

1. Basics of kallisto single cell — introduction

Overview of kallisto single cell: <https://pachterlab.github.io/kallisto/singlecell.html>

Quick start of kallisto: <https://pachterlab.github.io/kallisto/starting>

Presentation materials: <https://github.com/caltech-bioinformatics-resource-center/bioinformatics-symposium-2019>

2. Hands-on exercise

2.1 Software installation and configuration

This exercise covers installation of **kallisto**, **bustools**, **Jupyter Notebook**, **R-studio**, **Scanpy** and **Seurat** software tools and download of demo datasets.

Bioinformatics server users, please open a terminal window (install putty on Windows <https://www.putty.org>)

```
ssh user@bioinformatics.caltech.edu
```

Mac and Linux laptop users, please open a terminal window.

2.1.1 Installation of **kallisto**

Visit <http://pachterlab.github.io/kallisto/download> from any web browser

Download **Mac** or **Linux** executables from the version: **v0.45.0**

Mac users

```
cd /home/user  
wget https://github.com/pachterlab/kallisto/releases/download/v0.45.0/kallisto_mac-v0.45.0.tar.gz  
tar -xvzf kallisto_mac-v0.45.0.tar.gz
```

Linux (server and laptop) users

```
cd /home/user  
wget https://github.com/pachterlab/kallisto/releases/download/v0.45.0/kallisto_linux-v0.45.0.tar.gz  
tar -xvzf kallisto_linux-v0.45.0.tar.gz
```

CBRC 2019 Bioinformatics Symposium
kallisto single-cell RNA-Seq: how to run it

Test executable **kallisto**

```
export PATH="/home/user/kallisto_linux-v0.45.0:$PATH"  
kallisto bus
```

2.1.2 Installation of **bustools**

```
git clone https://github.com/BUSTools/bustools.git  
cd bustools  
mkdir build  
cd build  
cmake ..  
make
```

Test executable bustools

```
cd /home/user  
export PATH="/home/user/bustools/build/src:$PATH"  
bustools
```

2.1.3 Installation of **Anaconda** (<https://repo.continuum.io/archive/>)

We have tested **Anaconda3-5.2.0-Linux-x86_64.sh** (Linux)

Bioinformatics Server Users:

```
export PATH="/home/software/anaconda3.5.3/bin:$PATH"
```

Linux laptop users:

```
cd /home/user  
wget https://repo.continuum.io/archive/Anaconda3-5.3.0-Linux-x86.sh  
bash Anaconda3-5.3.0-Linux-x86.sh
```

2.1.4 Installation of **R** / **R-studio** (not necessary for Bioinformatics Server Users)

R (<https://www.r-project.org/>) R-studio (<https://www.rstudio.com/>)

2.1.5 Installation of **Scanpy** (not necessary for Bioinformatics Server Users)

Scanpy <https://scanpy.readthedocs.io/en/latest/installation.html>

2.1.6 Installation of **Seurat** (not necessary for Bioinformatics Server Users)

Seurat <https://satijalab.org/seurat/install.html>

CBRC 2019 Bioinformatics Symposium
kallisto single-cell RNA-Seq: how to run it

2.2 Bash command lines for data processing (let's try a small PBMC data with 400k reads)

2.2.1 Download demo data

```
cd /home/user  
wget http://bioinformatics.caltech.edu/events/pbmc4k_S1_L001_R1_001_400k.fastq.gz  
wget http://bioinformatics.caltech.edu/events/pbmc4k_S1_L001_R2_001_400k.fastq.gz
```

2.2.2 Download GRCh38 transcript annotation file

```
wget ftp://ftp.ensembl.org/pub/current_fasta/homo_sapiens/cdna/Homo_sapiens.GRCh38.cdna.all.fa.gz
```

2.2.3 Generate kallisto index file

```
kallisto index -i Homo_sapiens.GRCh38.cdna.all.kidx  
Homo_sapiens.GRCh38.cdna.all.fa.gz
```

2.2.4 Run kallisto bus

```
kallisto bus -i Homo_sapiens.GRCh38.cdna.all.kidx -x 10Xv2 -t 2 -o kallisto_out  
pbmc4k_S1_L001_R1_001_400k.fastq.gz pbmc4k_S1_L001_R2_001_400k.fastq.gz
```

2.2.5 Run bustools to sort output binary file

```
bustools sort -t 2 -o kallisto_out/output.sort.bus kallisto_out/output.bus
```

2.2.6 Run bustools to convert sorted binary file to human readable text file

```
bustools text -o kallisto_out/output.sort.txt kallisto_out/output.sort.bus
```

Please feel free to ask the tutors if you need troubleshooting!