

Lab 2, Comp 3353

Question #1:

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannguy$ samtools view -H sample.sorted.bam | head
@HD      VN:1.5   GO:none  SO:coordinate
@SQ      SN:1     LN:249250621
@SQ      SN:2     LN:243199373
@SQ      SN:3     LN:198022430
@SQ      SN:4     LN:191154276
@SQ      SN:5     LN:180915260
@SQ      SN:6     LN:171115067
@SQ      SN:7     LN:159138663
@SQ      SN:8     LN:146364022
@SQ      SN:9     LN:141213431
```

S0: **Coordinate**, Means the file is sorted.

Question #2:

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannguy$ samtools view sample.sorted.bam | wc -l
5026086
```

There are 5026086 alignments in sample.sorted.bam.

Question #3:

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannnguy$ samtools view sample.sorted.bam 1:1239892-12399392 | wc -l
36495
```

There are 36495 alignments that overlap chromosome 1 from position 12398392 to 12399392

Question #4



```
HWI-ST354R:351:C0UPMACXX:5:1308:15815:90293      163      1      861361      60      100M      =      861426      165      GACTGCCCGGGCT
GCCGAATATCTCCCGCGTGGTGAGATCGCGGGCTCGGTTGGGGCTGGGAGTTACTCTCCCTCGCGGAGCGTGTTCCTCGGGTTTC      ?=@DDDFFFDDFABGGHG@BGGGGGGGGHIG@;e=0
F=EAGAAD;8;A55+5~/9@05<BB<@23{>:;>@>3?B?##### MC:Z:100M      MD:Z:83T3C3G3T4      RG:Z:1719PC0017_51      NM:i:
4      MQ:i:60      AS:i:83      XS:i:0
HWI-ST354R:351:C0UPMACXX:6:2210:3183:62401      99      1      861362      60      100M      =      861401      139      ACTGCCCGGGCTG
CCGAATATCTCCCGCGTGGTGAGATCGCGGGCTCGGTTGGGGCTGGGAGTTACTCTCCCTCGCGAGCTGTTCCTCGGGTTTTCA      @CCFFEFHFFHHIIIGIJJIIJJJJJJJJJ6G@GHE
GIA@EHFBD?BDD75BDBBDB@BD(8+9>:::A:AC?<A<9-<@(>?<ACC@:15<55<?3:      MC:Z:100M      MD:Z:100      RG:Z:1719PC0017_51      NM:i:
0      MQ:i:60      AS:i:100      XS:i:0
HWI-ST354R:351:C0UPMACXX:6:2103:12417:96772      163      1      861397      60      100M      =      861420      123      AGATCGCGGGCTC
GGTTGGGGCTGGGAGTTACTCTCCCTCGCGAGCTGTTCCTCGCGGTTTTCAGGGTTTTCAGGATCGAGAGTCTTAACCTCACCCCT      CCCFFFFFHNNHHJJHJJJJJJJJJJHCHFFFFFIE
EDCBDBB;@BBDDBCDACDD05@BCCDD<8?BACCD<CC>@DDBA:>>CBBDDC2?BB0      MC:Z:100M      MD:Z:100      RG:Z:1719PC0017_51      NM:i:
0      MQ:i:60      AS:i:100      XS:i:0
HWI-ST354R:351:C0UPMACXX:6:2210:3183:62401      147      1      861401      60      100M      =      861362      -139      GCGGGCTCGGTT
GGGGCTGGGAGTTACTCTCCCTCGCGAGCTGTTCCTCGCGGTTTTCAGGGTTTTCAGGATCGAGAGTCTTAACCTCACCCCTCGCG      #####DDB<@B>8(8@DBA@CCA>4BBB>5B@;DB
BCCDDCDBDEEBHE@1GGEEBEGEHCIHGHHGFGFC@BGJIIJIIJIFTHFFDDDDFFFF@MC:Z:100M      MD:Z:100      RG:Z:1719PC0017_51      NM:i:
0      MQ:i:60      AS:i:100      XS:i:0
HWI-ST354R:351:C0UPMACXX:6:2301:2151:15436      147      1      861414      60      100M      =      861307      -207      GGGGCTGGGAGTT
ACTCTCCCTCGCGAGCTGTTCCTCGCGGTTTTCAGGGTTTTCAGGATCGAGAGTCTTAACCTCACCCCTCGCGGTGTGCTGGAGGG      BDDBBCDC:CC>:+B@DBDDDBDDDDCDBDB?DDD
DDDDDEFFFEFFHHHHJJJJJJJJJJJJJJJTHFTHD1)JJJJJJJJHHHHHHFFFFFCCC      MC:Z:100M      MD:Z:100      RG:Z:1719PC0017_51      NM:i:
0      MQ:i:60      AS:i:100      XS:i:0
HWI-ST354R:351:C0UPMACXX:6:2103:12417:96772      83      1      861420      60      100M      =      861397      -123      GGGAGTTACTCTC
CCCTCGCGGAGCTGTTCCTCGCGGTTTTCAGGGTTTTCAGGATCGAGAGTCTTAACCTCACCCCTCGCGGTGTGCTGGAGGGAGCCTC      ?DCDC@::DBDDDDDDDDDDDDDDDBDBDBB=DD
DDEEDDDFFFFHHHHJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJHHHHHHFFFFFCCC      MC:Z:100M      MD:Z:100      RG:Z:1719PC0017_51      NM:i:
0      MQ:i:60      AS:i:100      XS:i:0
```

It is not often the case that the TLEN is positive for one end of the pair and negative for the other end.

TLEN: signed observed Template LENgth. If all segments are mapped to the same reference, the unsigned observed template length equals the number of bases from the leftmost mapped base to the rightmost mapped base. The leftmost segment has a plus sign and the rightmost has a minus sign. The sign of segments in the middle is undefined. It is set as 0 for single-segment template or when the information is unavailable.

Question #5

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannguy$ samtools view sample.sorted.bam | cut -f 9 | sort -n | head -1  
-234029071
```

Min TLEN is -234029071.

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannguy$ samtools view sample.sorted.bam | cut -f 9 | sort -n | tail -1  
216773490
```

Max TLEN is 216773490.

MOST FREQUENT TLEN is -155.

Step 1. Find the Most Frequent Value.

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannguy$ samtools view sample.sorted.bam | cut -f 9 | sort | uniq -c | awk '$1=$1  
' | cut -d' ' -f1 | sort -n | tail  
16955  
16958  
16973  
16975  
16988  
16991  
17054  
17057  
17069  
17072
```

Step.2 Find Most Frequent Value's associated TLEN

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannguy$ samtools view sample.sorted.bam | cut -f 9 | sort | uniq -c | awk '$1=$1  
' | awk -F " " ' $1==17072{print $0}'  
17072 -155
```

Question #6

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannguy$ samtools view -f 0x2 sample.sorted.bam | wc -l  
5003529
```

```
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannguy$ samtools view -F 0x2 sample.sorted.bam | wc -l  
22557
```

$5003529 + 22557 = 5,026,086$ Properly and Not Properly Paired (TOTAL)

$(5003529/5,026,086) \times 100 = 99.55\%$

Yes, this is reasonably high quality because of the high percentage of properly aligned paired.

Question #7

```
[Beagans-MacBook-Pro-2:IGV_2.4.14 beagannGuy$ samtools view -f 0x10 sample.sorted.bam | wc -l  
2513067
```

2513067 alignments are on the negative strand.

```
[Beagans-MacBook-Pro-2:IGV_2.4.14 beagannGuy$ samtools view -F 0x10 sample.sorted.bam | wc -l  
2513019  
Beagans-MacBook-Pro-2:IGV_2.4.14 beagannGuy$
```

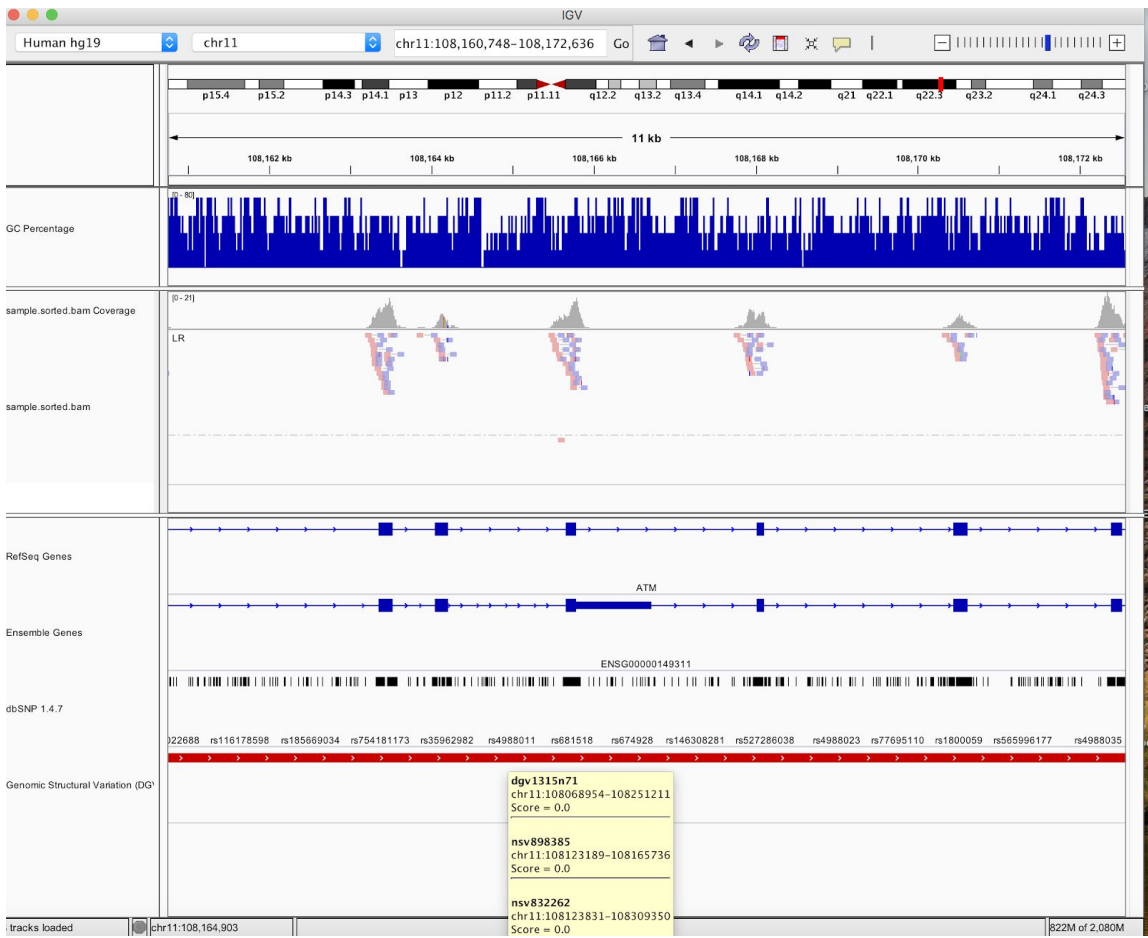
2513019 alignments are on the positive strand.

Question #8



Display read ends

Question #9



Display ATM region.