

# Assignment 1: Alignment

## Alignment Strategy

For this assignment Command Line tool HISAT2 version 2.1.0 (which is a spliced aligner) was used to align the reads to the (b37)hg19 human reference genome.

### Download raw data:

Data (SRA files) downloaded with the help of a simple Command Line Script from the AWS location of SRA link on NCBI website for each biosample.

Only files with SRR1\* accessions for both fetal and adult samples(1st run/replicate for each biosample) were used for downstream analysis.

Fetal samples:	Adult Samples:
SRR1554537	SRR1554535
SRR1554538	SRR1554536
SRR1554541	SRR1554539
SRR1554566	SRR1554556
SRR1554567	SRR1554561
SRR1554568	SRR1554534

### Extract FASTQ files:

The tool fasterq-dump (command line tool-sratoolkit version 2.11.0) was used to extract FASTQ files from the SRA accessions above using Command Line.

This resulted in 3 fastq files for each of the above samples

SRR1\*.fastq  
SRR1\*\_1.fastq  
SRR1\*\_2.fastq

Check quality of reads: Run fastqc on reads.

### Align to Reference Human Genome with Hisat2 (command line):

Step 1: The reference human genome (b37)hg19 was downloaded.

Step 2: hisat2 built in command for creating index was used to build an index for the reference human genome.

Step 3: hisat2 was used to align mate1 files SRR1\*\_1.fastq and mate2 files SRR1\*\_2.fastq to the reference human genome index and the output was saved to SRR1\*.sam files.

## Alignment Summary/Results

Summary files were generated for the alignments and flagstat command of samtools was used to get basic statistics for the alignments.

Average overall alignment rate for fetal samples was 99.74 and for adult samples was 99.75

### Alignments for Fetal Samples:

1) SRR1554537

Reads mapped(aligned) = 118101807 (99.75%)

Total reads (QC passed reads) = 118396356

Number of paired reads: 55133946 \* 2 = 110267892 (paired in sequencing)

Of these

Concordant alignments 0 times = 476664 \* 2 = 953328 (0.86%)

Concordant alignments (exactly 1 time): 52688947 \* 2 = 105377894 (95.75%)

Concordant alignments > 1 times = 1968335 \* 2 = 3936670 (3.57%)

Overall alignment rate: 99.73%

2) SRR1554538

Reads mapped (aligned) = 145893064 (99.77%)

Total reads (QC passed reads) = 146224905

Number of paired reads: 68026190 \* 2 = 136052380 (paired in sequencing)

Of these

Concordant alignments 0 times = 546149 \* 2 = 1092298 (0.80%)

Concordant alignments (exactly 1 time): 65073871 \* 2 = 130147742 (95.66%)

Concordant alignments > 1 times = 2406170 \* 2 = 4812340 (3.54%)

Overall alignment rate: 99.76%

3) SRR1554541

Reads mapped (aligned) = 147568454 (99.75%)

Total reads (QC passed reads) = 147940909

Number of paired reads: 69278357 \* 2 = 138556714 (paired in sequencing)

Of these

Concordant alignments 0 times = 562879 \* 2 = 1125758 (0.81%)

Concordant alignments exactly 1 time: 66406349 \* 2 = 132812698 (95.85%)

Concordant alignments > 1 times = 2309129 \* 2 = 4618258 (3.33%)

Overall alignment rate: 99.73%

4) SRR1554566

Reads mapped (aligned) = 115275781 (99.77%)

Total reads (QC passed reads) = 115536367

Number of paired reads = 53161501 \* 2 = 106323002 (paired in sequencing)

Of these:

Concordant alignments 0 times =  $386090 * 2 = 772180$  (0.73%)

Concordant alignments exactly 1 time:  $50660231 * 2 = 101320462$  (95.29%)

Concordant alignments > 1 times =  $2115180 * 2 = 4230360$  (3.98%)

Overall alignment rate: 99.75%

5) SRR1554567

Reads mapped (aligned) = 132306889 (99.77%)

Total reads (QC passed reads) = 132611027

Number of paired reads =  $619229351 * 2 = 123845870$  (paired in sequencing)

Of these

Concordant alignments 0 times =  $484693 * 2 = 969386$  (0.78%)

Concordant alignments exactly 1 time:  $59358910 * 2 = 118717820$  (95.86%)

Concordant alignments > 1 times =  $2079332 * 2 = 4158664$  (3.36%)

Overall alignment rate: 99.75%

6) SRR1554568

Reads mapped (aligned) = 103235507 (99.76%)

Total reads (QC passed reads) = 103480316

Number of paired reads =  $48184702 * 2 = 96369404$  (paired in sequencing)

Of these

Concordant alignments 0 times =  $391341 * 2 = 782682$  (0.81%)

Concordant alignments exactly 1 time =  $46139291 * 2 = 92278582$  (95.76%)

Concordant alignments > 1 times =  $1654070 * 2 = 3308140$  (3.43%)

Overall alignment rate: 99.75%

### Alignments for Adult Samples:

1) SRR1554535

Reads mapped (aligned) = 80203609 (99.74%)

Total reads (QC passed reads) = 80414830

Number of paired reads:  $38063721 * 2 = 76127442$  (paired in sequencing)

Of these

Concordant alignments 0 times =  $530732 * 2 = 1061464$  (1.39%)

Concordant alignment exactly 1 time =  $36356543 * 2 = 72713086$  (95.51%)

Concordant alignments > 1 times =  $1176446 * 2 = 2352892$  (3.09%)

Overall alignment rate: 99.72%

2) SRR1554536

Reads mapped (aligned) = 45114554 (99.87%)

Total reads (QC passed reads) = 45172764

Number of paired reads = 21450348 \* 2 = 42900696 (paired in sequencing)

Of these

Concordant alignments 0 times = 141860 \* 2 = 283720 (0.66%)

Concordant alignments exactly 1 time: 20545796 \* 2 = 41091592 (95.78%)

Concordant alignments > 1 times = 762692 \* 2 = 1525384 (3.56%)

Overall alignment rate: 99.86%

3) SRR1554539

Reads mapped (aligned) = 70483001 (99.73%)

Total reads (QC passed reads) = 70670921

Number of paired reads = 33742728 \* 2 = 67485456 (paired in sequencing)

Of these

Concordant alignments 0 times = 501866 \* 2 = 1003732 (1.49%)

Concordant alignments exactly 1 time: 32426206 \* 2 = 64852412 (96.10%)

Concordant alignments > 1 times = 814656 \* 2 = 1629312 (2.41%)

Overall alignment rate: 99.72%

4) SRR1554556

Reads mapped (aligned) = 104055748 (99.79%)

Total reads (QC passed reads) = 104271763

Number of paired reads = 49480779 \* 2 = 98961558 (paired in sequencing)

Of these

Concordant alignments 0 times = 433190 \* 2 = 866380 (0.88%)

Concordant alignments exactly 1 time = 47657728 \* 2 = 95315456 (96.32%)

Concordant alignments > 1 times = 1389861 \* 2 = 2779722 (2.81%)

Overall alignment rate: 99.78%

5) SRR1554561

Reads mapped (aligned) = 82851562 (99.71%)

Total reads (QC passed reads) = 83093446

Number of paired reads = 39272751 \* 2 = 78545502 (paired in sequencing)

Of these

Concordant alignments 0 times =  $593768 * 2 = 1187536$  (1.51%)  
Concordant alignments exactly 1 time =  $37506560 * 2 = 75013120$  (95.50%)  
Concordant alignments > 1 times =  $1172423 * 2 = 2344846$  (2.99%)  
Overall alignment rate: 99.69%

6) SRR1554534

Reads mapped (aligned) = 60071030 (99.73%)

Total reads (QC passed reads) = 60234971

Number of paired reads =  $28181772 * 2 = 56363544$  (paired in sequencing)

Of these

Concordant alignments 0 times =  $414771 * 2 = 829542$  (1.47%)  
Concordant alignments exactly 1 time =  $26727333 * 2 = 53454666$  (94.84%)  
Concordant alignments > 1 times =  $1039668 * 2 = 2079336$  (3.69%)  
Overall alignment rate: 99.71%