Supernova on HPC

What is Supernova?

Supernova is a software tool developed by 10x Genomics for de novo genome assembly of complex organisms. It uses a novel approach called linked-read sequencing, which involves barcoding and partitioning of large DNA molecules into smaller fragments to generate linked-reads. These linked-reads enable the reconstruction of long-range information that is critical for the assembly of complex genomes. The Supernova pipeline consists of several modules, including read preprocessing, barcode demultiplexing, read clustering, and assembly. The output of the pipeline is a high-quality genome assembly in the form of contiguous sequences (contigs) and scaffolds. Supernova has been used for the assembly of various complex genomes, including the human genome, and is known for its high accuracy, completeness, and scalability.

Links:

[Official Website and Manual]

Versions Available:

The following versions are available on the cluster:

- `Supernova 2.0.0`

How to load Supernova?

To load Supernova, use the following commands:

```
#Load the Supernova module
module load bio/supernova/2.0.0
```

To verify if the module is loaded correctly, use the following command,

```
# List all the module loaded in the environment
module list
```

How to use Supernova?

To use Supernova for de novo genome assembly, the first step is to prepare the input DNA samples using the 10x Genomics Chromium Controller to generate linked-read sequencing libraries. The user can then submit the raw sequencing data to the Supernova pipeline for analysis. The Supernova pipeline consists of several modules, including read preprocessing, barcode demultiplexing, read clustering, and assembly. The user can specify the desired output genome assembly parameters, such as minimum contig length or N50 value, and the pipeline will generate a high-quality genome assembly in the form of contiguous sequences (contigs) and scaffolds

Examples are available in the following websites:

Supernova Example

Here is a sample demonstration of generating FASTQs with supernova,

```
# Download the input files in a convenient location
wget https://cf.10xgenomics.com/supp/assembly/tiny-bcl-2.0.0.tar.gz
wget https://cf.10xgenomics.com/supp/assembly/tiny-bcl-simple-
2.1.0.csv
wget https://cf.10xgenomics.com/supp/assembly/tiny-bcl-samplesheet-
2.1.0.csv

#Unzip files
tar -xvf tiny-bcl-2.0.0.tar.gz
```

Here is a sample slurm script to generate run the supernova on the input file,

```
#!/bin/bash
#SBATCH --job-name=supernova_mkfastq
#SBATCH -p main
#SBATCH -qos main
#SBATCH --output=supernova_mkfastq.log
#SBATCH --time=12:00:00
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --cpus-per-task=32
#SBATCH --mem=32G

# Load Supernova module
module load bio/supernova/2.0.0

# Set paths
input_dir=/path/to/extracted_tiny_bcl_directory
output_dir=/path/to/output/

# Run mkfastq
supernova mkfastq --id=tiny-blc\
                  --run=${input_dir} \
                  --reads-per-fastq=40000000 \
                  --localcores=$SLURM_CPUS_PER_TASK \
                  --output-dir=${output_dir}
```

Where to find help?

If you are confused or need help at any point, please contact OIT at the following address.

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