

**Institute of Primate Research**

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

**STANDARD OPERATING PROCEDURE (SOP) DOCUMENT**

## **Bioinformatics Pipelines (From Raw Sequence Data to Analysis)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Approvals** |  |  |  |
|  | **Name** | **Signature** | **Date** |
| **Developed by:** | \_Patrick Waweru Mwaura\_ | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_6th October; 2025\_** |
|  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
|  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **Reviewed by:** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **Approved by:** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

**Table of Contents**

[1. PURPOSE 4](#_Toc144316958)

[2. SCOPE 4](#_Toc144316959)

[3. PERSONS RESPONSIBLE: 4](#_Toc144316960)

[4. FREQUENCY 4](#_Toc144316961)

[5. MATERIALS 4](#_Toc144316962)

[6. PROCEDURE 4](#_Toc144316963)

[7. REFERENCES 7](#_Toc144316964)

# 

# PURPOSE

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

# SCOPE

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

# PERSONS RESPONSIBLE:

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

# FREQUENCY

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

# MATERIALS

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

# 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.

# REFERENCES