Robust performance of single-cell multiomic analysis (CITE-Seq) demonstrated across multiple locations, replicates, and donors on challenging cryopreserved PBMC and BMMC samples

Q² Solutions® BioLegend®

Josh Croteau[1], Joseph Modarelli[2], Alexandra Corella[1], Priya Halvorsen[2], Anna Sheydina[1], Doug Wilson[2], Santosh Putta[1], Jon Lee[2], Kevin Taylor[1]

BioLegend Inc., San Diego, California[1], Q² Solutions, Durham, North Carolina[2]

Abstract

Emerging single-cell multiomic techniques like CITE-Seq[1] (combined gene expression and cell surface protein profiling by sequencing) require best practices, optimized parameters, and an understanding of practical limitations when comparing data from different labs, especially when analyzing clinically relevant samples. Here, we designed a study across two organizations using matched donor cryopreserved human peripheral blood mononuclear cells (cPBMCs) and bone marrow mononuclear cells (cBMMCs) to examine performance and reproducibility of CITE-Seq on the 10x Genomics Chromium platform. Within cPBMCs, reduced sample quality was achieved by "cold-shocking" one set of matched samples during cell-thaw to simulate cPBMCs of clinical relevance, and compare to normally processed samples. Within cBMMCs, identical cell-thawing and processing procedures were used to investigate reproducibility across study locations with challenging clinical samples.

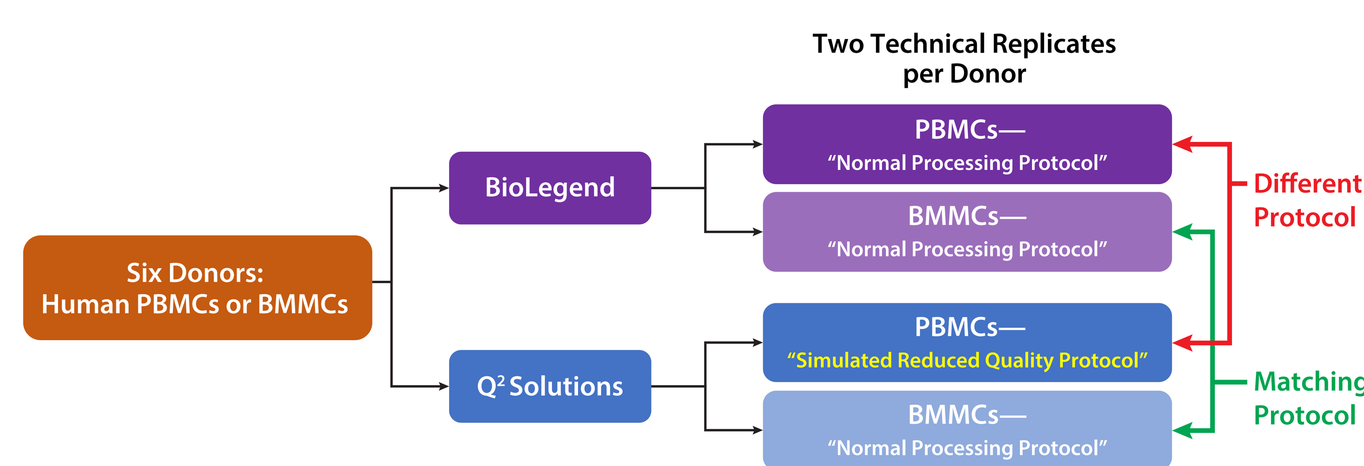
Over 140 surface proteins (ADTs) were examined in addition to >1100 RNA transcripts. 33 canonical PBMC and BMMC cell types were identified by TotalSeq™-C antibody staining profile. We found a high correlation across cell type, ADT, and RNA metrics across all samples, however, low correlation was observed across conditions (shocked vs non-shocked) as expected. Additionally, cell type metrics, ADT, and RNA counts between study sites were highly correlated indicating robustness of the CITE-Seq technology. Our data demonstrates feasibility of phenotypically rich and complex data using CITE-Seq across multiple locations on challenging sample types; specifically, cryopreserved clinically relevant cells. Additionally, meaningful guidance and considerations are provided when samples of variable quality are analyzed with CITE-Seq.

Methods

Samples and Experimental Design

Several vials of cPBMCs and cBMMCs from three different donors were procured from Stem Cell Technologies. Vials of each sample were sent to both study locations (BioLegend and Q² Solutions). Unless otherwise noted, all methods, materials, equipment, and analyses were matched between study locations whenever possible. cBMMCs were prepared using matched sample processing. cPBMCs at Q² Solutions were processed using an SOP (standard operating procedure) modification described below to simulate clinically-relevant, reduced sample quality (cold-shock) in matched samples.

1 Experimental design, sample processing



Sample Processing, Staining and Single-Cell Sequencing

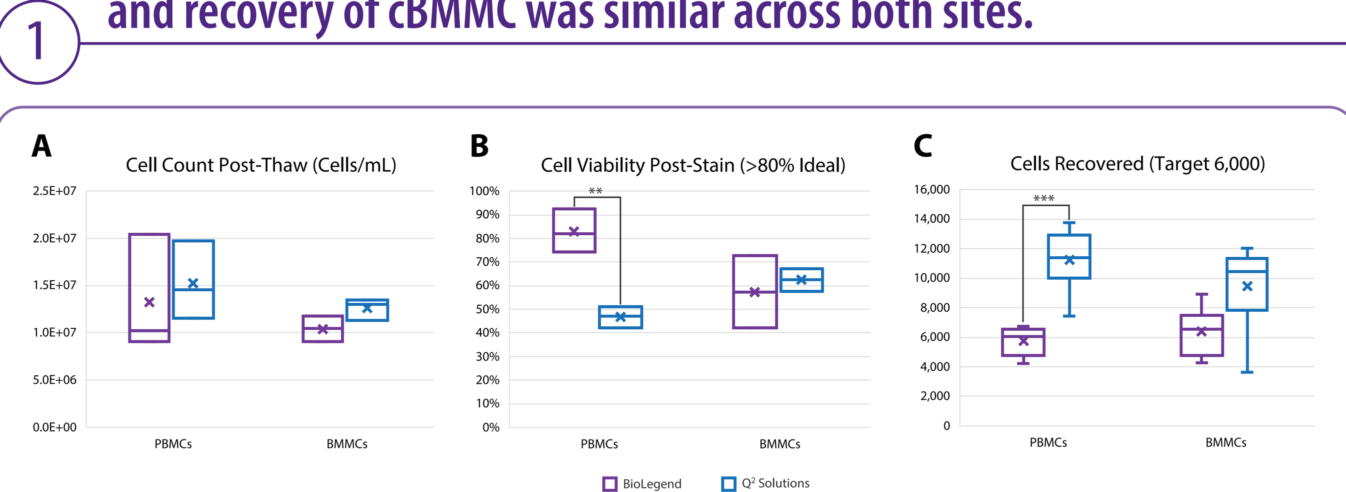
Sample thawing and processing prior to cell staining were performed using matched SOPs from Q² Solutions. Briefly, vials were thawed until ice was nearly melted, then, rapidly diluted in 10 mL of warm complete cell media (CM). For cPBMCs at Q² Solutions to simulate reduced quality donor-matched samples, a "cold-shock" technique was used with ice-cold media. Cells were then washed twice with CM, centrifuged at 4°C, resuspended in CM, counted and viability assessed. For staining with the TotalSeq™-C Universal cocktail (BioLegend, Cat. No. 399905) cells were washed with a Curiox Mini laminar washing system using custom settings. The TotalSeq™-C cocktail was used following the manufacturer's recommendations. Antibody tagged samples were processed following the 10x Genomics 5' Immune Profiling Solution User Manual. Samples were loaded into a 10x Genomics chip with a targeted cell capture of 6,000 cells per replicate. Cell loading at the Q² Solutions site was increased based on measured viability. All libraries were sequenced on Illumina's NovaSeq platform. Targeted sequencing depth was 50,000 reads/cell for gene expression library (GEX), 10,000 reads/cell for TotalSeq antibodies (ADT).

Data Processing, Bioinformatics and Statistical Analysis

Sequencing data was quantified into feature count matrix file sets using 10x Cell Ranger or 10x Cell Ranger Multi and sequencing metrics were extracted for review. Sample file sets were then imported into BioLegend Multiomics Analysis Software (MAS, biolegend.com/en-us/totalseq/mas) for final data analysis. Data was filtered to exclude low UMI counts, high mitochondrial reads, as well as antibody isotype background and biologically irrelevant events. A set of 33 cell types were established for each sample type using a set of ~30 canonical surface proteins(2-4) for identification of mature leukocyte populations and limited "stem/progenitor" like cells for cBMMCs only. Visit our website using the QR code below for representative hierarchal gating used to establish cell types. P-values shown for DE analysis were calculated using Wilcoxon rank-sum test. Pair-wise comparisons illustrated on box-and-whisker figures were made using paired t-test. Correlations shown are calculated using Pearson's correlation (R). Statistically significant differences are represented as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Dimensionality reductions shown are a novel cell type weighted t-distributed Stochastic Neighbor (weighted tSNE) Embedding plot. Briefly, this method uses the conventional tSNE algorithm with additional weighting applied to ADT dimensions used to define cell types established using bi-axial hierarchal gating. This approach allows for higher order to be applied to the tSNE according to cell types defined using known protein phenotype allowing for easier visualization of biologically complex samples without relying on clustering.

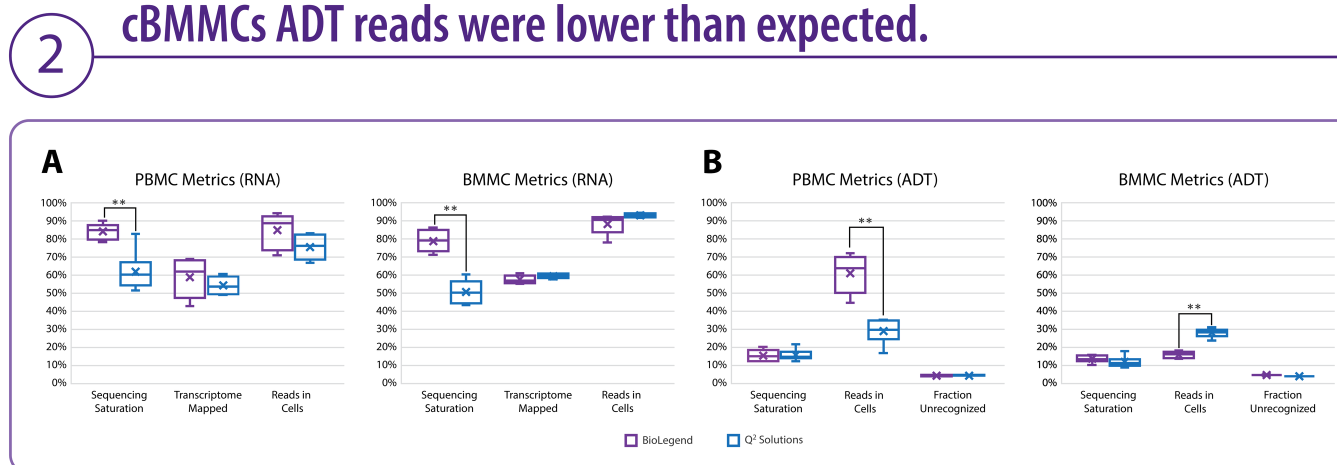
Results

1 cBMMCs show a lower recovery and cold-shocked cPBMCs had significantly lower viability whereas cell count, viability, and recovery of cBMMC was similar across both sites.



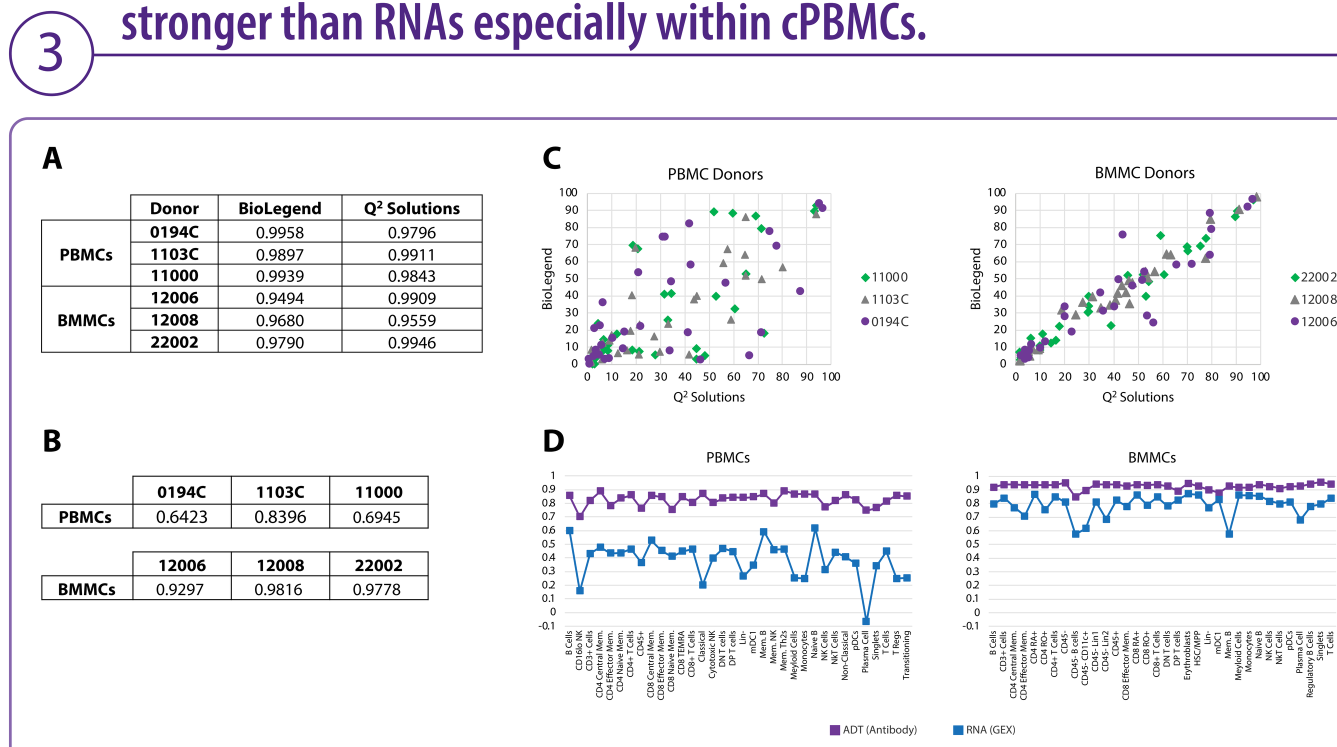
Cells were counted and viability was assessed at several points during sample processing and staining. Select metrics are shown from these data. (A) There were no significant differences between study locations in initial recovered cells. Fewer cells were recovered for cBMMCs, which is expected. (B) Cell viability measured at the end of the cell staining and washing protocol immediately before loading the 10x Chromium was significantly lower in the simulated reduced quality cPBMCs (cold-shocked). (C) Cells recovered per sample from the 10x Chromium were variable. Differences were due to loading formula differences between study sites.

2 RNA metrics were similar across sample types whereas ADT reads in cells were significantly lower in cold-shocked cPBMCs and cBMMCs ADT reads were lower than expected.



During sequencing data processing and quantification, 10x Cell Ranger generates a summary set of statistics on cell barcodes, UMIs, and sequencing metrics. These metrics are helpful to assess overall experimental quality; however, individual metrics may not always indicate usability of data. Selected RNA or ADT metrics are shown. (A) RNA (GEX) library sequencing saturation was notably higher in BioLegend samples. Cell recovery was lower in BioLegend samples; thus, each cell received a higher proportion of total sequencing. Saturation >50% is generally acceptable for GEX libraries. (B) ADT reads in cells were significantly lower in "cold-shocked" cPBMCs samples run at Q² Solutions. BMMC antibody reads in cells was low for both study sites, with the BioLegend samples being significantly lower than Q² Solutions. Understanding the impact of reduced metrics such as reads in cells was an objective of these experiments.

3 Cell type percentages are reproducible within replicates and more variable in cPBMCs than cBMMCs using different vs matched procedures between study sites. Count correlations of ADTs were stronger than RNAs especially within cPBMCs.



A

	Donor	BioLegend	Q² Solutions
PBMCs	0194C	0.9958	0.9796
	1103C	0.9897	0.9911
	11000	0.9939	0.9843
BMMCs	12006	0.9494	0.9909
	12008	0.9680	0.9559
	22002	0.9790	0.9946

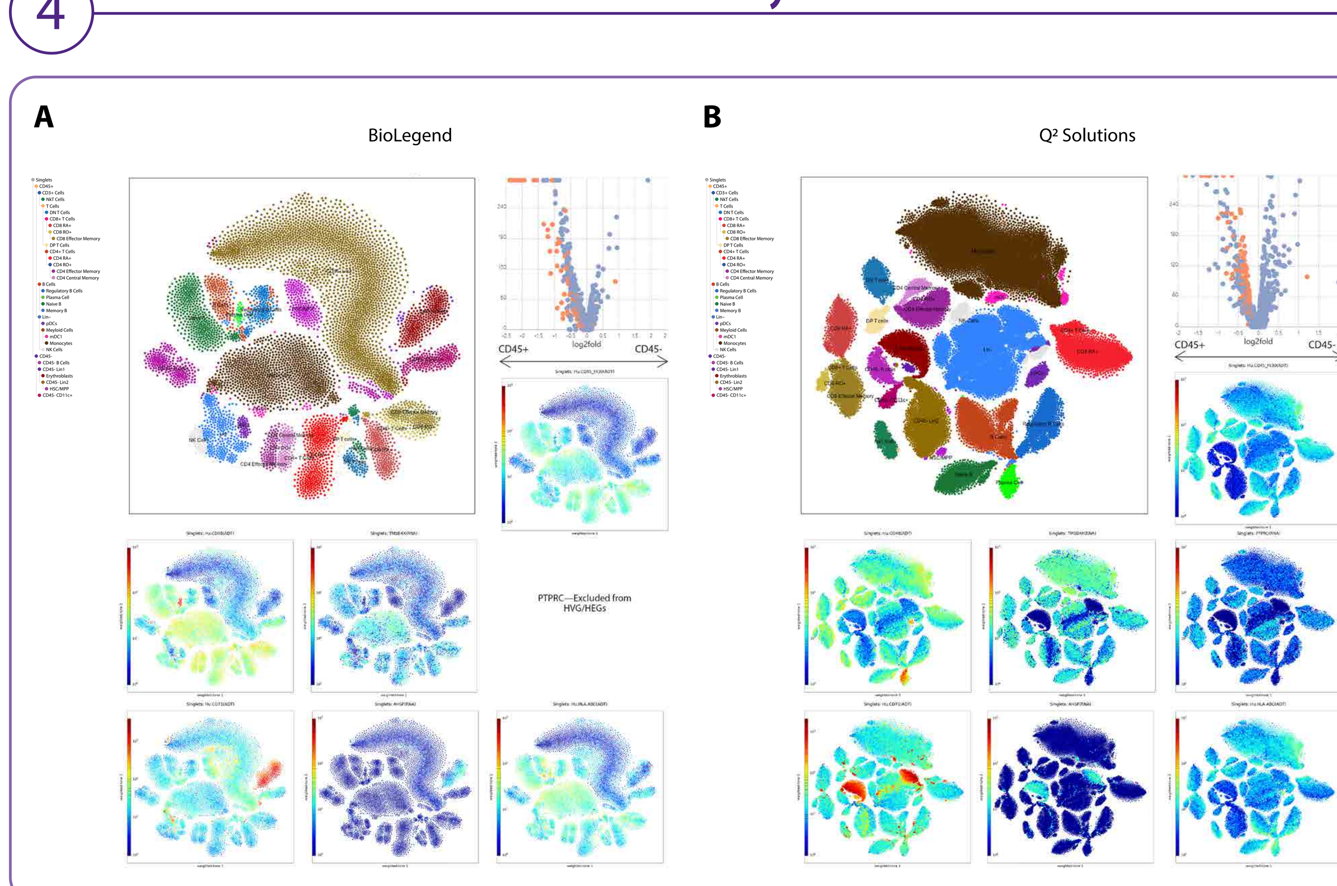
B

	0194C	1103C	11000
PBMCs	0.6423	0.8396	0.6945

	12006	12008	22002
BMMCs	0.9297	0.9816	0.9778

Technical reproducibility and differences in cell type and ADT/RNA (GEX) metrics between study sites was investigated. (A) Pearsons R values were calculated on cell type percentages of parent populations across technical replicates for each sample donor at each site. Technical reproducibility was very high across donors, sites and conditions. (B) Pearsons R values were calculated on cell type percentages of parent populations for each sample donor matched. Correlations were much stronger for all donors with cBMMCs, which had matched sample handling procedures as compared with cPBMCs using purposefully different procedures to artificially reduce quality in one sample set. (C) Percent of parent populations for each cell type within each donor when visualized have a much stronger relationship in cBMMCs compared to cPBMCs. (D) Median counts for ADT and RNA features were extracted for each cell type and Pearsons R calculated. cPBMCs had markedly lower concordance between study sites (normal vs lower quality donor matched) with RNA exhibiting especially low correlations compared with cBMMCs.

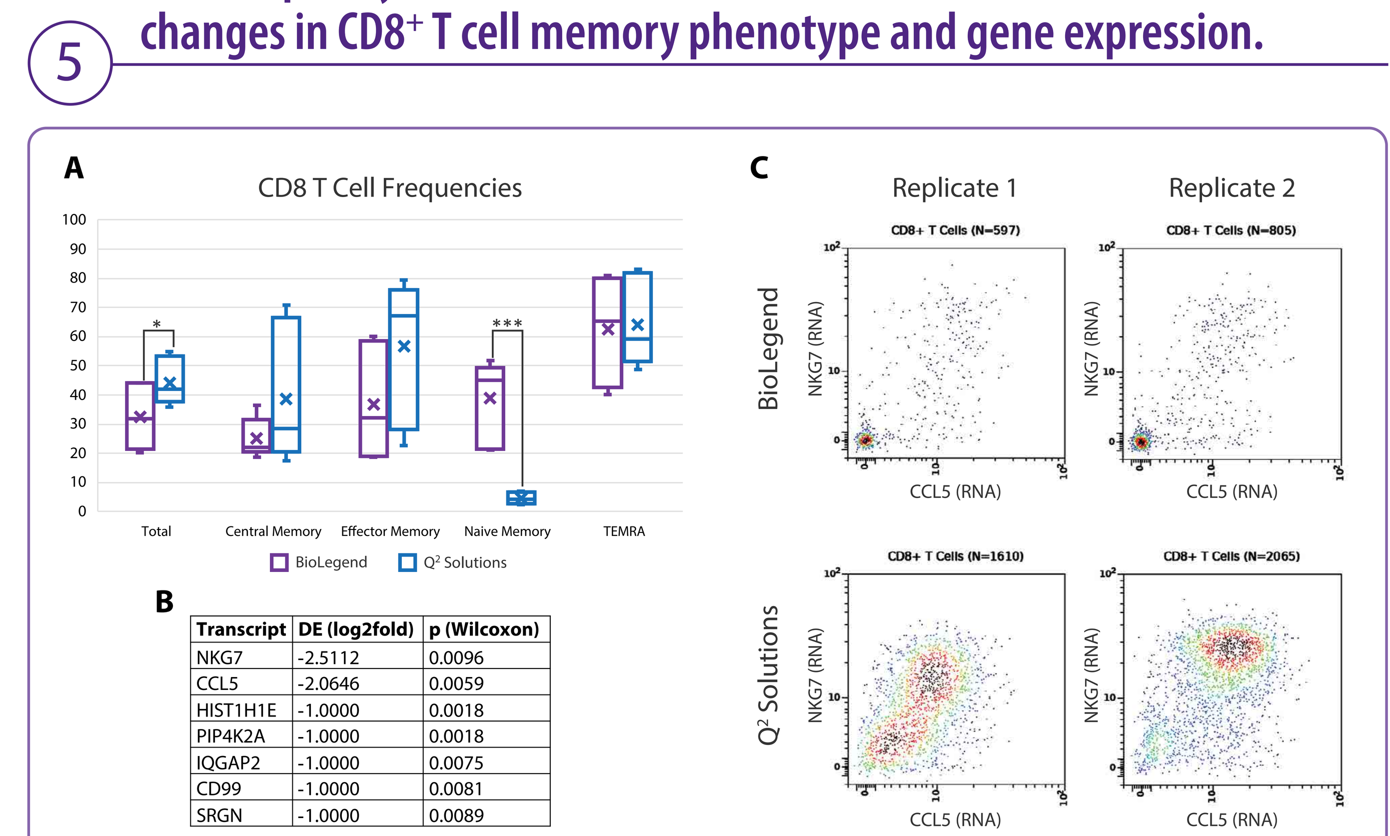
4 Relative cell type frequency, protein, and RNA expression patterns in cBMMCs are consistent across study sites.



Representative weighted tSNEs show colorized cell types, as well as specific proteins or transcripts patterns as indicated. Volcano plots demonstrating magnitude and number of altered proteins and transcripts between CD45- and CD45+ cells are also shown for cBMMCs processed using identical procedures at BioLegend (A) and Q² Solutions (B). Some differences in cell frequency between replicates can be observed within cBMMCs especially in CD45 populations. Protein and RNA features defining cells within each group are relatively consistent across sites with CD45+ cells being predominantly defined by markers associated with mature immune cell populations (MHC, CD48, TMSB4X) and CD45- having small populations of immature leukocytes, erythroid progenitors (CD71, AHSP), and a notable group of lineage marker negative cells (other stromal cells).

Explore our BioLegend and Q² Solutions provider page.

5 CITE-Seq analysis reveals that cold-shock treatment induces changes in CD8+ T cell memory phenotype and gene expression.

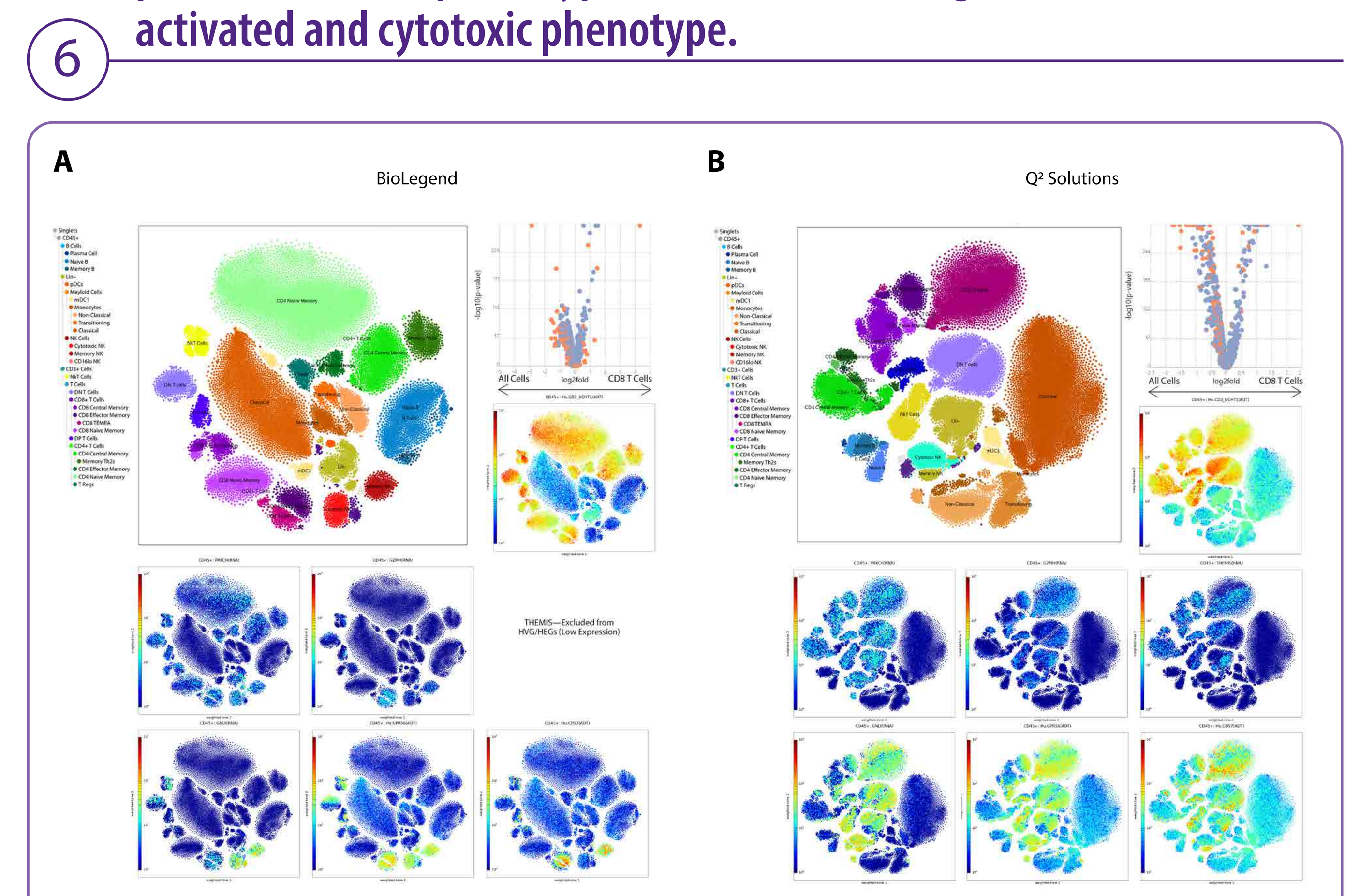


B

Transcript	DE (log2fold)	p (Wilcoxon)
NKG7	-2.5112	0.0096
CCL5	-2.0646	0.0059
HIST1H1E	-1.0000	0.0018
PIP4K2A	-1.0000	0.0018
IQGAP2	-1.0000	0.0075
CD99	-1.0000	0.0081
SRGN	-1.0000	0.0089

CITE-Seq analysis highlights that T cell populations were particularly different between cPBMCs processed using "normal" and "cold-shocked" procedures. CD8+ T cells were investigated further as a representative population. (A) Frequency of total CD8+ T cells in Q² Solutions cPBMCs (cold-shocked) were slightly higher, however, CD8+ T cells with "naïve memory" phenotypes were markedly lower. (B) Seven transcripts had significantly higher expression in Q² Solutions CD8+ T cells from cPBMCs with the two highest being CCL5 (RANTES) and NKG7, both associated with CD8+ T cell activation. (C) Dot plots from a representative donor show a large fraction of double positive CCL5 and NKG7 CD8+ T cells in "cold-shocked" cPBMCs (top panels) compared with non-shocked cells (bottom panels).

6 Cold-shocked cPBMCs show an altered cell type distribution, protein, and RNA phenotypes in T cells exhibiting an activated and cytotoxic phenotype.



Alterations in relative frequencies of several cell populations and proteins/transcripts associated with T cell activation are observed in simulated reduced quality cPBMCs (cold-shocked). Representative weighted tSNEs colorized by cell types, and specific proteins or transcripts were generated. Volcano plots demonstrating magnitude and number of altered proteins and transcripts between total CD45+ cells and CD8+ T cells are also shown for cPBMCs processed at BioLegend (normal) (A), and Q² Solutions (simulated reduced quality, "cold-shocked") (B). Substantially more cells in cPBMCs processed at Q² Solutions expressing either RNAs or proteins associated with activation (PRKCH, GPR56, CD57) and cytotoxicity (GNLY, GZMH) can be observed especially within T cell populations. Notably the T cell transcription factor THEMIS was detected only in the cold-shocked T cells and in a large proportion of T cells. THEMIS is not normally detected in blood T cells.

Conclusions

- Technical reproducibility for all sites, conditions, and sample types was high, indicating a high degree of CITE-Seq consistency using TotalSeq™ reagents by experienced users.
- Cold-shocking cPBMCs during sample processing resulted in significantly reduced cell viability and 10x Cell Ranger metrics, however, CITEseq data was still highly usable and revealed many alterations to cell types, protein, and RNA metrics when compared with donor-matched normally processed samples.
- Cell type and biologically relevant analyte differences could be observed within several cell populations such as CD8+ T cells for simulated reduced-quality cPBMCs.
- cBMMCs had a similar viability to cold-shocked cPBMCs, however, measured cell types, protein, and RNA metrics were highly correlated between study sites with matched donor samples and matching sample processing procedures.

Optimally, users should strive for the highest quality samples possible to ensure the best data. While researchers should objectively evaluate experimental results, suboptimal sequencing metrics may not necessarily indicate poor quality CITE-Seq data. With cryopreserved and/or biologically challenging samples types, reduced sample quality may be overcome to yield informative and consistent results. Care should be taken to consider optimal sequencing depth in the context of sample and sequencing data quality for ADT libraries, as too much or too little sequencing can both have negative impacts on data quality. Sample processing procedures used may impact data quality significantly as was demonstrated with the comparison of cold-shocked cPBMCs with normally processed cPBMCs. Users are encouraged to consult with investigators, service providers, and manufacturers experienced with CITE-Seq especially when working with challenging, clinically relevant samples.

References

1. Stoeckius, M. *et al*. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017; 14, 865–868.
2. Wood, B.L. *et al*. 2006 Bethesda International Consensus recommendations on the immunophenotypic analysis of hematolymphoid neoplasia by flow cytometry: Optimal reagents and reporting for the flow cytometric diagnosis of hematopoietic neoplasia. *Cytometry*, 72B: S14–S22.
3. Wang SR, *et al*. OMIP 071: A 31-Parameter Flow Cytometry Panel for In-Depth Immunophenotyping of Human T-Cell Subsets Using Surface Markers. *Cytometry*. 2021; 99: 273–277.
4. Boesch M, *et al*. Definition of all principal human leukocyte populations using a broadly applicable 14-color panel. *Cytometry*. 2022; 101: 15–20.

04-0154-00