

## Homework 2 Screenshots & Commands

### Ornit Bhonkar

#### Prerequisites:

```
qsub -l -l select=1:ncpus=2:mem=4gb -l walltime=01:00:00 /bin/bash
```

```
cp /home/stavnaky/HW1_Input_Files_2025_2026/* $HOME
```

```
-rw----- 1 bornit users 68M Dec 27 17:18 PDX_P0.chr17.R1.fastq.gz
-rw----- 1 bornit users 73M Dec 27 17:18 PDX_P0.chr17.R2.fastq.gz
-rw----- 1 bornit users 68M Dec 27 17:18 PDX_P2.chr17.R1.fastq.gz
-rw----- 1 bornit users 73M Dec 27 17:18 PDX_P2.chr17.R2.fastq.gz
```

#### Q1:

```
REF=/home/stavnaky/ref_grch38/GRCh38.d1.vd1.fa
```

```
bwa mem $REF PDX_P0.chr17.R1.fastq.gz PDX_P0.chr17.R2.fastq.gz > PDX_P0.sam
```

```
bwa mem $REF PDX_P2.chr17.R1.fastq.gz PDX_P2.chr17.R2.fastq.gz > PDX_P2.sam
```

```
-rw----- 1 bornit users 1.1G Dec 27 17:31 PDX_P0.sam
-rw----- 1 bornit users 1.1G Dec 27 17:43 PDX_P2.sam
```

#### Q2:

```
samtools view -b -o PDX_P0.bam PDX_P0.sam
```

```
samtools view -b -o PDX_P2.bam PDX_P2.sam
```

```
ls -lh PDX_P0.sam PDX_P0.bam
```

```
ls -lh PDX_P2.sam PDX_P2.bam
```

```
[bornit@aluf ~]$ ls -lh PDX_P0.sam PDX_P0.bam
-rw-r--r-- 1 bornit users 168M Dec 27 17:49 PDX_P0.bam
-rw----- 1 bornit users 1.1G Dec 27 17:31 PDX_P0.sam
[bornit@aluf ~]$ ls -lh PDX_P2.sam PDX_P2.bam
-rw-r--r-- 1 bornit users 165M Dec 27 17:50 PDX_P2.bam
-rw----- 1 bornit users 1.1G Dec 27 17:43 PDX_P2.sam
```

For P0 932 MB was saved. For P2, 935 MB was saved.

### Q3:

```
samtools sort -o PDX_P0.sorted.bam PDX_P0.bam
```

```
samtools sort -o PDX_P2.sorted.bam PDX_P2.bam
```

```
-rw----- 1 bornit users 168M Dec 27 14:32 PDX_P0.sorted.bam
-rw----- 1 bornit users 165M Dec 27 14:33 PDX_P2.sorted.bam
```

### Q4:

```
samtools index PDX_P0.sorted.bam
```

```
samtools index PDX_P2.sorted.bam
```

```
PDX_P0.bam
PDX_P0.chr17.R1.fastq.gz
PDX_P0.chr17.R2.fastq.gz
PDX_P0.sam
PDX_P0.sorted.bam
PDX_P0.sorted.bam.bai
PDX_P2.bam
PDX_P2.chr17.R1.fastq.gz
PDX_P2.chr17.R2.fastq.gz
PDX_P2.sam
PDX_P2.sorted.bam
PDX_P2.sorted.bam.bai
```

```
-rw----- 1 bornit users 168M Dec 27 14:32 PDX_P0.sorted.bam
-rw----- 1 bornit users 2.2M Dec 27 14:35 PDX_P0.sorted.bam.bai
-rw----- 1 bornit users 165M Dec 27 14:33 PDX_P2.sorted.bam
-rw----- 1 bornit users 2.1M Dec 27 14:36 PDX_P2.sorted.bam.bai
```

### Q5:

```
bcftools mpileup -f "$REF" PDX_P0.sorted.bam PDX_P2.sorted.bam >
PDX_P0_P2.pileup.vcf
```

```
PDX_P0_P2.pileup.vcf
PDX_P0.sorted.bam
PDX_P0.sorted.bam.bai
PDX_P2.sorted.bam
PDX_P2.sorted.bam.bai
```

```
-rw----- 1 bornit users 7.4G Dec 27 15:16 PDX.P0_P2.pileup.vcf
```

## Q6:

```
bcftools call -mv -Ov -o PDX.variants.vcf PDX.P0_P2.pileup.vcf
```

```
PDX.P0_P2.pileup.vcf  
PDX_P0.sorted.bam  
PDX_P0.sorted.bam.bai  
PDX_P2.sorted.bam  
PDX_P2.sorted.bam.bai  
PDX.variants.vcf
```

```
-rw----- 1 bornit users 22M Dec 27 15:20 PDX.variants.vcf
```

```
bcftools view -v snps -Ov -o PDX.snps.vcf PDX.variants.vcf
```

```
-rw----- 1 bornit users 17M Dec 27 15:23 PDX.snps.vcf
```

## Q7:

```
bcftools view -i 'GT[0]!="0/0" && GT[0]!="./."' -Ov -o PDX.germline.vcf PDX.snps.vcf
```

```
PDX.germline.vcf  
PDX.P0_P2.pileup.vcf  
PDX_P0.sorted.bam  
PDX_P0.sorted.bam.bai  
PDX_P2.sorted.bam  
PDX_P2.sorted.bam.bai  
PDX.snps.vcf  
PDX.variants.vcf
```

```
-rw----- 1 bornit users 12M Dec 27 15:27 PDX.germline.vcf
```

Count SNPs before and after:

```
grep -vc '^#' PDX.snps.vcf
```

```
grep -vc '^#' PDX.germline.vcf
```

```
[bornit@aluf3 ~]$ grep -vc '^#' PDX.snps.vcf
116517
[bornit@aluf3 ~]$ grep -vc '^#' PDX.germline.vcf
77472
```

## Q8:

```
bcftools view -i 'INFO/BQBZ==0 && INFO/DP>8 && INFO/MQ>=60 && INFO/AN==4' -
Ov -o PDX.Q8.filtered.vcf PDX.germline.vcf
-rw----- 1 bornit users 1.2M Dec 27 16:22 PDX.Q8.filtered.vcf
```

```
bcftools stats PDX.germline.vcf > stats.Q8.before.txt
```

```
bcftools stats PDX.Q8.filtered.vcf > stats.Q8.after.txt
```

```
stats.Q8.after.txt
stats.Q8.before.txt
```

```
grep '^SN' stats.Q8.before.txt | head -n 30
```

```
grep '^SN' stats.Q8.after.txt | head -n 30
```

```
[bornit@aluf3 ~]$ grep '^SN' stats.Q8.before.txt | head -n 30
SN      0      number of samples:      2
SN      0      number of records:    77472
SN      0      number of no-ALTs:      0
SN      0      number of SNPs: 77472
SN      0      number of MNPs: 0
SN      0      number of indels:      0
SN      0      number of others:      0
SN      0      number of multiallelic sites: 44
SN      0      number of multiallelic SNP sites: 44
[bornit@aluf3 ~]$ grep '^SN' stats.Q8.after.txt | head -n 30
SN      0      number of samples:      2
SN      0      number of records:    6352
SN      0      number of no-ALTs:      0
SN      0      number of SNPs: 6352
SN      0      number of MNPs: 0
SN      0      number of indels:      0
SN      0      number of others:      0
SN      0      number of multiallelic sites: 3
SN      0      number of multiallelic SNP sites: 3
```

### Further checking:

```
bcftools view -h PDX.germline.vcf | grep 'BQBZ'
```

```
[bornit@aluf3 ~]$ bcftools view -h PDX.germline.vcf | grep 'BQBZ'  
##INFO=<ID=BQBZ,Number=1,Type=Float,Description="Mann-Whitney U-z test of Base Quality Bias (closer to 0 is better)">
```

```
bcftools query -f '%BQBZ\n' PDX.Q8.filtered.vcf | sort | uniq -c
```

```
[bornit@aluf3 ~]$ bcftools query -f '%BQBZ\n' PDX.Q8.filtered.vcf | sort | uniq -c  
6352 0
```

```
grep -A10 '^DP' stats.Q8.before.txt | head
```

```
[bornit@aluf3 ~]$ grep -A10 '^DP' stats.before.txt | head  
DP 0 1 0 0.000000 3639 4.697181  
DP 0 2 0 0.000000 9956 12.851095  
DP 0 3 0 0.000000 5147 6.643691  
DP 0 4 0 0.000000 9599 12.390283  
DP 0 5 0 0.000000 3037 3.920126  
DP 0 6 0 0.000000 6072 7.837670  
DP 0 7 0 0.000000 2213 2.856516  
DP 0 8 0 0.000000 4965 6.408767  
DP 0 9 0 0.000000 1818 2.346654  
DP 0 10 0 0.000000 4147 5.352902
```

```
grep -A10 '^DP' stats.Q8.after.txt | head
```

```
[bornit@aluf3 ~]$ grep -A10 '^DP' stats.after.txt | head  
DP 0 9 0 0.000000 7 0.110202  
DP 0 10 0 0.000000 952 14.987406  
DP 0 11 0 0.000000 6 0.094458  
DP 0 12 0 0.000000 851 13.397355  
DP 0 13 0 0.000000 2 0.031486  
DP 0 14 0 0.000000 766 12.059194  
DP 0 15 0 0.000000 2 0.031486  
DP 0 16 0 0.000000 624 9.823678  
DP 0 18 0 0.000000 536 8.438287  
DP 0 19 0 0.000000 1 0.015743
```

```
bcftools query -f '%MQ\n' PDX.Q8.filtered.vcf | sort | uniq -c
```

```
bcftools query -f '%MQ\n' PDX.germline.vcf | sort | uniq -c
```

```
[bornit@aluf3 ~]$ bcftools query -f '%MQ\n' PDX.Q8.filtered.vcf | sort | uniq -c
 6352 60
[bornit@aluf3 ~]$ bcftools query -f '%MQ\n' PDX.germline.vcf | sort | uniq -c
 60 0
 4 1
 16 10
 6 11
 29 12
 35 13
 35 14
 48 15
 36 16
 42 17
 41 18
 30 19
 3 2
 94 20
 74 21
 55 22
 93 23
 85 24
 67 25
 74 26
 114 27
 131 28
 115 29
 2 3
 245 30
 128 31
 114 32
 132 33
 159 34
 142 35
 168 36
 130 37
 135 38
 159 39
 4 4
 502 40
 229 41
 250 42
 241 43
 189 44
 427 45
 383 46
 393 47
 417 48
 329 49
 3 5
 636 50
 542 51
 532 52
 679 53
 672 54
 772 55
 884 56
 1144 57
 1468 58
 1884 59
 5 6
 62052 60
 7 7
 12 8
 15 9
```

```
bcftools query -f '%AN\n' PDX.germline.vcf | sort | uniq -c
[bornit@aluf3 ~]$ bcftools query -f '%AN\n' PDX.germline.vcf | sort | uniq -c
 19212 2
 58260 4
```

```
bcftools query -f '%AN\n' PDX.Q8.filtered.vcf | sort | uniq -c
```

```
[bornit@aluf3 ~]$ bcftools query -f '%AN\n' PDX.Q8.filtered.vcf | sort | uniq -c
 6352 4
```

Q9:

```
bcftools view -i '(GT[0]=="0/1" || GT[0]=="1/0") && (GT[1]=="0/0" || GT[1]=="1/1")' -Ov -o
PDX.Q9.LOH.vcf PDX.Q8.filtered.vcf
```

```
bcftools view -i '(GT[0]=="0/1" || GT[0]=="1/0") && GT[1]=="0/0"' -Ov -o
PDX.Q9.LOH_ref.vcf PDX.Q8.filtered.vcf
```

```
bcftools view -i '(GT[0]=="0/1" || GT[0]=="1/0") && GT[1]=="1/1"' -Ov -o PDX.Q9.LOH_alt.vcf
PDX.Q8.filtered.vcf
```

```
-rw----- 1 bornit users 148K Dec 27 16:38 PDX.Q9.LOH_alt.vcf
-rw----- 1 bornit users 1.1M Dec 27 16:38 PDX.Q9.LOH_ref.vcf
-rw----- 1 bornit users 1.1M Dec 27 16:34 PDX.Q9.LOH.vcf
```

```
echo "LOH total:" $(grep -vc '^#' PDX.Q9.LOH.vcf)
```

```
echo "LOH to ref:" $(grep -vc '^#' PDX.Q9.LOH_ref.vcf)
```

```
echo "LOH to alt:" $(grep -vc '^#' PDX.Q9.LOH_alt.vcf)
```

```
[bornit@aluf3 ~]$ echo "LOH total:" $(grep -vc '^#' PDX.Q9.LOH.vcf)
echo "LOH to ref:" $(grep -vc '^#' PDX.Q9.LOH_ref.vcf)
echo "LOH to alt:" $(grep -vc '^#' PDX.Q9.LOH_alt.vcf)
LOH total: 5633
LOH to ref: 5588
LOH to alt: 45
```

Q10:

```
bcftools view -H -i 'CHROM="chr17" && POS=78392462' PDX.variants.vcf
```

```
[bornit@aluf ~]$ bcftools view -H -i 'CHROM="chr17" && POS=78392462' PDX.variants.vcf
chr17 78392462 . A G 10.1829 . DP=18;SGB=-0.516033;R
PBZ=-1.36698;MQBZ=0;MQSBZ=0;BQBZ=0;SCBZ=1.54919;FS=0;MQ0F=0;AC=1;AN=4;DP4=2,6,0,1;MQ=
60 GT:PL 0/1:45,0,197 0/0:0,12,212
```

bcftools view -H -i 'CHROM="chr17" && POS=78392462' PDX.Q9.LOH.vcf

```
[bornit@aluf3 ~]$ bcftools view -H -i 'CHROM="chr17" && POS=78392462' PDX.Q9.LOH.vcf
chr17 78392462 . A G 10.1829 . DP=18;SGB=-0.516033;R
PBZ=-1.36698;MQBZ=0;MQSBZ=0;BQBZ=0;SCBZ=1.54919;FS=0;MQ0F=0;AC=1;AN=4;DP4=2,6,0,1;MQ=
60 GT:PL 0/1:45,0,197 0/0:0,12,212
```

The answer is A -> G

## Q11:

bcftools stats PDX.snps.vcf > stats.Q11.snps.txt

grep '^ST' stats.Q11.snps.txt

```
[bornit@aluf ~]$ bcftools stats PDX.snps.vcf > stats.Q11.snps.txt
[bornit@aluf ~]$ grep '^ST' stats.Q11.snps.txt
ST 0 A>C 4030
ST 0 A>G 17504
ST 0 A>T 3350
ST 0 C>A 7966
ST 0 C>G 4938
ST 0 C>T 20237
ST 0 G>A 20962
ST 0 G>C 4967
ST 0 G>T 7759
ST 0 T>A 3488
ST 0 T>C 17443
ST 0 T>G 3920
```

bcftools stats PDX.Q9.LOH.vcf > stats.Q11.loh.txt

grep '^ST' stats.Q11.loh.txt

```
[bornit@aluf ~]$ bcftools stats PDX.Q9.L0H.vcf > stats.Q11.loh.txt
[bornit@aluf ~]$ grep '^ST' stats.Q11.loh.txt
ST      0      A>C      85
ST      0      A>G      652
ST      0      A>T      128
ST      0      C>A      639
ST      0      C>G      168
ST      0      C>T      1148
ST      0      G>A      1191
ST      0      G>C      162
ST      0      G>T      567
ST      0      T>A      122
ST      0      T>C      685
ST      0      T>G      86
```

The most common substitution is G>A.

### All files:

```
[bornit@aluf ~]$ ll -h PDX*
-rw----- 1 bornit users 12M Dec 27 15:27 PDX.germline.vcf
-rw-r--r-- 1 bornit users 168M Dec 27 17:49 PDX_P0.bam
-rw----- 1 bornit users 68M Dec 27 17:18 PDX_P0.chr17.R1.fastq.gz
-rw----- 1 bornit users 73M Dec 27 17:18 PDX_P0.chr17.R2.fastq.gz
-rw----- 1 bornit users 7.4G Dec 27 15:16 PDX_P0_P2.pileup.vcf
-rw----- 1 bornit users 1.1G Dec 27 17:31 PDX_P0.sam
-rw----- 1 bornit users 168M Dec 27 14:32 PDX_P0.sorted.bam
-rw----- 1 bornit users 2.2M Dec 27 14:35 PDX_P0.sorted.bam.bai
-rw-r--r-- 1 bornit users 165M Dec 27 17:50 PDX_P2.bam
-rw----- 1 bornit users 68M Dec 27 17:18 PDX_P2.chr17.R1.fastq.gz
-rw----- 1 bornit users 73M Dec 27 17:18 PDX_P2.chr17.R2.fastq.gz
-rw----- 1 bornit users 1.1G Dec 27 17:43 PDX_P2.sam
-rw----- 1 bornit users 165M Dec 27 14:33 PDX_P2.sorted.bam
-rw----- 1 bornit users 2.1M Dec 27 14:36 PDX_P2.sorted.bam.bai
-rw----- 1 bornit users 1.2M Dec 27 16:22 PDX_Q8.filtered.vcf
-rw----- 1 bornit users 148K Dec 27 16:38 PDX_Q9.L0H_alt.vcf
-rw----- 1 bornit users 1.1M Dec 27 16:38 PDX_Q9.L0H_ref.vcf
-rw----- 1 bornit users 1.1M Dec 27 16:34 PDX_Q9.L0H.vcf
-rw----- 1 bornit users 17M Dec 27 15:23 PDX_snps.vcf
-rw----- 1 bornit users 22M Dec 27 15:20 PDX_variants.vcf
```