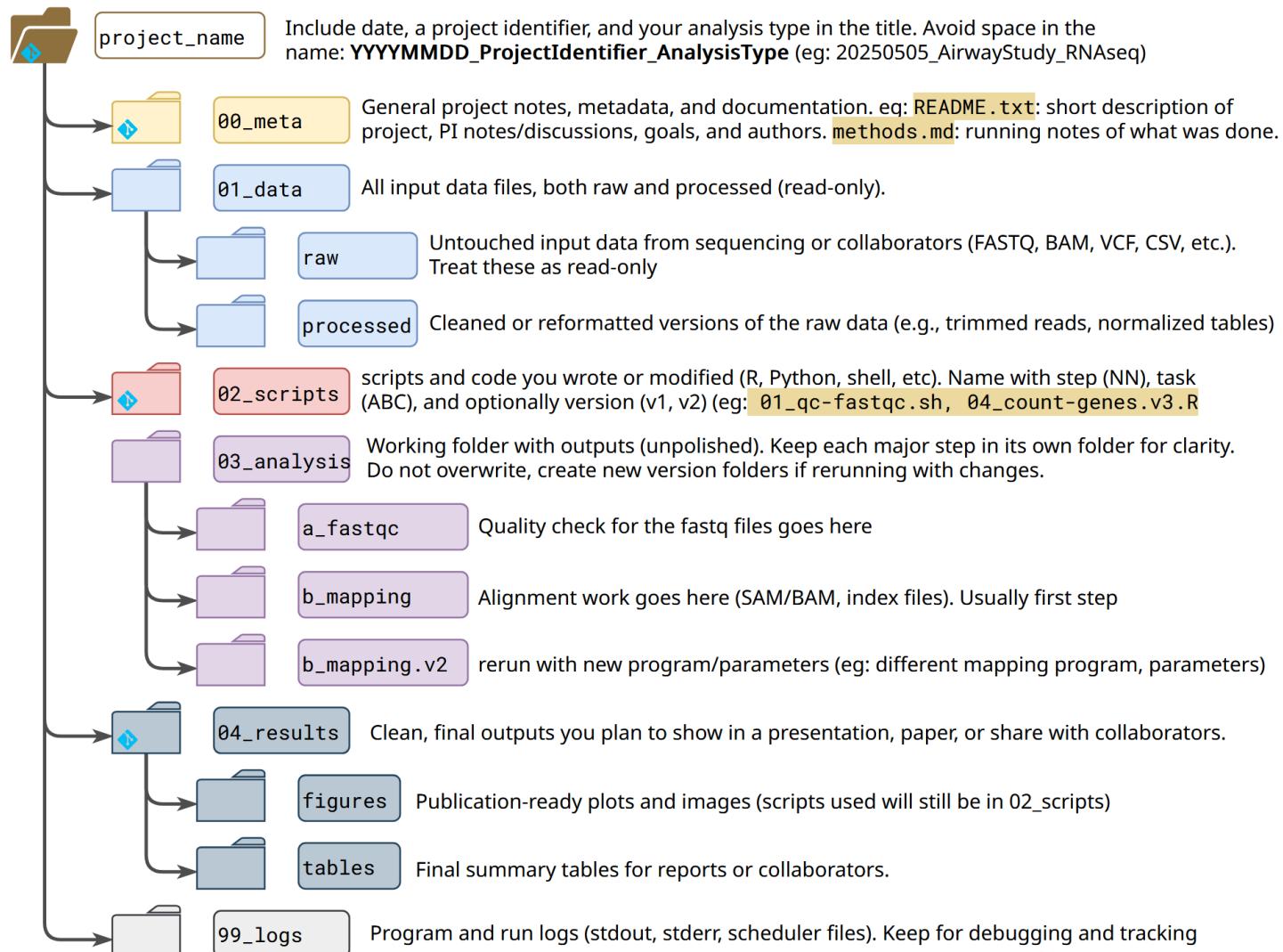


A simple folder structure for bioinformatics projects



Git* commands for revision control

Command	Purpose	Example
<code>git init</code>	Start tracking a project (run once inside project_name/)	<code>git init</code>
<code>git add <file></code>	Stage files to be tracked (use <code>.</code> for all)	<code>git add 02_scripts/</code>
<code>git commit -m "message"</code>	Save a snapshot with a short message	<code>git commit -m "added QC script"</code>
<code>git remote add origin <url></code>	Connect local project to a remote repository (GitHub, GitLab, etc.)	<code>git remote add origin https://github.com/user/project.git</code>
<code>git push origin main</code>	Upload your commits to the remote	<code>git push origin main</code>
<code>git pull origin main</code>	Bring down the latest version from the remote	<code>git pull origin main</code>

*Use Git only for small text files (scripts, notes, configs). Do not put raw data, results, or large files under version control.

Getting started workflow (step-by-step example)

1. Start a project

- Make the folder: `mkdir 20250916_AirwayStudy_RNAseq`
- Add subfolders: `mkdir 00_meta 01_data/raw 01_data/processed 02_scripts 03_analysis 04_results 99_logs`

2. Add notes

- Create `00_meta/README.txt`:

Project: Airway Study Demo
 Date: 2025-09-16
 PI: Dr. Example
 Goal: Demonstration of project organization

3. Save your first script

- Put into `02_scripts/01_qc-fastqc.sh`
- Make it executable: `chmod +x 01_qc-fastqc.sh`

4. Run and capture outputs

- Save QC reports in `03_analysis/a_fastqc`
- Move polished plots into `04_results/figures`

5. Back up with Git

- `git add 00_meta/ 02_scripts/`
- `git commit -m "Initial setup + QC script"`

Best practices

Do's	Don'ts
Keep raw data read-only.	Don't rename or overwrite raw files.
Number scripts logically (01, 02, ...).	Don't mix results into the scripts/ folder.
Commit scripts + notes regularly.	Don't commit large FASTQ/BAM files to Git.
Always document what changed.	Don't keep everything in your home directory.
Use relative paths in your scripts.	Don't use spaces in file or folder names.
Log computational outputs and errors.	Don't manually edit intermediate results.
Have a backup copy of your raw data	Don't/minimize hardcoding parameters in scripts.

Keep your raw data safe, keep your scripts under version control, and keep your results organized, so anyone (including future you) can understand and reproduce your work.