

## Lab 16: Analyzing sequencing data in the cloud

Patric Young

```
ubuntu@ip-172-31-1-79:~$ # Download
ubuntu@ip-172-31-1-79:~$ curl -O
https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz
  % Total    % Received % Xferd Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 82.3M  100 82.3M    0     0 19.3M    0 0:00:04 0:00:04 --:--:-- 20.4M
ubuntu@ip-172-31-1-79:~$
ubuntu@ip-172-31-1-79:~$ # Unzip and Untar
ubuntu@ip-172-31-1-79:~$ tar -zxvf sratoolkit.current-ubuntu64.tar
tar (child): sratoolkit.current-ubuntu64.tar: Cannot open: No such file or directory
tar (child): Error is not recoverable: exiting now
tar: Child returned status 2
tar: Error is not recoverable: exiting now
ubuntu@ip-172-31-1-79:~$ ls sratoolkit.3.0.1-ubuntu64/
ls: cannot access 'sratoolkit.3.0.1-ubuntu64/': No such file or directory
ubuntu@ip-172-31-1-79:~$ ls
sratoolkit.current-ubuntu64.tar.gz
ubuntu@ip-172-31-1-79:~$ ls sratoolkit.3.0.1-ubuntu64/
ls: cannot access 'sratoolkit.3.0.1-ubuntu64/': No such file or directory
ubuntu@ip-172-31-1-79:~$ cd ls sratoolkit.3.0.1-ubuntu64/bin/
-bash: cd: too many arguments
ubuntu@ip-172-31-1-79:~$ cd sratoolkit.3.0.1-ubuntu6
-bash: cd: sratoolkit.3.0.1-ubuntu6: No such file or directory
ubuntu@ip-172-31-1-79:~$ cd sratoolkit.3.0.1-ubuntu64/bin/
-bash: cd: sratoolkit.3.0.1-ubuntu64/bin/: No such file or directory
ubuntu@ip-172-31-1-79:~$ ls
sratoolkit.current-ubuntu64.tar.gz
ubuntu@ip-172-31-1-79:~$ cd sratoolkit.3.0.1-ubuntu64/bin
-bash: cd: sratoolkit.3.0.1-ubuntu64/bin: No such file or directory
ubuntu@ip-172-31-1-79:~$ pwd
/home/ubuntu
ubuntu@ip-172-31-1-79:~$ $ ~/sratoolkit.3.0.1-ubuntu64/bin/prefetch --version
$: command not found
ubuntu@ip-172-31-1-79:~$
ubuntu@ip-172-31-1-79:~$ /home/ubuntu/sratoolkit.3.0.1-ubuntu64/bin/prefetch : 3.0.1
-bash: /home/ubuntu/sratoolkit.3.0.1-ubuntu64/bin/prefetch: No such file or directory
```

```
ubuntu@ip-172-31-1-79:~$ ls
sratoolkit.current-ubuntu64.tar.gz
ubuntu@ip-172-31-1-79:~$ cd sratoolkit.3.0.1-ubuntu64/bin
-bash: cd: sratoolkit.3.0.1-ubuntu64/bin: No such file or directory
ubuntu@ip-172-31-1-79:~$ pwd'
> # Download
> curl -O https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz
>
> # Unzip and Untar
> tar -zxvf sratoolkit.current-ubuntu64.tar
>
Display all 1530 possibilities? (y or n)
>
> n
>
>
>
>
> #
> Download
> curl -O https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz
>
> # Unzip and Untar
> tar -zxvf sratoolkit.current-ubuntu64.tar.cz
> tar -zxvf sratoolkit.current-ubuntu64.tar.gz
> ^C
ubuntu@ip-172-31-1-79:~$ ls
sratoolkit.current-ubuntu64.tar.gz
ubuntu@ip-172-31-1-79:~$ gzip -d sratoolkit.current-ubuntu64.tar.gz
ubuntu@ip-172-31-1-79:~$ tar -xvf sratoolkit.current-ubuntu64.tar
sratoolkit.3.0.1-ubuntu64/
sratoolkit.3.0.1-ubuntu64/README.md
sratoolkit.3.0.1-ubuntu64/schema/
sratoolkit.3.0.1-ubuntu64/schema/insdc/
sratoolkit.3.0.1-ubuntu64/schema/insdc/sra.vschema
sratoolkit.3.0.1-ubuntu64/schema/insdc/insdc.vschema
sratoolkit.3.0.1-ubuntu64/schema/insdc/seq.vschema
sratoolkit.3.0.1-ubuntu64/schema/align/
sratoolkit.3.0.1-ubuntu64/schema/align/refseq.vschema
sratoolkit.3.0.1-ubuntu64/schema/align/qstat.vschema
```

sratoolkit.3.0.1-ubuntu64/schema/align/pileup-stats.vschema  
sratoolkit.3.0.1-ubuntu64/schema/align/mate-cache.vschema  
sratoolkit.3.0.1-ubuntu64/schema/align/align.vschema  
sratoolkit.3.0.1-ubuntu64/schema/align/seq.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/  
sratoolkit.3.0.1-ubuntu64/schema/sra/helicos.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/454.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/pacbio.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/generic-fastq.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/nanopore.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/pevents.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/ion-torrent.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/illumina.vschema  
sratoolkit.3.0.1-ubuntu64/schema/sra/abi.vschema  
sratoolkit.3.0.1-ubuntu64/schema/vdb/  
sratoolkit.3.0.1-ubuntu64/schema/vdb/built-in.vschema  
sratoolkit.3.0.1-ubuntu64/schema/vdb/vdb.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/sra.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/seq-graph.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/spotname.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/wgs-contig.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/pnbrdb.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/varloc.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/clip.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/seq.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/ncbi.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/stats.vschema  
sratoolkit.3.0.1-ubuntu64/schema/ncbi/trace.vschema  
sratoolkit.3.0.1-ubuntu64/schema/csra2/  
sratoolkit.3.0.1-ubuntu64/schema/csra2/csra2.vschema  
sratoolkit.3.0.1-ubuntu64/schema/csra2/read.vschema  
sratoolkit.3.0.1-ubuntu64/schema/csra2/reference.vschema  
sratoolkit.3.0.1-ubuntu64/schema/csra2/stats.vschema  
sratoolkit.3.0.1-ubuntu64/CHANGES  
sratoolkit.3.0.1-ubuntu64/bin/  
sratoolkit.3.0.1-ubuntu64/bin/vdb-decrypt.3  
sratoolkit.3.0.1-ubuntu64/bin/vdb-encrypt  
sratoolkit.3.0.1-ubuntu64/bin/fastq-dump.3  
sratoolkit.3.0.1-ubuntu64/bin/sra-sort-cg.3

sratoolkit.3.0.1-ubuntu64/bin/fasterq-dump-orig.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-validate.3  
sratoolkit.3.0.1-ubuntu64/bin/srapath.3  
sratoolkit.3.0.1-ubuntu64/bin/cg-load  
sratoolkit.3.0.1-ubuntu64/bin/illumina-load.3  
sratoolkit.3.0.1-ubuntu64/bin/abi-load  
sratoolkit.3.0.1-ubuntu64/bin/kdbmeta  
sratoolkit.3.0.1-ubuntu64/bin/vdb-unlock.3  
sratoolkit.3.0.1-ubuntu64/bin/srf-load.3  
sratoolkit.3.0.1-ubuntu64/bin/abi-dump.3  
sratoolkit.3.0.1-ubuntu64/bin/srf-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/dump-ref-fasta.3  
sratoolkit.3.0.1-ubuntu64/bin/vdb-config  
sratoolkit.3.0.1-ubuntu64/bin/vdb-copy.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-search.3  
sratoolkit.3.0.1-ubuntu64/bin/test-sra  
sratoolkit.3.0.1-ubuntu64/bin/abi-dump.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-lock.3  
sratoolkit.3.0.1-ubuntu64/bin/srapath.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-dump  
sratoolkit.3.0.1-ubuntu64/bin/vdb-dump.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/align-info.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/cg-load.3  
sratoolkit.3.0.1-ubuntu64/bin/kar.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/srapath-orig.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-pileup  
sratoolkit.3.0.1-ubuntu64/bin/dump-ref-fasta  
sratoolkit.3.0.1-ubuntu64/bin/sra-pileup.3  
sratoolkit.3.0.1-ubuntu64/bin/fasterq-dump.3  
sratoolkit.3.0.1-ubuntu64/bin/helicos-load  
sratoolkit.3.0.1-ubuntu64/bin/fastq-dump  
sratoolkit.3.0.1-ubuntu64/bin/sam-dump.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/kar.3  
sratoolkit.3.0.1-ubuntu64/bin/align-info.3  
sratoolkit.3.0.1-ubuntu64/bin/sam-dump.3  
sratoolkit.3.0.1-ubuntu64/bin/sff-dump.3  
sratoolkit.3.0.1-ubuntu64/bin/vdb-config.3  
sratoolkit.3.0.1-ubuntu64/bin/bam-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-copy  
sratoolkit.3.0.1-ubuntu64/bin/bam-load.3

sratoolkit.3.0.1-ubuntu64/bin/helicos-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-pileup.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-stat  
sratoolkit.3.0.1-ubuntu64/bin/vdb-lock  
sratoolkit.3.0.1-ubuntu64/bin/illumina-dump  
sratoolkit.3.0.1-ubuntu64/bin/pacbio-load.3  
sratoolkit.3.0.1-ubuntu64/bin/fastq-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-copy.3  
sratoolkit.3.0.1-ubuntu64/bin/dump-ref-fasta.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/cache-mgr.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/pacbio-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/abi-dump  
sratoolkit.3.0.1-ubuntu64/bin/vdb-encrypt.3  
sratoolkit.3.0.1-ubuntu64/bin/cg-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sff-load  
sratoolkit.3.0.1-ubuntu64/bin/abi-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-stat.3  
sratoolkit.3.0.1-ubuntu64/bin/vdb-unlock.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/srapath  
sratoolkit.3.0.1-ubuntu64/bin/sff-load.3  
sratoolkit.3.0.1-ubuntu64/bin/kar  
sratoolkit.3.0.1-ubuntu64/bin/sra-sort.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-stat.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/ncbi/  
sratoolkit.3.0.1-ubuntu64/bin/ncbi/default.kfg  
sratoolkit.3.0.1-ubuntu64/bin/ncbi/vdb-copy.kfg  
sratoolkit.3.0.1-ubuntu64/bin/ncbi/certs.kfg  
sratoolkit.3.0.1-ubuntu64/bin/kdbmeta.3  
sratoolkit.3.0.1-ubuntu64/bin/fastq-dump-orig.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-config.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-search  
sratoolkit.3.0.1-ubuntu64/bin/sam-dump-orig.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/illumina-dump.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/illumina-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/abi-load.3  
sratoolkit.3.0.1-ubuntu64/bin/test-sra.3  
sratoolkit.3.0.1-ubuntu64/bin/test-sra.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/fastq-load.3  
sratoolkit.3.0.1-ubuntu64/bin/prefetch-orig.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sratools.3.0.1

sratoolkit.3.0.1-ubuntu64/bin/cache-mgr.3  
sratoolkit.3.0.1-ubuntu64/bin/fasterq-dump  
sratoolkit.3.0.1-ubuntu64/bin/illumina-dump.3  
sratoolkit.3.0.1-ubuntu64/bin/fastq-load  
sratoolkit.3.0.1-ubuntu64/bin/helicos-load.3  
sratoolkit.3.0.1-ubuntu64/bin/latf-load  
sratoolkit.3.0.1-ubuntu64/bin/vdb-validate  
sratoolkit.3.0.1-ubuntu64/bin/vdb-lock.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/srf-load  
sratoolkit.3.0.1-ubuntu64/bin/kdbmeta.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/cache-mgr  
sratoolkit.3.0.1-ubuntu64/bin/sff-dump.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/rcexplain.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-sort.3  
sratoolkit.3.0.1-ubuntu64/bin/sra-search.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/fastq-dump.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/bam-load  
sratoolkit.3.0.1-ubuntu64/bin/sra-pileup-orig.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-decrypt  
sratoolkit.3.0.1-ubuntu64/bin/sra-sort-cg  
sratoolkit.3.0.1-ubuntu64/bin/vdb-decrypt.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-unlock  
sratoolkit.3.0.1-ubuntu64/bin/vdb-encrypt.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-validate.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/align-info  
sratoolkit.3.0.1-ubuntu64/bin/sff-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/vdb-dump-orig.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/prefetch  
sratoolkit.3.0.1-ubuntu64/bin/rcexplain.3  
sratoolkit.3.0.1-ubuntu64/bin/sam-dump  
sratoolkit.3.0.1-ubuntu64/bin/prefetch.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sff-dump  
sratoolkit.3.0.1-ubuntu64/bin/rcexplain  
sratoolkit.3.0.1-ubuntu64/bin/fasterq-dump.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/illumina-load  
sratoolkit.3.0.1-ubuntu64/bin/vdb-dump.3  
sratoolkit.3.0.1-ubuntu64/bin/latf-load.3.0.1  
sratoolkit.3.0.1-ubuntu64/bin/sra-sort  
sratoolkit.3.0.1-ubuntu64/bin/prefetch.3  
sratoolkit.3.0.1-ubuntu64/bin/sra-sort-cg.3.0.1

```

sratoolkit.3.0.1-ubuntu64/bin/pacbio-load
sratoolkit.3.0.1-ubuntu64/bin/latf-load.3
sratoolkit.3.0.1-ubuntu64/example/
sratoolkit.3.0.1-ubuntu64/example/perl/
sratoolkit.3.0.1-ubuntu64/example/perl/base-stats.pl
sratoolkit.3.0.1-ubuntu64/example/perl/mismatch-stats.pl
sratoolkit.3.0.1-ubuntu64/example/perl/gene-lookup.pl
sratoolkit.3.0.1-ubuntu64/example/perl/splitfastq.pl
sratoolkit.3.0.1-ubuntu64/example/perl/quality-stats.pl
sratoolkit.3.0.1-ubuntu64/example/perl/simplefastq.pl
sratoolkit.3.0.1-ubuntu64/example/perl/dump-reference.pl
sratoolkit.3.0.1-ubuntu64/README-blastn
sratoolkit.3.0.1-ubuntu64/README-vdb-config
ubuntu@ip-172-31-1-79:~$ ls
sratoolkit.3.0.1-ubuntu64 sratoolkit.current-ubuntu64.tar
ubuntu@ip-172-31-1-79:~$ export PATH=$PATH:/home/ubuntu/sratoolkit.3.0.1-ubuntu64/bin
ubuntu@ip-172-31-1-79:~$ prefetch SRR600956

```

2023-03-02T18:28:45 prefetch.3.0.1: Current preference is set to retrieve SRA Normalized Format files with full base quality scores.

2023-03-02T18:28:46 prefetch.3.0.1: 1) Downloading 'SRR600956'...

2023-03-02T18:28:46 prefetch.3.0.1: SRA Normalized Format file is being retrieved, if this is different from your preference, it may be due to current file availability.

2023-03-02T18:28:46 prefetch.3.0.1: Downloading via HTTPS...

2023-03-02T18:29:02 prefetch.3.0.1: HTTPS download succeed

2023-03-02T18:29:04 prefetch.3.0.1: 'SRR600956' is valid

2023-03-02T18:29:04 prefetch.3.0.1: 1) 'SRR600956' was downloaded successfully

ubuntu@ip-172-31-1-79:~\$ fastq-dump SRR600956

\$ head SRR600956.fastq

```

@SRR600956.1 HWI-EAS486_0002:3:1:1382:1342 length=38
GTGTTCCAAATGCTGCAAATGGGTGTCAATGTATGTTA
+SRR600956.1 HWI-EAS486_0002:3:1:1382:1342 length=38
D?BCCA?BDBDBACD@=??BAAC>CBBBBBCBBBD?%%
@SRR600956.2 HWI-EAS486_0002:3:1:1382:5487 length=38
GATGATAGTTTCTTTTGCCGTTAGCACAAATTTTCCAA
+SRR600956.2 HWI-EAS486_0002:3:1:1382:5487 length=38
DCEECEAECEFFDFECEEEFFFFDB?BBADEEEEE??
@SRR600956.3 HWI-EAS486_0002:3:1:1382:4694 length=38
TGTAGGCTCCACCTCTGGGGGCAGGGCACAGACAAACA

```

```
$ grep -c "@SRR600956" SRR600956.fastq
```

```
25849655
```

```
Read 25849655 spots for SRR600956
```

```
Written 25849655 spots for SRR600956
```

```
ubuntu@ip-172-31-1-79:~$ $ head SRR600956.fastq
```

```
$: command not found
```

```
ubuntu@ip-172-31-1-79:~$
```

```
ubuntu@ip-172-31-1-79:~$ @SRR600956.1 HWI-EAS486_0002:3:1:1382:1342 length=38
```

```
@SRR600956.1: command not found
```

```
ubuntu@ip-172-31-1-79:~$ GTGTTCCAAATGCTGCAAATGGGTGTCAATGTATGTTA
```

```
GTGTTCCAAATGCTGCAAATGGGTGTCAATGTATGTTA: command not found
```

```
ubuntu@ip-172-31-1-79:~$ +SRR600956.1 HWI-EAS486_0002:3:1:1382:1342 length=38
```

```
+SRR600956.1: command not found
```

```
ubuntu@ip-172-31-1-79:~$ D?BCCA?BDBDBACD@=?BAAC>CBBBBBCBBBD?%%%
```

```
D?BCCA?BDBDBACD@=?BAAC: command not found
```

```
ubuntu@ip-172-31-1-79:~$ @SRR600956.2 HWI-EAS486_0002:3:1:1382:5487 length=38
```

```
@SRR600956.2: command not found
```

```
ubuntu@ip-172-31-1-79:~$ GATGATAGTTTCTTTTGCCGTTAGCACAATTTTCCAA
```

```
GATGATAGTTTCTTTTGCCGTTAGCACAATTTTCCAA: command not found
```

```
ubuntu@ip-172-31-1-79:~$ +SRR600956.2 HWI-EAS486_0002:3:1:1382:5487 length=38
```

```
+SRR600956.2: command not found
```

```
ubuntu@ip-172-31-1-79:~$ DCEECEAECEFFDFECEEEFFFFDB?BBADEEEEE???
```

```
DCEECEAECEFFDFECEEEFFFFDB?BBADEEEEE???: command not found
```

```
ubuntu@ip-172-31-1-79:~$ @SRR600956.3 HWI-EAS486_0002:3:1:1382:4694 length=38
```

```
@SRR600956.3: command not found
```

```
ubuntu@ip-172-31-1-79:~$ TGTAGGCTCCACCTCTGGGGGCAGGGCACAGACAAACA
```

```
TGTAGGCTCCACCTCTGGGGGCAGGGCACAGACAAACA: command not found
```

```
ubuntu@ip-172-31-1-79:~$ $ grep -c "@SRR600956" SRR600956.fastq
```

```
$: command not found
```

```
ubuntu@ip-172-31-1-79:~$
```

```
ubuntu@ip-172-31-1-79:~$ 25849655
```

```
25849655: command not found
```

```
ubuntu@ip-172-31-1-79:~$ $ head SRR600956.fastq
```

```
$: command not found
```

```
ubuntu@ip-172-31-1-79:~$ head SRR600956.fastq
```

```
@SRR600956.1 HWI-EAS486_0002:3:1:1382:1342 length=38
```

```
GTGTTCCAAATGCTGCAAATGGGTGTCAATGTATGTTA
```

```
+SRR600956.1 HWI-EAS486_0002:3:1:1382:1342 length=38
```

```
D?BCCA?BDBDBACD@=?BAAC>CBBBBBCBBBD?%%%
```



```
@SRR600956.2 HWI-EAS486_0002:3:1:1382:5487 length=38
GATGATAGTTTCTTTTGCCGTTAGCACAATTTTCCAA
+SRR600956.2 HWI-EAS486_0002:3:1:1382:5487 length=38
DCEECEAECEFFDFECEEEFFFFDB?BBADEEEEE??
@SRR600956.3 HWI-EAS486_0002:3:1:1382:4694 length=38
TGTAGGCTCCACCTCTGGGGGCAGGGCACAGACAAACA
ubuntu@ip-172-31-1-79:~$ grep -c "@SRR600956" SRR600956.fastq
25849655
ubuntu@ip-172-31-1-79:~$ prefetch SRR2156849 SRR2156850 SRR2156851
```

2023-03-02T18:40:16 prefetch.3.0.1: Current preference is set to retrieve SRA Normalized Format files with full base quality scores.

2023-03-02T18:40:17 prefetch.3.0.1: 1) Downloading 'SRR2156849'...

2023-03-02T18:40:17 prefetch.3.0.1: SRA Normalized Format file is being retrieved, if this is different from your preference, it may be due to current file availability.

2023-03-02T18:40:17 prefetch.3.0.1: Downloading via HTTPS...

2023-03-02T18:40:25 prefetch.3.0.1: HTTPS download succeed

2023-03-02T18:40:25 prefetch.3.0.1: 'SRR2156849' is valid

2023-03-02T18:40:25 prefetch.3.0.1: 1) 'SRR2156849' was downloaded successfully

2023-03-02T18:40:25 prefetch.3.0.1: 'SRR2156849' has 0 unresolved dependencies

2023-03-02T18:40:25 prefetch.3.0.1: Current preference is set to retrieve SRA Normalized Format files with full base quality scores.

2023-03-02T18:40:26 prefetch.3.0.1: 2) Downloading 'SRR2156850'...

2023-03-02T18:40:26 prefetch.3.0.1: SRA Normalized Format file is being retrieved, if this is different from your preference, it may be due to current file availability.

2023-03-02T18:40:26 prefetch.3.0.1: Downloading via HTTPS...

2023-03-02T18:40:33 prefetch.3.0.1: HTTPS download succeed

2023-03-02T18:40:34 prefetch.3.0.1: 'SRR2156850' is valid

2023-03-02T18:40:34 prefetch.3.0.1: 2) 'SRR2156850' was downloaded successfully

2023-03-02T18:40:34 prefetch.3.0.1: 'SRR2156850' has 0 unresolved dependencies

2023-03-02T18:40:34 prefetch.3.0.1: Current preference is set to retrieve SRA Normalized Format files with full base quality scores.

2023-03-02T18:40:34 prefetch.3.0.1: 3) Downloading 'SRR2156851'...

2023-03-02T18:40:34 prefetch.3.0.1: SRA Normalized Format file is being retrieved, if this is different from your preference, it may be due to current file availability.

2023-03-02T18:40:34 prefetch.3.0.1: Downloading via HTTPS...

2023-03-02T18:40:41 prefetch.3.0.1: HTTPS download succeed

2023-03-02T18:40:41 prefetch.3.0.1: 'SRR2156851' is valid

```
2023-03-02T18:40:41 prefetch.3.0.1: 3) 'SRR2156851' was downloaded successfully
2023-03-02T18:40:41 prefetch.3.0.1: 'SRR2156851' has 0 unresolved dependencies
ubuntu@ip-172-31-1-79:~$ fastq-dump --split-3 SRR2156849 SRR2156850 SRR2156851
Read 2985576 spots for SRR2156849
Written 2985576 spots for SRR2156849
Read 2669778 spots for SRR2156850
Written 2669778 spots for SRR2156850
Read 2369745 spots for SRR2156851
Written 2369745 spots for SRR2156851
ubuntu@ip-172-31-1-79:~$ head SRR2156848 1.fastq
head: cannot open 'SRR2156848' for reading: No such file or directory
head: cannot open '1.fastq' for reading: No such file or directory
ubuntu@ip-172-31-1-79:~$ ls
'CBBBBBCBBBD?%%'  SRR2156850_1.fastq  SRR600956
SRR2156849        SRR2156850_2.fastq  SRR600956.fastq
SRR2156849_1.fastq SRR2156851          sratoolkit.3.0.1-ubuntu64
SRR2156849_2.fastq SRR2156851_1.fastq  sratoolkit.current-ubuntu64.tar
SRR2156850        SRR2156851_2.fastq
ubuntu@ip-172-31-1-79:~$ head SRR2156848 1.fastq
head: cannot open 'SRR2156848' for reading: No such file or directory
head: cannot open '1.fastq' for reading: No such file or directory
ubuntu@ip-172-31-1-79:~$ wget
https://github.com/pachterlab/kallisto/releases/download/v0.44.0/kallisto_linux-v0.44.0.tar.gz
--2023-03-02 18:50:28--
https://github.com/pachterlab/kallisto/releases/download/v0.44.0/kallisto_linux-v0.44.0.tar.gz
Resolving github.com (github.com)... 192.30.255.113
Connecting to github.com (github.com)|192.30.255.113|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location:
https://objects.githubusercontent.com/github-production-release-asset-2e65be/26562905/b78619
ae-0542-11e8-908e-6777f5455785?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Crede
ntial=AKIAIWNJYAX4CSVEH53A%2F20230302%2Fus-east-1%2Fs3%2Faws4_request&X-A
mz-Date=20230302T185017Z&X-Amz-Expires=300&X-Amz-Signature=32fc958bd371361484
1645058be79d59e8e901bda47e3580583bccd01e2a8258&X-Amz-SignedHeaders=host&actor_id
=0&key_id=0&repo_id=26562905&response-content-disposition=attachment%3B%20filename
%3Dkallisto_linux-v0.44.0.tar.gz&response-content-type=application%2Foctet-stream
[following]
--2023-03-02 18:50:28--
https://objects.githubusercontent.com/github-production-release-asset-2e65be/26562905/b78619
ae-0542-11e8-908e-6777f5455785?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Crede
```

```
ntial=AKIAIWNJYAX4CSVEH53A%2F20230302%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20230302T185017Z&X-Amz-Expires=300&X-Amz-Signature=32fc958bd3713614841645058be79d59e8e901bda47e3580583bccd01e2a8258&X-Amz-SignedHeaders=host&actor_id=0&key_id=0&repo_id=26562905&response-content-disposition=attachment%3B%20filename%3Dkallisto_linux-v0.44.0.tar.gz&response-content-type=application%2Foctet-stream
Resolving objects.githubusercontent.com (objects.githubusercontent.com)... 185.199.111.133,
185.199.108.133, 185.199.109.133, ...
Connecting to objects.githubusercontent.com
(objects.githubusercontent.com)[185.199.111.133]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6668961 (6.4M) [application/octet-stream]
Saving to: 'kallisto_linux-v0.44.0.tar.gz'
```

```
kallisto_linux-v0.44. 100%[=====>] 6.36M 22.9MB/s in 0.3s
```

```
2023-03-02 18:50:29 (22.9 MB/s) - 'kallisto_linux-v0.44.0.tar.gz' saved [6668961/6668961]
```

```
ubuntu@ip-172-31-1-79:~$ tar -zxvf kallisto_linux-v0.44.0.tar.gz
kallisto_linux-v0.44.0/
kallisto_linux-v0.44.0/license.txt
kallisto_linux-v0.44.0/kallisto
kallisto_linux-v0.44.0/test/
kallisto_linux-v0.44.0/test/chrom.txt
kallisto_linux-v0.44.0/test/transcripts.gtf.gz
kallisto_linux-v0.44.0/test/reads_2.fastq.gz
kallisto_linux-v0.44.0/test/transcripts.fasta.gz
kallisto_linux-v0.44.0/test/README.md
kallisto_linux-v0.44.0/test/Snakefile
kallisto_linux-v0.44.0/test/reads_1.fastq.gz
kallisto_linux-v0.44.0/README.md
ubuntu@ip-172-31-1-79:~$ ~/kallisto_linux-v0.44.0/kallisto
kallisto 0.44.0
```

Usage: kallisto <CMD> [arguments] ..

Where <CMD> can be one of:

index	Builds a kallisto index
quant	Runs the quantification algorithm
pseudo	Runs the pseudoalignment step

h5dump	Converts HDF5-formatted results to plaintext
inspect	Inspects and gives information about an index
version	Prints version information
cite	Prints citation information

Running kallisto <CMD> without arguments prints usage information for <CMD>

```
ubuntu@ip-172-31-1-79:~$ export PATH=$PATH:/home/ubuntu/kallisto_linux-v0.44.0
ubuntu@ip-172-31-1-79:~$ echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/home/ubuntu/srtoolkit.3.0.1-ubuntu64/bin:/home/ubuntu/kallisto_linux-v0.44.0
ubuntu@ip-172-31-1-79:~$ wget
ftp://ftp.ensembl.org/pub/release-67/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh37.67.cdna.all.fa.gz
--2023-03-02 19:00:05--
ftp://ftp.ensembl.org/pub/release-67/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh37.67.cdna.all.fa.gz
=> 'Homo_sapiens.GRCh37.67.cdna.all.fa.gz'
Resolving ftp.ensembl.org (ftp.ensembl.org)... 193.62.193.139
Connecting to ftp.ensembl.org (ftp.ensembl.org)|193.62.193.139|:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done. ==> PWD ... done.
==> TYPE I ... done. ==> CWD (1) /pub/release-67/fasta/homo_sapiens/cdna ... done.
==> SIZE Homo_sapiens.GRCh37.67.cdna.all.fa.gz ... 59979785
==> PASV ... done. ==> RETR Homo_sapiens.GRCh37.67.cdna.all.fa.gz ... done.
Length: 59979785 (57M) (unauthoritative)

Homo_sapiens.GRCh37.6 100%[=====>] 57.20M 13.8MB/s in
5.4s

2023-03-02 19:00:12 (10.7 MB/s) - 'Homo_sapiens.GRCh37.67.cdna.all.fa.gz' saved
[59979785]

ubuntu@ip-172-31-1-79:~$ gunzip Homo_sapiens.GRCh37.67.cdna.all.fa.gz
ubuntu@ip-172-31-1-79:~$ kallisto index -i hg19.ensembl
Homo_sapiens.GRCh37.67.cdna.all.fa

[build] loading fasta file Homo_sapiens.GRCh37.67.cdna.all.fa
[build] k-mer length: 31
[build] warning: clipped off poly-A tail (longer than 10)
```



(optional for --genomebam, but recommended)  
ubuntu@ip-172-31-1-79:~\$ prefetch SRR2156848

2023-03-02T19:12:28 prefetch.3.0.1: Current preference is set to retrieve SRA Normalized Format files with full base quality scores.

2023-03-02T19:12:28 prefetch.3.0.1: 1) Downloading 'SRR2156848'...

2023-03-02T19:12:28 prefetch.3.0.1: SRA Normalized Format file is being retrieved, if this is different from your preference, it may be due to current file availability.

2023-03-02T19:12:28 prefetch.3.0.1: Downloading via HTTPS...

2023-03-02T19:12:37 prefetch.3.0.1: HTTPS download succeed

2023-03-02T19:12:38 prefetch.3.0.1: 'SRR2156848' is valid

2023-03-02T19:12:38 prefetch.3.0.1: 1) 'SRR2156848' was downloaded successfully

2023-03-02T19:12:38 prefetch.3.0.1: 'SRR2156848' has 0 unresolved dependencies

ubuntu@ip-172-31-1-79:~\$ fastq-dump --split-3 SRR2156848

Read 2959900 spots for SRR2156848

Written 2959900 spots for SRR2156848

ubuntu@ip-172-31-1-79:~\$ ls

'CBBBBBCBBBD?%%%' SRR2156850\_2.fastq

Homo\_sapiens.GRCh37.67.cdna.all.fa SRR2156851

SRR2156848 SRR2156851\_1.fastq

SRR2156848\_1.fastq SRR2156851\_2.fastq

SRR2156848\_2.fastq SRR600956

SRR2156848\_quant SRR600956.fastq

SRR2156849 hg19.ensembl

SRR2156849\_1.fastq kallisto\_linux-v0.44.0

SRR2156849\_2.fastq kallisto\_linux-v0.44.0.tar.gz

SRR2156850 sratoolkit.3.0.1-ubuntu64

SRR2156850\_1.fastq sratoolkit.current-ubuntu64.tar

ubuntu@ip-172-31-1-79:~\$ kallisto quant -i hg19.ensembl -o SRR2156848\_quant

SRR2156848\_1.fastq SRR2156848\_2.fastq

[quant] fragment length distribution will be estimated from the data

[index] k-mer length: 31

[index] number of targets: 176,981

[index] number of k-mers: 100,753,348

[index] number of equivalence classes: 671,976

[quant] running in paired-end mode

[quant] will process pair 1: SRR2156848\_1.fastq

SRR2156848\_2.fastq

[quant] finding pseudoalignments for the reads ... done

[quant] processed 2,959,900 reads, 2,563,611 reads pseudoaligned  
[quant] estimated average fragment length: 190.486  
[ em] quantifying the abundances ... done  
[ em] the Expectation-Maximization algorithm ran for 1,057 rounds

```
ubuntu@ip-172-31-1-79:~$ ls SRR2156848_quant
abundance.h5 abundance.tsv run_info.json
ubuntu@ip-172-31-1-79:~$ # You complete these steps
ubuntu@ip-172-31-1-79:~$ $ kallisto quant -i ____ -o SRR2156849_quant SRR2156849_1.fastq
SRR2156849_2.fastq
$: command not found
ubuntu@ip-172-31-1-79:~$
ubuntu@ip-172-31-1-79:~$ $ kallisto quant -i hg19.ensembl -o SRR2156850_quant _____.fastq
_____.fastq
$: command not found
ubuntu@ip-172-31-1-79:~$
ubuntu@ip-172-31-1-79:~$ kallisto quant -i hg19.ensembl -o SRR2156849_quant
SRR2156849_1.fastq SRR2156849_2.fastq
```

[quant] fragment length distribution will be estimated from the data  
[index] k-mer length: 31  
[index] number of targets: 176,981  
[index] number of k-mers: 100,753,348  
[index] number of equivalence classes: 671,976  
[quant] running in paired-end mode  
[quant] will process pair 1: SRR2156849\_1.fastq  
SRR2156849\_2.fastq  
[quant] finding pseudoalignments for the reads ... done  
[quant] processed 2,985,576 reads, 2,600,800 reads pseudoaligned  
[quant] estimated average fragment length: 188.479  
[ em] quantifying the abundances ... done  
[ em] the Expectation-Maximization algorithm ran for 1,046 rounds

```
ubuntu@ip-172-31-1-79:~$ kallisto quant -i hg19.ensembl -o SRR2156850_quant
SRR2156850_1.fastq SRR2156850_2.fastq
```

[quant] fragment length distribution will be estimated from the data  
[index] k-mer length: 31  
[index] number of targets: 176,981  
[index] number of k-mers: 100,753,348

```
[index] number of equivalence classes: 671,976
[quant] running in paired-end mode
[quant] will process pair 1: SRR2156850_1.fastq
                        SRR2156850_2.fastq
[quant] finding pseudoalignments for the reads ... done
[quant] processed 2,669,778 reads, 2,372,309 reads pseudoaligned
[quant] estimated average fragment length: 186.747
[ em] quantifying the abundances ... done
[ em] the Expectation-Maximization algorithm ran for 969 rounds
```

```
ubuntu@ip-172-31-1-79:~$ kallisto quant -i hg19.ensembl -o SRR2156851_quant
SRR2156851_1.fastq SRR2156851_2.fastq
```

```
[quant] fragment length distribution will be estimated from the data
[index] k-mer length: 31
[index] number of targets: 176,981
[index] number of k-mers: 100,753,348
[index] number of equivalence classes: 671,976
[quant] running in paired-end mode
[quant] will process pair 1: SRR2156851_1.fastq
                        SRR2156851_2.fastq
[quant] finding pseudoalignments for the reads ... done
[quant] processed 2,369,745 reads, 2,111,474 reads pseudoaligned
[quant] estimated average fragment length: 191.177
[ em] quantifying the abundances ... done
[ em] the Expectation-Maximization algorithm ran for 1,058 rounds
```

```
ubuntu@ip-172-31-1-79:~$ client_loop: send disconnect: Broken pipe
Patrics-MacBook-Air:class16 patricyoung$
```