

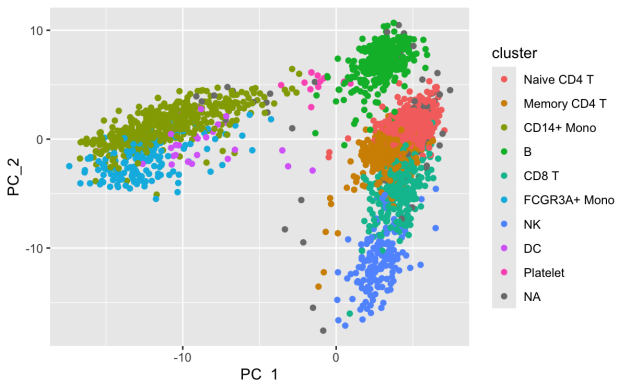
o4-mini PC Interpretations Compared to Graphical Visualization.

Arianna Aalami
Burns Life Sciences Consulting GmbH
July 24, 2025

Principal component 1

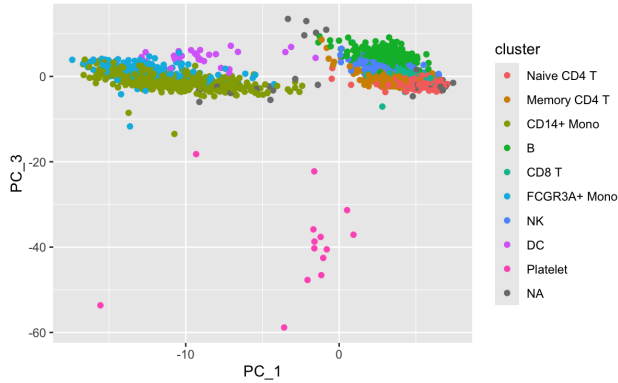
Taken together, a cell with a high PC1 score will express high levels of these phagocytic, complement, iron-handling, inflammasome and acute-phase genes—i.e. it’s a classical inflammatory monocyte/macrophage. Cells with low PC1 scores will be the lymphoid (T/B/NK) or non-monocyte compartments. So biologically PC1 = “classical monocyte / innate-immune activation” versus “rest of PBMCs

Principal component 2

PC1 vs. PC2	LLM interpretation of top 30 genes
 <p>A PCA plot showing PC1 on the x-axis (ranging from -10 to 0) and PC2 on the y-axis (ranging from -10 to 10). Data points are colored by cluster: Naive CD4 T (red), Memory CD4 T (orange), CD14+ Mono (yellow-green), B (green), CD8 T (teal), FCGR3A+ Mono (light blue), NK (blue), DC (purple), Platelet (pink), and NA (grey). The plot shows a clear separation of B cells (green) at the top right from other cell types, and a separation of NK and CD8 T cells (blue/teal) at the bottom right from the rest of the PBMCs.</p>	<p>PC2 chiefly contrasts “cytotoxic lymphocytes” vs “B-lineage/APC” cells. \n– Cells with positive loadings on PC2 are enriched for NK and CD8+ T-cell effector genes (high perforin/granzyme, chemokines). \n– Cells with negative loadings (or vice versa, depending on your sign) express B-cell markers and MHC-II. \n– In other words, PC2 is not a cell-cycle or stress axis but a lineage/functional axis separating cytotoxic (innate/adaptive cellular immunity) from humoral/APC programs.</p>
<p>Correlation? PC 2 clearly separates the B cells from the rest of the PBMCs, being farthest from the NK and CD8 T cells. This aligns well with the LLM interpretation as PC 2 separating cytotoxic cells from APC cells.</p>	

Principal component 3

PC1 vs. PC3	LLM interpretation of top 30 genes
-------------	------------------------------------



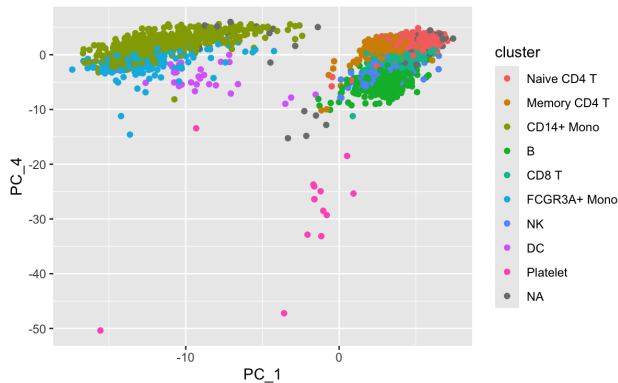
In practical terms, cells with high positive PC 3 scores are either pure platelets or cells with bound/ambient platelet fragments. Cells with strongly negative PC 3 scores are the MHC-II-high antigen presenters (B-cells, monocytes/DCs). \n\nSo PC 3 in your dataset is not a subtle metabolic or signaling axis so much as a cell-type/technical axis: it separates droplets dominated by platelet RNA from droplets dominated by APC (B-cell/DC/monocyte) RNA. If you are primarily interested in T-cell biology, you will probably want to regress out or remove the platelet signal (e.g. by filtering on PPBP or PF4 expression) or exclude doublets.

Correlation?

PC3 clearly separates out the platelet population along the y axis, which goes aligns well with the LLM’s interpretation of high PC 3 scores being pure platelets or cells with bound platelet fragments, versus all other PBMCs

Principal component 4

PC1 vs. PC4



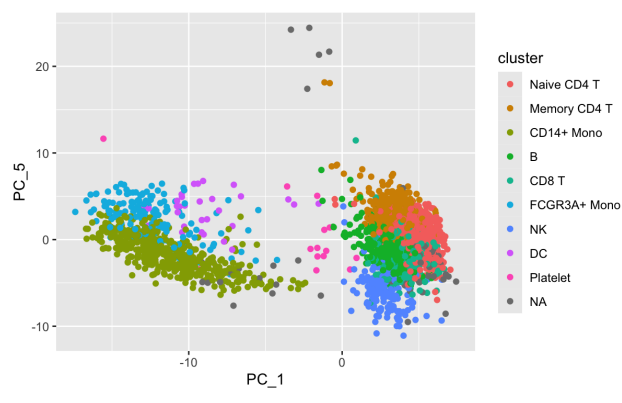
LLM interpretation of top 30 genes

In practical terms this PC may be driven by: \n – true biological variation in the B-cell ↔ platelet abundance ratio across your PBMC samples, or \n – technical “contamination” (platelets or platelet–cell doublets sticking to mononuclear cells). \n\nSo rather than reflecting, say, a T-cell/resident-memory axis or cell-cycle drive, PC4 is best understood as an antigen-presentation/B-cell vs platelet signature.

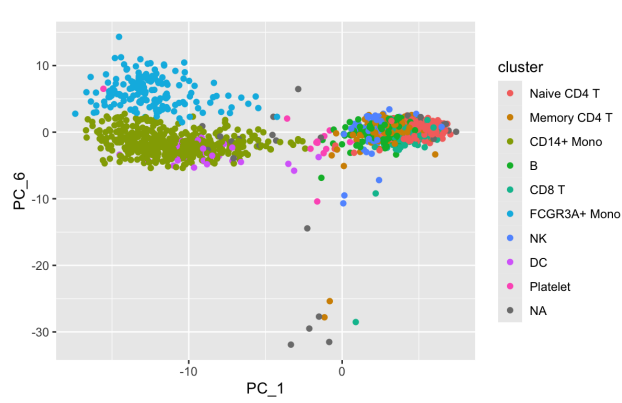
Correlation?

PC 4 only separates out the platelets across the y axis, and the LLM interpretation says that platelet signature is likely one extreme of the PC 4 scores. However, the LLM interpretation says that the other side of PC 4 would be B cell/ antigen presentation, not mentioning all of the other PBMCs that seems to be clustered at the top as well.

Principal component 5

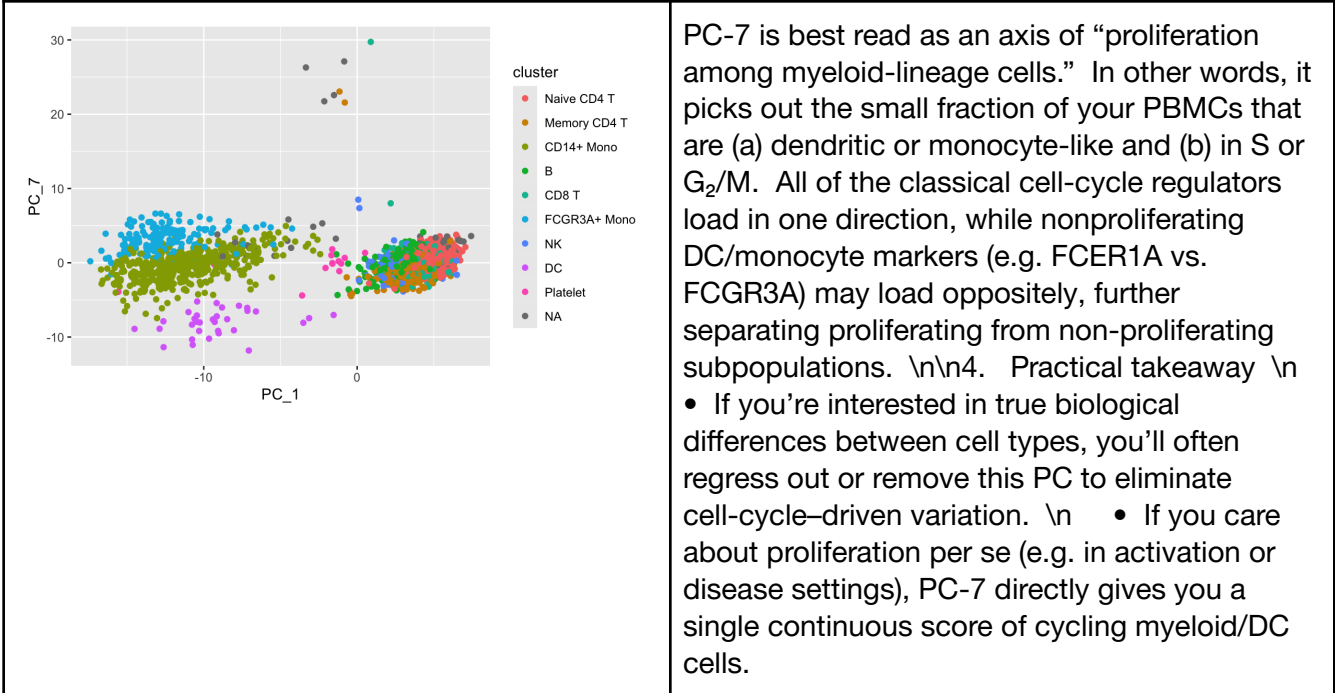
PC1 vs. PC5	LLM interpretation of top 30 genes
	<p>PC-5 is not a generic “cell-cycle” or “monocyte–B cell” axis at all, but rather an axis of cytotoxic-effector function. Cells scoring highly on PC-5 are almost certainly NK cells or CD8+ T cells that have up-regulated the perforin/granzyme machinery plus inflammatory chemokines. Low-PC-5 cells are everything else (naïve T, B, monocyte, dendritic). \n\nIn practice you can overlay PC-5 on your UMAP or t-SNE to highlight the cytotoxic-effector compartment and distinguish it from non-cytotoxic populations.</p>
<p>Correlation? The LLM response would suggest that we have a separation of cytotoxic cells such as NKs or T cells, which is not very well represented in the plot.</p>	

Principal component 6

PC1 vs. PC6	LLM interpretation of top 30 genes
	<p>Putting it all together, PC6 appears to pick out a small population of myeloid cells (likely CD16⁺ monocytes) that are actively cycling or in an “emergency” activation state. In other words, PC6 largely reflects a cell-cycle-driven axis within a monocyte-derived subset in your PBMC dataset.</p>
<p>Correlation? The only separation we can see along the y axis is amongst the monocytes, and the LLM interpretation does suggest a “cell-cycle-driven axis” within the monocyte population.</p>	

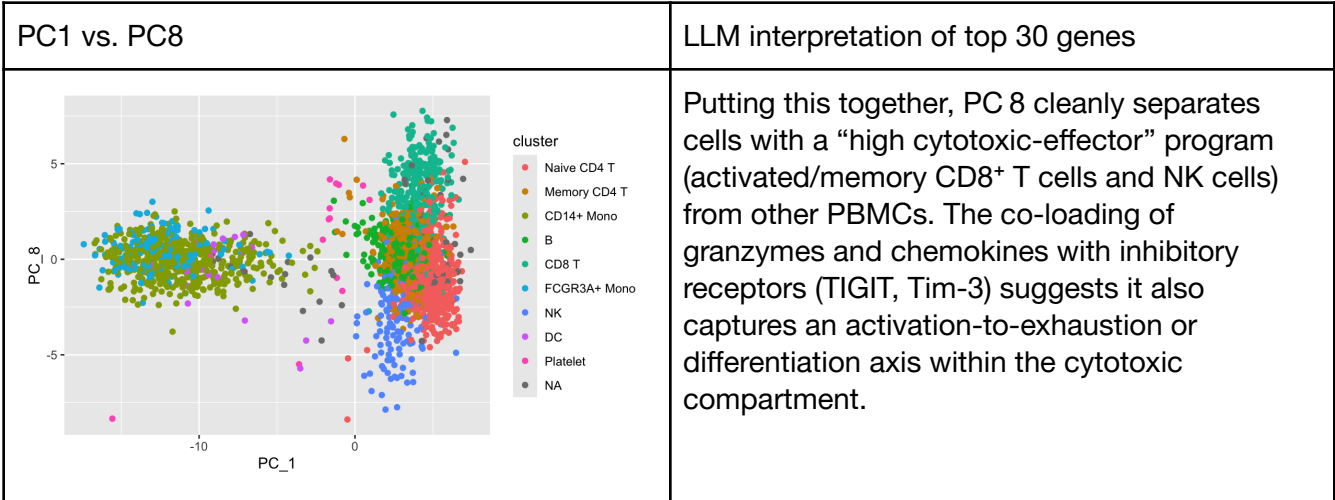
Principal component 7

PC1 vs. PC7	LLM interpretation of top 30 genes
-------------	------------------------------------



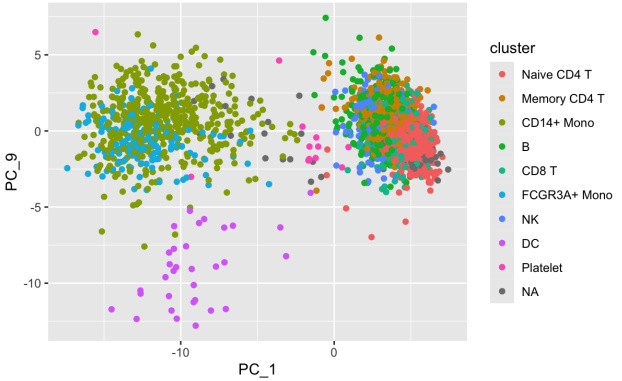
<p>Correlation?</p> <p>The most separated population in the plot are the dendritic cells, which the LLM does interpret as being separated by the other myeloid cells.</p>
--

Principal component 8

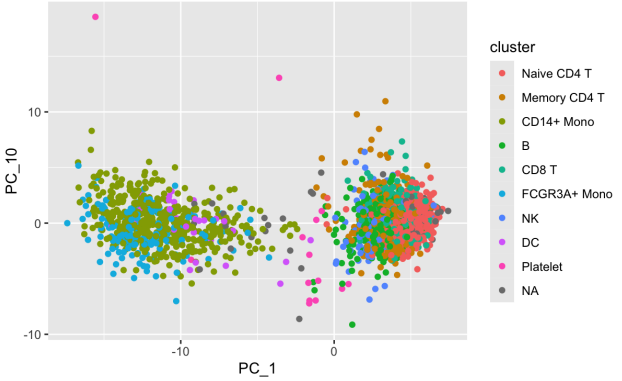


<p>Correlation?</p> <p>Any sort of separation is difficult to see in the plot, and CD8 and NK cells are on opposite sides of the</p>

Principal component 9

PC1 vs. PC9	LLM interpretation of top 30 genes
	<p>PC9 is almost certainly capturing the “who’s-a-myeloid/DC-cell vs who’s-a-B-cell” axis in your joint tSNE/UMAP. Cells with high positive PC9 scores look like inflammatory monocytes or cDC2s, whereas cells with strongly negative PC9 scores look like mature B lymphocytes.</p> <p>Why it shows up in PC9 rather than PC1–PC2</p> <p>The big splits (T cells vs NK vs monocytes vs B cells) usually dominate PC1–PC3 in PBMC. Once you’ve accounted for those, a finer distinction—“inflammatory monocyte/DC activation state” vs “mature B-cell identity”—often emerges further down, here at PC9.</p> <p>In short, PC9 reflects an antigen-presenting/inflammatory axis that discriminates myeloid-lineage cells (monocytes + cDC2s) from B-lymphoid cells.</p>
<p>Correlation?</p> <p>The dendritic cells are separated from the other cell populations, but the rest is quite muddled. The LLM response does capture the separation of DCs.</p>	

Principal component 10

PC1 vs. PC10	LLM interpretation of top 30 genes
	<p>Pathway enrichment: “Defense response to virus,” “Type-I interferon signaling,” “ISGylation,” “nucleic acid sensing.”</p> <ul style="list-style-type: none"> Cell-type/state: rather than marking a single PBMC subset, PC10 seems to capture the degree to which each cell has engaged a cell-intrinsic antiviral/IFN program—monocytes, pDCs, T/NK cells and even differentiating B/plasmablasts all boost these genes under IFN. Physiology: could reflect in vivo or ex vivo activation by type-I IFN (e.g. low-level viral exposure, TLR stimulation during processing, donor variability in IFN-α/β tone). <p>In short, PC10 is not a “cell-type” axis so much as an</p>

	“antiviral/interferon response” axis across your PBMC atlas.
Correlation? It is difficult to see any patterns of population separation in the plot. The LLM response also captures the ambiguity of PC 10 because it seems to separate cells by intrinsic properties that can be shared by many different kinds of PBMCs.	