



Institute of Primate Research

STANDARD OPERATING PROCEDURE (SOP) DOCUMENT

**Bioinformatics pipelines (from raw sequence data to
analysis)**

SOP No.	Issue Number	Issue Date	Revision Status	Revision Date
SOP/KIPRE/RPD/DSAS/3.1.76	Version 01	October 2025	-	-

Approvals

	Name	Signature	Date
Developed by:	<u>Patrick Waweru Mwaura</u>	<u></u>	<u>6th October; 2025</u>
	<u></u>	<u></u>	<u></u>
	<u></u>	<u></u>	<u></u>
Reviewed by:	<u></u>	<u></u>	<u></u>
Approved by:	<u></u>	<u></u>	<u></u>

Table of Contents

1. PURPOSE.....	4
2. SCOPE	4
3. PERSONS RESPONSIBLE:	4
4. FREQUENCY.....	4
5. MATERIALS.....	4
6. PROCEDURE.....	4
7. REFERENCES	4

1. PURPOSE

To define standardized workflows for building and running bioinformatics pipelines that ensure reproducibility, accuracy, and efficiency.

2. SCOPE

Covers raw sequence data (DNA/RNA) through processing, alignment, annotation, and downstream analysis.

3. PERSONS RESPONSIBLE:

- **Bioinformatician:** Designs and runs pipelines.
- **Computational Biologist:** Interprets results.
- **Head of DS&AS:** Approves workflows before deployment.

4. FREQUENCY

- Pipeline validation prior to first use.
- Updates whenever new tools, methods, or reference builds are introduced.

5. MATERIALS

- Workflow management systems (Nextflow, Snakemake, Galaxy).
- Analysis tools (BWA, GATK, DESeq2, BLAST).
- HPC/Cloud infrastructure.
- Version control (Git).

6. PROCEDURE

1. **Pipeline Design:** Define steps (QC, alignment, variant calling, annotation).
2. **Implementation:** Build reproducible workflows using workflow managers.
3. **Testing:** Validate pipeline on benchmark datasets.
4. **Execution:** Run pipelines on HPC/Cloud, log outputs systematically.
5. **Version Control:** Maintain pipeline scripts in Git repositories.
6. **Documentation:** Archive pipeline details, parameter settings, and results in DS&AS repository.

7. REFERENCES