FastQC on sequence datasets for Genome Assembly

```
#!/bin/bash
#SBATCH --job-name=fastqc_all
#SBATCH --partition=msc_appbio
#SBATCH --cpus-per-task=4
#SBATCH --mem=12G
#SBATCH --time=02:00:00
#SBATCH --output=fastqc_all_%j.out

module load fastqc

# This script performs quality control for both yeast strains (BY4742 and SY14) using FastQC.
# The results are used to assess read quality prior to downstream
# analyses such as polishing, mapping, and variant calling.

# Define project directories
BASE_DIR=/scratch/grp/msc_appbio/Group2_ABCC/Gene_Assembling
DATA_DIR=${BASE_DIR}/data
QC_DIR=${BASE_DIR}/qc_reports

# Create output directories for FastQC reports
# (separate folders are used for each strain for clarity)
mkdir -p ${QC_DIR}/BY4742_fastqc
mkdir -p ${QC_DIR}/SY14_fastqc

# Run FastQC on Illumina reads from the BY4742 strain
# These reads are later used for genome polishing and validation
fastqc ${DATA_DIR}/BY4742/illumina/*.fastq.gz \
    -o ${QC_DIR}/BY4742_fastqc \
    -t 4

# Run FastQC on Illumina reads from the SY14 strain
# These reads support polishing and downstream comparative analyses
fastqc ${DATA_DIR}/SY14/illumina/*.fastq.gz \
    -o ${QC_DIR}/SY14_fastqc \
    -t 4

echo "FastQC analysis completed for both strains!"
```