

Bioinformatics Analysis Reproducibility Checklist

A practical checklist for organizing data, metadata, software, code and reporting so an analysis can be repeated.

Project reproducibility checklist

Area	Checklist
Data provenance	Record repository, accession, download date and original filenames.
Metadata	Keep a machine-readable sample sheet and define all groups/covariates.
Reference files	Record genome build, annotation version and source URLs.
Software	Record tool names, versions and key parameters.
Environment	Save conda/environment files, container tags or session information.
Code	Keep scripts under version control and avoid manual undocumented edits.
Randomness	Set and record seeds for stochastic analyses.
Intermediate files	Document which files can be regenerated and which must be archived.
Quality control	Save QC reports and explain filtering decisions.
Statistics	Document model design, contrasts, thresholds and multiple-testing methods.
Figures	Generate figures from scripts, not manual spreadsheet edits.
Logs	Capture command logs, software messages and failed steps when useful.
README	Describe project structure and the order needed to reproduce the workflow.
Sharing	Check licenses, consent and controlled-access restrictions before sharing data.

Goal: A collaborator should be able to understand where the data came from, how it was processed, which software was used and how the main results were generated.