptome capture chemistry (Chromium Fixed RNA Profiling) with the splint-oligo-hybridized antibody technology previously established in SCITO-seq [27]. This probe-based chemistry employs hybridization-based RNA capture using pre-designed probe sets, enabling high-efficiency transcript detection while allowing the incorporation of PBC for molecule-level deconvolution. The combination of these techniques enables ultra-high-throughput

deconvolution of RNA and surface protein reads, where PBC-based cellular origin identification allows scalable multiplexing by maintaining precise molecular correspondence within multiplets (Fig. 1A).

To test the validity of our approach, we analyzed peripheral blood from healthy donors. PBMCs were divided into four replicate pools, each treated with a custom 209-antibody panel prepared by conjugating antibodies to splint oligos (Additional file 2: Table S1) and commercial RNA probes, enabling verification of data consistency across different pools derived from identical samples. After pool-specific barcoding, we combined the cells from each pool and loaded ~200,000 cells into a single channel for gel bead-in-emulsion (GEM) generation.

To assess the technical performance, we conducted droplet generation simulations based on our previous work, incorporating a fixed number of unique PBCs and cell loading parameters [27]. We then compared the simulated and experimental pool collision rates, defined as the proportion of droplets that encapsulate multiple cells originating from the same pool. Such *collided multiplets* cannot be resolved by pool barcode (PBC) information alone, as all constituent cells share the same PBC and thus remain as multiplets after deconvolution. In datasets incorporating cell hashtags, these collided multiplets can additionally be recognized by mixed hashtag signals, providing an extra layer of QC. To infer the overall frequency, we quantified cells that were removed during QC and doublet filtering due to exhibiting multiplet-like characteristics—such as abnormally high total UMI counts, elevated feature numbers, or mixed phenotypic marker expression. Using this approach, the experimental collision rate was estimated at 10.69%, which was in close agreement with the simulation-based estimate (11.95%) (Fig. 1B). Notably, our simulation revealed that increasing the number of PBCs reduced collision rates even at high cell loading densities, enabling more efficient single-cell recovery. Under our experimental setup, loading ~200,000 cells yielded 93,782 final analyzable cells (Fig. 1B). On the basis of this predictive model, we estimated that with 16 PBCs, loading 580,000 cells would yield more than a 15-fold increase in cell recovery (~290,000) compared with the single-plexed configuration (~19,000) while maintaining 50% recovery efficiency (Additional file 1: Fig. S1A).

Next, to trace pool-assigned cells to their original droplets, we performed parallel processing of data with and without pool barcode separation, enabling quantification of the number of cells per droplet. Our analysis revealed that 69.94% of the final 93,782 cells originated from multiplets, which were successfully deconvoluted into single-cell data through our pool barcoding strategy (Fig. 1C). Hence, this approach enabled an analysis that would otherwise be limited to only 28,189 singlet cells to include 65,593 additional cells, demonstrating the effectiveness of our multiplexing strategy in maximizing cell recovery.

We further examined the molecular capture efficiency by comparing gene expression (GEX) and protein expression (antibody-derived tag, ADT) UMI counts with SCITO-seq, conventional CITE-seq, and AbSeq on BD Rhapsody platform [26, 27, 32, 33]. SCITO-seq2 demonstrated robust GEX UMI detection with improved yield (median log₁₀-transformed UMI count 3.35 [IQR: 3.18–3.66]) compared to previous ligation-based SCITO-seq (1.64 [IQR: 1.36–1.98]) while maintaining ADT detection efficiency,

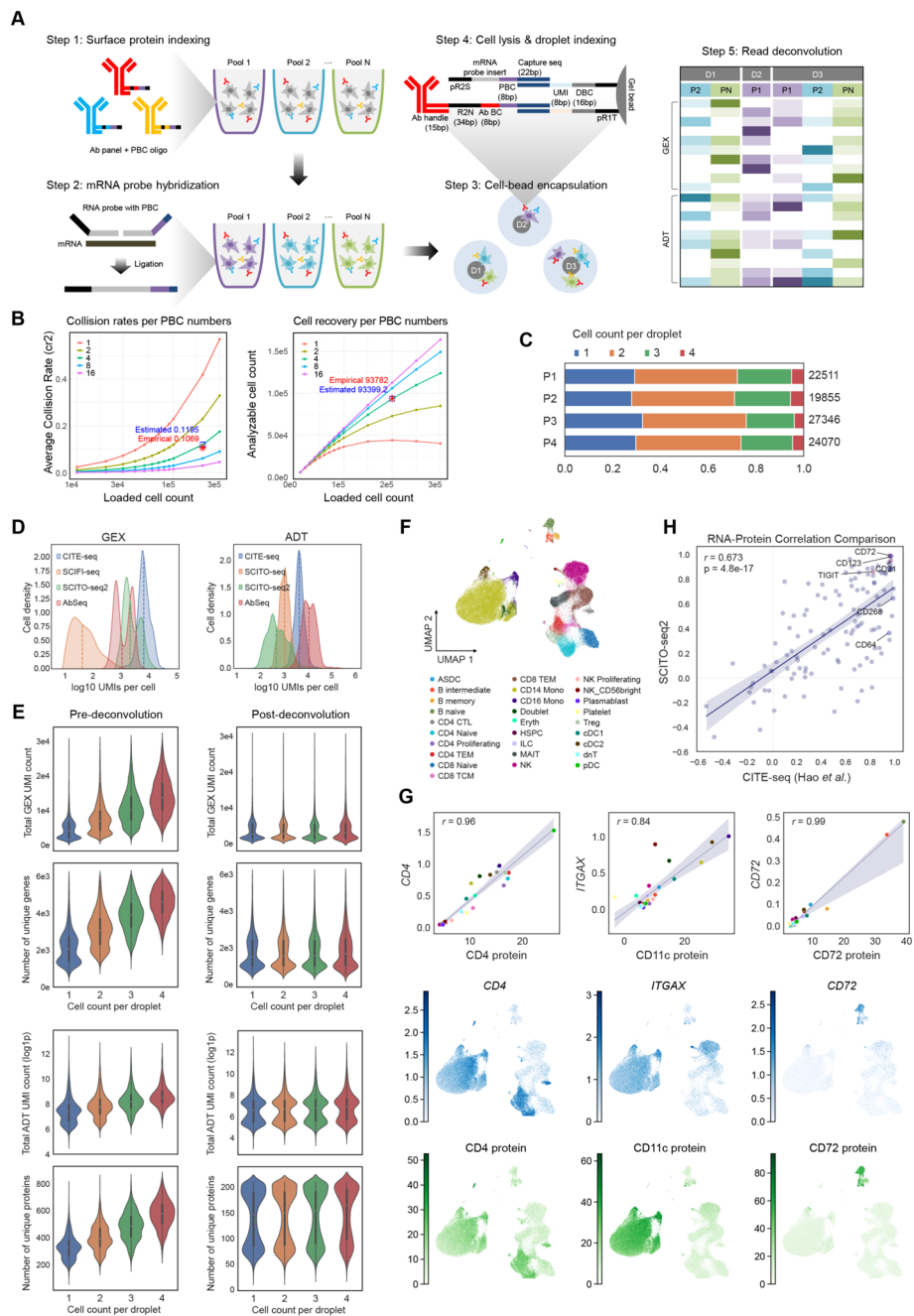


Fig. 1 Development and technical validation of SCITO-seq2 workflow. **A** Schematic overview of the SCITO-seq2 workflow. Pool-specific barcoding of both RNA and proteins enables the deconvolution of individual cells from multiplexed droplets. **B** Collision rates and cell recovery by loaded cell count at different PBC numbers. The plot shows simulated values (estimated; blue) based on the variables used in the experiment (cells loaded = 200,000, PBC number = 4) and the actual observed values (empirical; red). **C** Cell count per droplet from deconvoluted multiplets categorized by pool. **D** Comparison of CITE-seq, SCITO-seq, SCITO-seq2, and AbSeq. Density histogram of the total GEX log₁₀ UMI count per cell (left). ADT log₁₀ UMI count per cell (right). **E** Total UMI count and number of detected features by cell count per droplet before and after deconvolution. **F** UMAP visualization of quality control-filtered cells colored by annotated cell type. **G** Pearson correlation between RNA and protein expression levels (top), shown as cell-type averages for selected markers, with corresponding UMAP visualizations of RNA (middle) and protein (bottom) expression. **H** Correlation analysis between SCITO-seq2 and CITE-seq methods, comparing their respective RNA–protein correlations for proteins detected by antibodies from identical clones

although direct comparisons are limited by the variations in antibody panels and sequencing depth across studies (Fig. 1D).

To address potential concerns regarding reverse transcription efficiency in overloaded droplets, where multiple cells share reaction reagents within GEMs and bead oligos, we assessed quality control metrics in relation to the number of cells within a single droplet (Fig. 1E). After computational deconvolution, we observed no significant differences in the GEX and ADT metrics regardless of the number of cells per droplet, demonstrating successful deconvolution of multiplexed cells while preserving data quality.

To evaluate the fidelity of molecular correspondence following multiplet deconvolution, we performed correlation analysis between representative protein markers and RNA using cell type-averaged expression values [5], with reference-based annotation (Fig. 1F). This analysis demonstrated high concordance between RNA and corresponding protein expression levels, with strong correlations for key immune markers (CD4: $r=0.96$, CD11c: $r=0.79$, and CD72: $r=0.99$) (Fig. 1G). When we compared protein–RNA correlations between our dataset and the reference CITE-seq data, we observed consistent agreement for proteins detected by antibodies with shared clones (Fig. 1H and Additional file 1: Fig. S1B and C), although some proteins exhibited dataset-specific correlation biases (Additional file 1: Fig. S1D and E). Similarly, comparison with publicly available AbSeq data also revealed comparably high correlations (Additional file 1: Fig. S1F).

To gain additional insight into antibody signal characteristics, we examined finer-scale variation across markers (Additional file 1: Fig. S2). Specifically, we evaluated the separation of ADT signals between marker-positive and marker-negative cell populations for key lineage markers. For CD3 and CD8, ADT signal separation between marker-positive and marker-negative populations was reduced in SCITO-seq2 compared with other platforms, reflecting platform-specific differences in antibody signal characteristics. For the remaining markers, SCITO-seq2 showed signal separation comparable to that of the other methods.

Collectively, these results establish SCITO-seq2 as a robust platform for large-scale single-cell multi-omics analysis, demonstrating its ability to generate high-fidelity gene expression and surface protein measurements at single-cell resolution while effectively resolving multiplets.

Patient sample demultiplexing with multi-modal measurements via SCITO-seq2

Building upon the enhanced cell recovery achieved through pool-specific barcoding, we implemented additional sample multiplexing within each pool to fully leverage the high-throughput capability while maintaining a sufficient cell number per sample. Since probe-based chemistry employs indirect RNA capture for improved efficiency rather than direct capture methods that enable genetic variant-based demultiplexing, we utilized antibody-based cell hashing for sample multiplexing within pools (Fig. 2A and Additional file 1: Fig. S3A). To demonstrate the clinical utility of this approach, we analyzed PBMCs from patients with two autoimmune diseases: childhood systemic lupus erythematosus (SLE) and CTLA4 haploinsufficiency with autoimmune infiltration (CHAI), a rare genetic disorder (Additional file 2: Table S2). Samples from four SLE patients and four healthy controls were labeled with unique hashtag oligo antibodies and

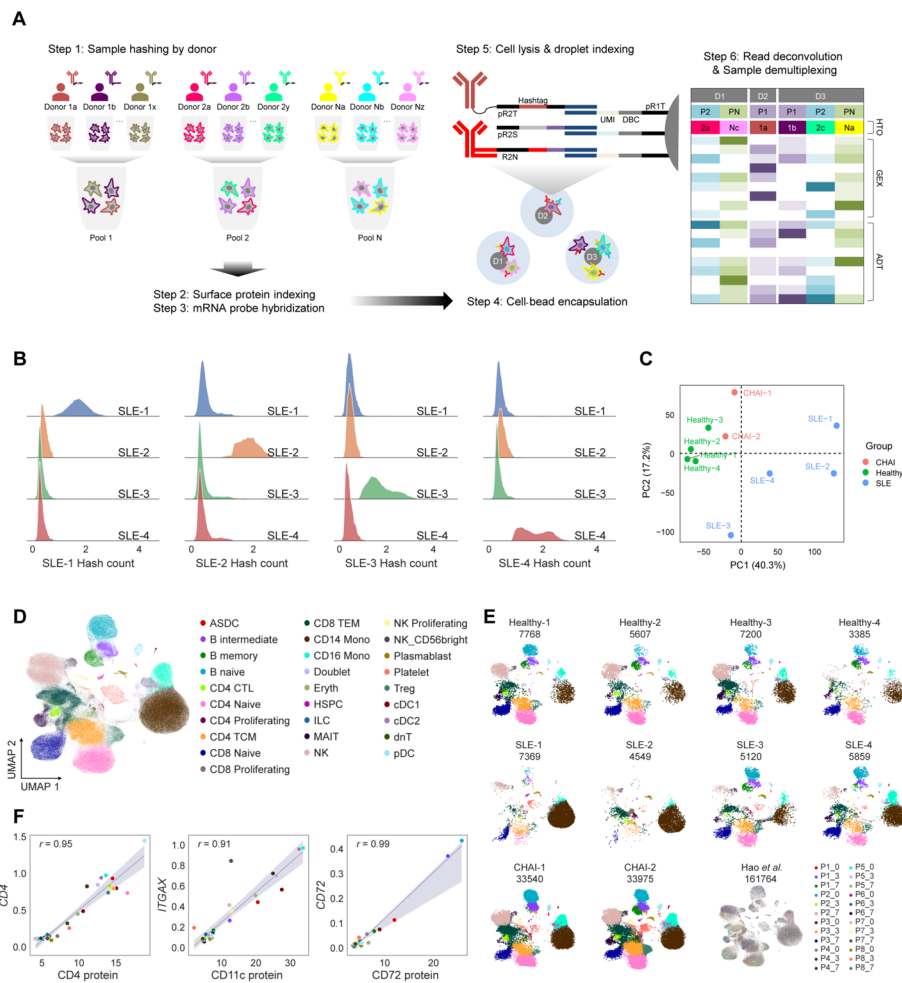


Fig. 2 Application and validation of SCITO-seq2 multiplexing strategy in clinical samples. **A** Overview of the donor multiplexing strategy in SCITO-seq2, which combines pool barcoding with antibody-based cell hashing technology. **B** HTO count (central log ratio-normalized) distribution in cells assigned to each SLE donor. Cells with negative or multiple donor assignments were excluded from further analysis. **C** PCA of pseudobulk RNA expression data showing clustering of individual donors. **D** and **E** UMAP visualization of co-embedded SCITO-seq2 and public CITE-seq data (Hao et al.) after integration. **F** Pearson correlation of cell type-averaged RNA and protein expression for cell type markers in SCITO-seq2 data

pooled within their respective groups, while the two CHAI samples were processed as separate pools. A detailed step-by-step protocol covering sample preparation through library construction is provided in the [Methods](#) section and on protocols.io (<https://doi.org/10.17504/protocols.io.14egn1z6pv5d>).

To evaluate the robustness of our multiplexing strategy, we processed the samples following our established workflow to generate three types of libraries: GEX, ADT, and hashtag oligos (HTO). Analysis of the pool barcode distribution confirmed their high specificity, with GEX barcodes showing clear separation between pools and ADT barcodes demonstrating minimal cross-pool contamination (Additional file 1: Fig. S3B and C). This dual-barcoding approach enabled stringent filtering of potential background signals, ensuring the reliable cell-to-sample assignment essential for downstream clinical sample analysis.

Following pool-level separation, we next deconvoluted individual donor identities within the SLE and healthy pools through HTO-based demultiplexing (Fig. 2B and Additional file 1: Fig. S3D). Principal component analysis (PCA) of the demultiplexed samples revealed clustering of samples within their respective expected groups while maintaining clear separation between donors within the same pool (Fig. 2C), confirming the robust compatibility of the hashtag system with the SCITO-seq2 platform. After rigorous quality control, including doublet removal and unassigned HTO filtering, we obtained a final dataset of 114,372 cells.

We next evaluated the compatibility of the SCITO-seq2 data through integration with existing CITE-seq datasets (Fig. 2D). The results confirmed high compatibility, with cell type classifications consistent with the public reference data while preserving sample-specific distributions (Fig. 2E). Data quality was further evidenced by consistently high protein-RNA correlations for key markers (Fig. 2F).

Finally, having established the technical robustness and data quality of our multiplexing strategy, we evaluated its cost-effectiveness (Additional file 1: Fig. S3E). Cost analysis comparing SCITO-seq2 with a standard 3' sequencing workflow demonstrated that its shared library preparation substantially reduces per-cell costs as sample numbers increase. Consequently, SCITO-seq2 achieves comparable per-cell costs to the conventional approach when analyzing around nine samples, and becomes increasingly more cost-effective when analyzing twelve or more samples, with greater advantages for larger-scale studies. Together, these results establish SCITO-seq2 as a robust, compatible, and cost-efficient platform for large-scale multi-omics analysis.

Improvement of cell clustering by inclusion of ADT

To assess the benefit of including ADT during analysis for aberrant genetic features, we plotted all quality-filtered cells and directly compared the cell clustering patterns obtained using GEX data only (the “GEX-only” set) or including ADT data with the GEX data (the “GEX + ADT” set). The overall clustering pattern and cell type composition appeared largely similar, but pronounced differences were observed (Fig. 3A and B). More specifically, the UMAP pattern in the GEX + ADT set was better separated, providing higher resolution among clusters of the same group, such as T and B cell groups. In addition, clusters that were not detected in the GEX-only set emerged in the GEX + ADT set. For example, the CD56 “bright” natural killer (NK) cell (NK_CD56bright) cluster, a subset of NK cells mainly detected in tissues, was detected in the GEX + ADT set, which became possible by measuring CD56 protein expression (Fig. 3C and D). Similarly, the plasmablast (PB) cluster with high expression of *MZB1* and *JCHAIN* diverged from the intermediate B cell (B intermediate) cluster (Fig. 3C and Additional file 1: Fig. S4A), and the cDC1 cluster was found to be separated from the cDC cluster with strong expression of *CLEC9A* and *DNASE1L3* (Fig. 3C and Additional file 1: Fig. S4B). Additionally, a dnT cluster characterized by high *PTPN3* gene expression and low CD4 and CD8 ADT expression was identified (Fig. 3C and Additional file 1: Fig. S4C). Finally, the hematopoietic stem and

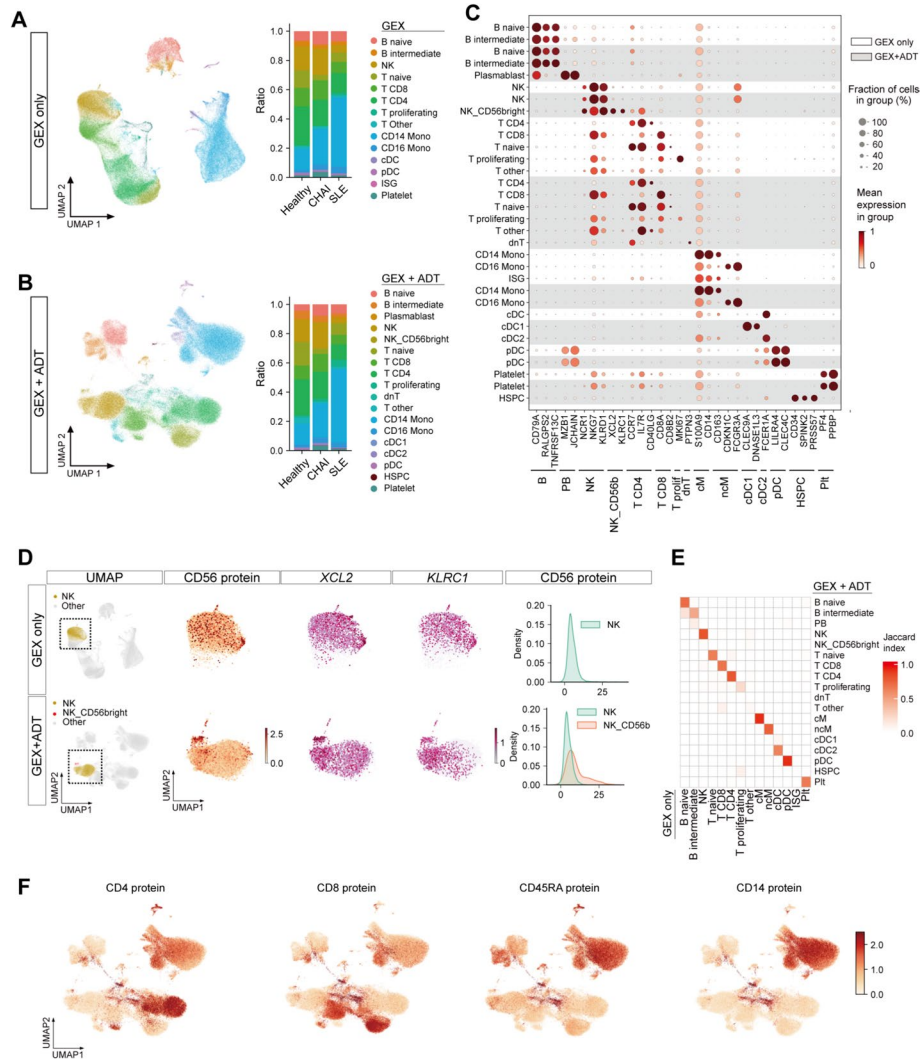


Fig. 3 The incorporation of ADT increased the clustering ability. **A** and **B** Cell cluster profiles by UMAP clustering and the cell type ratio using GEX data alone [“GEX-only”, (A)] and GEX and ADT together [“GEX + ADT”, (B)]. **C** Gene markers distinguishing each cluster. To facilitate direct comparisons of the two sets, cell types are grouped by expression pattern and presented by dataset in an alternating order. **D** Separation of the NK_CD56bright cluster with consideration of CD56 protein expression and the expression intensity of CD56 protein, *XCL2*, and *KLRC1*. The top row shows the GEX-only set, and the bottom row shows the GEX + ADT set. Plots at the far right display the distribution of CD56 protein expression intensity among NK cells. **E** Heatmap depicting pairwise Jaccard index scores from the comparison of the GEX-only and GEX + ADT sets. **F** Expression of selected proteins projected into the UMAP plot

progenitor cell (HSPC) cluster, characterized by high expression of *CD34*, represents a group with erythroid potential, potentially poised for differentiation into erythroid cells (Fig. 3C). These observations were summarized in the form of pairwise comparisons between the two sets (Fig. 3E), and the expression of cluster-specific proteins is shown in Fig. 3F. These findings are in line with prior multi-omic studies [34], which demonstrated that integrating transcriptomic and surface-protein measurements improves resolution and enables identification of additional immune

subsets, such as cytotoxic CD4⁺ T cells and mesenchymal stem cells, that are not readily distinguished using transcriptomic data alone.

Identification of disease-specific protein markers for SLE and CHAI

We next sought to identify new biological features that ADT may reveal in the SLE and CHAI patient cells. To address SLE samples, we co-clustered our cells with publicly available cells from adult and childhood SLE patients, comprising data from 5 adult healthy donors, 7 adult SLE (aSLE) patients, 11 child healthy donors, and 33 child SLE (cSLE) patients [35]; the results revealed a comparable clustering pattern (Fig. 4A). In terms of the overall composition of major cell groups, our patient set showed higher enrichment in monocytes, probably due to a greater range of case severity (Fig. 4B). A number of proteins showed markedly higher expression in SLE samples compared to healthy samples (Fig. 4C). In addition, SLE samples presented increased interferon (IFN) and cytotoxic scores for the major cell types, similar to the aSLE and cSLE samples from the previous study (Fig. 4D and E) [35]. Assessment of SCENIC-inferred IRF

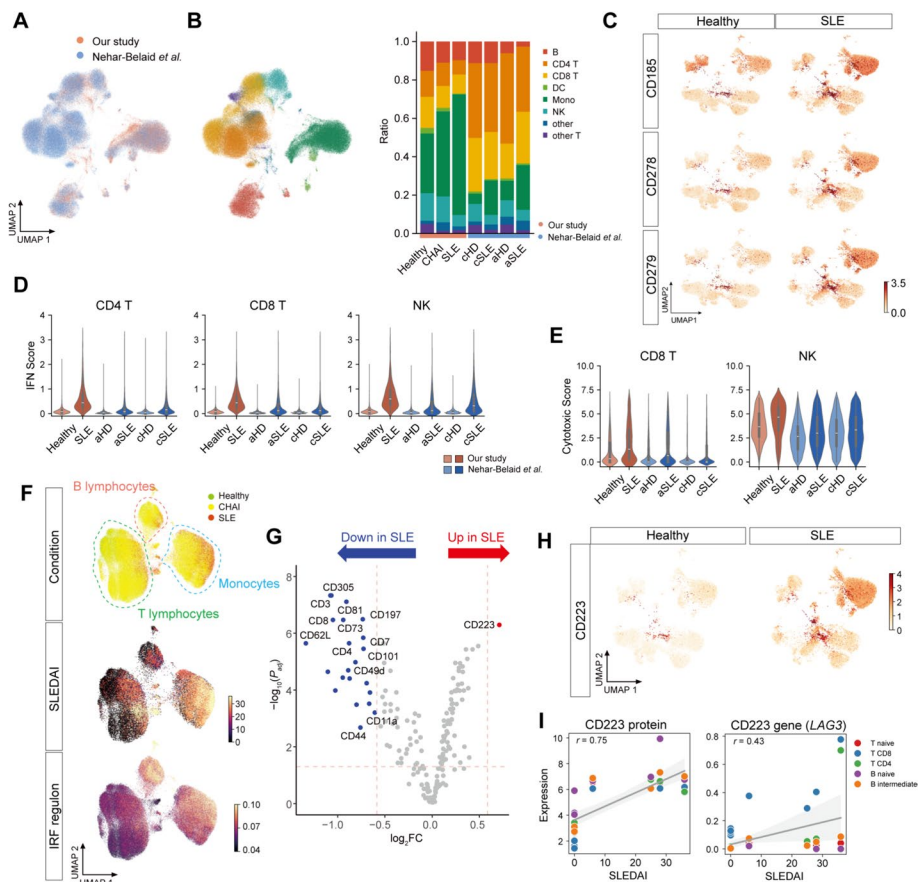


Fig. 4 Characteristics of child SLE cells resolved with GEX and ADT. **A** Co-clustering of cells from our SLE patients and Neihar-Belaid et al. **B** Combined UMAP and the ratio of the major cell clusters from the two datasets. **C** Expression of selected proteins projected into the UMAP. **D** and **E** IFN scores (**D**) and cytotoxic scores (**E**) in each disease group for the designated cell types. **F** UMAP plots displaying combined regulon activities, annotated by patient status (top), SLEDAI score (middle), and IRF regulon (bottom). **G** Volcano plot displaying DEPs between the SLEDAI patients and healthy individuals. **H** Expression of CD223 protein in UMAP clusters. **I** Correlations of the SLEDAI score with CD223 protein (left) and its transcript, *LAG3* (right)

transcriptional regulon activity revealed a stratified distribution in each major cell type, which aligned well with the SLE disease activity index (SLEDAI), a symptom severity score (Fig. 4F). Differentially expressed protein (DEP) analysis led to the identification of a single protein, CD223, that was significantly increased in the SLE samples, along with 21 proteins that were significantly decreased (Fig. 4G and H).

Additionally, we observed that protein expression provides better prediction of SLEDAI score than does RNA expression (155/180, or 86.1% for all RNA–protein pairs and 94/105, or 89.5% for pairs with positive correlations in both) (Additional file 1: Fig. S5). Among these, we noted that SLEDAI score correlation with CD223 protein or *LAG3* gene expression showed the score to be more strongly explained by protein expression. CD223 is an inhibitory receptor encoded by the *LAG3* gene, and its upregulation is associated with SLE [36] (Fig. 4I). This finding allows us to emphasize the functional role of CD223 in SLE due to enhanced capture rate of protein expression and the utility of the ADT-based approach in the evaluation of SLE severity. In addition to the single-marker correlations, we evaluated modality-level informativeness for disease activity using patient-level leave-one-patient-out (LOPO) cross-validation. While limited by the small cohort and the targeted scope of the antibody panel, ADT outperformed RNA across all linear and nonlinear models (Additional file 2: Tables S3 and S4).

Next, we investigated whether ADT data could shed additional insights on the molecular changes in CHAI patient samples. DEP analysis of CHAI versus healthy samples demonstrated increased expression of CD103 protein in the monocyte cluster (Additional file 1: Fig. S6A to C). Visualization by a heatmap further indicated CHAI- and SLE-specific expression of CD103 and CD223, respectively (Fig. S6D). Finally, DEG analysis in pseudobulk cells revealed a number of regulated genes, with significant enrichment of ontology terms related to immune activation and inflammation (Additional file 1: Figs. S6E to G and S7). Taken together, these results highlight the increased resolution of our method and its utility in analyzing patient samples, thereby facilitating the discovery of disease-associated biomarkers.

Discussion

The advent of single-cell technologies has revolutionized biological discovery by enabling high-resolution analyses. Consequently, providing robust evidence through large-scale, multi-modal investigations has become increasingly crucial. Here, we demonstrate that SCITO-seq2 enables precise transcriptome and protein analysis at an unprecedented scale. Through an optimized experimental design, we achieved recovery of >100,000 analyzable cells from a single microfluidic channel while maintaining high-quality GEX and ADT data across multiple groups and samples. Our platform demonstrates significant technical advantages over its predecessor, SCITO-seq. Probe-based RNA molecular capture efficiency enables comprehensive multi-modal analysis at scale, with droplet formation dynamics that closely mirror theoretical predictions. This improved predictability empowers researchers to design large-scale studies with greater precision in terms of the expected outcomes.

It is worth noting that, during the course of this work, 10× Genomics introduced a multiplexed transcriptome and antibody profiling protocol based on Flex chemistry. However, this platform employs a ligation-based TotalSeqC design that is

chemistry-specific, whereas SCITO-seq2 implements a splint-oligo hybridization strategy, resulting in a chemistry-independent configuration that forms the conceptual basis of SCITO-seq2's flexibility. Specifically, the splint oligo hybridizes to the feature barcode region of antibody-conjugated oligos, enabling compatibility with any antibody format (TotalSeqA, B, or C). In addition, by altering the PCR handle, the same framework can be extended to additional modalities, as demonstrated by its integration with cell-hashing antibodies in this study. This modular design not only provides potential compatibility with modalities such as sgRNA capture and ATAC barcoding but also enables immune-receptor profiling through substitution of the capture sequence with a template-switch oligo sequence. Consequently, SCITO-seq2 offers an open and customizable framework distinct from commercial kits optimized for standardized use, providing researchers with the adaptability required to integrate emerging single-cell modalities and scale their studies efficiently.

Importantly, we show that the recovered cells, mostly derived from deconvoluted multiplets, exhibit comparable data quality to singlets, as validated by protein-transcript correlations, demonstrating strong concordance with CITE-seq measurements. In our marker-level analysis of antibody profiles, we aimed to assess variability in signal quality, an important consideration in antibody-based protein detection. We found that several cell type-specific markers that were distinctly separated in other platforms, such as CD3, showed poorer signal separation SCITO-seq2. Given that SCITO-seq2 and SCITO-seq share the same antibody-oligo chemistry, these observations may reflect the effects of the fixation process introduced for RNA profiling. Such effects could arise from enhanced retention of unbound antibodies within droplets or from reduced antibody binding efficiency due to epitope alteration or masking during fixation.

The robustness and scalability of SCITO-seq2 are further validated through its successful integration with cell hashing, enabling simultaneous analysis of multiple donors and conditions. We demonstrated this capability through comprehensive analyses of two autoimmune disorders: SLE and CHAI. The platform's performance is evidenced by the clear manifestation of pathological phenotypes in our SLE dataset, effectively capturing the disease state when analyzed alongside published datasets. In this study, samples from donors belonging to the same disease group were processed within the same pool and distinguished by sample-level hashtags. Designing future experiments in which different biological groups are mixed across pools—while still maintaining unique hash barcodes for individual donors—could help minimize potential batch effects and improve comparability between groups.

Of particular note, SCITO-seq2's high-resolution protein measurements revealed cellular features that were undetectable through transcriptomic analysis alone. The integrated analysis identified minor populations, including NK_CD56bright cells and PBs, generally indistinguishable when using GEX-based clustering. These populations showed disease-state-dependent variations: the NK_CD56bright/NK ratio was markedly elevated in SLE patients (11.8%) compared to healthy controls (2.9%) or CHAI patients (1.7%), with similar trends observed in the PB/B cell ratios (SLE: 1.4%, healthy: 0.3%, CHAI: 0.2%). These ratios were positively correlated with the SLEDAI score in our SLE patients, consistent with previous reports (Additional file 1: Fig. S8) [37, 38]. Additionally, DEP analysis identified CD223 as an SLE-specific immune marker, demonstrating

a stronger positive correlation with the SLEDAI score than its corresponding transcript (*LAG3*). Consistent with our findings, independent flow cytometric data from an external SLE cohort also demonstrate elevated CD223 protein [39].

In CHAI, despite moderate disease signatures due to genetic background and ongoing clinical treatments (as evidenced by their PCA proximity to healthy controls, Fig. 2C), we detected consistent immunological features, including significant upregulation of CD103 protein and genes involved in immune cell migration pathways. These tissue residency-associated features provide novel insights into disease-specific immune states, which is particularly valuable for CHAI, which has not been previously characterized at single-cell resolution. While the identified markers warrant further validation, SCITO-seq2 represents an important advancement in our ability to conduct comprehensive multi-modal analyses of complex biological systems at scale.

Conclusion

SCITO-seq2 establishes a robust technological framework for large-scale single-cell RNA and surface protein analysis, demonstrating significant improvements in scalability and data quality. The platform's capabilities enabled novel biological insights into well-studied conditions such as SLE while establishing new research directions for previously unexplored rare genetic disorders like CHAI.

Methods

Single cell multi-ome experiment

PBMC sample preparation

All experiments were performed using primary human PBMCs, and no established cell lines were used. Therefore, cell line authentication and mycoplasma testing were not applicable. For PBMC isolation, fresh blood samples were processed on the day of collection. The blood was diluted with an equal volume of PBS and carefully layered over an equivalent volume of Histopaque. The samples were then centrifuged at $400 \times g$ for 30 min at room temperature (RT) with minimal acceleration and deceleration. Following centrifugation, the opaque PBMC layer at the interface was carefully transferred to a new tube using a transfer pipette. The isolated PBMCs were washed twice with pre-cooled 2% FBS in PBS and centrifuged at $250 \times g$ for 10 min at 4 °C. The washed PBMCs were then resuspended in Cell Banker 2, transferred to cryovials, and stored in liquid nitrogen for long-term preservation. To prepare cryopreserved PBMCs for sequencing, the vials were rapidly thawed in a 37 °C water bath. The thawed cells were immediately mixed dropwise with warm thawing buffer (TB; 20% FBS in RPMI) to minimize osmotic shock. The cells were then centrifuged at $300 \times g$ for 10 min at RT, resuspended in fresh TB, and centrifuged again. After counting, the PBMCs were resuspended in labeling buffer (LB; 10% BSA in PBS) for subsequent experiments.

Cell hashing

In the Healthy-SLE-CHAI group experiment, cells from healthy and SLE samples were subjected to cell hashing prior to pooling to facilitate subsequent sample demultiplexing. Each sample was resuspended in labeling buffer (LB), followed by a blocking step with

5 μl of FcX for ten minutes on ice. After blocking, 1 μg of TotalSeq-B hashing antibody was added, and incubated for 30 min on ice. The cells were then washed twice with LB and centrifuged at $400 \times g$ for five minutes at 4 °C to remove unbound antibodies.

Oligonucleotide and antibody panel hybridization and surface protein binding

The custom TotalSeq-A antibody cocktail was prepared according to the manufacturer's instructions outlined in BioLegend's "TotalSeq™ Universal Cocktails Instructions for Use." Briefly, the antibody vials were equilibrated to RT, briefly centrifuged, and reconstituted in 27.5 μl (55 μl for pilot studies) of LB. The reconstituted antibody solutions were incubated for 5 min at RT. The antibody solutions were subsequently transferred to fresh microcentrifuge tubes and centrifuged at $14,000 \times g$ for 10 min at 4 °C.

Oligonucleotide pools were designed to include antigen-specific antibody binding sites, antigen barcodes, pool barcodes, and capture sequences. For Read 2 primer capture sequences, those used in conjunction with hashtags were designed with Nextera, whereas those not associated with hashtags were designed using TruSeq. A volume of 1.5 μl of 1 μM oligonucleotide, corresponding to each target protein within a given pool, was added to the respective antibody solutions for hybridization. The reaction mixtures were incubated for 15 min at RT, after which unhybridized oligonucleotides were removed by diluting the mixture with PBS and filtering it through a 50 kDa Amicon filter.

The hybridized antibodies were then concentrated by centrifugation at $14,000 \times g$ for 5 min, with the supernatant retained inside the filter. The recovered antibodies were stored on ice until further use. Each oligo-hybridized antibody panel was applied to a pool of multiplexed samples from the same group and incubated on ice for 30 min. The vials containing each pool were finally washed twice with labeling buffer (LB) and centrifuged at $400 \times g$ for 5 min at 4 °C after each wash.

Cell fixation and RNA probe hybridization

Following antibody binding, the cells were fixed according to the $10 \times$ Genomics protocol (CG000478, Rev C). Each pool was resuspended in 1 ml of fixation buffer (1X Fix & Perm Buffer [PN-2000517] in 4% formaldehyde in nuclease-free water) and incubated for one hour at RT or overnight at 4 °C. After fixation, the cells were centrifuged at $850 \times g$ for 5 min at RT and resuspended in 1 ml of cold quenching buffer (1X Quench Buffer [PN-2000516] in nuclease-free water). RNA probe hybridization was performed according to the $10 \times$ Genomics protocol (CG000527, Rev D). The fixed cells were centrifuged at $850 \times g$ for 5 min at RT, resuspended in Hyb mix and treated with Human WTA Probes (PN-2000495–8). After overnight incubation at 42 °C, Post-Hyb wash buffer (WB) was added, and the cells were pooled into a single vial to reach the desired cell count. The pooled sample underwent three rounds of washing with WB, including incubation at 42 °C and centrifugation at $850 \times g$ for 5 min at RT. Following the final wash, the cells were resuspended in Post-Hyb resuspension buffer (RB), filtered through a 30 μm filter, and prepared for gel bead-in-emulsion (GEM) generation.

Single-cell multi-ome sequencing

In both experiments described in this study, 200,000 cells were loaded onto a Chromium Next GEM Chip Q, and GEMs were generated via the Chromium X system. Library construction was carried out according to the 10× Genomics fixed RNA profiling protocol (CG000477). During the pre-amplification step, 1 µl of TotalSeq C-additive primer (100 µM) was spiked in to amplify reads containing the Read 2 Nextera sequence. The completed libraries were sequenced on an Illumina NovaSeq 6000 platform using 150 paired-end reads.

Computational analysis pipeline

Raw data preparation

The gene expression (GEX) data was split into four files based on the WTA barcode sequences. For comparison, an additional processing pipeline was applied to the unseparated file. Meanwhile, all the surface protein expressions (antibody-derived tag, ADT) and hashtag oligo (HTO) data were merged for uniform reference alignment and counting. For each of the four GEX files, read counting was performed using the Cell Ranger multi pipeline with the 10× Genomics-provided reference (Chromium Human Transcriptome Probe Set v1.0.1 GRCh38-2020-A). ADT and HTO data were processed using custom barcode reference. Sample assignment within each dataset was conducted via HTODemux based on the resulting hashtag counts. Subsequently, the denoised and scaled by background method was applied to normalize the ADT counts [40], leveraging unfiltered data. Finally, the four processed GEX files, sample assignment information, and normalized ADT counts were merged into a single dataset.

Quality control

To exclude damaged cells unsuitable for downstream analysis, obsolete cells and genes were first removed. The cells were subsequently filtered based on criteria determined by the distance from the trimmed mean in units of trimmed standard deviation (div), which was calculated by excluding the top and bottom 5% of values. The filtering thresholds were set as follows: mitochondrial gene count proportion ($< 10 * \text{div}$), hemoglobin gene count proportion ($< 10 * \text{div}$), number of detected genes ($\geq -1.75 * \text{div}$ and $< 5 * \text{div}$), and total gene count ($\geq -1.25 * \text{div}$ and $< 7.5 * \text{div}$).

Doublet removal

The data were first processed to remove clusters with high doublet prediction scores identified by DoubletDetection [41] and Scrublet [42]. RNA count normalization was performed, and 2,000 highly variable genes were identified using the Seurat_v3 method, with each pool treated as a batch key. After principal component analysis (PCA), batch correction was applied via Harmony across 50 components [43] and across individual donors. The corrected components were then used for nearest-neighbor analysis and Leiden clustering [44], identifying 45 distinct clusters. Doublet prediction was subsequently performed on each pool, and clusters exceeding both the Scrublet score

threshold of 0.275 and the DoubletDetection score threshold of 50 were removed. These thresholds were established after confirming the actual doublet status through hashtag assignment and verifying the co-expression of markers from different cell types within each cluster.

Cell type prediction using GEX data

A semi-supervised method was employed for cell type prediction based on RNA expression. The model was established using scVI [45] trained on a CITE-seq reference dataset comprising 211,000 human PBMCs [33]. The model utilized 2,981 common highly variable genes shared between the reference and query datasets, with individual donors treated as batch keys. Model training was conducted over 300 epochs. Subsequently, scANVI was applied to train both the reference and query datasets (GEO ID: GSE164378) [46]. The reference dataset's cell types were used as labels for supervised training performed over 40 epochs. This trained model was then used for the query dataset, which underwent 500 epochs of training to predict cell types.

Collision rate simulation

We developed a three-step computational model to simulate the cell distribution in droplets and estimate collision rates. Let N_{drops} denote the total number of droplets, N_{cells} the number of loaded cells, N_{pools} the number of pool barcodes, and η the droplet recovery rate. We assumed that $\eta = 0.6$ and $N_{\text{drops}} = 2 \times 10^5$ per microfluidic reaction. The probability of observing k cells in a droplet follows a Poisson distribution:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where $\lambda = \frac{N_{\text{cells}}}{N_{\text{drops}}}$ represents the mean number of cells per droplet. The recovered droplets ($D_{\text{recovered}}$) are sampled from the total droplet population (D_{total}):

$$D_{\text{recovered}} = \text{Sample}(D_{\text{total}}, \eta \cdot |D_{\text{total}}|).$$

For each droplet i containing k cells, pool barcodes (BC_{pool}^i):

$$BC_{\text{pool}}^i = \{b_1, \dots, b_k\},$$

where $b_j \sim \text{Uniform}(1, N_{\text{pools}})$. The number of unique barcodes (U_i) in each droplet is then calculated as:

$$U_i = \left| \text{unique} \left(BC_{\text{pool}}^i \right) \right|.$$

The pool collision count (M_i) for each droplet i is computed as:

$$M_i = \sum_{j=1}^{N_{\text{pools}}} \mathbb{1}_{[\text{count}(j, BC_{\text{pool}}^i) > 1]},$$

where $\mathbb{1}$ is an indicator function that equals 1 if the count of barcode j is greater than 1 and 0 otherwise. The pool collision rate (CR_p) is then computed as:

$$CR_p = \frac{\sum_i U_i}{\sum_i M_i}.$$

Finally, the total number of rescuable cells (Y_{total}) is estimated as:

$$Y_{\text{total}} = N_{\text{cells}} \cdot \eta \cdot (1 - CR_p).$$

Cost calculation

For SCITO-seq2, per-cell costs included library preparation (Chromium kit, indices, and chip), sequencing, and antibody/oligo reagents. The latter also covered additional hybridization oligos and hashtag antibodies when the number of pooled samples exceeded the barcode limit of 16. Library preparation costs were scaled according to the number of GEMs required, assuming that up to 58 pooled samples could be processed within a single GEM, based on an input of 10,000 cells per sample and considering 50% recovery efficiency as the acceptable lower limit (corresponding to ~580,000 cells per GEM). Additional GEMs were used when this capacity was exceeded. Hashtag antibody costs were distributed proportionally across all loaded cells, based on the number of samples that required hashing. For CITE-seq, per-cell costs included library preparation, sequencing, and antibody reagents, with multiplexing assumed to rely on natural genetic variation rather than hashtag antibodies. All costs were normalized by the total number of loaded cells per experiment to obtain per-cell estimates. The assumptions and calculation workflow used to generate these estimates are implemented in the public Jupyter notebook (`cost_calculation.ipynb`) available in our GitHub repository (see *Data availability* section).

Unsupervised clustering

Cell clustering based on GEX was performed using the Seurat v5 package [47]. The raw counts were normalized using the SCTransform function [48], specifying individual donors as a batch key to correct for batch effects. Data integration was performed using canonical correlation analysis. The nearest-neighbor graph was constructed using the integrated dimensions, and clustering was performed at a resolution of 0.4 using the Leiden algorithm [44]. For analyses integrating both GEX and ADT data, the totalVI package was utilized [49]. Raw RNA and ADT counts were jointly modeled using totalVI, with individual donors specified as a batch key. The integrated latent representations learned by totalVI were used for unsupervised clustering using the Leiden algorithm [44]. UMAP visualization was carried out based on the totalVI latent space. To compare the clustering patterns from the two methods, the PairWise-JaccardSetsHeatmap function from the scclusteval package was employed [50], which allowed for pairwise comparisons of clustering sets. Jaccard index scores were represented as a heatmap to visualize the degree of overlap between clusters from the GEX-only set and the GEX-ADT combined set.

Public data integration

Publicly available SLE scRNA-seq data was downloaded from the Gene Expression Omnibus (GSE135779) [51]. Missing protein expression was imputed using the

totalVI framework [49]. Harmony batch correction was applied, and downstream analyses were subsequently performed, including unsupervised clustering and visualization with UMAP. Cell annotation was conducted using Azimuth Level 1 [33].

Downstream analysis

Gene regulatory network analysis was conducted using the standard SCENIC workflow [49] to infer regulons and assess transcription factor activity. In brief, three main steps are involved: co-expression network inference, candidate regulon detection, and regulon activity scoring. Raw gene expression counts were used as input. Motif enrichment analysis utilized the cisTarget databases generated using the 2022 SCENIC + motif collection [49]. Differential gene expression (DEG) analysis was conducted using the memento package [49], with the capture rate set to 0.105. DEGs were defined based on the criteria of $de_pval < 0.05$ and $de_coef > 0.5$. Differential protein expression (DEP) analysis was performed using the limma package [52] based on protein pseudobulk counts. Normalization was carried out using the trimmed mean of the M-values method, followed by the voom transformation to estimate the mean–variance trend and ensure that the data met the assumptions of the linear model. DEPs were defined based on the criteria of adjusted P -value < 0.05 and \log_2 fold change $> \log_2(1.5)$.

Patient-level prediction of disease activity

To assess modality-level informativeness, we predicted the SLE Disease Activity Index (SLEDAI) using leave-one-patient-out (LOPO) cross-validation. For linear models (Ridge), we constructed donor \times cell-type pseudo-bulk features, performed grouped hyper-parameter tuning on training donors only, and aggregated test clusters to the patient level. For XGBoost, we trained on per-cell matrices with donor-level LOPO and averaged test-cell predictions to the patient level. We report patient-level R^2 , RMSE, MAE, and Spearman ρ .

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-026-03954-x>.

Additional file 1: Supplementary figures S1–S8.

Additional file 2: Supplementary tables S1–S4.

Acknowledgements

Not applicable.

Peer review information

Claudia Feng and Andrew Cosgrove were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions

B.H. and M.C. conceived the project and supervised the research. S.L. and B.J. contributed equally to this work. S.L. performed the experiments and technical validation and developed the data processing pipelines. B.J. performed downstream analysis and biological data interpretation. C.L. and S.P. collected healthy blood samples and provided immunological insights. D.R.K., A.S., Y.K., and S.K. contributed patient blood samples and clinical interpretations for CHAI and SLE patients, respectively. All the authors reviewed and approved the final manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00276271, RS-2022-M3A9B6082674, RS-2025-17992968, and RS-2025-02214844), the Samsung Science and Technology Foundation (SSTF-BA2301-01), and a faculty research grant of Yonsei University College of Medicine (6-2022-0181).

Data availability

Both the raw and processed sequencing data are available in the Gene Expression Omnibus (GEO) repository, GSE282881 [53]. These datasets are publicly available and can be accessed without restrictions. The source code is available at <https://github.com/Yonsei-Hwang-Lab/SCITO-seq> under the MIT License [54], and the version used in this manuscript is archived at Zenodo (DOI: <https://doi.org/10.5281/zenodo.17446995>) [55]. Publicly available datasets reanalyzed in this study include the SCITO-seq dataset (GSE147808) [56], a CITE-seq reference dataset (GSE164378) [57], an AbSeq reference dataset deposited on Figshare (<https://doi.org/10.6084/m9.figshare.13398065>) [32], and an SLE dataset (GSE135779) [51].

Declarations**Ethics approval and consent to participate**

This study was approved by the Institutional Review Board of Samsung Medical Center (SMC 2021-04-096) and the Institutional Review Board of Seoul National University Hospital (2302-082-1404 and 2306-111-1439) and was conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent to participate in this study.

Consent for publication

All participants provided written informed consent for publication of the study findings.

Competing interests

The authors declare no competing interests.

Received: 12 February 2025 Accepted: 12 January 2026

Published online: 27 January 2026

References

- 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;35:936–9.
- Gerlach JP, van Buggenum JAG, Tanis SEJ, Hogeweg M, Heuts BMH, Muraro MJ, et al. Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells. *Scientific Rep.* 2019;9:1–10.
- Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat Methods.* 2019;16:409–12.
- Chung H, Parkhurst CN, Magee EM, Phillips D, Habibi E, Chen F, et al. Joint single-cell measurements of nuclear proteins and RNA in vivo. *Nat Methods.* 2021;18:1204–12.
- 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;14:865–8.
- Rotem A, Ram O, Shores N, Sperling RA, Goren A, Weitz DA, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol.* 2015;33:1165–72.
- Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 2016;26:304–19.
- 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;361:1380–5.
- Clark SJ, Argelaguet R, Kapourani CA, Stubbs TM, Lee HJ, Alda-Catalinas C, et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun.* 2018;9:1–9.
- Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol.* 2019;37:1452–7.
- 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.
- 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;39:1246–58.
- Pan L, Ku WL, Tang Q, Cao Y, Zhao K. scPCOR-seq enables co-profiling of chromatin occupancy and RNAs in single cells. *Commun Biol.* 2022;5:1–9.
- Zhang X, Song B, Carlino MJ, Li G, Ferchen K, Chen M, et al. An immunophenotype-coupled transcriptomic atlas of human hematopoietic progenitors. *Nat Immunol.* 2024;25:703–15.
- Stephenson E, Reynolds G, Botting RA, Calero-Nieto FJ, Morgan MD, Tuong ZK, et al. Single-cell multi-omics analysis of the immune response in COVID-19. *Nat Med.* 2021;27:904–16.
- Ren X, Wen W, Fan X, Hou W, Su B, Cai P, et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell.* 2021;184:1895-1913.e19.
- Reed AD, Pensa S, Steif A, Stenning J, Kunz DJ, Porter LJ, et al. A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. *Nat Genet.* 2024;56:652–62.

18. Kumar V, Ramnarayanan K, Sundar R, Padmanabhan N, Srivastava S, Koima M, et al. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. *Cancer Discov.* 2022;12:670–91.
19. Anderson AG, Kulkarni A, Konopka G. A single-cell trajectory atlas of striatal development. *Sci Rep.* 2023;13:1–11.
20. Cao S, Feng H, Yi H, Pan M, Lin L, Zhang YS, et al. Single-cell RNA sequencing reveals the developmental program underlying proximal–distal patterning of the human lung at the embryonic stage. *Cell Res.* 2023;33:421–33.
21. Replogle JM, Saunders RA, Pogson AN, Hussmann JA, Lenail A, Guna A, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale perturb-seq. *Cell.* 2022;185:2559–2575.e28.
22. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science.* 2017;357:661–7.
23. Martin BK, Qiu C, Nichols E, Phung M, Green-Gladden R, Srivatsan S, et al. Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nat Protocols.* 2022;18:188–207.
24. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature.* 2019;566:496–502.
25. Lareau CA, Duarte FM, Chew JG, Kartha VK, Burkett ZD, Kohlway AS, et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol.* 2019;37:916–24.
26. Datlinger P, Rendeiro AF, Boenke T, Senekowitsch M, Krausgruber T, Barreca D, et al. Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. *Nat Methods.* 2021;18:635–42.
27. Hwang B, Lee DS, Tamaki W, Sun Y, Ogorodnikov A, Hartoularos GC, et al. SCITO-seq: single-cell combinatorial indexed cytometry sequencing. *Nat Methods.* 2021;18:903–11.
28. Lobato-Moreno S, Yildiz U, Claringbould A, Servaas NH, Vlachou EP, Arnold C, et al. Scalable ultra-high-throughput single-cell chromatin and RNA sequencing reveals gene regulatory dynamics linking macrophage polarization to autoimmune disease. *bioRxiv.* 2024;2023.12.26.573253.
29. Zhu C, Yu M, Huang H, Juric I, Abnousi A, Hu R, et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat Struct Mol Biol.* 2019;26:1063–70.
30. Li Y, Huang Z, Zhang Z, Wang Q, Li F, Wang S, et al. FIPRESCI: droplet microfluidics based combinatorial indexing for massive-scale 5'-end single-cell RNA sequencing. *Genome Biol.* 2023;24:1–32.
31. 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;19:1–12.
32. Velten L, Triana S, Haas S, Vonficht D, Jopp-Saile L, Paulsen M. Expression of 197 surface markers and 462 mRNAs in 15281 cells from blood and bone marrow from a young healthy donor. 2021. Datasets Figshare. <https://doi.org/10.6084/m9.figshare.13398065>.
33. 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.
34. Triana S, Vonficht D, Jopp-Saile L, Raffel S, Lutz R, Leonce D, et al. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat Immunol.* 2021;22:1577–89.
35. Nehar-Belaid D, Hong S, Marches R, Chen G, Bolisetty M, Baisch J, et al. Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat Immunol.* 2020;21(9):1094.
36. Wang B, Zhang B, Wu M, Xu T. Unlocking therapeutic potential: targeting lymphocyte activation Gene-3 (LAG-3) with fibrinogen-like protein 1 (FGL1) in systemic lupus erythematosus. *J Transl Autoimmun.* 2024;9:100249.
37. Peng Y, Guo F, Liao S, Liao H, Xiao H, Yang L, et al. Altered frequency of peripheral B-cell subsets and their correlation with disease activity in patients with systemic lupus erythematosus: a comprehensive analysis. *J Cell Mol Med.* 2020;24:12044.
38. Schepis D, Gunnarsson I, Eloranta ML, Lampa J, Jacobson SH, Kärre K, et al. Increased proportion of CD56bright natural killer cells in active and inactive systemic lupus erythematosus. *Immunology.* 2009;126:140–6.
39. Chen K, Li X, Shang Y, Chen D, Qu S, Shu J, et al. FGL1-LAG3 axis impairs IL-10-producing regulatory T cells associated with Systemic lupus erythematosus disease activity. *Heliyon.* 2023;9:e20806.
40. Mulè MP, Martins AJ, Tsang JS. Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nat Commun.* 2022;13:1–12.
41. Gayoso A, Shor J. JonathanShor/DoubletDetection: doubletdetection v4.2. Zenodo. 2022. <https://doi.org/10.5281/zenodo.6349517>.
42. Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* 2019;8:281–291.e9.
43. 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;16:1289–96.
44. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep.* 2019;9:5233.
45. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods.* 2018;15:1053–8.
46. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol Syst Biol.* 2021;17:9620.
47. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multi-modal, and massively scalable single-cell analysis. *Nat Biotechnol.* 2023;42:293.
48. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 2019;20:296.
49. Gayoso A, Steier Z, Lopez R, Regier J, Nazor KL, Streets A, et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat Methods.* 2021;18:272.
50. Tang M, Kaymaz Y, Logeman BL, Eichhorn S, Liang ZS, Dulac C, et al. Evaluating single-cell cluster stability using the Jaccard similarity index. *Bioinformatics.* 2020;37:2212.

51. Nehar-Belaid D, Flynn WF, Banchereau J, Pascual V, Robson P. A single cell approach to map cellular subsets involved in Systemic Lupus Erythematosus (SLE) heterogeneity. Datasets. Gene Expression Omnibus. 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135779>.
52. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
53. Lee S, Jin B, Choi M, Hwang B. Ultra high-throughput single-cell transcriptome and epitope profiling of healthy human PBMCs using SCITO-seq2. Datasets. Gene Expression Omnibus. 2026. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE282881>.
54. Lee S, Hwang B. Yonsei-Hwang-Lab/SCITO-seq2. GitHub. 2025. <https://github.com/Yonsei-Hwang-Lab/SCITO-seq2>.
55. Lee S, Hwang B. Yonsei-Hwang-Lab/SCITO-seq2: v1.0.1_archived. Zenodo. 2025. <https://doi.org/10.5281/zenodo.17446995>.
56. Hwang B. SCITO-seq: single-cell combinatorial indexed cytometry sequencing. Datasets. Gene Expression Omnibus. 2021. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147808>.
57. Hao Y. Integrated analysis of multimodal single-cell data. Datasets. Gene Expression Omnibus. 2021. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164378>.

Publisher's Note

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