Qualification of Gene and Protein Signatures from CITE-seq Data for Deep Immunophenotyping in Human Clinical Trials

Matt Clutter[1], Sarah Asbury[1], Jean-Philippe Goulet[1], Pieter Mestdagh[2], Pieter-Jan van Dam[2], Emmanuel Rivière[2], Zach Bjornson[1], Eustache Paramithiotis[1]

[1]CellCarta, Montreal, QC, Canada [2]CellCarta, Antwerp, Belgium

CellCarta

NOVEL ASPECTS

Cellular indexing of transcriptomes and epitopes (CITE-seq)[1] is the most comprehensive tool for immune monitoring today but identifying relevant and robust features from such vast datasets has been challenging. We investigated the repeatability of multimodal immune signatures at the single cell level, which to our knowledge has not been previously published or presented in the field. In addition, we evaluated the repeatability of novel signatures of T cell exhaustion.

INTRODUCTION

Rapidly evolving methods and decreasing costs are accelerating the use of single cell sequencing for deep immune profiling. CITE-seq generates highly multiplexed protein and gene expression readouts. The complexity of such data requires new analysis tools to extract insights. We demonstrate how signatures from CITE-seq data can be visualized, interrogated, and assessed for repeatability as foundational steps for their application in clinical trial monitoring.

METHODS

Peripheral blood mononuclear cells (PBMCs) from a single healthy donor were thawed and stained in 4 replicates with 137 antibodies from the Biolegend Total-Seq-C universal cocktail v1. 14,000 cells per replicate were captured on the 10X Genomics Chromium X using the 5' HT v2 workflow. RNA and Antibody libraries were sequenced on an Illumina NovaSeq 6000 to obtain ~28,000 RNA reads/cell and ~12,000 Antibody reads/cell. Cell Ranger (10x Genomics) was used to generate feature-barcode H5 files, then a CellCarta pipeline leveraging aspects of Seurat[2] generated FCS files with the following parameters per cell:

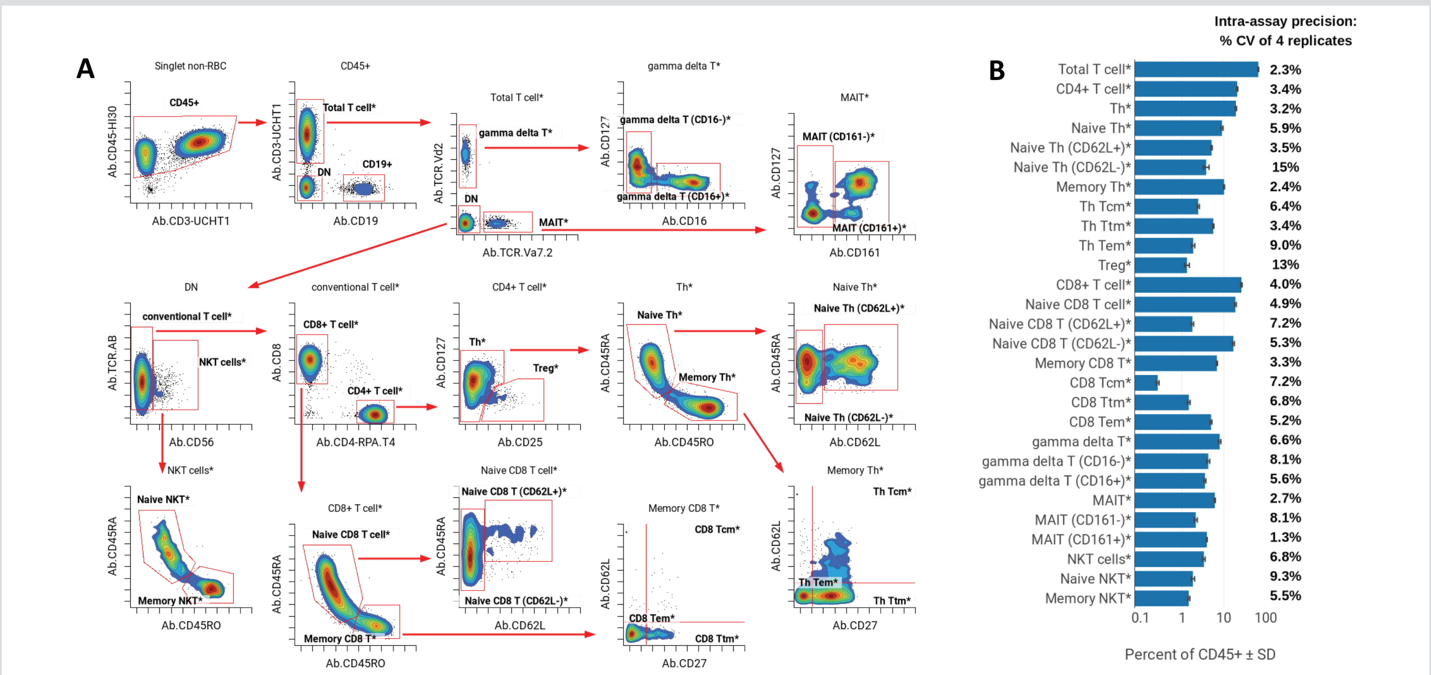
- 137 protein features from TotalSeq antibody staining
- 300 sctype marker genes of various immune cell populations[3]
- 1,000 highly variable genes across all cells of the experiment
- 487 enrichment scores computed by scGSEA[4] from various MSigDB[5] gene set collections (Hallmark, C5 GO Biological Processes, C7 Immune, C2 Biocarta and Reactome)
- 122 custom gene set enrichment scores created from literature search and targeted publications[6,7,8,9,10]

The resulting ~2,000-parameter FCS files were uploaded to CellCarta's cloud-based CellEngine software for analysis.

RESULTS

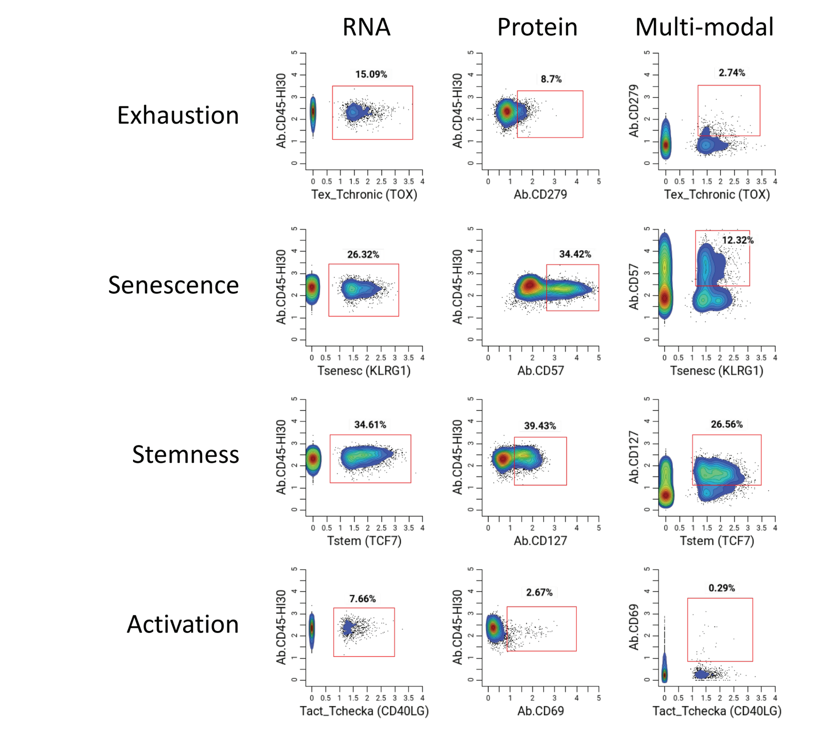
Cells from 4 replicate samples were gated using canonical protein markers to identify 17 unique T cell subset populations (Fig 1A). Frequencies of gated populations ranged from 0.26% (CD8 Tcm) to 67% (Total T cells) of CD45+ leukocytes. Intra-assay precision ranged from 1.3% to 15% CV (Fig 1B).

Figure 1: Gating T cell subsets in CITE-seq data


A
B
Intra-assay precision: % CV of 4 replicates
Percent of CD45+ ± SD

Expression of RNA and Protein features of T cell function (exhaustion, senescence, stemness, and activation) were visualized for Total T cells (Fig 2). Cells with positive expression scores were gated. Intra-assay precision based on Mean and % Positive statistics ranged from 0.77% to 12% CV. Multi-modal readouts were defined using both RNA and Protein expression levels. Intra-assay precision for multi-modal features ranged from 6% to 14% CV.

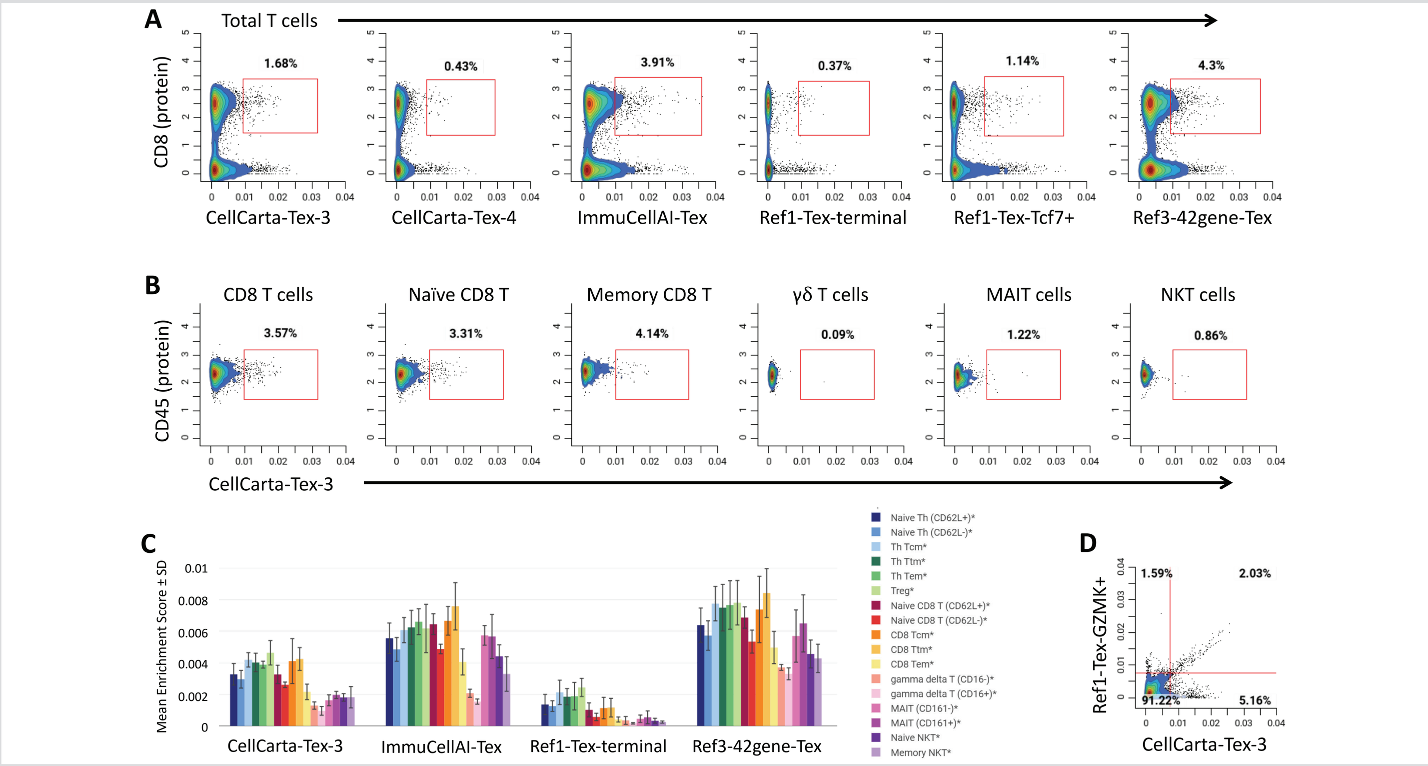
Figure 2: RNA, Protein, and Multi-modal features


RNA
Protein
Multi-modal
Exhaustion
Senescence
Stemness
Activation

Readout	Mean expression	% Positive
TOX (RNA)	0.24 ± 0.027 12% CV	15% ± 1.5% 10% CV
PD-1 (Protein)	0.89 ± 0.018 2.0% CV	8.7% ± 0.38% 4.4% CV
TOX+ PD-1+	–	2.6% ± 0.37% 14% CV
KLRG1 (RNA)	0.45 ± 0.011 2.4% CV	28% ± 1.0% 3.8% CV
CD57 (Protein)	2.5 ± 0.032 1.3% CV	36% ± 1.5% 4.0% CV
KLRG1+ CD57+	–	14% ± 0.87% 6.4% CV
TCF7 (RNA)	0.64 ± 0.041 6.5% CV	35% ± 2.1% 6.1% CV
CD127 (Protein)	1.1 ± 0.017 1.5% CV	40% ± 0.80% 2.0% CV
TCF7+ CD127+	–	28% ± 1.7% 6.0% CV
CD40LG (RNA)	0.12 ± 0.0067 5.5% CV	8.4% ± 0.59% 7.0% CV
CD69 (Protein)	0.29 ± 0.0022 0.77% CV	2.6% ± 0. 093% 3.6% CV
CD40LG+ CD69+	–	0.31% ± 0.023% 7.3% CV

Gene enrichment scores for 6 T cell exhaustion signatures were visualized for Total T cells to identify exhausted T cells (Fig 3A). The fraction of exhausted cells ranged from 0.4% to 4% depending on the composition of the gene set. The T cell exhaustion signatures were also visualized across different T cell populations (Fig 3B and 3C). Leveraging the single cell aspect of enrichment scores, cells enriched with multiple signatures were identified simply using standard flow cytometry-based gating methods (Fig 3D).

Figure 3: Visualizing single cell gene signatures


A
Total T cells
CD8 (protein)
CellCarta-Tex-3
CellCarta-Tex-4
ImmuCellAI-Tex
Ref1-Tex-terminal
Ref1-Tex-Tcf7+
Ref3-42gene-Tex
B
CD8 T cells
Naïve CD8 T
Memory CD8 T
γδ T cells
MAIT cells
NKT cells
CD45 (protein)
C
Mean Enrichment Score ± SD
D
Ref1-Tex-GZMK+

Gene sets for calculating T cell exhaustion scores are listed. Intra-assay precision for the signatures ranged from 8% to 33% CV for Total T cells and 16% to 44% CV on average across the 17 unique T cell subsets.

Table 1: Composition of T cell exhaustion gene signatures

Signature	Gene list	%CV in Total T cells	%CV in 17 T cell pops. (Avg)
CellCarta-Tex-2	PDCD1 (PD-1), LAG3, HAVCR2 (TIM-3), TOX, ENTPD1 (CD39)	22%	29%
CellCarta-Tex-3	PDCD1 (PD-1), LAG3, HAVCR2 (TIM-3), TOX, ENTPD1 (CD39), CTLA4, TIGIT	8%	19%
ImmuCellAI-Tex[6]	ADGRG1, AFAP1L2, CCND2, CD38, CD8A, CD8B, CTLA4, EOMES, FUT8, LAG3, PDCD1 (PD-1), +8 others	9%	16%
Ref1-Tex-terminal[7]	CXCL13, TNFRSF9 (4-1BB), ENTPD1 (CD39), HAVCR2 (TIM-3), CTLA4, TNFRSF18 (GITR), TIGIT, DUSP4, +42 others	33%	44%
Ref3-Tex-42genes[8]	HAVCR2 (TIM-3), ENTPD1 (CD39), PDCD1 (PD-1), TNFSF4 (OX40L), GZMA, MKI67 (Ki67), TOX, TIGIT, TNFRSF9 (4-1BB), LAG3, CTLA4, +30 others	15%	18%
Ref4-Tex[9]	GZMB, CXCL13, LAG3, VCAM1, HAVCR2 (TIM-3), CCL5, PDCD1 (PD-1), DUSP4, TNFRSF9 (4-1BB), CTLA4, GZMH, ENTPD1 (CD39), +48 others	14%	16%
Ref5-Tex[10]	CCL3, CXCL13, IFNG, GZMB, CCL5, CD8A, CD63, XCL1, XCL2, PDCD1 (PD-1), GZMH, HAVCR2 (TIM-3), LAG3, DUSP4, +82 others	15%	25%

24 gene signatures primarily focused on T cell function were evaluated in Total T cells. 79% of the signature enrichment scores had a CV < 25% (Fig 4A).

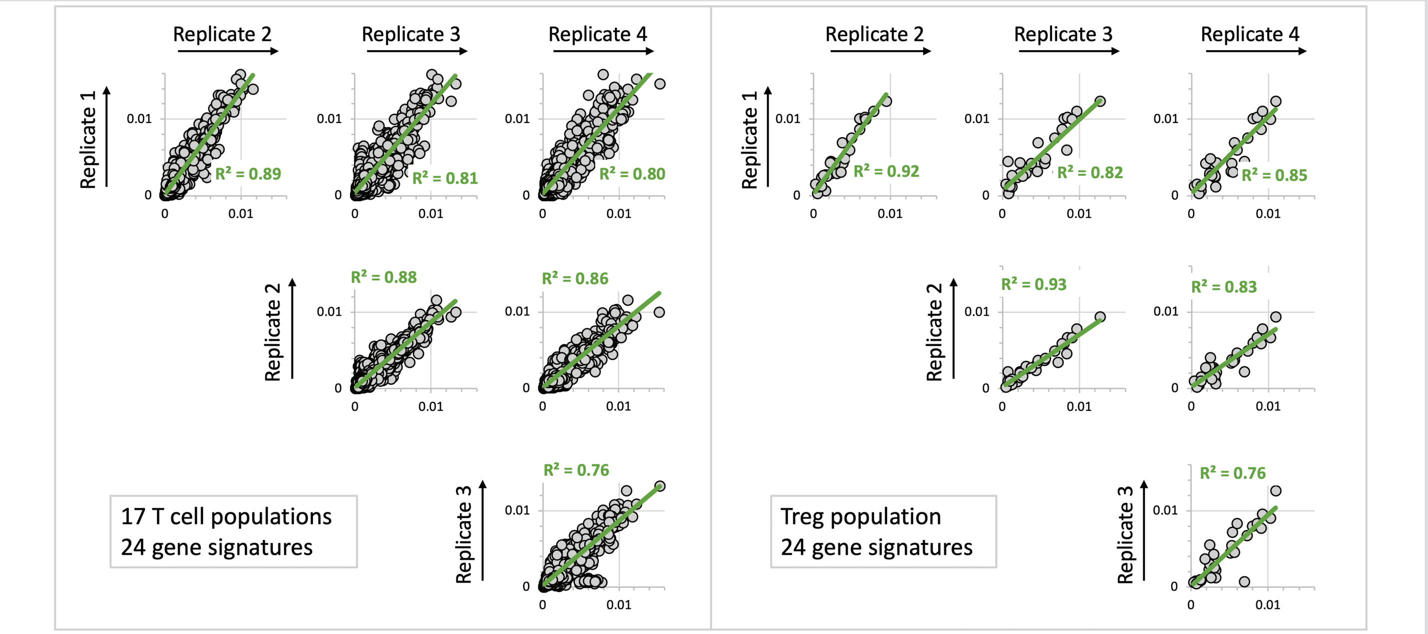
Figure 4: Repeatability of single cell gene signatures


A
Gene Enrichment Score Repeatability
Total T cells - 24 gene signatures analyzed
CV < 25%: 79% of signatures
Coefficient of Variation
Average Enrichment Score
B
Gene Enrichment Score Repeatability
17 T cell populations - 24 gene gene signatures analyzed
CV < 25%: 53% of signatures
C
Cytotoxicity
Proliferation
T chronic stim.
T exhaustion 3
T senescence
R² = 0.17
R² = 0.35
R² = 0.28
R² = 0.37
R² = 0.11
Cell number per T cell subpopulation

When these 24 gene signatures were evaluated in 17 T cell subsets, including rare memory populations with as few as ~30 cells, 53% of the signature enrichment scores had a CV < 25% (Fig 4B). The enrichment score values correlated with CVs; on average features having scores below ~0.005 showed higher CVs than those with higher enrichment scores. In addition, most gene signatures showed a correlation where on average the T cell populations with lower event counts had higher CVs (Fig 4C).

CITE-seq measures 1000's of genes, proteins, and signatures across dozens of cell populations, so qualifying every readout is impractical. We analyzed the correlation between all pairs of replicates as a high-level measure of assay repeatability. R^2 for Total T cells ranged from 0.76 to 0.89 (Fig 5). These values were unchanged in a relatively rare T cell subset, Treg cells.

Figure 5: Correlation of average enrichment scores for gene signatures across replicates


Replicate 1
Replicate 2
Replicate 3
Replicate 4
R² = 0.89
R² = 0.81
R² = 0.80
R² = 0.88
R² = 0.86
R² = 0.76
17 T cell populations
24 gene signatures
R² = 0.92
R² = 0.82
R² = 0.85
R² = 0.93
R² = 0.83
R² = 0.76
Treg population
24 gene signatures

KEY TAKEAWAYS

- RNA, Protein, and Multi-modal features of T cell function - including several T cell exhaustion signatures - showed consistent intra-assay precision with CVs < 25% (Fig 2, Table 1).
- Single cell gene set enrichment scores can be visualized and gated like flow cytometry data (Fig 3).
- Correlation coefficients between sample replicates may help evaluate repeatability for the large-scale assays such as CITE-seq (Fig 5).
- Understanding how the repeatability of biological signatures is affected by cell population size, number of features chosen, and feature abundance will help enable the use of novel signatures as biomarkers in clinical trial monitoring.

REFERENCES

1. Stoeckius et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 14:865-8 (2017)
2. Hao and Hao et al. Integrated analysis of multimodal single-cell data. Cell 184:3573 (2021)
3. Ianevski et al. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat Commun 13, 1246 (2022).
4. Franchini et al. Single-cell gene set enrichment analysis and transfer learning for functional annotation of scRNA-seq data. NAR Genom Bioinform. 5(1):lqad024 (2023).
5. https://www.gsea-msigdb.org/gsea/msigdb
6. Miao et al. ImmuCellAI: A unique method for comprehensive T cell subsets abundance prediction and its application in cancer immunotherapy. Adv. Sci. 7(7):1902880 (2020).
7. Zheng et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science. 374(6574):abe6474 (2021).
8. Zhang et al. Prioritizing exhausted T cell marker genes highlights immune subtypes in pan-cancer. iScience. 26(4) 106484 (2023).
9. Tietscher et al. A comprehensive single cell map of T cell exhaustion-associated immune environments in human breast cancer. Nat Commun. 6;14(1) 98 (2023).
10. Li et al. Novel T cell exhaustion gene signature to predict prognosis and immunotherapy response in thyroid carcinoma from integrated RNA-seq analysis. Sci Rep. 14:8375 (2024).

Please send inquiries to
Matt Clutter, PhD
Director, Global R&D
matt.clutter@Cellcarta.com

POSTER NUMBER: Day2-P15

www.cellcarta.com