## **scRNA-seq Cluster Snapshot (Markers: CD3D, MS4A1, LYZ, MKI67)**

You have 12 single cells, each quantified for 4 marker genes. Cells are manually labeled into Tcell, Bcell, and Myeloid clusters. You will verify that marker trends match cluster labels.

Questions:

1. str(cells), summary() for numeric columns.
2. Compute per-cell total expression across the 4 genes (apply), store as total.
3. Label cells as "Active" if total > median total, else "Resting"; add to cells.
4. Using tapply(), compute cluster-wise mean for each gene. Which cluster is CD3D-high?
5. Which single cell has the highest MKI67 (proliferation marker)? Return cell\_id and value.
6. For each cell, compute max gene (the marker with the highest value). Produce a vector of marker names per cell.
7. Count how many cells per cluster and how many Active vs Resting (table).
8. Create a matrix expr\_mat (12×4) from the gene columns only (rownames = cell\_id). Confirm identical (rownames(expr\_mat), cells$cell\_id).
9. Compute column means on expr\_mat; which marker has the highest overall mean?
10. Compute row ranges (max–min) , which cell is most uneven across markers?
11. Subset cells to only Tcell and recompute mean of CD3D; compare to overall mean.
12. Rename column MS4A1 to MS4A1\_B and verify.
13. Reorder cells by descending MKI67 without using order(-x) (hint: get the descending index manually).
14. Create a list call scRNA containing cells,expr\_mat and markers(CD3D, MS4A1\_B, LYZ, MKI67)).
15. From scRNA, extract the LYZ values for all Myeloid cells (practice nested subsetting with a logical mask).
16. Create a simple barplot of cluster-wise means for MKI67
17. Use sapply() on gene columns to return standard deviations per gene.
18. Which cluster is most active on average?
19. Zero all expression values < 1 (simulate thresholding) on a copy of expr\_mat; report how many entries changed
20. Write a concise interpretation: do marker patterns support cluster labels?

## Note: Here is the Grading rubric that will be used to grade your submission:

**Correctness (40%)**: Each question answered accurately with working base R code.

**Code Quality (20%)**: Clear variable names, consistent style, comments explaining *why*.

**Reasoning (20%)**: Report explains what the numbers mean (not just outputs).

**Reproducibility (20%)**: Clean script, you can use Rmd if you want