Analyzing CITE-seq data like an Immunologist with a hybrid hierarchical gating approach in CellEngine software

Matt Clutter[1], Sarah Asbury[2], David Riley[1], Pamela White[1], Jean-Philippe Goulet[2], Eustache Paramithiotis[2], Zach Bjornson[1]

[1]CellCarta, Fremont, CA, USA, [2]CellCarta, Montreal, QC, Canada

POSTER NUMBER: 163

CellCarta

BACKGROUND

The use of single cell sequencing for deep immune monitoring in both discovery and clinical settings is accelerating. Cellular indexing of transcriptomes and epitopes (CITE-seq) is a version of single cell sequencing that generates highly multiplexed protein and gene expression readouts[1]. Dimensional reduction and clustering are the standard methods used to define immune cell populations in CITE-seq analyses. In contrast, hierarchical gating has been the approach favored by Immunologists for decades to identify cell populations in flow cytometry single cell datasets. Here we evaluated the use of hierarchical gating to identify immune cell types in CITE-seq data.

METHODS

Two lab operators thawed two PBMC samples from a single healthy donor. Samples were stained with the Biolegend TotalSeq-C universal cocktail, then 14,000 cells per sample were captured on the 10X Genomics Chromium X using the 5' HT v2 workflow. RNA and Antibody libraries were generated and sequenced on an Illumina NovaSeq 6000 S2 flow cell to obtain ~28,000 RNA reads/cell and ~12,000 Antibody reads/cell. 10X Genomics Cell Ranger was used to generate feature-barcode matrix H5 files. A CellCarta pipeline leveraging aspects of Seurat[2] was used to generate customizable FCS files from H5 files containing the following parameters for each cell:

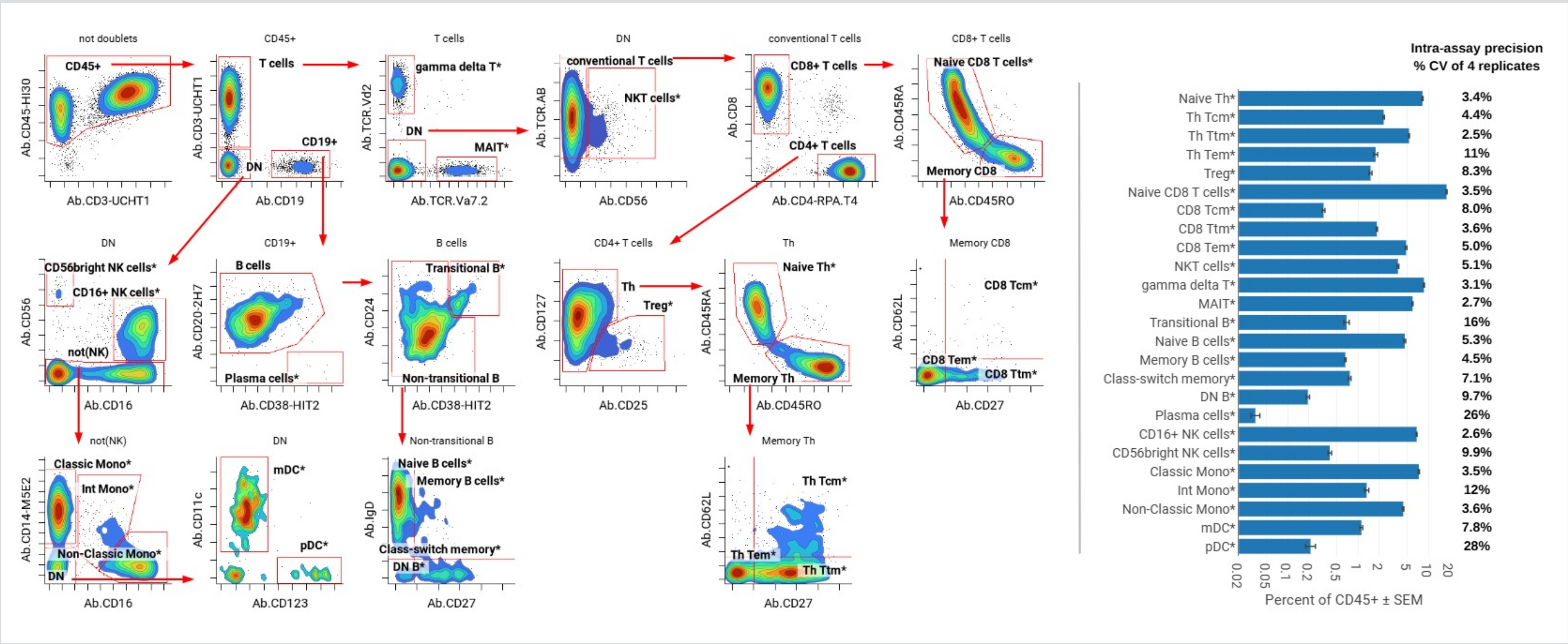
- 137 protein features from TotalSeq antibody staining
- 295 sctype marker genes associated with various immune cell populations[3]
- 1,000 of the most highly variable genes selected across all cells of the experiment
- 164 gene set enrichment scores defined by the MSigDB Hallmark and GO Biological Processes collections, computed by scGSEA[4]
- 4 dimensionality reduction parameters (UMAP1, UMAP2, t-SNE1, t-SNE2)

The resulting 1,600-parameter FCS files were uploaded to CellCarta's cloud-based CellEngine software for analysis.

RESULTS

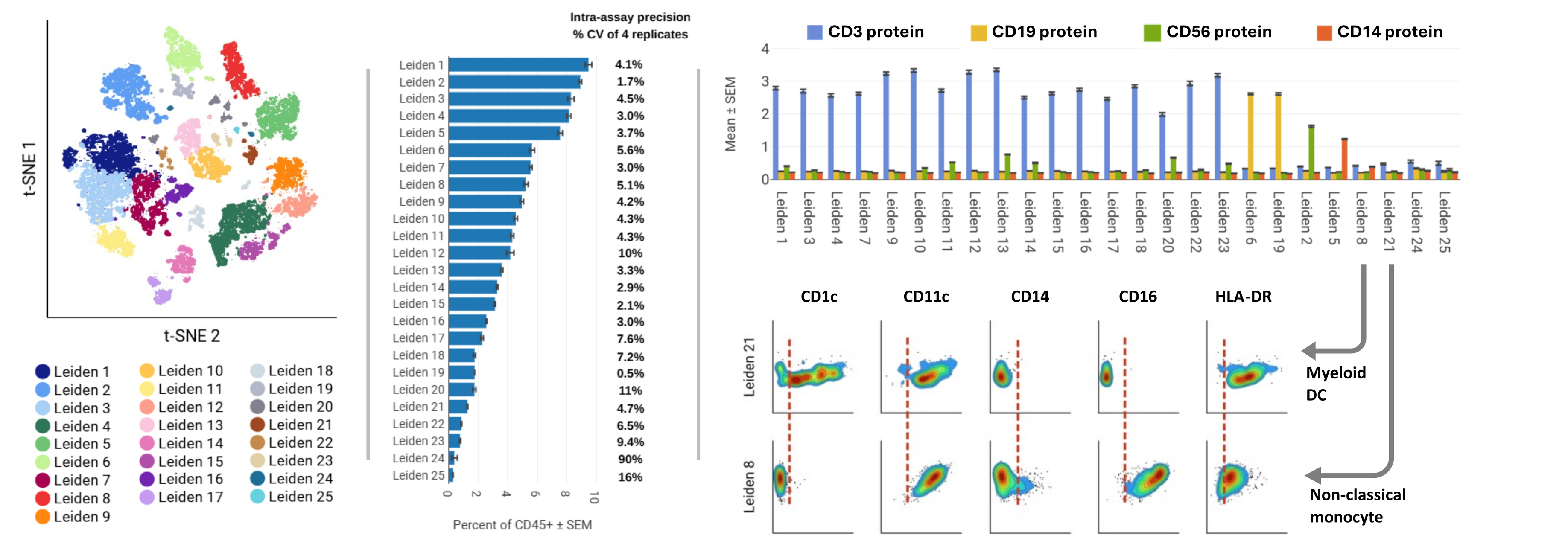
Hierarchical gating consistently identified cell populations in CITE-seq data

Cells from 4 replicate samples were gated using 23 canonical protein markers from the CITE-seq data to identify 25 immune populations. The frequencies of the gated populations showed intra-assay precision with %CVs ranging from 2.5% to 28%.

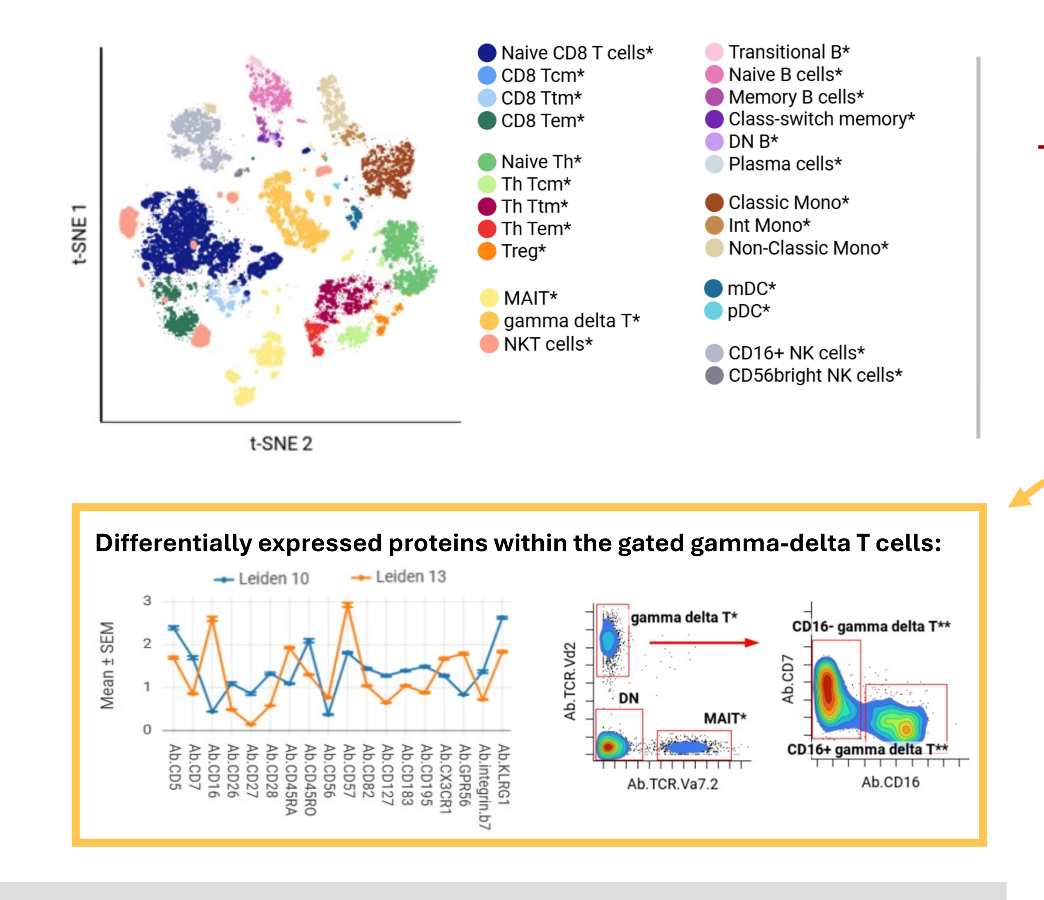

Intra-assay precision % CV of 4 replicates
Naive Th* 3.4%
Th Tcm* 4.4%
Th Ttm* 2.5%
Th Tem* 11%
Treg* 8.3%
Naive CD8 T cells* 3.5%
CD8 Tcm* 8.0%
CD8 Ttm* 3.6%
CD8 Tem* 5.0%
NKT cells* 5.1%
gamma delta T* 3.1%
MAIT* 2.7%
Transitional B* 16%
Naive B cells* 5.3%
Memory B cells* 4.5%
Class-switch memory* 7.1%
DN B* 9.7%
Plasma cells* 26%
CD16+ NK cells* 2.6%
CD56bright NK cells* 9.9%
Classic Mono* 3.5%
Int Mono* 12%
Non-Classic Mono* 3.6%
mDC* 7.8%
pDC* 28%
Percent of CD45+ ± SEM

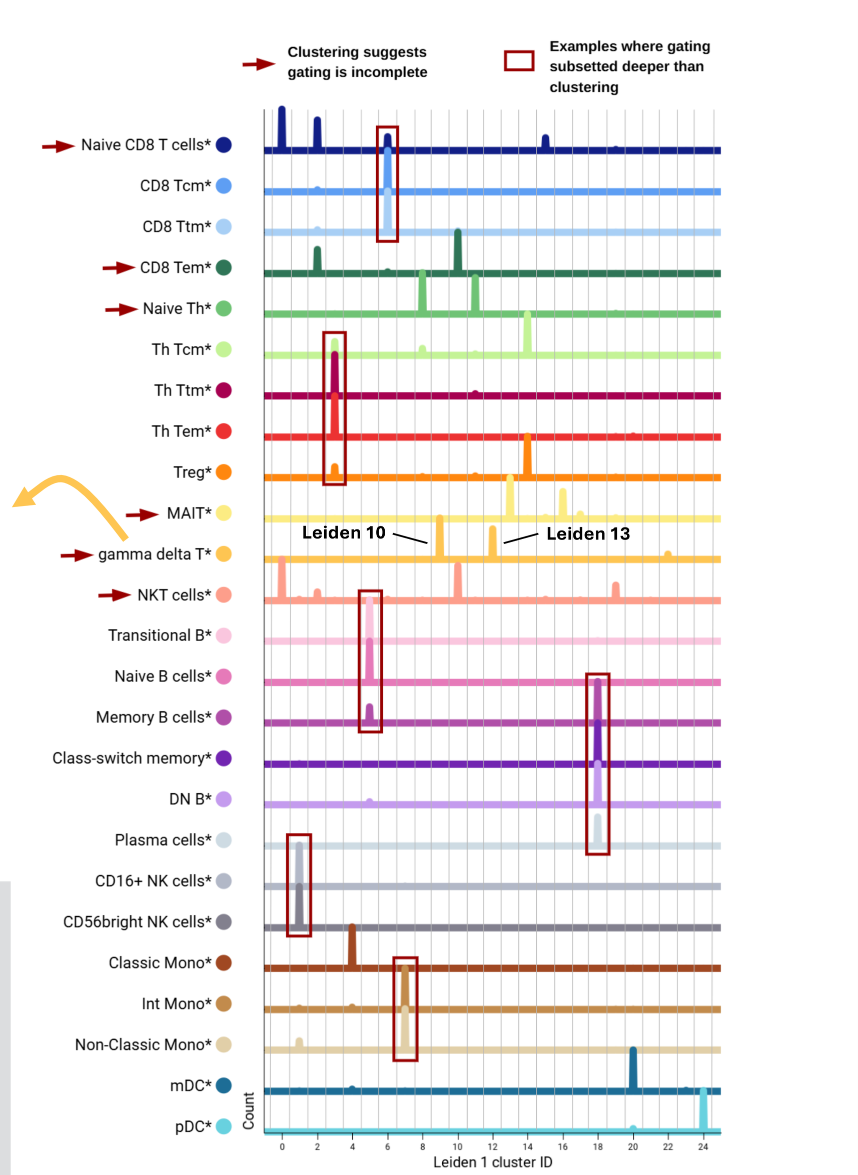
Unbiased clustering generated consistent clusters across replicates

Cells from the 4 replicate samples were clustered by the 23 protein features used for hierarchical gating. t-SNE dimensionality reduction and Leiden Community Detection algorithms implemented in CellEngine resulted in consistent clustering across replicates. Clusters were interrogated in CellEngine for expression of protein and gene markers of interest, including common phenotypic markers like CD3 (T cells), CD19 (B cells), CD56 (NK cells), and CD14 (monocytes).


t-SNE 1
t-SNE 2
Intra-assay precision % CV of 4 replicates
Leiden 1 4.1%
Leiden 2 1.7%
Leiden 3 4.5%
Leiden 4 3.0%
Leiden 5 3.7%
Leiden 6 5.6%
Leiden 7 3.0%
Leiden 8 5.1%
Leiden 9 4.2%
Leiden 10 4.3%
Leiden 11 4.3%
Leiden 12 10%
Leiden 13 3.3%
Leiden 14 2.9%
Leiden 15 2.1%
Leiden 16 3.0%
Leiden 17 7.6%
Leiden 18 7.2%
Leiden 19 0.5%
Leiden 20 11%
Leiden 21 4.7%
Leiden 22 6.5%
Leiden 23 9.4%
Leiden 24 90%
Leiden 25 16%
Percent of CD45+ ± SEM
CD3 protein
CD19 protein
CD56 protein
CD14 protein
Mean ± SEM
CD1c
CD11c
CD14
CD16
HLA-DR
Leiden 21
Leiden 8
Myeloid DC
Non-classical monocyte

Clustering revealed several gated populations that could be further subsetted

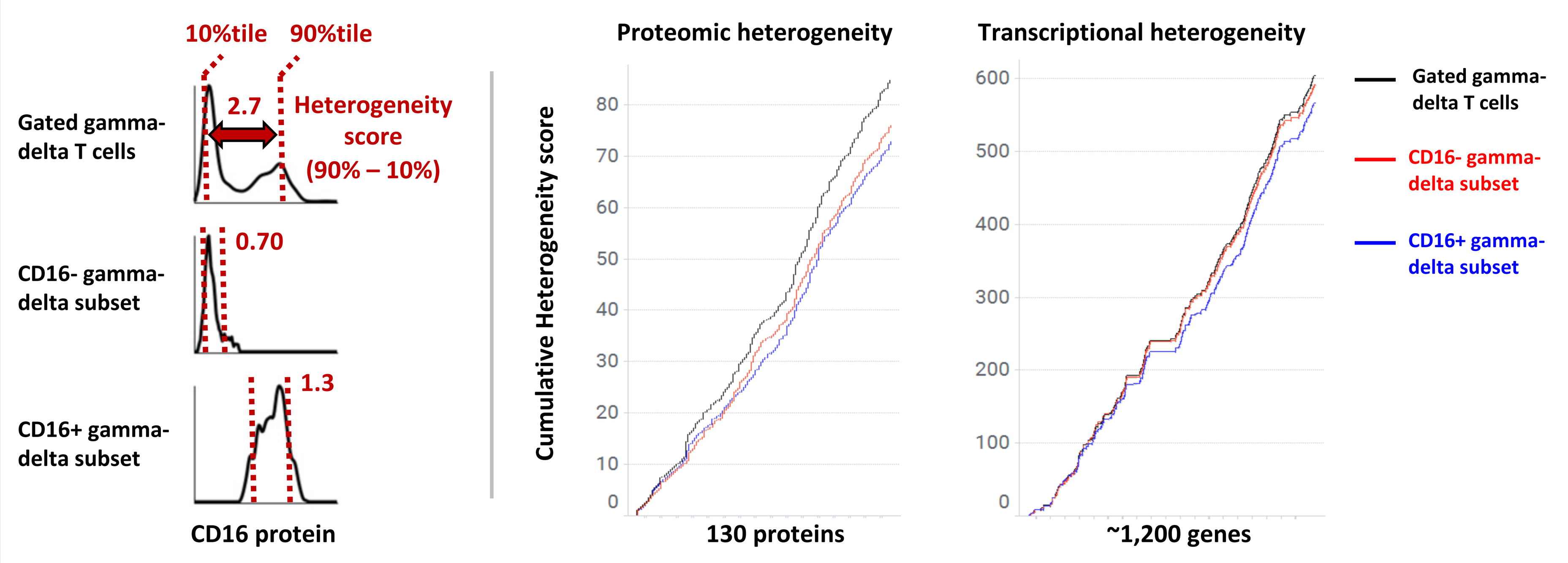

t-SNE 1
t-SNE 2
Naive CD8 T cells*
CD8 Tcm*
CD8 Ttm*
CD8 Tem*
Naive Th*
Th Tcm*
Th Ttm*
Th Tem*
Treg*
MAIT*
gamma delta T*
NKT cells*
Transitional B*
Naive B cells*
Memory B cells*
Class-switch memory*
DN B*
Plasma cells*
Classic Mono*
Int Mono*
Non-Classic Mono*
mDC*
pDC*
CD16+ NK cells*
CD56bright NK cells*
Differentially expressed proteins within the gated gamma-delta T cells:
Leiden 10
Leiden 13
Mean ± SEM
gamma delta T*
DN
MAIT*
Ab.TCR.Vd2
Ab.TCR.Va7.2
CD16- gamma delta T**
CD16+ gamma delta T**
Ab.CD7
Ab.CD16


Clustering suggests gating is incomplete
Examples where gating subsetted deeper than clustering
Leiden 10
Leiden 13
Count
Leiden 1 cluster ID

Gated populations were displayed in t-SNE space for comparison to the clustering results (upper left). By examining the fraction of cells from each population that were assigned to each Leiden cluster (right), we identified gated populations with heterogeneity that warranted further cell subsetting. For example, the gated gamma-delta T cells were split between Leiden clusters 10 and 13. Markers such as CD16 enabled the identification of two distinct gamma-delta T subsets.

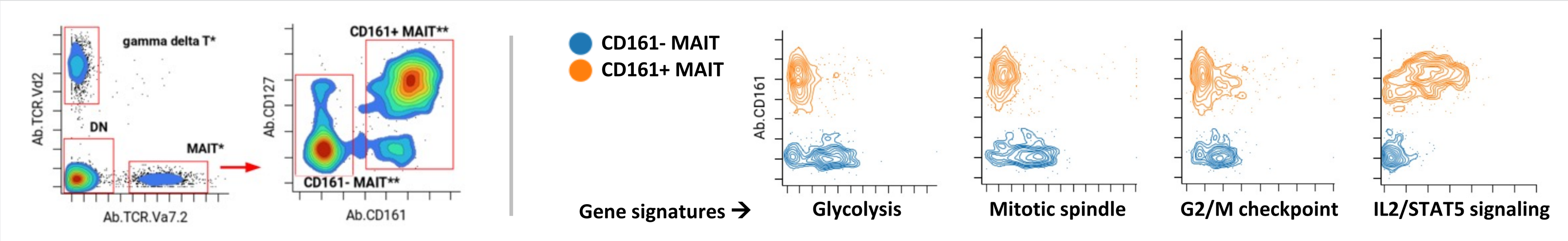
Hybrid gated populations were more uniform than the original gated populations and the unbiased cell clusters

A heterogeneity score was devised as the spread between the 90th and 10th percentiles of the expression of each protein or gene in each cell population. Lower scores indicate a more molecularly uniform population. By looking at cumulative heterogeneity across all proteins and genes, we concluded that immune cell populations identified by hybrid gating with input from clustering were more uniform than the originally gated populations (illustrated for gamma-delta T cell subsets, below).


10%tile
90%tile
Gated gamma-delta T cells
2.7
Heterogeneity score (90% – 10%)
CD16- gamma-delta subset
0.70
CD16+ gamma-delta subset
1.3
CD16 protein
Proteomic heterogeneity
Cumulative Heterogeneity score
130 proteins
Transcriptional heterogeneity
~1,200 genes
Gated gamma-delta T cells
CD16- gamma-delta subset
CD16+ gamma-delta subset

Hybrid gated subpopulations had unique gene signatures

Hybrid gated subsets of the original immune cell populations showed different gene signature scores related to metabolism, cell cycle, and signaling pathways (illustrated for MAIT cell subsets, below).


gamma delta T*
DN
MAIT*
Ab.TCR.Vd2
Ab.TCR.Va7.2
CD161+ MAIT**
CD161- MAIT**
Ab.CD127
Ab.CD161
CD161- MAIT
CD161+ MAIT
Gene signatures →
Glycolysis
Mitotic spindle
G2/M checkpoint
IL2/STAT5 signaling

KEY TAKEAWAYS

- Gating is intuitive to Immunologists and defines cell populations using established markers.
- Unbiased clustering can uncover additional cell subsets within gated populations.
- A hybrid approach combining hierarchical gating with clustering enables a more comprehensive analysis of CITE-seq data anchored in uniform, immunologically relevant cell populations.
- CellEngine enables hybrid gating with clustering in a single analysis environment.

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 184:3573-87 (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).

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

163

www.cellcarta.com