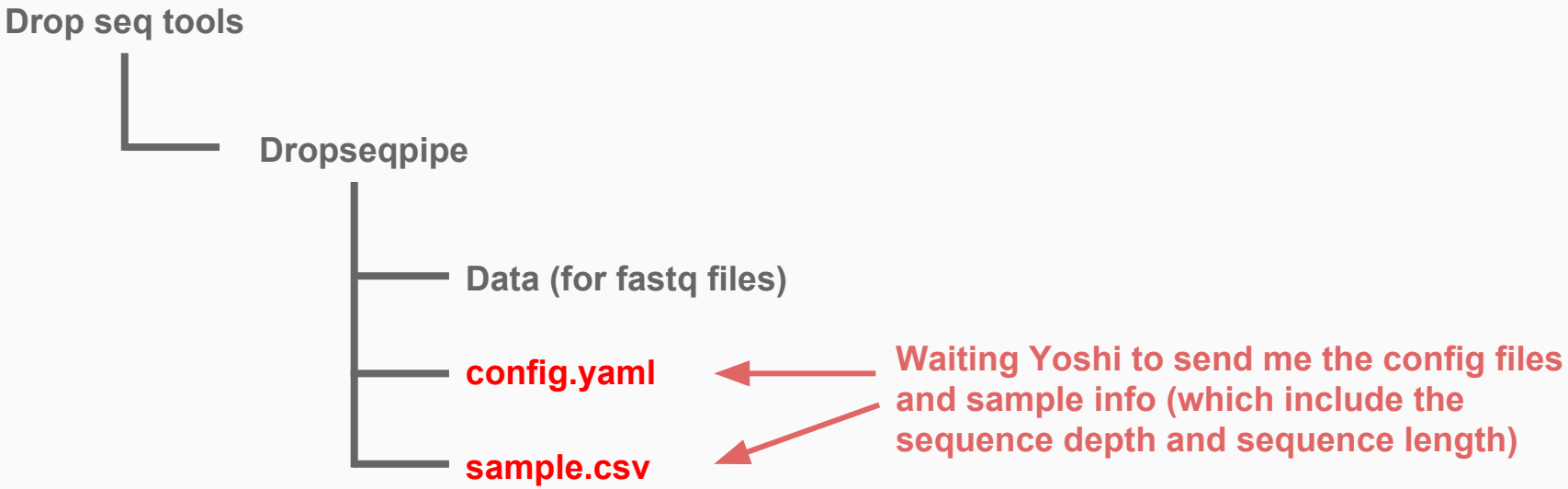
Progress Report

Kuei-Yueh (Clint) Ko

Drop-seq



Yesterday after asking Yoshi some details of drop seq analysis, I have finished set up the file directories to run the drop seq pipeline.

Drop-seq

Gland_10x.txt (normal lung)

Airway_3dpi_10x.txt (injured lung)

Yoshi has also send me the final gene count matrix using 10X genomics (cell ranger) and hisat2 alignment.

I can use it to compare the results of drop-seq pipeline and STAR alignment.

(There will be some differences because of different alignment software.)

Question: data visualization of drop seq results

YP

Yoshihiko Kobayashi, Ph.D.

Wed 5/16, 3:33 AM

Kuei Yueh Ko

Reply all

Hi Kuei,

Here is the good example which we want to have: http://dropviz.org
The problem of this website is currently limited only for brain data set which they uploaded. But it looks like there are source code of this database here: https://github.com/broadinstitute/dropviz

I am guessing following this source code allows us to get to there quickly although I'm not sure if we can easily modify this codes for our purpose.

Best regards,
Yoshi

http://dropviz.org/

HTS

Fastq -> trimmed fastq (using fastq-mcf)
Trimmed fastq -> aligned bam (using tophat2)
Aligned bam to sorted bam (using samtools)
Sorted bam to counts (using htseq)

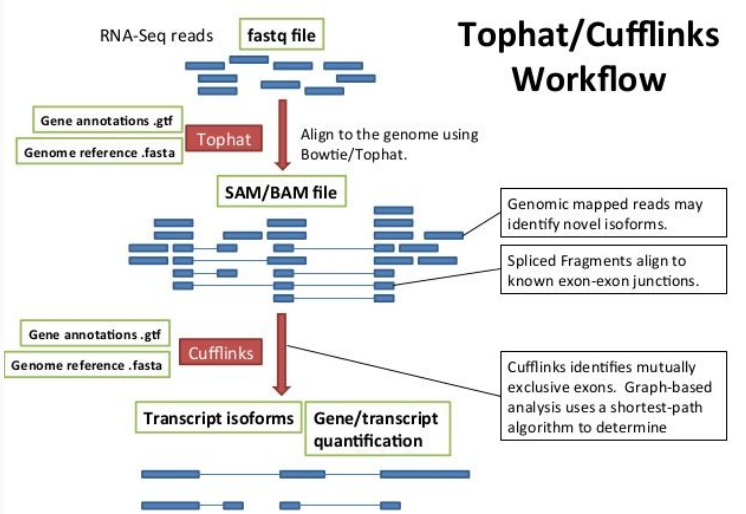
The tools are not installed on Bubbles yet. (I have sent the list to Scott already; however, I did not know which versions or specific sources that should be installed.)

In order to give it a try first, I have installed the commands

- *Fastq-mcf*
- *Tophat2*
- *Samtools*
- *Htseq-count*

on my desktop (https://98.26.42.234:8000)

Question:
in the material, the scripts also includes bowtie.
Is it required?



Another question is: I did not understand the files Korous sent in the email

- adapter_sequence.fastq
- GCF_000149245.1_CNA3_genomic.fna
- GCF_000149245.1_CNA3_genomic.gff
- Cryptococcus_neoformans.ASM9104v1.39.gff3
- Cryptococcus_neoformans.ASM9104v1.dna.toplevel.fa

I suggest that you do a few samples to start with. After review of the pipeline, you can process all of the files.

Annotation:
Option 1
wget http://mirrors.vbi.vt.edu/mirrors/ftp.ncbi.nih.gov/genomes/refseq/fungi/Cryptococcus_neoformans/latest_assembly_versions/GCF_000149245.1_CNA3/GCF_000149245.1_CNA3_genomic.fna.gz

wget http://mirrors.vbi.vt.edu/mirrors/ftp.ncbi.nih.gov/genomes/refseq/fungi/Cryptococcus_neoformans/latest_assembly_versions/GCF_000149245.1_CNA3/GCF_000149245.1_CNA3_genomic.gff.gz

Option 2
wget ftp://ftp.ensemblgenomes.org/pub/fungi/release-39/gff3/cryptococcus_neoformans/Cryptococcus_neoformans.ASM9104v1.39.gff3.gz
wget ftp://ftp.ensemblgenomes.org/pub/fungi/release-39/fasta/cryptococcus_neoformans/dna/Cryptococcus_neoformans.ASM9104v1.dna.toplevel.fa.gz

Adapter sequence fasta (see attached)

Note: Discussion during the meeting on 180531

- Project:
  - Work on the parametric tSNE
  - If everything works well, we will implement a parametric UMAP on Tensorflow
- Data+
  - Work on the tutorial files (Scikit Learn and Scanpy) for the undergrads
  - The data visualization interface for tata's lab will be part of the project
  - Learn the plotly dash
- HTS
  - Stop digging into constructing the pipeline. Wait until the meeting and see what jobs are assigned