**Standard Operating Procedure (SOP) for pre-processing Single-Cell RNA-seq Data Analysis**

**1. Downloading Instrument Data and Verifying Integrity**

First, download the sequencing data from the specified source. Replace your username and your password with your actual credentials and adjust the URL to your data source:

wget --user=your username --password=your password --recursive --no-parent --continue --reject='index.html\*' -e robots=off <https://your_data_source_url/>

Verify the integrity of the downloaded files using MD5 checksums:

md5sum -c \*.md5

**2. Retrieving Reference Sequences**

Download the reference genome and annotations for your organism of interest from a public database. Here, Ascaris suum is used as an example:

wget ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/current\_release/species/ascaris\_suum/your\_project\_id/ascaris\_suum.your\_project\_id.current\_release.genomic.fa.gz

wget <ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/current_release/species/ascaris_suum/your_project_id/ascaris_suum.your_project_id.current_release.annotations.gff3.gz>

Decompress the downloaded files:

gunzip ascaris\_suum.your\_project\_id.current\_release.genomic.fa.gz

gunzip ascaris\_suum.your\_project\_id.current\_release.annotations.gff3.gz

Convert GFF3 to GTF format using gffread:

gffread ascaris\_suum.your\_project\_id.current\_release.annotations.gff3 -T -o ascaris\_suum.your\_project\_id.current\_release.annotations.gtf

**3. Preparing the Cell Ranger Reference**

Filter the GTF file for specific gene biotypes and prepare the Cell Ranger reference:

cellranger mkgtf ascaris\_suum.your\_project\_id.current\_release.annotations.gtf filtered\_annotations.gtf --attribute=gene\_biotype:protein\_coding

Create a Cell Ranger reference:

cellranger mkref --genome=GenericGenomeName --fasta=ascaris\_suum.your\_project\_id.current\_release.genomic.fa --genes=filtered\_annotations.gtf --ref-version=your version

**4. Running Cell Ranger Count**

Run cellranger count for each sample. Adjust the path to your FASTQ files and sample names as necessary:

cellranger count --id=SampleID\_count --fastqs=path\_to\_fastqs --sample=SampleName --localcores=24 --localmem=128 --transcriptome=path\_to\_reference

Repeat for each sample, modifying the --id and --sample parameters accordingly.

**5. Analyzing Data with Seurat**

Following Cell Ranger processing, analyze the single-cell data using Seurat. Ensure Seurat and its dependencies are installed in R:

# Example code to start a Seurat analysis

library(Seurat)

seurat\_object <- CreateSeuratObject(counts = your\_data\_matrix)

# Save RDS file of your data (save per sample, or in aggregate)

saveRDS(seurat\_object, file = “Seurat\_object.rds”)

# Continue with your Seurat analysis workflow

This SOP describes how to take single-cell sequencing data and run it through CellRanger to prepare it for analysis using Seurat. Ensure to replace placeholder values with actual data pertinent to your analysis, such as project IDs, genome versions, sample names, and file paths. Adjustments may be necessary based on specific software updates or database changes.