**Protocol (QC → Normalization → Batch handling → Validation → DE-ready)**

1. **Assemble inputs**

* Raw gene-level counts matrix (HTSeq/STAR; genes × samples).
* Metadata with: sample\_id, cohort, batch, purity, RIN, necrosis\_frac, age, sex, IDH\_status, MGMT\_status (+ optional alignment metrics like mapped\_rate, ribo\_frac, total\_reads, genes\_detected).

1. **Sample QC (hard gates + adaptive)**

* Hard gates (apply when available): purity ≥ 0.80, RIN ≥ 6.8, necrosis ≤ 0.50.
* Adaptive outliers: 5×MAD on log10(library size) and log10(genes detected); also MAD on mapped\_rate and ribo\_frac if present.
* Save a QC table and keep only qc\_pass samples.

1. **Gene filtering**

* Drop trivially low genes (e.g., total counts < 10).
* Use edgeR::filterByExpr() (respects group sizes) for robust prevalence filtering.

1. **Normalization artifacts for different tasks**

* **DE-ready counts**: keep raw counts; model batch as covariate in DESeq2/edgeR.
* **Log-normalized** (log2(counts/size\_factor + 1)) for correlations/plots.
* **VST** (DESeq2) for PCA/RLE and general visualization.

1. **Batch effects**

* Don’t pre-correct for DE; include ~ batch + condition in the model.
* For visualization only: optional ComBat-seq on counts, then VST → PCA (expect batch PCs to drop).
* Re-validate that biology > batch in PCA.

1. **Validation checks**

* PCA colored by batch and biological covariates (IDH/MGMT/subtype).
* RLE plot: medians near 0, similar spread.
* Sample–sample correlations within groups high and not batch-driven.

1. **Differential expression (template)**

* Use DESeq2 with design = ~ batch + condition (or add covariates).
* Shrink LFC (e.g., type="ashr") and export full results.

**What I prepared for you (download & run)**

* **Bulk RNA-seq pipeline** (end-to-end):

[Download bulk\_rnaseq\_qc\_pipeline.R](sandbox://mnt/data/bulk_rnaseq_qc_pipeline.R?_chatgptios_conversationID=68de9ce6-5958-8324-ab0a-af71b7cdb39b&_chatgptios_messageID=dca9537a-e266-4804-a781-ad5f872ae628)

* **Optional scRNA-seq QC + Harmony integration** (if you need single-cell):

[Download scRNAseq\_qc\_integration.R](sandbox://mnt/data/scRNAseq_qc_integration.R?_chatgptios_conversationID=68de9ce6-5958-8324-ab0a-af71b7cdb39b&_chatgptios_messageID=dca9537a-e266-4804-a781-ad5f872ae628)

* **README quick start**:

[Download README\_DGAT\_QC.txt](sandbox://mnt/data/README_DGAT_QC.txt?_chatgptios_conversationID=68de9ce6-5958-8324-ab0a-af71b7cdb39b&_chatgptios_messageID=dca9537a-e266-4804-a781-ad5f872ae628)