Unsupervised clustering reveals cell population in melanoma patients treated with anti-PD-1 therapy

Gage Black PhD, Hannah Selken, Justin Jarrell PhD, Li-Chun Cheng PhD, Ramji Srinivasan

Teiko

Background

Gating has been the status quo to analyze cytometry datasets for decades. The challenge with this approach is the inability to scale to high-parameter (20-40+) panels. For instance, a 43-marker panel theoretically yields billions of potential cell types, requiring ~239 hours to fully gate.

Unsupervised computational methods overcome these limitations by enabling unbiased exploration of the immune landscape. Algorithms such as FlowSOM for clustering and Uniform Manifold Approximation and Projection (UMAP) for visualization can automatically identify distinct cell populations that may be overlooked by traditional gating.

These tools can identify distinctive cell populations that would be missed by manual gating. We developed an unsupervised clustering pipeline that reflects biological structure and enables high-resolution identification of immune cell states. Applying this workflow to 70 samples from 29 melanoma patients treated with anti-PD-1 therapy revealed a previously uncharacterized immune cell population associated with immune-related adverse events (IRAEs) and response, which would have been overlooked by conventional gating approaches.

Methods

We analyzed a 29 patient, 70 PBMC specimen melanoma dataset from the Huntsman Cancer Institute using a 43-marker mass cytometry panel.

Gating

We performed manual gating of FCS files using CellEngine (CellCarta, Montreal, Canada) following the gating strategy found on app.teiko.bio/projects/HCI001/overview.

FlowSOM

We clustered cells using FlowSOM, which employs a self-organizing map (SOM) algorithm to organize cells with similar marker expression patterns into a grid structure. The algorithm first maps all cells onto a grid, where each grid node represents cells with comparable expression profiles through an iterative learning process that minimizes the distance between cells and their assigned nodes. Following SOM training, the algorithm applies hierarchical consensus meta-clustering to group the SOM nodes into final clusters, identifying cell populations that share similar phenotypic characteristics. Here, we used a 12x12 SOM grid and a total of 70 meta clusters.

Naming and Quality Control

We applied our naming algorithm to assign biologically relevant names to cell clusters by evaluating their marker expression. The algorithm assigned biologically relevant labels (e.g., "CD8+ T Memory" vs. "CD8+ T Naive" based on CD27 and CD45RA), refined by marker-specific distinctions (e.g., CD11B). To ensure clusters were correct, we generated a cluster versus marker heatmap and series of UMAP plots, one for each cluster and marker (coloring only the selected cluster or marker).

Acknowledgements

We are grateful for the support from Dr. Siwen Hu-Lieskovan and colleagues at the Huntsman Cancer Institute for providing the PBMC samples analyzed in this study.

Conclusions

Unsupervised clustering revealed a distinct CD161⁺ T-cell population (Cluster 10) that was not captured by the conventional gating scheme. This subset showed heterogeneous CD8 expression, low CD45RA, and variable CD27, consistent with a memory-like phenotype, and expressed very high levels of CD161 compared to other T cells.

The frequency of this cluster was significantly higher in responders to anti–PD-1 therapy and was also elevated during on-treatment timepoints in patients who developed IRAEs. This enrichment suggests that CD161⁺ T cells may expand alongside therapeutic activation of the immune system, potentially contributing to both anti-tumor immunity and immune toxicity.

Published studies have linked CD161⁺ and MAIT-like T cells to Th17 cytokine programs, IL-17/IL-23–driven inflammation, and checkpoint inhibitor–associated toxicity. Our findings build on this by providing direct evidence from high-dimensional cytometry that a CD161-rich T-cell compartment is associated with clinical outcomes in melanoma.

Together, these findings highlight the value of computational clustering approaches in uncovering rare, biologically meaningful immune subsets that may contribute to both therapeutic response and autoimmune toxicity in melanoma patients treated with PD-1 blockade.

Cluster Naming

(1) Cluster cells with FlowSOM.

Each cell is assigned a cluster.

Clusters [70 X 43]

	CD3	CD19	CD56	
[1]	4.2	0.1	0.3	# Cluster 1
[2]	0.2	3.8	0.2	# Cluster 2
[3]	0.3	0.2	4.1	# Cluster 3

(3) Calculate cosine similarity * between the cluster expression table and the gated expression table.

Label cluster with cell type that has highest value.

Similarity matrix [34 X 70]

	Cluster1	Cluster2	Cluster3
T_CELL	0.98	0.15	0.20
B_CELL	0.12	0.99	0.14
NK_CELL	0.18	0.13	0.97

	CD3	CD19	CD56
T_CELL	4.0	0.2	0.3
B_CELL	0.3	3.9	0.2
NK_CELL	0.2	0.1	4.0

Median of MCV for Gated cell populations [34 X 43]

New Cluster Names
1_T_CELL
2_B_CELL
3_NK_CELL

Manually QC clusters

(2) Manually gate cells.

Each cell is assigned a gate.

(4) Look at UMAP expression plots vs UMAP plots with labeled clusters (see right panel for example)

We run FlowSOM to generate clusters. For every identified cluster, we calculate the median channel value (MCV) across all markers.

Separately, we manually gate populations of cells. We gather the MCV for all markers across all cells in each population.

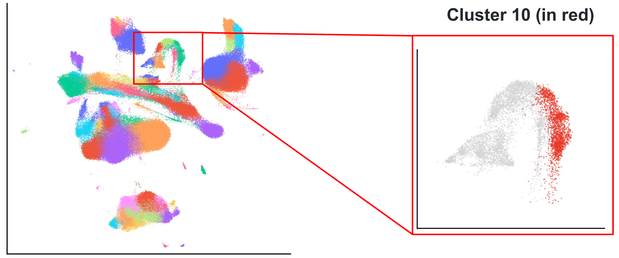
We calculate a cosine similarity between the cluster cells and the gated cells matrices. This results in a matrix where the columns are clusters and the rows are gated cell populations. Each cluster population has a similarity score for every gated cell population. We identify the gated cell population with the highest score for each cluster. We use this to rename the cluster.

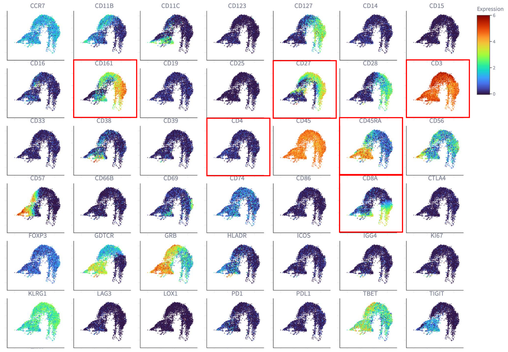
* Cosine similarity is the dot product of two vectors divided by the product of their lengths, i.e. (A * B) / (|| A || * || B ||)

CD161⁺ T Cell Subset Discovered with High-Dimensional Cytometry, FlowSOM, & UMAP

Start with ~11M cells

Cluster 10 (in red)





After clustering ~11 million cells and performing automated cluster annotation, we examined each cluster to verify that its assigned identity aligned with the expected marker expression profiles.

Cluster 10 (highlighted in red) is an example where the automated label did not align with any known T-cell type identified by manual gating. We visualized its position within the T-cell region of the UMAP, with cells from Cluster 10 colored red and all other cells in grey.

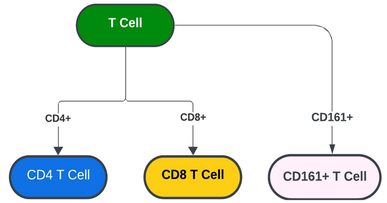
To evaluate marker expression, we generated UMAP expression plots for each measured protein marker, where blue indicates low expression and red indicates high expression.

The algorithm annotated Cluster 10 as a CD8⁺ central memory T (TCM) cluster based on partial CD8 and CD27 expression and the absence of CD45RA.

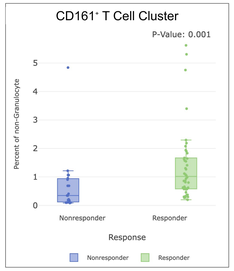
Closer inspection of CD3 and CD8 showed that only about half of the cells expressed CD8, indicating that this cluster is not a typical CD8⁺ T-cell subset. Instead, its distinct separation was driven by strong CD161 expression, which was much brighter than in other T-cell populations. The cluster also showed low CD45RA and variable CD27, consistent with a memory-like phenotype.

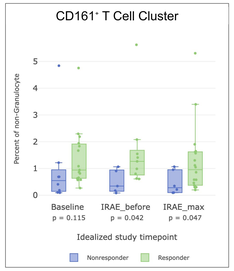
Taken together, these findings suggest that Cluster 10 represents a CD161⁺ T-cell subset that is not captured by traditional gating strategies. This population may correspond to an innate-like or MAIT-like T-cell subset with distinct functional properties relevant to immune-related inflammation.

The Frequency of the CD161⁺ T Cells Show a Significant Difference Between Responders and Nonresponders



The CD161⁺ T cell cluster (cluster 10) falls outside of the standard gating convention, where T cells are initially split into CD4 and CD8 subsets.





When comparing cluster frequencies between responders and nonresponders, Cluster 10 (CD161⁺ T cells) was significantly enriched in responders.

This difference was evident both when all timepoints were analyzed together and when comparing response within each timepoint, where the frequency of CD161⁺ T cells remained higher among responders.

This dataset is available on our dashboard at: app.teiko.bio/projects/HCI001 | hannah@teiko.bio | SMS at (910) 585-2023