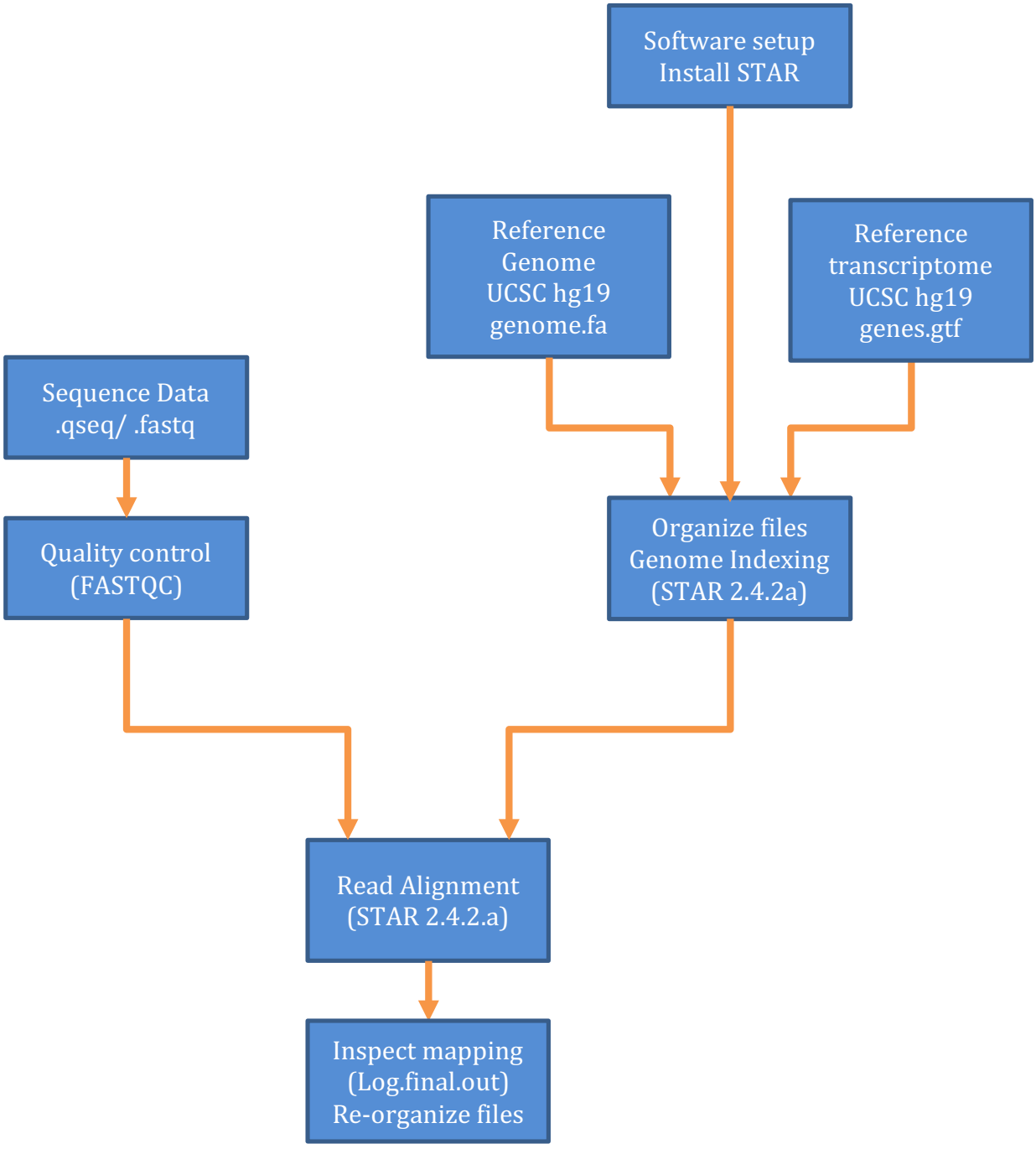
STAR ALIGNMENT WORKFLOW DIAGRAM



Herman Pappoe
07/06/2016
Edit Teia Noel 072916

Programs Required

- Hoffman account and Terminal or Cygwin
- STAR version desired preferably 2.4.2a or above
- For additional reference view current STAR manual
  https://github.com/*alexdobin*/*STAR*/blob/master/doc/STARmanual.pdf

Installing STAR in Hoffman

- Go to https:// github.com/alexdobin/STAR/releases
- Choose the version of STAR you prefer (I have installed version 2.4.2a)
- Once you have chosen the version of STAR desired you can simply right click the Downloads link (either the zip file or tar.gz file) and select copy link
- Open up preferred command-line interphase Cygwin/PuTTY or Terminal for Mac users.
- Install in your $HOME directory or folder, but run all your samples and outputs in your $SCRATCH directory
- . Enter the command:
  cd $SCRATCH
- For latest version and step by step instructions from github go to:
  https://github.com/alexdobin/STAR, or see summary below:

  Commands for Installing and Building STAR in LINUX

  ```
  # Get latest STAR source from releases
  wget https://github.com/alexdobin/STAR/archive/STAR_2.5.2a.tar.gz
  tar -xzf STAR_2.5.2a.tar.gz
  cd STAR_2.5.2a

  # Build STAR
  make STAR
  ```

  Commands for Installing and Building STAR in Mac OS X

  ```
  # Get latest STAR source from releases
  wget https://github.com/alexdobin/STAR/archive/STAR_2.4.2a.tar.gz
  tar -xzf STAR_2.4.2a.tar.gz
  cd STAR_2.4.2a

  # Build STAR
  cd source
  make STARforMacStatic
  ```

Analysis Scripts

Herman Pappoe
07/06/2016
Edit Teia Noel 072916

##Make sure to request at least 32G of memory for indexing.
##You should mkdir a genome directory (genomeDir or any other name you prefer) where the indices will be stored and indicate the path to it in the command as shown below

STAR --runMode genomeGenerate --genomeDir /path/to/genomeDir --genomeFastaFiles /path/to/genome/fasta --sjdbGTFfile /path/to/annotations.gtf --sjdbOverhang ReadLength-1

###To Run STAR mapping with basic options

STAR --genomeDir /path/to/STAR_genome_indices --readFilesIn data.fastq

##Additional options

#Be aware that STAR outputs the files with standard names. Use the following option to change file prefixes

--outFileNamePrefix /path/to/output/dir/prefix.

#Use following option to generate a fastq file with unmapped reads

--outReadsUnmapped Fastx

#Use following option to generate tab file with number of reads per gene.

--quantMode GeneCounts

My examples of commands for real data, which includes some advanced options are shown below

Also please make sure you are requesting at least 2G more than the memory necessary for genome indexing which is 30G on average

##Take note that above we have the basic template to run indexing and mapping with STAR but because STAR is costing in terms of memory you should consider applying to the command line similar parameters as the ones used

###MyIndexingExample

qsub -cwd -l h_data=36G,h_rt=24:00:00 -N Star_gnm_INDX -b y "./STAR --runMode genomeGenerate --genomeDir ~/scratch/Workshop4/STAR_example/genome_index_output_dir --genomeFastaFiles ~/scratch/Homo_sapiens/UCSC/hg19/Sequence/Bowtie2Index/genome.fa --

sjdbOverhang 99 --sjdbGTFfile ~/project-mcdb/Homo_sapiens/UCSC/hg19/Annotation/Genes/genes.gtf --genomeSAindexNbases 14 --runThreadN 16 --limitGenomeGenerateRAM 30000000000"

###MyAlignmentExample
qsub -cwd -l h_data=32G,h_rt=24:00:00 -N G0-1 -b y "./STAR --runThreadN 1 --outSAMtype BAM Unsorted --genomeDir ~/scratch/Workshop4/STAR_example/genome_index_output_dir --readFilesIn G0-1_S4_L003_R1_001.fastq G0-1_S4_L003_R2_001.fastq"

Recommended Options (under construction)

Sample test files (under construction)

Quality control (under construction)

Herman Pappoe
07/06/2016
Edit Teia Noel 072916