Kallisto on HPC

What is Kallisto?

Kallisto is a program for quantifying the abundance of transcripts from RNA-Seq data. It is a very fast and accurate method for estimating transcript abundance, and it is particularly well-suited for quantifying abundances of transcripts from large genomes, such as the human genome. The basic idea behind Kallisto is to use a reference transcriptome and a set of RNA-Seq reads to estimate the abundance of transcripts in the sample. Kallisto uses a novel approach called "pseudoalignment" to align the reads to the reference transcriptome, which allows it to perform the alignment very quickly without the need for a computationally intensive alignment step. This makes Kallisto a useful tool for researchers who need to perform RNA-Seq analysis on large datasets or on a high-throughput basis.

Links:

Official Website

Manual

Versions Available:

The following versions are available on the cluster:

- kallisto 0.43.1

How to load Kallisto?

To load Kallisto, use the following commands:

```
#Load the KALLISTO module
module load bio/kallisto/0.43.1
```

To verify if the module and dependencies are loaded correctly, use the following command.

```
#Show all the modules loaded
module list
```

This should list all the KALLISTO dependencies that are loaded- HDF5 and GCC.

How to use Kallisto?

To demonstrate the usage of this program, download the test files using this command,

```
# Get test files
svn checkout https://github.com/pachterlab/kallisto/trunk/test/

# Go to the downloaded directory
cd ./test
# Make a script
touch slurm.script
```

Use the following threaded slurm script template for kallisto ,

```
#!/bin/bash
#SBATCH --job-name=kallisto_job
#SBATCH --output=kallisto_job.out
#SBATCH --error=kallisto_job.err
#SBATCH -p main
#SBATCH --qos main
#SBATCH --mem=10GB
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=10

# Load Kallisto module
module load bio/kallisto/0.43.1
```

```
# Build a index
kallisto index -i transcripts.idx transcripts.fasta.gz

# Single end reads
kallisto quant --threads $SLURM_CPUS_PER_TASK -i transcripts.idx -o
output -b 100 --single -l 180 -s 20 reads_1.fastq.gz
```

The output should be in the same directory as the script.

Where to find help?

If you are stuck on some part or need help at any point, please contact OIT at the following address.

https://ua-app01.ua.edu/researchComputingPortal/public/oitHelp