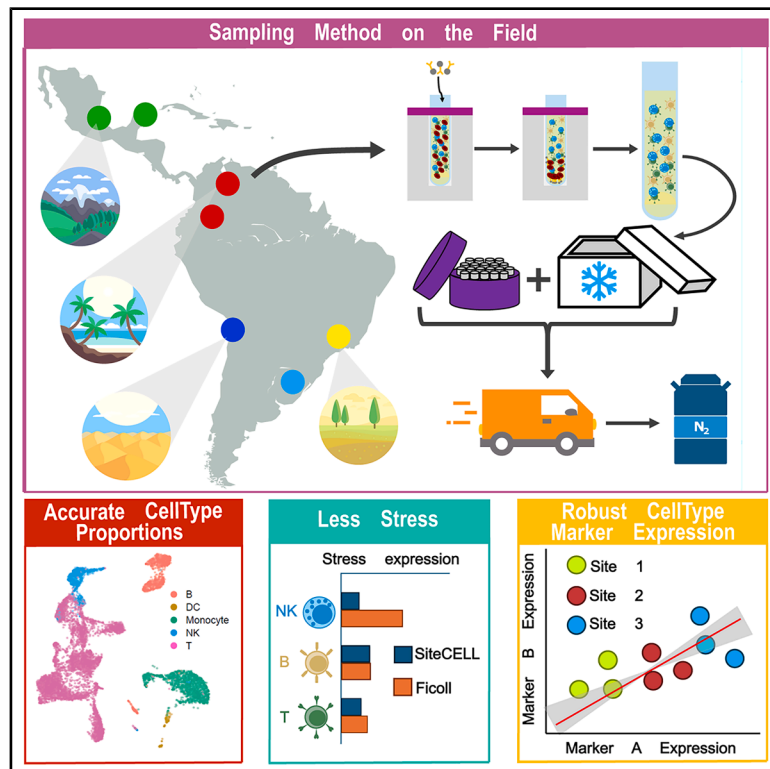


SiteCELL enables *on-site* PBMCs purification and cryopreservation for immune single-cell profiling of underrepresented ancestries

Graphical abstract



Authors

Aarón Espinosa-Jaime,
Octavio Zambada-Moreno,
Jose Antonio Corona-Gómez, ...,
Patricia Severino,
Andrés Moreno-Estrada,
Carlos Ortiz-Ramírez

Correspondence

carlos.ortiz@cinvestav.mx

In brief

Immunological methods; Genomics

Highlights

- SiteCELL enables inclusion of underrepresented ancestries in immune atlases
- Method delivers high purity and accurate cell type proportions in field conditions
- SiteCELL samples showed reduced levels of stress-related gene activation
- Expression of cell type marker genes is highly robust across sampling sites



Article

SiteCELL enables *on-site* PBMCs purification and cryopreservation for immune single-cell profiling of underrepresented ancestries

Aarón Espinosa-Jaime,¹ Octavio Zambada-Moreno,^{1,2} Jose Antonio Corona-Gómez,² María de Jesús-Ortega,¹ Marcela Hernandez-Coronado,³ Barbara Dos Santos Dias,⁴ Larissa Alvez Diniz,⁴ Laura Leaden,⁴ Guilherme Martelossi Cebinelli,⁴ Domenica Marchese,⁵ Adolfo Rojas-Hidalgo,⁵ Emiliano Vicencio,⁵ Diego Pérez-Stuardo,⁵ Sebastián Urquiza-Zurich,⁵ Marisol Espitia Fajardo,^{2,14} Alicia Colombo,^{6,7} Gerardo Donoso,^{6,7} Tamara Méndez,⁸ Carla Gallo,⁹ Hugo Guerrero-Cázares,¹⁰ Carla Daniela Robles-Espinoza,¹¹ Patricia A. Possik,^{12,13} Guillermo Barreto,¹⁴ Lucía Spangenberg,^{15,16} Ricardo A. Verdugo,⁶ Vinicius Maracaja-Coutinho,^{5,17,18} Patricia Severino,^{4,19} Andrés Moreno-Estrada,² and Carlos Ortiz-Ramírez^{1,20,*}

¹Cell Evolution and Development Laboratory, Unidad de Genómica Avanzada, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Irapuato, Guanajuato, México

²Human Evolutionary and Population Genomics Laboratory, Centro de Investigación sobre el Envejecimiento, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Mexico City, México

³Unidad Irapuato, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Guanajuato, México

⁴Experimental Research Center, Einstein Hospital Israelita, São Paulo, Brazil

⁵Unidad de Genómica Avanzada - UGA, Advanced Center for Chronic Diseases - ACCDiS, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile

⁶Departamento de Oncología Básico Clínica, Facultad de Medicina, Universidad de Chile, Santiago, Chile

⁷Center for Cancer Prevention and Control (CECAN), Santiago, Chile

⁸Multidisciplinary Agroindustry Research Laboratory, Instituto de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Cinco Poniente #1670 Talca and Región del Maule, Talca 3467987, Chile

⁹Laboratorios de Investigación y Desarrollo, Facultad de Ciencias e ingeniería, Universidad Peruana Cayetano Heredia, Lima 15102, Peru

¹⁰Mayo Clinic, Jacksonville, FL, USA

¹¹Laboratorio Internacional de Investigación Sobre el Genoma Humano, Universidad Nacional Autónoma de México, Santiago de Querétaro, México

¹²Division of Basic and Experimental Research, Brazilian National Cancer Institute, Rio de Janeiro, Brazil

¹³Wellcome Sanger Institute, Hinxton, UK

¹⁴Human Molecular Genetics Laboratory, Departamento de Biología, Universidad del Valle, Cali, Colombia

¹⁵Bioinformatics Unit, Institut Pasteur de Montevideo, Montevideo, Uruguay

¹⁶Departamento Básico de Medicina, Hospital de Clínicas, Facultad de Medicina, Montevideo, Uruguay

¹⁷Laboratório de Medicina e Saúde Pública de Precisão - MESP2, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz - Fiocruz, Salvador, Brazil

¹⁸Instituto Nacional de Ciência e Tecnologia em Saúde Digital (INCT-DigiSaúde), Salvador, Brazil

¹⁹Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan

²⁰Lead contact

*Correspondence: carlos.ortiz@cinvestav.mx

<https://doi.org/10.1016/j.isci.2026.115961>

SUMMARY

Population-level immune single-cell profiling has the potential to enable the association of cellular function with human ancestry, demographic, and environmental variables. However, methods for collecting and processing peripheral blood mononuclear cells (PBMCs) for single-cell sequencing are difficult to implement in remote and rural settings, resulting in a lack of representation of underserved communities across the global south. We developed SiteCELL, a method that enables purification and cryopreservation of PBMCs from whole blood at the site of collection using minimal laboratory equipment and without electricity. By comparing matched samples of purified PBMCs, we showed that SiteCELL performs as well as ficoll density gradient (FDG), both in laboratory and rural settings. This method ensures accurate recovery of cell type proportions and excels in reducing stress and minimizing variability in sampling batches across countries. Hence, SiteCELL is ideal for multinational initiatives targeting the inclusion of underrepresented ancestries in cellular atlases.



INTRODUCTION

The advent of single-cell technologies has enabled a deeper characterization of the immune system by disentangling its highly heterogeneous cellular and molecular responses in the context of health and disease. This higher resolution in immune profiling has already proven useful for the identification of new cellular subtypes and developmental checkpoints, as well as more precise biomarkers for disease states.^{1–4}

Although single-cell studies were initially done by sourcing peripheral blood mononuclear cells (PBMCs) from relatively small cohorts, several initiatives now exist that aim to build comprehensive cellular atlases including information from hundreds of individuals and millions of cells. These initiatives have opened up the possibility of correlating cell types and functions to biological conditions, as well as assessing the effect of demographics and ancestry on gene expression.^{5,6} However, populations living in low to middle income or remote areas are largely absent from those projects due to limited access to specialized equipment and the trained personnel necessary to isolate and preserve PBMCs *on site*.⁷ This is also true for many clinical settings in which samples cannot be processed promptly to avoid changes in gene expression.

In most studies, PBMCs are isolated from whole blood and then cryopreserved for future downstream analyses, including single-cell profiling. However, the standard ficoll density gradient (FDG) method for PBMC isolation consists of separating cells into distinct layers and recovering the PBMCs containing fraction.⁸ This can be time-consuming and requires specialized equipment not available in remote settings. Shipping fresh blood to a laboratory is not always an option because some sampling locations can be days away from the nearest city, and changes in gene expression have been detected in PBMCs after only 4 h post blood draw.^{9–11} In addition, variability across sampling sites is common, especially regarding cell composition and the introduction of unwanted cells like platelets and erythrocytes.

Alternatives that avoid isolating PBMCs at the point of collection by either sample fixation or whole blood cryopreservation exist.^{12–14} However, fixation is not compatible with most single-cell sequencing technologies, and isolating PBMCs from blood that has undergone a cycle of freezing and thawing is challenging and can introduce biases. For example, it has been reported that cell type proportions change after PBMCs isolation from thawed whole blood. In particular, fewer myeloid cells are recovered because they tend to attach to stressed or dying granulocytes, which are particularly sensitive to freezing and are targeted for removal during purification.¹² A way to circumvent this issue was recently reported, in which the use of either FACS or enrichment of live PBMCs after initial purification using dead cell removal beads together with PE conjugated antibodies can render good quality cells for single-cell downstream processing.¹³ However, this means extra processing steps and time, increased costs, and the need for FACS equipment, which is not available or affordable for many research institutions. In addition, isolation of PBMCs post-thawing means that they cannot be frozen again and stored in biobanks, a significant drawback for large-scale and cross-purpose studies.

We report an optimized method for *on-site* preservation of PBMCs, called SiteCELL, which allows simple and consistent PBMC purification and cryopreservation in remote or challenging settings. Using immunomagnetic negative selection and rate-controlled freezing in dry ice, PBMCs can be purified and cryopreserved at the site of collection in less than 1 h for a single batch, without the need of specialized or electrically powered equipment. Purified PBMCs are cryopreserved *on-site* and transported in dry ice, which avoids transcriptional perturbations. The use of standardized checklists, the deliberate avoidance of electrical equipment, and the harmonization of reagent usage are key improvements that enhance the protocol's robustness. Collectively, these optimizations facilitate the recovery of high-quality cells suitable for downstream single-cell analysis. Furthermore, this method supports biobanking of PBMCs, which is particularly advantageous when clinical assays are conducted in parallel.¹⁵

Here, we first tested SiteCELL under laboratory conditions and showed it performs equally well compared to the FDG gold standard in terms of cell viability, cell type relative proportion capture, and reduced gene expression changes. We then implemented SiteCELL in remote and rural settings to profile indigenous communities across Latin America, with the goal of increasing representation of Native Latin American ancestries in the generation of an immune single-cell atlas by the LatinCells consortium (www.latincells.org). We show that SiteCELL maintains remarkable performance in challenging purification conditions. It excels in reduced activation of cell stress genes, cell purity (absence of platelets and erythrocytes), and minimal technical variability across batches, making it an ideal method for multi-national initiatives and for increasing representation of underserved communities. Furthermore, a detailed “bench-ready” (Data S1) user manual is provided including standardized checklists, logistics, and technical steps to ensure successful implementation of SiteCELL in the field.

RESULTS

Analyses of matched samples demonstrate similar performance between FDG and SiteCELL

To evaluate our method's performance, we first collected blood samples from healthy individuals at Hospital Israelita Albert Einstein (Sao Paulo, Brazil) and processed aliquots in parallel using both SiteCELL and FDG. In brief, samples were split into two and used to purify and cryopreserve PBMCs, followed by thawing and single-cell profiling using the 10× Genomics Next GEM single-cell 3'v3.1 HT kit. In the case of SiteCELL, negative selection of red blood cells, platelets and granulocytes was done with immunomagnetic beads, followed by addition of cryopreservation media (see detailed protocol) and controlled freezing (1 °C/min) in dry ice using a cool cell container (Figure 1). For the rest of the blood sample, purification was done by gradient centrifugation, followed by the addition of cryopreservation media and controlled freezing in a –80°C ultrafreezer. Post thaw, processing steps were the same for both methods and included multiplexing cells from four individuals in a single Next GEM chip well. Genotype-based demultiplexing was applied to sequenced data, and analyses were performed to estimate data quality.

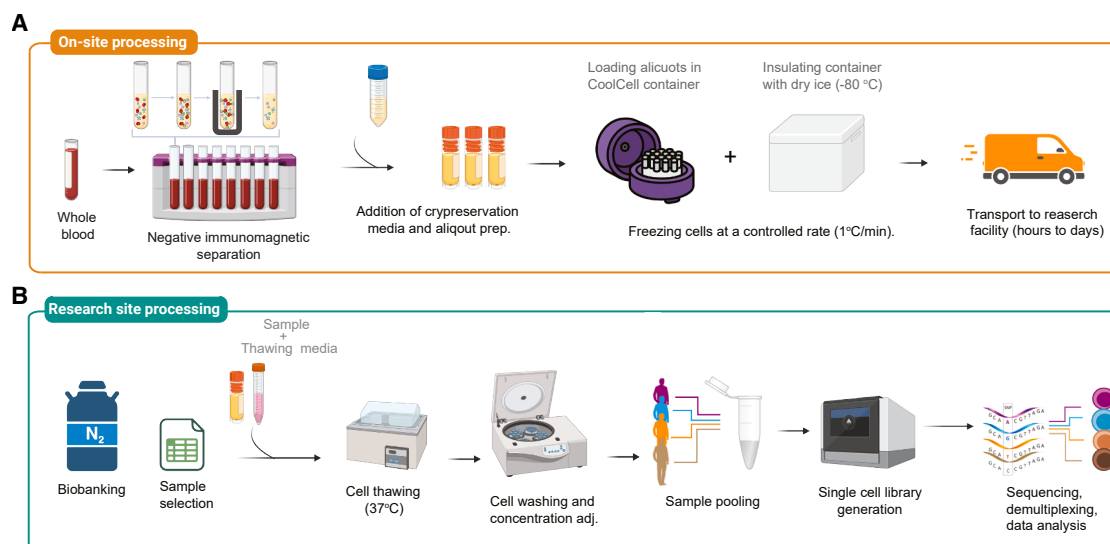


Figure 1. SiteCELL protocol workflow

(A) Diagrams illustrating the individual steps performed at the site of collection to purify, cryopreserve, and transport frozen PBMCs to the nearest research facility. (B) Workflow showing individual steps performed at the laboratory to thaw, wash, and prepare single-cell libraries. With SiteCELL, PBMCs are transported frozen at -80°C and readily stored in biobanks at the nearest research facility.

Cell viability upon thawing was approximately 90% for both methods, with slightly higher values observed for samples processed using SiteCELL. ($93\% \pm 2.6$ SiteCELL and $86\% \pm 3.1$ FDG) (Figure 2A). The difference between methods was statistically significant ($p = 0.006$, t test). In total, we sequenced 13,271 cells processed with gradient centrifugation and 10,620 processed with SiteCELL, each containing cells from the same four individuals. Both protocols produced high-quality data. Comparable numbers of unique molecular identifiers (UMIs) were detected with both SiteCELL (avg. $7,056 \pm 4,598$) and FDG (UMIs avg. $6,426 \pm 3,957$), as well as comparable numbers of genes per cell (SiteCELL, avg. $2,484 \pm 946$; FDG, avg. $2,366 \pm 868$) (Figures 2B and 2C). We estimated the percentage of mitochondrial RNA, as it is known to be high ($>15\%$) in lysed cells or those undergoing apoptosis. Cells from both protocols showed a low average proportion of mitochondrial transcripts ($\approx 4\%$) (Figure 2D). However, FDG samples had a slightly higher number of cells with a mitochondrial transcript content of above 15% (FDG 139 cells, SiteCELL 26 cells).

Biases in recovered cell type proportions have been reported for FDG alternative methods. Thus, to compare relative cell type proportions between matched samples, we clustered cells according to their transcriptome similarity using the uniform manifold approximation and projection (UMAP) analysis from the Seurat package. Cells from different individuals were integrated using Seurat's CCA algorithm and annotated using reference-based mapping by Azimuth with l2 resolution. All cell types were identified in our PBMC samples with a similar distribution (Figure S1). We calculated the mean relative proportion of major cell types (T cells, monocytes, natural killer cells, dendritic cells, and B cells) and found that they were in the expected range of what has been previously reported by single-cell studies¹⁶ (Figures 2E–2G). Corroboration of this result was done by

employing flow cytometry analysis to quantify antibody-labeled cell types (Figure S2). We quantitatively determined cell type proportion similarity between protocols by calculating Pearson correlation across parallel-processed samples. Correlation was very high ($r = 0.997$, $p < 0.001$), showing that proportions were almost identical and no cell capture bias was introduced by SiteCELL (Figure 2I). Notably, we clearly detected the presence of platelets and red blood cells in samples processed by FDG but not in SiteCELL samples, indicating the latter clearly performs better in terms of purity (Figure 2H).

It has been shown that moderate differences in sample handling and processing time can induce transcriptional changes even if cell viability seems unaffected.⁹ Hence, we analyzed the activity of the top marker genes for each cell subtype (a total of 219 genes) and performed correlation analyses between matched samples to determine if processing differences attributed to SiteCELL could result in expression biases. Correlation between protocols was very high ($r = 0.99$, $p = 0.0001$), showing that SiteCELL reliably conserves expression signatures of cell identity genes (Figure 2J). Furthermore, we estimated induction of cell stress genes due to processing by quantifying the expression levels of eight genes that are known to be upregulated in PBMCs in response to stress conditions and processing delays.¹⁷ Most genes showed slightly higher expression values across cell types in samples processed by FDG compared to SiteCELL, except for *NEAT1* (Figure 2K). At the cell type level, natural killer cells displayed more differences between protocols, presenting more stress when subjected to FDG. We also noted that regardless of the protocol, monocytes showed the strongest evidence of stress, and thus, seem to be the most susceptible cell type to handling (Figure 2K). Overall, when quantifying normalized expression of all stress-related genes across all cell types, SiteCELL processed cells showed

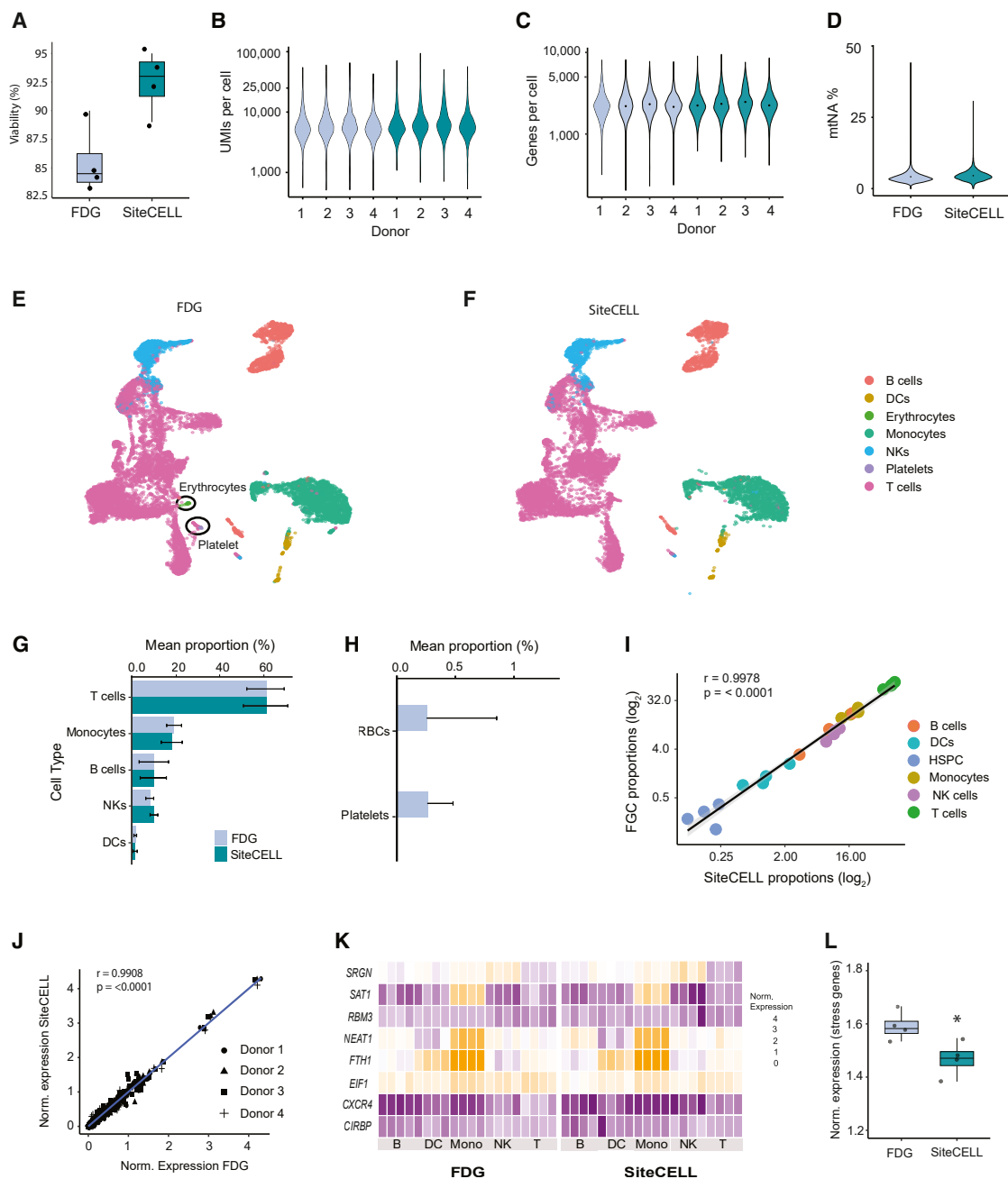


Figure 2. Comparison of matched PBMC samples processed with SiteCELL and FDG

(A) Cell viability for samples processed by FDG (light gray) and SiteCELL (green) was quantified using trypan blue-stained dead cells.
 (B) Number of UMIs per cell for each donor calculated by the Seurat package for FDG (light gray) and SiteCELL (green) samples.
 (C) Number of genes per cell for each donor calculated by the Seurat package for FDG (light gray) and SiteCELL (green) samples.
 (D) Percentage of mitochondrial transcripts from the total number of transcripts per cell across donors for FDG (light gray) and SiteCELL (gray) samples.
 (E) UMAP showing cells colored by major immune cell types for FDG samples. The presence of platelets and erythrocytes is highlighted by dark circles.
 (F) UMAP showing cells colored by major immune cell types for SiteCELL samples.
 (G) Comparison of relative cell proportions of major immune cell types for FDG (light gray) and SiteCELL (green) samples.
 (H) Comparison of the proportion of contaminating cell types, platelets, and erythrocytes per protocol.
 (I) Correlation of cell type proportions between matched samples processed by FDG and SiteCELL. Each dot from each individual color represents a donor.
 (J) Correlation of gene expression levels for cell identity genes between matched FDG and SiteCELL samples. Each dot represents a gene per donor. The blue line represents a linear regression fit.

(legend continued on next page)

significantly lower values ($p = 0.003$, t test) (Figure 2L). Since processing times were similar for both protocols under laboratory conditions, this result suggests that the extra centrifugation steps, combined with the exposure to ficoll necessary for PBMCs isolation with FDG, might induce some stress.

SiteCELL delivers accuracy in cell type proportions and optimal quality when implemented in remote settings

We showed that SiteCELL performs as well as FDG in the laboratory setting using matched samples. However, since the main goal of SiteCELL is to enable PBMC collection and purification in remote and rural locations, we next evaluated the performance and reproducibility of the protocol in the field. We analyzed three independent batches containing samples collected from six indigenous communities across Latin America, in which PBMCs were purified and cryopreserved *on-site* (Figure 1A). Libraries were prepared by two independent processing hubs in México (two batches) and Chile (one batch). As a reference, we analyzed in parallel five publicly available PBMC datasets generated by different research teams using FDG and the same single-cell library preparation technology (10 $\times$ genomics).^{18–22} We assessed sample quality focusing on three key criteria: transcript capture, relative cell proportion accuracy, and cell stress gene expression. Ancestry and sampling details are presented in Tables S1 and S2.

SiteCELL relies on dry ice and coolers to achieve cryopreservation in the field. Hence, we tested sample's freezing rates under variable ambient temperatures (simulating field conditions) by placing an ultra-low wireless temperature sensor inside a fully loaded CoolCell container. No significant differences were found between cryopreservation done under warm (36°C) or cool (23°C) ambient conditions. Importantly, a rate no faster than $-1^{\circ}\text{C}/\text{min}$ was maintained during the entire cryopreservation processes, and a low freezing fusion point was observed, showing that SiteCELL ensured optimal cryopreservation under variable field conditions (Figure S3).

Regarding cell quality, most SiteCELL samples had above 90% post-thaw cell viability, except for a single Colombian sample (Table S1). A total of 63,992 cells were sequenced in batches of 2–4 individuals. Analyses showed samples had a high average of genes ($2,664 \pm 340$) and unique UMIs per cell ($7,796 \pm 1,487$). In the case of FDG datasets, viability was not reported for most batches. Each dataset had cells from three to five individuals, and a total of 220,604 cells were analyzed. A significantly lower average of genes ($1,388 \pm 260$) and unique UMIs ($4,424 \pm 1,143$) per cell were detected compared with SiteCELL (Figures 3B and 3C). This was not due to differences in sequencing depth, because FDG samples had a higher number of reads per cell (Table S3). However, we note that although both SiteCELL and FDG libraries were prepared using 10 $\times$ genomics single-cell RNA sequencing technology, some of the latter were processed using older versions of that kit, which

might have reduced transcript capture efficiency. Detailed information for both datasets can be found in Table S3.

The percentage of mitochondrial transcripts per cell was evaluated for all datasets. Most had an average below 10%, which is considered acceptable, except for the FDG batch 2. Overall, there was a higher variability across the FDG datasets, with mitochondrial transcripts averages ranging from 2.9% to 16% (Figure 3D). Values for SiteCELL datasets ranged from 5.0% to 5.43%. We also calculated the number of cells with more than 15% of mitochondrial RNAs, which are considered to be severely stressed or dying, and observed a difference between protocols. For SiteCELL, we found that 1.7% of all cells were above this threshold (1,105 total cells), while for FDG datasets, 5.1% of all cells had more than 15% of mitochondrial transcripts (11,219 total cells). There was substantial variability across FDG batches, with one and two contributing most cells with a high proportion of mitochondrial transcripts (>15%).

To evaluate whether our protocol introduced any bias in cell type capture efficiency when applied to challenging settings (fieldwork conditions), we identified cell type populations in all datasets as described before. We clustered cells and integrated different batches using UMAP and CCA algorithms from Seurat, followed by annotation with Azimuth (Figures 3E and 3F). Subsequently, cell type relative proportions were calculated for each major group across all datasets and protocols. Relative proportions for both SiteCELL and FDG samples were similar and matched previously reported values,¹⁶ with T cells comprising 65%–70%, B cells 5%–22%, NK 5%–14%, and monocytes 4%–12% of the total number of cells (Figure 3G). We also confirmed that SiteCELL samples had almost no platelet (5 platelet total counts) and no red blood cell contamination compared with FDG samples (Figure 3G).

We further assessed if SiteCELL was able to accurately capture and report rare or low-abundant immune cells. We focused on T cells because they group some of the rarest PBMC subtypes, like CD8 TCM and CD4 CTL. For performing comparative analyses, T cells from both protocols were re-clustered to annotate subtypes, and their proportions were calculated relative to the total number of T cells (major group) (Figures 4A and 4B). SiteCELL samples had the same number of T cell subtypes and in equal proportions (no statistical differences) compared to FDG samples, including the ones with a lower abundance: proliferating T, CD4 CTL, Treg, MAIT, CD8 TCM, gdT, and dnT (Figure 4C). The subtype that showed the biggest difference between protocols, in terms of total cell number, was CD8 TCM. However, most of the cells reported for FDG were contained in only one batch; hence, this was not statistically significant. Overall, no bias was found in rare subtype recovery when SiteCELL was applied.

Regarding induction of cell stress genes, we quantified the normalized expression of established stress-responsive genes as done previously. In this case, PBMC purification and cryopreservation were done in rural settings where sampling can

(K) Heatmap showing normalized expression values for major immune cell types of eight stress-responsive genes. FDG samples are grouped at the left and SiteCELL samples at the right.

(L) Average values of the same eight stress-responsive genes per donor for FDG (light gray) and SiteCELL (gray) samples. Bars represent the mean and SD ($n = 4$ donors per method). Boxplots show the median, first and third quartile. Asterisks indicate statistical significant difference (t test, $p = 0.003$).

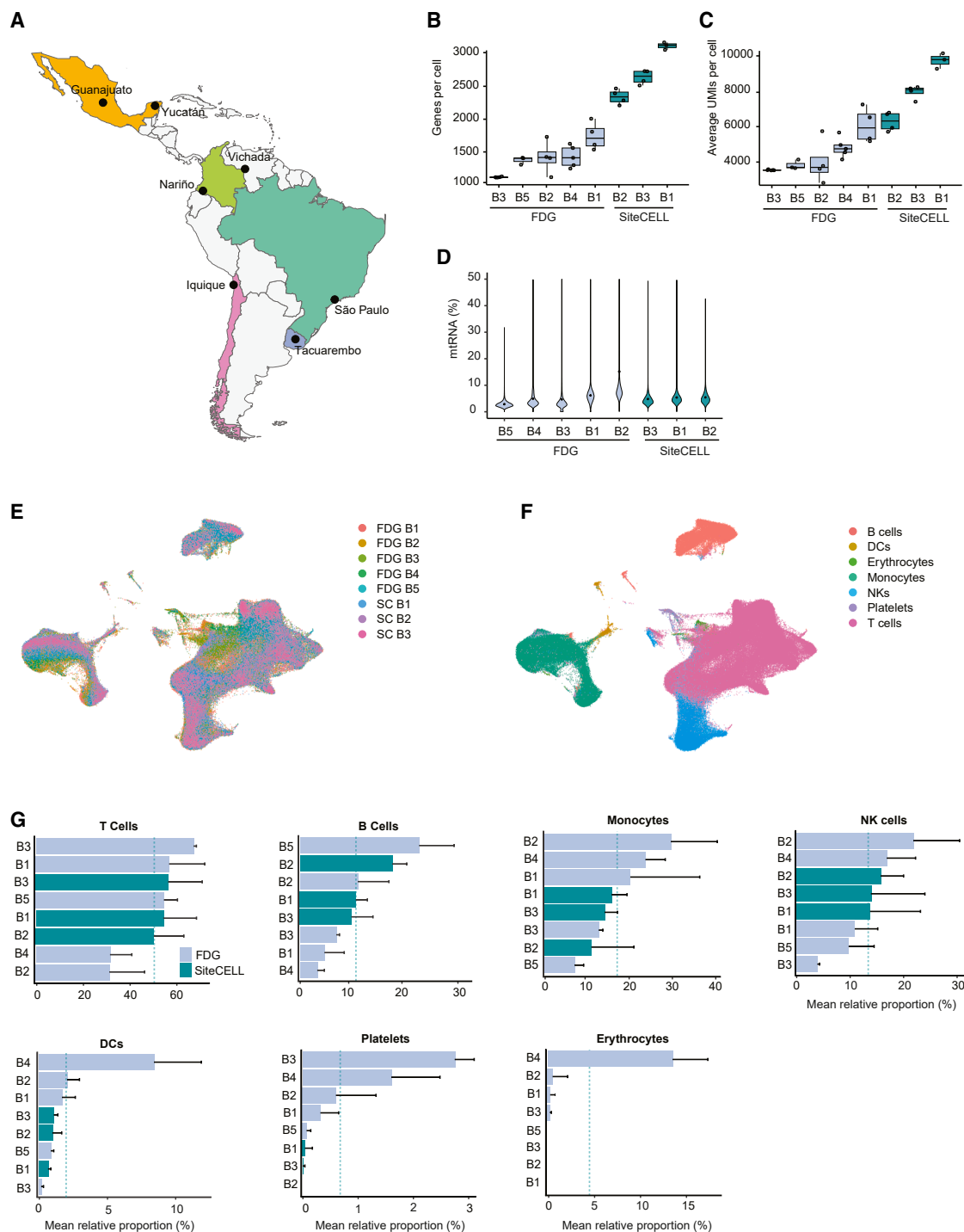


Figure 3. SiteCELL performance for PBMCs isolation in remote and rural settings

(A) Map showing sampling sites across LATAM.

(B) Average values of genes per cell for individual samples and across batches calculated with the Seurat package for FDG (light gray) and SiteCELL (green).

(C) Average values of UMIs per cell for individual samples and across batches calculated with the Seurat package for FDG (light gray) and SiteCELL (green).

(D) Percentage of mitochondrial transcripts from the total number of transcripts per cell across batches for FDG (light gray) and SiteCELL (green) samples.

(E) UMAP showing cells colored by sample processing batches. FDG, ficoll density gradient; SC, SiteCELL.

(F) UMAP showing cells colored by major immune cell types for all integrated batches with the CCA algorithm from the Seurat package.

(legend continued on next page)

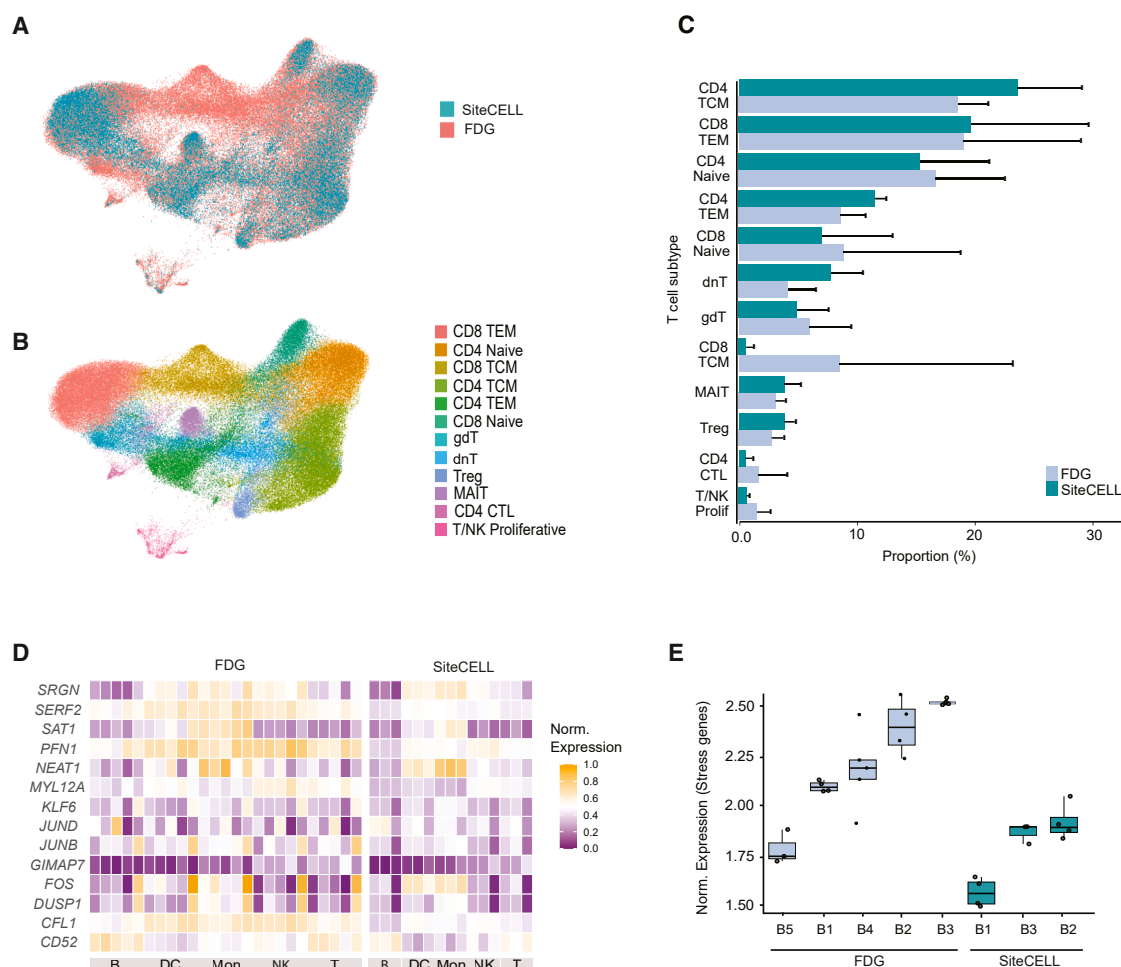


Figure 4. Recovery of low-abundant immune cell types and induction of stress genes

(A) UMAP showing re-clustering of T cells of both FDG (light gray) and SiteCELL (green) samples. Cells are colored by protocol. (B) UMAP shows re-clustered T cells from both protocols, in which cells are colored by identity subtypes. (C) Comparison of relative T cell subtype proportions for FDG (light gray) and SiteCELL (green) samples across batches. (D) Heatmap showing normalized expression values for major immune cell types of 12 stress-responsive genes. Samples processed by FDG are grouped at the left and SiteCELL samples at the right. (E) Average values of the same 12 stress-responsive genes across batches processed by FDG (light gray) and SiteCELL (gray). Bars represent the mean and SD ($n = 21$ donors FDG; $n = 11$ donors SC). Boxplots show the median, first and third quartile.

be unpredictable, hence we also included genes from the JUN and FOS families and from the DUSP and GIMAP families. These are known to be upregulated when there are considerable delays before processing.²³ SiteCELL samples showed low levels of stress-responsive gene expression compared to the FDG references (Figures 4D and 4E). Again, a cell type-dependent response was observed, with monocytes being the most affected regardless of the protocol, and NK cells showing more differences between protocols: very low or no expression in SiteCELL samples, to moderate expression in FDG samples. Overall, these findings highlight the suitability of SiteCELL to

process samples in remote locations without inducing significant cell stress gene expression.

SiteCELL improves consistency and robustness across sampling sites

Large-scale single-cell studies with hundreds of samples collected across geographically diverse settings are becoming more common due to their broad applications. These projects are usually developed by consortia of multinational teams that collect and process samples across countries at different time points. Thus, ensuring data reproducibility is a major challenge.

(G) Comparison of relative cell proportions of major immune cell types for FDG (light gray) and SiteCELL (green) samples across batches. Bars represent the mean and SD for FDG samples ($n = 4$ donors B1; $n = 4$ donors B2; $n = 5$ donors B3; $n = 5$ donors B4; $n = 3$ donors B5) and SC samples ($n = 3$ donors B1; $n = 4$ donors B2, $n = 4$ donors B3). Boxplots show the median, first, and third quartile.

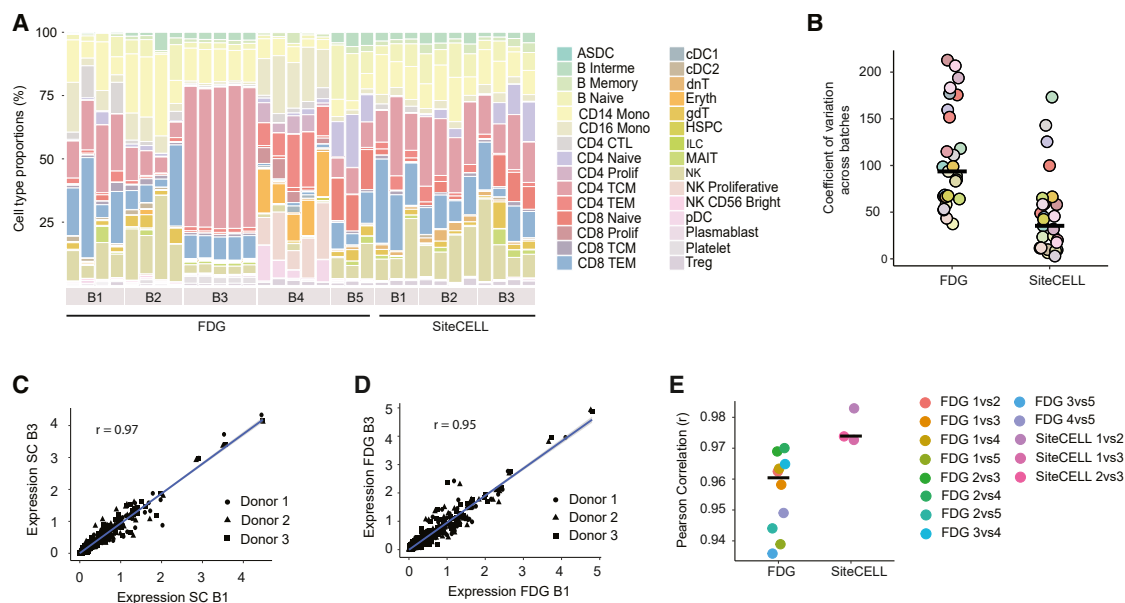


Figure 5. Cell subtype proportion and gene expression variability across sampling batches

(A) Stacked bar plot showing relative cell proportions colored by immune subtypes across batches. Each stacked bar corresponds to a single individual. (B) Coefficient of variation representing the relative dispersion of data (cell type proportion) for individual cell subtypes across batches. Higher values indicate higher variability across processing batches. Cell subtypes are colored as in (A). (C) Correlation of gene expression levels for cell identity genes between representative batches processed by SiteCELL (left) and FDG (right). Each dot represents a gene per donor. The blue lines represent linear regression fits. (E) Pearson correlation values for all pairwise comparisons of available batches for each protocol

In this context, a method that is highly robust and reproducible is necessary to reduce batch effects across sites.

We evaluated variability across sample batches collected and processed with SiteCELL at different locations and time points, focusing on changes in cell type proportions and gene expression, and compared them to the FDG reference. For these analyses, we calculated the relative proportions of 27 cell subtypes and evaluated the consistency in which these were recovered. Variability was observed at the level of individuals, which was expected since it has been reported that age, sex, and overall health can produce moderate differences in immune cell type proportions^{24,25} (Figure 5A). However, when assessing variability across batches, differences were apparent between protocols. To quantify this, we calculated the coefficient of variation from the average proportion of each cell subtype across batches. We observed that although the degree of variability is cell-type specific for both protocols, those cells processed with SiteCELL have an overall lower relative variability ($\approx 40\%$) compared to those purified with FDG ($\approx 95\%$), with one exception: B intermediate cells (Figure 5B). This suggests that implementation of SiteCELL reduces batch effects due to processing. Notably, some cell types had a relatively high variability regardless of the protocol (CD8 naive, CD4 naive, and CD4 CTL) (Figure 5B). Therefore, caution is advised when making conclusions about proportion differences of these subtypes due to ancestry or biological conditions. Altogether, these results support that SiteCELL preserves cellular composition even in challenging processing settings.

Variability in the expression levels of 219 cell identity genes (top markers for each cell subtype) across batches was

assessed. We did pairwise linear regression analyses and calculated Pearson correlation between all batches in each protocol. In general, gene expression correlation was high for both methods, suggesting that expression levels of cell identity genes are maintained even if batches are processed by different teams and at different time points. Although SiteCELL batches were collected and processed in remote and challenging settings, they had slightly higher correlation values (Figures 5C and 5D) compared to FDG samples. SiteCELL r values across dataset comparisons ranged from 0.972 to 0.985, while FDG r values across datasets ranged from 0.936 to 0.97 ($p = 0.001$). Average correlation was $r = 0.973$ for SiteCELL samples and $r = 0.961$ for FDG samples (Figure 5E).

In addition, we determined operator-to-operator reproducibility by selecting one sample collected in the field and splitting it into two independent processing and sequencing runs. Split replicates showed minimal variation and high correlation coefficients in all metrics analyzed (Figure S4). Hence, we conclude that SiteCELL is a robust and easy-to-implement method for PBMCs purification that reduces technical sample variability across sites. This makes it ideal for large-scale multinational single-cell genomics projects and enables the inclusion of under-represented human populations that live in remote and rural areas with no infrastructure.

DISCUSSION

Initiatives to build comprehensive immune cellular atlases are currently underway. However, most methods to purify and

cryopreserve PBMCs are difficult to implement in remote locations that lack basic infrastructure, resulting in a lack of representation of underserved communities and ancestries. So far, the alternative to FDG for sampling in challenging settings has been the cryopreservation of whole blood samples immediately after blood draw without purification.^{12,13} However, there are some drawbacks to freezing whole blood first and purifying PBMCs later. Mainly biases in cell type proportions, incompatibility with biobank storage, increased isolation complexity and costs, and induction of cell stress genes. The latter is especially relevant in the case of granulocytes, which become much harder to remove post-thawing due to loss of viability. SiteCELL overcomes many of these issues, enabling the purification and cryopreservation of PBMCs at the point of collection without electricity and minimal equipment. We employ a pre-freeze PBMC purification strategy that avoids cell stress gene expression and complicated granulocyte removal procedures after thawing. Cryopreserved PBMCs are transported frozen to the laboratory settings, where they can be immediately used for single-cell profiling or directly stored in biobanks.

We analyzed matched samples processed to generate single-cell data and demonstrated that SiteCELL performs as well as the gold standard FDG method in terms of cell quality and viability. Additionally, SiteCELL faithfully maintains PBMC cell type relative proportions as shown by the very high correlation scores of matched samples. This has been a drawback for some whole blood cryopreservation methods that report biases in relative cell type proportions due mainly to the fact that monocytes tend to attach to platelets or stressed granulocytes, which are targeted for removal post-thaw.¹² Granulocytes tend to present low viability in whole blood freezing protocols because they are very sensitive to freezing and thawing cycles; hence, their removal at the point of collection before freezing is a considerable advantage of the SiteCELL protocol.^{23,26} We also showed that SiteCELL is better at removing platelets and red blood cells, which is important because platelets can act as suppressors, inhibiting proliferative immune cells and affecting gene expression.^{27,28} In addition, their unwanted presence implies a waste of valuable resources, since they can end up taking expensive single-cell reagents and, in some cases, a considerable amount of sequencing reads.

In terms of gene expression, both methods show a very high correlation in the expression level of 219 cell subtype marker genes, showing that SiteCELL faithfully conserves expression profiles of genes related to cell identity. An important finding was that SiteCELL samples showed reduced levels of stress-responsive gene activation. This was especially evident for NK cells. Reduction was more significant when comparing multiple independent batches than in matched experiments. This could be due to a faster processing time with SiteCELL compared to other protocols. For example, cells can be immediately processed at the site of blood draw without the need to transfer them to any facility, and a batch can be purified and cryopreserved in less than an hour. This considerably reduces the time cells are exposed to *ex vivo* conditions, especially when sampling is conducted far from the laboratory. Another explanation could be that SiteCELL does not rely on centrifugation steps or ficoll plaque reagent, which can induce cell stress and toxicity when the gradient is not formed optimally. Absence of such

processing steps might decrease the chances of cells being exposed to harsh conditions.

As large-scale single-cell projects profiling hundreds or thousands of individuals and samples across contrasting settings become more common, simple but robust purification and cryopreservation methods are highly needed. We demonstrate that data generated from independent batches processed by SiteCELL at different rural locations was highly consistent. Two aspects were analyzed: cell type relative proportions and gene expression profiles of marker genes. In both cases, less variability was detected across batches with SiteCELL compared to FDG. However, we note that PMBC isolation with SiteCELL was highly harmonized across sampling sites, while FDG datasets used for comparison were generated by independent research groups which might have used slightly different versions of the FDG protocol. Hence, SiteCELL does not necessarily overperform FDG, rather our results show that it is comparable to the FDG gold standard even when deployed in rural and remote settings that do not allow PBMCs isolation by conventional methods. We also note that SiteCELL is easy to implement and requires minimal training, although this comes at a cost. It can be more expensive than available FDG protocols, since immunomagnetic purification kits have to be acquired, and some whole blood freezing methods like CryoSCAPE can achieve higher throughput per batch when combined with automatic liquid handlers. A detailed comparison of available protocols is provided in Table 1.

Overall, SiteCELL excels in reproducibility, portability, purity, and low activation of stress genes, and it is well-suited for deployment in remote and rural settings. We believe this method will facilitate the inclusion of underrepresented populations in biomedical research by enabling global, multi-ethnic, and single-cell transcriptomic studies.

Limitations of the study

We acknowledge that although our data show a significant reduction in the expression of stress related genes for SiteCELL processed samples, we do not know if this translates into functional differences or long-term viability. Comparative functional experiments (e.g., cell stimulation assays) will be needed to assess if reduced expression of stress genes actually impacts cellular responses. In addition, we show low variability across SiteCELL samples processed at different countries and timepoints. However, we cannot state that this variation is significantly lower compared to FDG processed samples as is suggested by results, because the latter were not collected using the same harmonized protocol. Hence, it is possible this accounts for the higher FDG variability. We can only state that variation with SiteCELL is indeed low and similar to that observed with FDG.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Carlos Ortiz-Ramirez (carlos.ortiz@cinvestav.mx).

Table 1. Comparison of available PBMCs isolation protocols

Method	Basis	Total processing time	Cost per sample ^a	Specialized equipment required at collection site	Laboratory specialized equipment	On-site PBMCs isolation	Replicability across processing sites	Yield	Expression of stress genes	Enable PBMCs biobanking
SiteCELL	Onsite Immuno-magnetic precipitation + freezing	60–80 min per batch (8–50 samples depending on magnet)	Medium (\$\$)	Magnetic rack, CoolCell, Pipettes	None – on site isolation	Yes – PBMCs purified directly on-site	Very high	1,868 cells/ μ L ^b	Low	YES
Ficoll density gradient (8)	Density gradient centrifugation +freezing	60–80 min per batch (8 samples)	Low (\$)	NA	Centrifuge	No – requires lab	Good, but operator-dependent	800–3,200 cells/ μ L	Low	YES
SENSE (12)	Whole blood freezing on-site + post-thaw Immuno-magnetic precipitation	65 min per sample or batch ($n = 8$), depending on magnet	Medium (\$\$)	Mr. Frosty, Pipettes, –80 ultra freezer	Centrifuge, magnetic rack	No – requires lab	Not reported	Not reported	Not reported	NO
CryoSCAPE (13)	Whole blood freezing on-site + post-thaw PBMC isolation	≈80–90 min ^c	Low to high (\$-\$\$\$), depending on throughput	CoolCell, Pipettes, –80 ultra freezer	Centrifuge, FACS, magnetic rack	No – requires lab	Not reported	Not reported	Low	NO

^a\$ ~ 5 USD; \$\$ ~ 20 USD; \$\$\$ ~ 40.USD.

^bAverage value from all included pools.

^cNot reported, calculated an approximated time from available information in [STAR Methods](#) section.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Single-cell RNA-seq data produced by SiteCELL protocol is available at the European Nucleotide Archive (ENA), ArrayExpress: E-MTAB-15719.
- Data produced by FDG is publicly available from GEO with the following accession numbers: FDG 1: GEO: GSE226896, FDG 2: GEO: GSE149689, FDG 3: GEO: GSE161918, FDG 4: GEO: GSE213516, FDG 5: GEO: GSE168732.
- All original code has been deposited at Github (<https://github.com/latincells/SiteCELL>) and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this study is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We thank all participants in this study as well as the community leaders for their support across all sampling sites. We acknowledge the Chan Zuckerberg Initiative grant 2021-240108 (5022) and to the Secretaria de Ciencia, Humanidades, Tecnología e Innovación (CECIHTI) for postgraduate grants. P.S. was supported in part by the MEXT Cooperative Research Project Program, Medical Research Center Initiative for High Depth Omics, and CUR-E:JPMXP1323015486 for MIB, Kyushu University. We thank Alanis Huañaco for her contribution to sample collection.

AUTHOR CONTRIBUTIONS

A.E.-J., O.Z.-M., and C.O.-R. collected and processed PBMC samples, analyzed the data, and helped write the manuscript. J.A.C.-G., M.d.J.-O., M.H.-C., B.D.S.D., L.A.D., L.L., G.M.C., D.M., A.R.-H., M.E.F., E.V., D.P.-S., A.C., G.D., T.M., A.R.-H., and S.U.-Z. collected and processed PBMC samples. C.G., H.G.-C., C.D.R.-E., P.A.P., G.B., L.S., R.A.V., V.M.-C., P.S., and A.M.-E. were involved in community engagement, sampling logistics, and manuscript edits. A.M.-E. and C.O.-R. conceived experiments and helped write the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Participant selection
 - Ethics approval and informed consent
 - Data collection
 - Sample collection
- **METHOD DETAILS**
 - Immunomagnetic PBMCs isolation
 - FDG PBMCs isolation
 - Cryopreservation
 - Measurement of freezing rates
 - DNA extraction
 - Microarray genotyping
 - Cell thawing and viability
 - Immunophenotyping with flow cytometer
 - Cellular multiplexing and library construction
 - VCF generation
 - Mapping and quality control
 - Genotype-based demultiplexing
 - scRNA-seq data analysis

- Collection of FDG-generated scRNAseq data
- Cell quality, proportion and quantification of stress gene expression
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2026.115961>.

Received: January 26, 2026

Revised: March 8, 2026

Accepted: April 28, 2026

REFERENCES

- Proserpio, V., and Mahata, B. (2016). Single-cell technologies to study the immune system. *Immunology* 147, 133–140. <https://doi.org/10.1111/imm.12553>.
- Gaublomme, J.T., Yosef, N., Lee, Y., Gertner, R.S., Yang, L.V., Wu, C., Pandolfi, P.P., Mak, T., Satija, R., Shalek, A.K., et al. (2015). Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. *Cell* 163, 1400–1412. <https://doi.org/10.1016/j.cell.2015.11.009>.
- Nestorowa, S., Hamey, F.K., Sala, B.P., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* 128, e20–e31. <https://doi.org/10.1182/blood-2016-05-716480>.
- Vegh, P., and Haniiffa, M. (2018). The impact of single-cell RNA sequencing on understanding the functional organization of the immune system. *Brief. Funct. Genomics* 17, 265–272. <https://doi.org/10.1093/bfpg/ely003>.
- Kock, K.H., Tan, L.M., Han, K.Y., Ando, Y., Jevapatarakul, D., Chatterjee, A., Lin, Q.X.X., Buyamin, E.V., Sonthalia, R., Rajagopalan, D., et al. (2025). Asian diversity in human immune cells. *Cell* 188, 2288–2306.e24. <https://doi.org/10.1016/j.cell.2025.02.017>.
- Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M.G., Andersen, S., Lu, Q., Rowson, A., Taylor, T.R.P., Clarke, L., et al. (2022). Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* 376, eabf3041. <https://doi.org/10.1126/science.abf3041>.
- Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The Missing Diversity in Human Genetic Studies. *Cell* 177, 26–31. <https://doi.org/10.1016/j.cell.2019.02.048>.
- Boyum, A. (1968). Separation of leukocytes from blood and bone marrow. *Introduction. Scand. J. Clin. Lab. Invest. Suppl.* 97, 7.
- Barnes, M.G., Grom, A.A., Griffin, T.A., Colbert, R.A., and Thompson, S.D. (2010). Gene Expression Profiles from Peripheral Blood Mononuclear Cells Are Sensitive to Short Processing Delays. *Biopreserv. Biobank.* 8, 153–162. <https://doi.org/10.1089/bio.2010.0009>.
- Massoni-Badosa, R., Iacono, G., Moutinho, C., Kulis, M., Palau, N., Marchese, D., Rodríguez-Ubreva, J., Ballestar, E., Rodríguez-Esteban, G., Marsal, S., et al. (2020). Sampling time-dependent artifacts in single-cell genomics studies. *Genome Biol.* 21, 112. <https://doi.org/10.1186/s13059-020-02032-0>.
- Olson, W.C., Smolkin, M.E., Farris, E.M., Fink, R.J., Czarkowski, A.R., Fink, J.H., Chianese-Bullock, K.A., and Slingluff, C.L., Jr. (2011). Shipping blood to a central laboratory in multicenter clinical trials: effect of ambient temperature on specimen temperature, and effects of temperature on mononuclear cell yield, viability and immunologic function. *J. Transl. Med.* 9, 26. <https://doi.org/10.1186/1479-5876-9-26>.
- Satpathy, S., Thomas, B.E., Pilcher, W.J., Bakhtiari, M., Ponder, L.A., Pacholczyk, R., Prahalad, S., Bhasin, S.S., Munn, D.H., and Bhasin, M.K. (2023). The Simple prEservationN of Single cells method for cryopreservation enables the generation of single-cell immune profiles from whole

- blood. *Front. Immunol.* 14, 1271800. <https://doi.org/10.3389/fimmu.2023.1271800>.
13. Heubeck, A.T., Phalen, C., Kaul, N., Wittig, P.J., Garber, J., Weiss, M., Genge, P.C., Thomson, Z., Gustafson, C., Reading, J., et al. (2025). Cryo-SCAPE: Scalable immune profiling using cryopreserved whole blood for multi-omic single cell and functional assays. *J. Transl. Med.* 23, 5. <https://doi.org/10.1186/s12967-024-06010-z>.
14. Serra, V., Orrù, V., Lai, S., Lobina, M., Steri, M., Cucca, F., and Fiorillo, E. (2022). Comparison of Whole Blood Cryopreservation Methods for Extensive Flow Cytometry Immunophenotyping. *Cells* 11, 1527. <https://doi.org/10.3390/cells11091527>.
15. Lafzi, A., Moutinho, C., Picelli, S., and Heyn, H. (2018). Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat. Protoc.* 13, 2742–2757. <https://doi.org/10.1038/s41596-018-0073-y>.
16. Chen, H., Schürch, C.M., Noble, K., Kim, K., Krutzik, P.O., O'Donnell, E., Tuig, J.V., Nolan, G.P., and McIlwain, D.R. (2020). Functional comparison of PBMCs isolated by Cell Preparation Tubes (CPT) vs. Lymphoprep Tubes. *BMC Immunol.* 21, 15. <https://doi.org/10.1186/s12865-020-00345-0>.
17. Savage, A.K., Gutschow, M.V., Chiang, T., Henderson, K., Green, R., Chaudhari, M., Swanson, E., Heubeck, A.T., Kondza, N., Burley, K.C., et al. (2021). Multimodal analysis for human ex vivo studies shows extensive molecular changes from delays in blood processing. *iScience* 24, 102404. <https://doi.org/10.1016/j.isci.2021.102404>.
18. Wang, Z., Xie, L., Ding, G., Song, S., Chen, L., Li, G., Xia, M., Han, D., Zheng, Y., Liu, J., et al. (2021). Single-cell RNA sequencing of peripheral blood mononuclear cells from acute Kawasaki disease patients. *Nat. Commun.* 12, 5444. <https://doi.org/10.1038/s41467-021-25771-5>.
19. Zhu, H., Chen, J., Liu, K., Gao, L., Wu, H., Ma, L., Zhou, J., Liu, Z., and Han, J.-D.J. (2023). Human PBMC scRNA-seq-based aging clocks reveal ribosome to inflammation balance as a single-cell aging hallmark and super longevity. *Sci. Adv.* 9, eabq7599. <https://doi.org/10.1126/sciadv.abq7599>.
20. Liu, C., Martins, A.J., Lau, W.W., Rachmaninoff, N., Chen, J., Imberti, L., Mostaghimi, D., Fink, D.L., Burbelo, P.D., Dobbs, K., et al. (2021). Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell* 184, 1836–1857.e22. <https://doi.org/10.1016/j.cell.2021.02.018>.
21. Sati, S., Huang, J., Kersh, A.E., Jones, P., Ahart, O., Murphy, C., Prouty, S.M., Hedberg, M.L., Jain, V., Gregory, S.G., et al. (2024). Recruitment of CXCR4+ type 1 innate lymphoid cells distinguishes sarcoidosis from other skin granulomatous diseases. *J. Clin. Investig.* 134, e178711. <https://doi.org/10.1172/JCI178711>.
22. Lee, J.S., Park, S., Jeong, H.W., Ahn, J.Y., Choi, S.J., Lee, H., Choi, B., Nam, S.K., Sa, M., Kwon, J.-S., et al. (2020). Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci. Immunol.* 5, eabd1554. <https://doi.org/10.1126/sciimmunol.abd1554>.
23. Knight, S.C., O'Brien, J.A., and Farrant, J. (1980). Injury to human granulocytes at low temperatures. *Cryobiology* 17, 273–281. [https://doi.org/10.1016/0011-2240\(80\)90034-6](https://doi.org/10.1016/0011-2240(80)90034-6).
24. Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. *Nat. Rev. Immunol.* 16, 626–638. <https://doi.org/10.1038/nri.2016.90>.
25. Li, W., Zhang, Z., Kumar, S., Botey-Bataller, J., Zoodsma, M., Ehsani, A., Zhan, Q., Alaswad, A., Zhou, L., Grondman, I., et al. (2025). Single-cell immune aging clocks reveal inter-individual heterogeneity during infection and vaccination. *Nat. Aging* 5, 607–621. <https://doi.org/10.1038/s43587-025-00819-z>.
26. Yang, H., Arnaud, F., and McGann, L.E. (1992). Cryoinjury in human granulocytes and cytoplasts. *Cryobiology* 29, 500–510. [https://doi.org/10.1016/0011-2240\(92\)90053-5](https://doi.org/10.1016/0011-2240(92)90053-5).
27. Gerdes, N., Zhu, L., Ersoy, M., Hermansson, A., Hjerdahl, P., Hu, H., Hansson, G.K., and Li, N. (2011). Platelets regulate CD4(+) T-cell differentiation via multiple chemokines in humans. *Thromb. Haemost.* 106, 353–362. <https://doi.org/10.1160/TH11-01-0020>.
28. Wang, Y., and Niu, J. (2008). Platelets inhibit in vitro response of lymphocytes to mitogens. *Immunol. Lett.* 119, 57–61. <https://doi.org/10.1016/j.imlet.2008.04.003>.
29. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
30. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
EasySep™ Direct Human PBMC Isolation kit	STEMCELL Technologies	19654
EasyEights™ EasySep™ Magnet	STEMCELL Technologies	18103
EasySep Buffer	STEMCELL Technologies	20144
DMSO	Sigma-Aldrich	D8418
Fetal Bovine Serum, Heat Inactivated	Sigma-Aldrich	F1435
CoolCell FTS30, Freezing container 30 × 1 mL or 2 mL cryogenic vials	CORNING	432006
Corning® 2 mL Internal Threaded Polypropylene Cryogenic Vial, Self- Standing with Round Bottom	CORNING	431386
BD Vacutainer(R) K2 EDTA 10.8 mg	Beckton Dickinson	367863
BD Vacutainer PrecisionGlide	Beckton Dickinson	360213
ORAclect DNA	DNA Genotek Inc	OCR-100
prepIT-L2P	DNA Genotek Inc	PT-L2P-45
Infinium Assay Kit Post 4 LV	Illumina	15023544
BeadChip Array Infinium Global Screening Array-24 v3.0 Assay: Infinium HTS	Illumina	20031595
Fetal Bovine Serum, Heat Inactivated	Sigma-Aldrich	F4135
RPMI Medium 1640 (1×)	Gibco	11875-093/11875-119
10× PBS Buffer pH7.4	Invitrogen	AM9624
Bovine Serum Albumin	Sigma-Aldrich	A9418
Next GEM Single Cell 3' HT GEM kit v3.1	10× Genomics	1000351
Next GEM Chip M & Gaskets/Next GEM HT Training Chip Kit	10× Genomics	1000349/1000385
Library construction kit	10× Genomics	1000352
Dual Index	10× Genomics	1000215
Chromium X	10× Genomics	1000331
Deposited data		
Single-cell RNA sequencing 3' libraries generated with SiteCELL	ArrayExpress (EBI)	ArrayExpress: E-MTAB-15719
Ficoll Gradient-based PBMC single-cell RNAseq datasets for comparison	Gene Omnibus Expression	GEO: GSE149689, GEO: GSE226896, GEO: GSE161918, GEO: GSE213516, GEO: GSE168732
Software and algorithms		
R (4.3.3) "Angel Food Cake"	The R Foundation	https://www.r-project.org
Seurat (5.1.0)	Stuart et al., 2019	https://satijalab.org/seurat/
Azimuth (0.5.0)	Hao et al. ²⁹	https://azimuth.hubmapconsortium.org/
Demuxlet (from the 'popscl' software suite)	Kang et al., 2018	https://github.com/statgen/popscl
Algorithms used for analyses	This manuscript	https://github.com/latincells/SiteCELL
Freemuxlet (v2, from the 'popscl' software suite)	Kang et al., 2018	https://github.com/statgen/popscl
PLINK (2.00 alpha)	Chang et al., 2015	https://www.cog-genomics.org/plink/2.0/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CellRanger (8.0.1)	10× Genomics	https://www.10xgenomics.com/support/software/cell-ranger/downloads/previous-versions
GenomeStudio (2.0.5)	Illumina	https://www.illumina.com/products/by-type/informatics-products/microarray-software/genomestudio.html
GraphPad Prism software v.8.02.	GraphPad software	https://www.graphpad.com/updates/prism-802-release-notes
Other		
Hg38 human genome reference for CellRanger	10× Genomics	https://cf.10xgenomics.com/supp/cell-exp/refdata-cellranger-hg19-1.2.0.tar.gz
GrCh38_1000Genomes_ExonFiltered_ChrEncoding.vcf	Demuxafy	https://www.dropbox.com/s/1akxxa7s9u485ka/GRCh38_1000G_MAF0.01_ExonFiltered_ChrEncoding.vcf

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participant selection

Participants were recruited in both urban and rural areas based on the following inclusion criteria: they were adults, had not experienced an infectious process within two weeks before participation, and had no prior diagnosis of autoimmune diseases. In some cases, self-reported indigenous ancestry was required. A total of 5 males and 7 females were recruited with an age ranging from 19 to 81 years old. See [Table S1](#) for more details. Written informed consent was obtained before participation in the study under protocol approved.

Ethics approval and informed consent

This study received ethical approval from Einstein Hospital Israelita Research Ethics Committee approval CAEE: 68371323.3.0000.0071, Comité de Ética de Investigación en Seres Humanos (CEISH), Facultad de Medicina, Universidad de Chile: CEISH/091–2023/041, Comité de Bioética para la Investigación en Seres Humanos (COBISH) Cinvestav, México: COBISH/085/2022, Comité de Ética en Investigación en Salud (CEIS), Facultad de Salud, Universidad del Valle. Colombia: CI 085-022, Comité de Ética Institucional Institut Pasteur de Montevideo (CEI-IP) Uruguay: CEI 006-2023).

Data collection

For all participants included in the project, the following data were collected.

1. Informed Consent: A signed document confirming the participant's voluntary and informed participation in the study.
2. Questionnaire: A comprehensive document used to gather general information about the participants. This included: personal details (full name, age, and contact information), family medical history, biometric data (weight, height, and blood type), prevalence of diseases, ancestry information.

These data were securely stored and used solely for this research, following ethical guidelines and ensuring participant confidentiality.

Sample collection

Two types of biological samples were collected from each participant: (1) a blood sample, at least one 6 mL blood sample was drawn using a lavender-top Vacutainer tube containing EDTA; (2) a saliva sample collected using Oracollect-DNA OCR-100 swab (DNA Genotek), designed for non-invasive collection of high-quality DNA from oral cells. The swab and blood tube were both labeled with the participants' identification number.

METHOD DETAILS

Immunomagnetic PBMCs isolation

Peripheral blood mononuclear cells (PBMCs) were isolated on-site from freshly collected blood samples (6 mL per participant in EDTA vacutainer tubes) within a maximum 4 h post blood draw using the EasySep Direct Human PBMC Isolation Kit (STEMCELL Technologies) following the manufacturer's instructions with modifications to adapt it to field conditions in cases of high-altitude

sampling. Two milliliters of whole blood were transferred to a 5 mL (12×75 mm) polystyrene tube and mixed with 100 μ L of Isolation Cocktail by pipetting 10–13 times (without vortexing or inversion to avoid red blood cell contamination). Samples were incubated for 5 min at room temperature (RT). Then, two milliliters of EasySep buffer were added, mixed gently, and combined with 100 μ L of RapidSpheres previously vortexed for 30 s. Tubes (uncapped) were placed in the EasySep magnet for 5 min at RT, and ~ 3 mL of enriched suspension was carefully transferred to a new tube. A second depletion step was performed by adding 100 μ L of RapidSpheres to the enriched fraction, mixing by pipetting, and repeating the magnetic incubation for 5 min, after which the clear supernatant was collected. A third magnetic incubation of the enriched fraction for 5 min yielded ~ 2.5 mL of highly purified PBMC suspension. When sampling on high altitude communities (>2000 m above sea level) an extra depletion and cleaning step was done to ensure complete depletion of red blood cells, which we observe were significantly more abundant on those communities (please refer to the bench-ready manual, [Data S1](#)). This negative selection approach, based on antibody-coated magnetic particles that bind and deplete red blood cells and platelets, avoided centrifugation and electrically powered equipment, yielding untouched and functionally intact PBMCs suitable for immediate cryopreservation and downstream applications.

FDG PBMCs isolation

For the Ficoll method, blood was diluted 1:1 with PBS1 $\times$ and carefully added to 15 mL conical tube containing 5 mL of Ficoll solution. After a 40 min centrifugation at $500\times g$ without break, the buffy coat containing the PBMCs were collected into a new 15 mL tube with 5 mL of PBS 1 $\times$ and centrifuged for 5 min at $500\times g$. The cell pellets were washed two more times as previously described. At the end of isolation protocols, the cells from both methods were frozen in inactivated Bovine Fetal Serum (Gibco, Waltham, Massachusetts, USA) with 10% DMSO (LGC, London, United Kingdom).

Cryopreservation

Following PBMC isolation, cells were cryopreserved in a solution consisting of 10% dimethyl sulfoxide (DMSO) and 90% fetal bovine serum (FBS), prepared by gently adding 0.5 volumes of 30% DMSO in FBS (1.25 mL) to 2.5 mL of purified PBMC suspension, keeping the mixture on ice at all times to minimize DMSO toxicity. The final suspension contained $0.5\text{--}10 \times 10^6$ cells/mL, and three pre-cooled cryovials (CORNING, 2 mL each) were prepared per donor. Cryovials were labeled and arranged in advance. Additional “calibration” vials containing 1 mL of cryomedium were included to ensure all wells of the CoolCell freezing container (BioCision) were filled, as the device is calibrated for use when full. The CoolCell container was pre-equilibrated on ice and immediately transferred into an insulated box filled with dry ice, allowing controlled-rate cooling ($\sim 1^\circ\text{C}/\text{min}$) without electricity. Samples remained fully submerged in dry ice during transport, with constant replenishment to prevent exposure above -70°C , and precautions were taken to avoid CO_2 accumulation in enclosed vehicles during handling of >5 kg dry ice. Once equilibrated at -80°C for ≥ 4 h, cryovials were transported either fully embedded in dry ice, transferred to smaller insulated boxes for shipment, or, when possible, moved into liquid nitrogen tanks for extended storage, where they were kept in vapor phase below -135°C until downstream use.

Measurement of freezing rates

An ultralow temperature sensor (ST10, Mocreco) connected wireless to a Mocreco hub was placed inside a CoolCELL container loaded with 30 cryotubes tubes (2 mL) filled with 1 mL cryopreservation media. Tubes were pre-cooled at 4°C by placing them in a icebox filled with ice to mimic processing procedures in the field. Once loaded with the sensor and tubes, the CoolCELL container was placed in a styrofoam cooler and completely covered with dry ice. Cooler and dry ice management were done to mimic those on the field. The cooler was sealed and maintained at the same ambient temperature until the end of each experiment. For obtaining freezing rates under cool ambient conditions the entire procedure was done in the laboratory at room temperature (23°C). Freezing rates under warm ambient conditions were obtained by performing the experiment inside a walk-in chamber at a constant temperature of 36°C .

DNA extraction

Saliva samples were collected using Oragene collection kits (DNA Genotek Inc.), which stabilize and preserve DNA at room temperature for up to six months. DNA was extracted following the manufacturer’s protocol with the prepIT-L2P Kit (DNA Genotek Inc.), optimized for high-yield recovery from saliva. DNA concentration was measured using a Qubit Fluorometer (Thermo Fisher Scientific), and purity was evaluated by NanoDrop spectrophotometry (Thermo Fisher Scientific), with acceptable A260/A280 ratios ranging from 1.7 to 2.0. Integrity was assessed by agarose gel electrophoresis to verify the predominance of high-molecular weight DNA. Samples exhibiting strong degradation or abnormal purity ratios were excluded from downstream applications. Only DNA extractions with concentrations >20 ng/ μ L, total yield >200 ng, and no evidence of contamination (e.g., protein carryover or RNA traces) were advanced to genotyping.

Microarray genotyping

Genotyping was carried out with the Infinium Global Screening Array v3.0 (GSAv3, Illumina), which interrogates $\sim 650,000$ SNPs representative of diverse global populations and enriched for clinically relevant loci. The Infinium workflow comprises DNA amplification, fragmentation, precipitation, hybridization to beadchips, single-base extension, enzymatic staining, and scanning, requiring

~2–3 days to complete. DNA samples passing initial quality thresholds (≥ 200 ng input, A260/A280 1.7–2.0, intact high-molecular-weight DNA) were processed according to the manufacturer's instructions.

Cell thawing and viability

Cells were recovered from liquid nitrogen storage and kept on dry ice during transport to the bench. Cryovials were immediately thawed in a 37 °C water bath for ~2 min, or until only small ice fragments remained, taking care not to submerge vials beyond the screw-cap to prevent contamination. Immediately after thawing, the cell suspension was transferred into a pre-labeled 15 mL Falcon tube containing 8 mL of prewarmed (37 °C) thawing buffer (RPMI supplemented with 10% FBS and 0.04% BSA), and centrifuged at 400 g for 5 min at room temperature. The supernatant was carefully aspirated with a pipette, avoiding disruption of the pellet, which typically appeared as an opaque ring at the bottom of the tube. The pellet was washed with 4 mL of washing buffer (1× PBS +5% FBS +0.04% BSA), centrifuged again under the same conditions, and the supernatant was discarded. Cells were then slowly resuspended in 0.4 mL of washing buffer and transferred into sterile pre-labeled Eppendorf tubes, yielding a suspension close to the required concentration for downstream single cell applications. Cell concentration and viability were assessed by Trypan Blue exclusion (0.4%), using either manual microscopy-based counting or automated counters, with a minimum viability threshold of 70% required and ~90% considered optimal. When necessary, concentrations were adjusted to fall within the optimal range of 800–1,200 cells/ μ L, aiming for 1,200 cells/ μ L and a final recovery of 20,000 PBMCs per sample pool (5,000 cells per individual). It was preferable to achieve a high concentration at the resuspension step and dilute with washing buffer as needed, rather than perform additional centrifugation steps that could compromise viability. Once the target concentration was achieved, cells were passed through a 40 μ m strainer to eliminate aggregates and maintained on ice until single cell encapsulation.

Immunophenotyping with flow cytometer

Immunophenotyping was performed with the following antibodies:

CD14 - PE (clone:63D3), CD3 - APC (Clone: OKT3), CD4 - BV605 (Clone OKT4), CD45 - APC-Cy7 (Clone: HI30), CD66b - Alexa Fluor 700 (Clone: G10F5), CD8a - PerCP-Cy5.5 (Clone: RPA-T8) (BioLegend, San Diego, California, USA), CD19 - PE-CF594 (Clone: HIB19), CD16 - PE-Cy7 (Clone: 3G8) and CD56 - PE-Cy5 (Clone: B159) and Zombie Aqua Fixable Viability Kit (BD Biosciences, Franklin Lakes, New Jersey, USA).

After cell thawing, each marker was mixed in a cocktail making the final dilution of the antibodies 1:100 and viability marker 1:1000 when added one μ L of the cocktail to each sample (5×10^5 cells in 100 μ L of PBS 1×) and incubated for 20 min at room temperature in the dark. In the same way, 10 single-color tubes with 1×10^5 cells in 100 μ L of PBS 1× were made by adding one μ L of each marker to each tube and one with the cells unstained. After incubation time, all cells were washed using PBS 1× and resuspended in 300 μ L of PBS 1×. After acquiring unstained and single-color control samples to calculate the compensation matrix, each sample in the two isolation conditions was acquired using Attune NxT cytometer (ThermoFisher, Waltham, Massachusetts, USA). The data analysis was performed using FlowJo software v.10.6.2 and statistical analysis was performed using the GraphPad Prism software v.8.02.

Cellular multiplexing and library construction

For single cell RNA sequencing, cells from different participants were pooled together based on equal cell ratios. To maximize genetic diversity and minimize processing costs, four individuals were selected for each multiplexed library construction. Cryovials from the selected individuals were thawed, and cells were washed before a viability assessment was performed using a microscope-based cell count. The Next GEM Single Cell 3' HT GEM Kit (10× Genomics) was used to generate single cell libraries. Each participant contributed approximately 5,000 cells per multiplex, targeting a recovery of around 20,000 cells per well on the 10× Genomics Chip M. After generating gel bead-in-emulsion (GEM) droplets (Chromium X), the reverse transcription and cDNA amplification was performed within the droplets, incorporating unique molecular identifiers (UMIs) and cell barcodes. Library preparation followed the standard 10× Genomics protocol, including fragmentation, end repair, A-tailing, adaptor ligation, and PCR amplification. Quality control of the libraries was performed using Qubit Fluorometry for quantification and Agilent 2100 Bioanalyzer or TapeStation 4150 capillary electrophoresis for size distribution analysis. Three processing HUBs were established across Latin America. The samples used in this study were processed in Mexico, Chile, and Brazil. Libraries were sequenced using the NovaSeqX Plus platform and the NovaSeq 6000 platform using S4 flowcells and Illumina-recommended protocols.

VCF generation

DNA sequencing data (from microarrays) was used to perform variant calling using GenomeStudio (v2.0). Variants with CallFreq <95% were removed and copy number variants (CNVs) annotated as [D/I] or [I/D] were also excluded. After recalculating the sample statistics, only donors with a CallRate greater than 95% were retained. We removed all individuals with >5% missing genotype data and all genotypes with >5% missing individuals. We limited the analysis to biallelic SNP and duplicated SNPs were removed. Subsequently, BIM files were generated with³⁰. The resulting VCF contains merged genotypes from all donors participating in LatinCells (~650,000 SNPs).

Mapping and quality control

Raw scRNA-seq reads were processed using Cell Ranger (v8.3.0, 10× Genomics) with default parameters against the GRCh38-2024-A human reference genome (Ensembl), resulting in the generation of gene-by-cell expression matrices. Mapping quality was assessed through standard Cell Ranger summary metrics, including the fraction of reads confidently mapped to the transcriptome, sequencing saturation, mean reads per cell, and mean genes detected per cell. BAM files were generated for each multiplexed pool for subsequent demultiplexing and downstream analyses.

Genotype-based demultiplexing

Demultiplexing of single cell data at the donor level was performed with Demuxlet (v1.0), which leverages genotype information from matched donors to assign each cell to its most likely individual of origin. For this, we filtered our genotype freeze using donor identifiers to generate a pool-specific VCF file for each multiplexed library, ensuring direct compatibility with the scRNA-seq data. All runs were conducted using default parameters as recommended by the developers. To increase confidence and efficiency in genotype-based demultiplexing, we employed Freemuxlet (Popsicle suite; <https://github.com/statgen/popsicle>), which performs demultiplexing without requiring matched donor genotypes. Instead, Freemuxlet infers donor identities *de novo* by clustering cells based on SNP information. For this purpose, we used a reference VCF file derived from the 1000 Genomes Project,³⁰ filtered to include only exonic variants with minor allele frequency (MAF) > 5%. This file can be found on the popsicle page. The expected number of donors per pool (3–4) was specified, and all analyses were run with default parameters. Finally, we integrated results from both Demuxlet and Freemuxlet to generate consensus donor assignments, maximizing robustness in donor identification across multiplexed pools.

scRNA-seq data analysis

We excluded low-quality cells using the following thresholds: >15% of mitochondrial RNA percentage, <500 genes per cell, and 800–20,000 UMIs. Gene expression was normalized and corrected using a 10,000 factor. 2,000 variable genes were calculated and scaled, and PCA matrix was inferred. Clusters were generated using the top 25–50 PC dimensions (according to the *ElbowPlot* function on each dataset) and visualized on a UMAP. We finally merged and integrated datasets from both protocols to eliminate batch effects using the CCA algorithm from Seurat.

Collection of FDG-generated scRNAseq data

Raw scRNA-seq data were obtained from Gene Omnibus Expression (Table 1) and was selected based on the following criteria: (1) FDG was the PBMC isolation method, (2) only healthy donors were considered, (3) each donor must have at least 5,000 raw cells, and (4) libraries were constructed using 10× Genomics RNA library kit.

Cell quality, proportion and quantification of stress gene expression

To compare the performance of both FDG and SiteCELL protocols, we utilized metrics such as (1) cell quality and changes in cell expression profile, (2) cellular proportion variation and cell contamination, and (3) cellular diversity captured. To assess cell quality we evaluated and calculated percentage of cell viability, UMIs and genes per cell, as well as mitochondrial percentage. Cellular proportions were calculated for both subtypes and major group level in all samples after cell type annotation using a PBMC reference in the Azimuth R library (v0.5.0).²⁹ We annotated cells using the second resolution which identifies 30 different cell types, including other blood cells such as red blood cells and platelets and calculate the relative proportion by dividing the number of cells in each subtype by the total number of cells and multiplying by 100. To evaluate changes in gene expression profiles of matched samples, we integrated all datasets included in the comparison and generated a pseudobulk by summing together the gene counts of all the cells grouped by dataset and major cell type using *AggregateExpression* and then comparing the normalized expression values. To evaluate induction of cell stress genes, we profiled a set of genes previously reported to change due to cellular stress (Massoni-Badosa et al.) (*SRGN*, *SAT1*, *RBM3*, *HEAT1*, *FTH1*, *EIF1*, *CXCR4*, *CIRBP*) and average their expression values per donor and per batch. We replicated this analysis for the experiment assessing performance of SiteCELL in rural settings vs. available FDG experiments using same aggregation strategy but including and expanded set of genes reported to vary due to time delay in processing (Savage et al.) (*SRGN*, *SERF2*, *SAT1*, *PFN1*, *NEAT1*, *MYL12A*, *KLF6*, *JUND*, *JUNB*, *GIMAP7*, *FOS*, *DUSP1*, *CFL1* and *CD52*). Lastly, to evaluate rare cell subtype capture efficiency we created a subset of all T cells and then re-clustered and annotated those cells. We re-ran *FindClusters*, *FindNeighbors* and *RunUMAP* on the subsetted object and calculated cell subtype relative proportions as before. Finally, to calculate the expression level correlation of cell identity genes, we averaged the normalized expression value of each gene across major cell types per donor (from pseudobulked counts) and calculated person correlation for the top 10 unique marker genes per cell subtype (219 in total) across donors from matched experiments.

QUANTIFICATION AND STATISTICAL ANALYSIS

Pearson correlation coefficients were used to assess linear relationships between SiteCELL and FDG isolation protocols. These analyses included comparisons of cell type proportions across samples and normalized aggregated expression of Azimuth marker genes in the matched MX0049 sample. *c.or.test()* function was employed.

The coefficient of variation of cell type proportions across batches was calculated to assess variability associated with each isolation protocol. Student's t test was used to assess statistical significant differences in stress related gene expression between FDG and SC.

Further statistical details of experiments can be found in figure legends where "n" represents number of human individuals.