

RECORDED MODEL BEHAVIOR

Assess the cell-type information carried by non-promoter transcription signatures

6/15 criteria passed

GPT-6-ASTRA · FRESH SOLVER AND GRADER

The selected publication contains a completed fresh solve and grade.

FAILED CRITERIA IN THE SAVED GRADE

C-006 - Bound the NFkB interpretation to the shown immune group: The report correctly maps NFkB1/2-REL/RELB to T/B-Cells and cautions against treating white cells as depletion or universal absence. However, it never compares those same columns with two named non-T/B groups, as required to demonstrate the restricted pattern.

C-007 - Explain why widespread activity is weak identity evidence: The report describes the JUN/FOS block as broadly shared, but does not compare it in two named, separately branching groups. It also omits the required explanation that equal encoded states contribute zero to Euclidean distance.

C-008 - Interpret a multi-factor signature rather than a single decisive marker: The bounded conclusion is correct, but the table only lists group-to-regulator annotations. It does not check each of two restricted regions in both named groups or compare their shared JUN/FOS activity. These required comparisons cannot be supplied by the general statement that signatures contain multiple features.

C-009 - Interpret the clustering input as activity calls: The report correctly states the 1/0/-1 encoding, Ward clustering, and Euclidean distance, and distinguishes inferred activity from TF abundance. It omits the required one-feature calculation: differences of 2 versus 1 and squared-distance contributions of 4 versus 1.

C-010 - Separate grouping of known samples from prediction of new ones: The report names mixed-study successes and says prediction on new samples and unrepresented cell types remains untested. However, it does not explain that both named studies helped construct the joint clustering and that neither was assigned using a fixed rule built without it. The criterion explicitly requires this explanation beyond noting missing validation.

C-011 - Use unfamiliar regulators as candidates, not proved causes: The report lists HeLa and K562 regulator sets and correctly warns that inferred activity does not prove causation. It also discusses K562's single-study limitation elsewhere. But it never compares the study support for these named regulator associations, which is required beyond a factor list and a general causal caveat.

C-013 - Keep the reported sample counts separate from plot membership: The report correctly states eight source cell types, 18 labs, and 351 factors, and recognizes composite labels. It omits the 39 labeled rows and eight displayed group count and does not explain that a plotted row is not a lab. The required distinction among these units is therefore incomplete.

C-014 - Keep the evidence about non-promoter sites separate from a promoter comparison: The required promoter-comparison assessment is absent. The report does not explain that the non-promoter grouping cannot establish superiority over promoter signatures because no matching promoter result is supplied.

C-015 - Limit regulator coverage to factors with a usable motif model: The report states the high-quality PWM requirement and the 351-factor display. It does not explain that this selected set is not a complete inventory of regulators or that a regulator outside the assessed set cannot be declared inactive. Mentioning motif coverage as a transfer assumption does not supply those conclusions.

Read runs/gpt-6-astra/evaluation/outputs/submission/, outputs/judgments.json, outputs/report.md and attempts/0001/ for the unedited answer, reasons and events. These are recorded grader judgments, not a new adjudication.