Testing for variation between LLM interpretations of principal components 1 and 20

Arianna Aalami
Burns Life Sciences Consulting GmbH
July 24, 2025

Introduction

Below is an experiment where we inputted a list of the top 30 loadings from PC1 and PC20, and prompted a LLM (o4-mini) to output a biological interpretation. The twist is that for each PC, we ran this prompt ten times, in order to check for consistency between the outputs. The results are below.
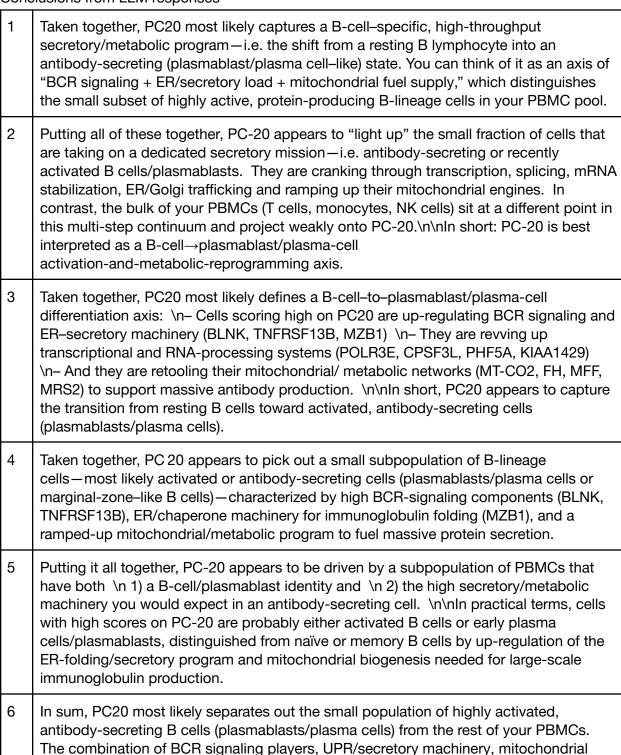
Principal component 1.

Conclusions from LLM responses

1	Taken together, cells that score high on PC-1 are classical or intermediate monocytes actively engaging their antimicrobial and antigen-presenting machinery, whereas cells with low PC-1 loadings are depleted for these myeloid programs (e.g. lymphoid cells, resting non-myeloid populations). In practical terms PC-1 is your "monocyte/innate activation" axis in PBMC.
2	Taken together, PC1 appears to track (1) the abundance of classical monocytes/macrophage-like cells in your dataset and/or (2) an activation/inflammatory program within those cells. Cells scoring high on PC1 are enriched for innate-immune effectors (phagocytosis, complement, iron metabolism, inflammasome activation), whereas cells scoring low on PC1 are likely lymphoid (T, B, NK) or less activated myeloid subsets. In short, PC1 is a "myeloid/innate immunity – monocyte activation" axis.
3	Taken together, cells scoring high on PC1 are strongly enriched for the classical/inflammatory monocyte programme—high phagocytic capacity, complement and pattern-recognition activity, inflammasome readiness, antigen processing/presentation, iron handling, etc.—whereas cells with low PC1 scores will be mainly lymphoid cells (T, B, NK). \n\nBiological summary: PC1 = "monocyte/innate immune activation" axis in your PBMC scRNA-seq.
4	Taken together these genes point to a continuum from "low PC1" cells (likely lymphoid

	or resting cells with low myeloid/innate-immune program) toward "high PC1" cells (classical CD14+ monocytes and perhaps activated/inflammatory monocyte-derived cells). Thus, PC1 can be interpreted as an axis of monocyte abundance and/or activation/inflammatory state in your PBMC dataset.
5	Taken together, PC1 is best understood as an axis of "myeloid (monocyte/macrophage) abundance and activation" versus other PBMC populations (e.g. lymphocytes). Cells with high PC1 scores express a coordinated program of phagocytosis, antigen processing, iron sequestration and pro-inflammatory effector functions, whereas cells with low PC1 scores are likely lymphoid (T, B, NK) or resting states.
6	Put simply: cells with a high PC1 score are enriched for a myeloid-innate-immune gene program (classical and inflammatory monocytes/macrophages), whereas cells with low PC1 scores are relatively depleted for those features (e.g. lymphocytes, dendritic subsets, etc.). This PC likely also captures variation in activation state within the monocyte compartment (phagocytic capacity, inflammasome activity, iron handling, S100 alarmin release), so it doubles as both a "cell type" axis and an "innate activation/inflammation" axis.
7	PC-1 is essentially "monocyte vs. non-monocyte" (or more broadly "myeloid vs. lymphoid") \n • Cells with high PC-1 scores are dominated by innate-immune/myeloid programs—especially classical CD14$^+$ monocytes \n • Cells with low PC-1 scores will be lymphoid (T, B, NK) or dendritic subsets that do not express these high-loading genes \n\nBiological interpretation: PC-1 captures the myeloid/innate-immunity axis in your PBMCs—i.e. the degree to which each cell expresses a monocyte-like, inflammatory, phagocytic program.
8	Putting it all together, PC1 appears to capture a "myeloid activation/inflammatory" axis. Cells with high PC1 scores are almost certainly monocytes (or monocyte-derived macrophages/dendritic cells) in an activated, phagocytic, acute-phase–competent state; whereas low-scoring cells would tend to be lymphoid (T, B, NK) or less-activated myeloid cells. In other words, PC1 is not a cell-cycle or generic housekeeping axis but rather a monocyte–innate-immunity/inflammation program.
9	PC1 effectively separates "innate" from "adaptive" cells in your PBMC dataset. High-PC1 cells express a coordinated program of monocyte/macrophage differentiation, phagocytic function, complement activity, iron handling, pro-inflammatory DAMPs, and in some cases antiviral effectors. In short, PC1 is an axis of myeloid lineage identity and activation (with a bias toward classical/inflammatory monocytes/neutrophils), versus lymphoid or less-activated cell types.
10	Putting it together, PC1 appears to capture the axis "monocytes/innate-immune activation" versus "other PBMC types." Cells with high PC1 scores are expressing a coordinated program of lysosomal enzymes, complement factors, scavenger receptors, inflammatory alarmins and interferon-responsive genes—hallmarks of classical/intermediate monocytes and their activated states—whereas cells with low PC1 scores are likely lymphoid (T/NK/B) or dendritic subsets

Principal component 20.

Conclusions from LLM responses

1	Taken together, PC20 most likely captures a B-cell–specific, high-throughput secretory/metabolic program—i.e. the shift from a resting B lymphocyte into an antibody-secreting (plasmablast/plasma cell–like) state. You can think of it as an axis of "BCR signaling + ER/secretory load + mitochondrial fuel supply," which distinguishes the small subset of highly active, protein-producing B-lineage cells in your PBMC pool.
2	Putting all of these together, PC-20 appears to "light up" the small fraction of cells that are taking on a dedicated secretory mission—i.e. antibody-secreting or recently activated B cells/plasmablasts. They are cranking through transcription, splicing, mRNA stabilization, ER/Golgi trafficking and ramping up their mitochondrial engines. In contrast, the bulk of your PBMCs (T cells, monocytes, NK cells) sit at a different point in this multi-step continuum and project weakly onto PC-20.\n\nIn short: PC-20 is best interpreted as a B-cell→plasmablast/plasma-cell activation-and-metabolic-reprogramming axis.
3	Taken together, PC20 most likely defines a B-cell–to–plasmablast/plasma-cell differentiation axis: \n– Cells scoring high on PC20 are up-regulating BCR signaling and ER–secretory machinery (BLNK, TNFRSF13B, MZB1) \n– They are revving up transcriptional and RNA-processing systems (POLR3E, CPSF3L, PHF5A, KIAA1429) \n– And they are retooling their mitochondrial/ metabolic networks (MT-CO2, FH, MFF, MRS2) to support massive antibody production. \n\nIn short, PC20 appears to capture the transition from resting B cells toward activated, antibody-secreting cells (plasmablasts/plasma cells).
4	Taken together, PC 20 appears to pick out a small subpopulation of B-lineage cells—most likely activated or antibody-secreting cells (plasmablasts/plasma cells or marginal-zone–like B cells)—characterized by high BCR-signaling components (BLNK, TNFRSF13B), ER/chaperone machinery for immunoglobulin folding (MZB1), and a ramped-up mitochondrial/metabolic program to fuel massive protein secretion.
5	Putting it all together, PC-20 appears to be driven by a subpopulation of PBMCs that have both \n 1) a B-cell/plasmablast identity and \n 2) the high secretory/metabolic machinery you would expect in an antibody-secreting cell. \n\nIn practical terms, cells with high scores on PC-20 are probably either activated B cells or early plasma cells/plasmablasts, distinguished from naïve or memory B cells by up-regulation of the ER-folding/secretory program and mitochondrial biogenesis needed for large-scale immunoglobulin production.
6	In sum, PC20 most likely separates out the small population of highly activated, antibody-secreting B cells (plasmablasts/plasma cells) from the rest of your PBMCs. The combination of BCR signaling players, UPR/secretory machinery, mitochondrial

	genes and interferon-stimulated factors fits exactly with what you'd expect in B cells that have received strong activation signals and are transitioning into high-rate immunoglobulin factories.
7	Taken together, PC20 is dominated by a handful of B-cell/plasma-cell markers plus genes required for massive secretory and metabolic reprogramming. In a PBMC dataset this usually points to the presence—and variation in abundance or activation state—of plasmablasts (or highly activated, interferon-responsive B cells). In other words, PC20 is not a generic "housekeeping" axis but rather a rare-cell signal driven by B-lineage cells gearing up for antibody secretion.
8	Taken together, PC20 can be interpreted as a spectrum from resting B cells (low scores) to activated/plasmablast-like B cells (high scores)—cells that have switched on the B-cell receptor/TACI axis, mounted an interferon/antiviral response, and dramatically ramped up their mitochondrial, secretory and RNA-processing machinery to support antibody production.
9	It appears to distinguish a B-cell subset (likely plasmablasts or highly activated/memory B cells) that has up-regulated its protein-secretory machinery (MZB1, chaperones) and mitochondrial metabolic capacity (MT-CO2, FH, MFF) to meet the bioenergetic cost of proliferation and antibody production. \n– It also carries an interferon-response signature (OAS1) and a general ramp-up of transcription/splicing factors needed for high transcript throughput. \n\nIn a PBMC atlas, you'd expect this PC to light up on plasmablasts or late-stage B-cell effectors rather than naïve or resting T cells, monocytes, etc.
10	Putting it all together, PC20 appears to pick out a small subpopulation of highly activated, antibody-secreting B cells or plasmablasts that are (a) engaged in intense protein/RNA processing and chromatin remodeling, (b) ramping up mitochondrial metabolism, and (c) carrying an interferon/inflammatory signature. In other words, PC20 is not a "generic" cell-cycle axis but rather marks a differentiation- and activation-specific program in the B-cell lineage.