

Module 1 Exam

17/17 points (100%)

Quiz, 17 questions

 **Congratulations! You passed!**[Next Item](#)1 / 1
points

1.

How many chromosomes are there in the genome?

Correct Response

You can find this by extracting and counting the lines representing fasta headers from the 'apple.genome' file ('%' denotes the terminal prompt):

```
1 % grep -c ">" apple.genome
```

1 / 1
points

2.

How many genes?

Correct Response

In the file 'apple.genes', each line represents a different transcript variant (n.b., this is the same as 'transcript'). The columns are:

```
1 geneID transcriptID chrom strand transcriptStart transcriptEnd
2 transcriptStructure
3 MDP0000303933 MDP0000303933 chr1 - 4276 5447 (4276-4368,4423-4542,4733-4911,5321
  -5447)
```

where 'transcriptStructure' gives the list of all exons' coordinates.

To determine the number of genes, one has to find the number

Module 1 Exam

17/17 points (100%)

Quiz, 17 questions

```
1 % cut -f1 apple.genes | sort -u | wc -l
```

and to determine the number of transcripts, one has to find the number of distinct names in column 2:

```
1 % cut -f2 apple.genes | sort -u | wc -l
```



1 / 1
points

3.

How many transcript variants?

5456

Correct Response

In the file 'apple.genes', each line represents a different transcript variant (n.b., this is the same as 'transcript'). The columns are:

```
1 geneID transcriptID chrom strand transcriptStart transcriptEnd
2 transcriptStructure
3 MDP0000303933 MDP0000303933 chr1 - 4276 5447 ((4276-4368,4423-4542,4733-4911,5321
-5447))
```

where 'transcriptStructure' gives the list of all exons' coordinates.

To determine the number of genes, one has to find the number of distinct names in column 1:

```
1 % cut -f1 apple.genes | sort -u | wc -l
```

and to determine the number of transcripts, one has to find the number of distinct names in column 2:

```
1 % cut -f2 apple.genes | sort -u | wc -l
```



1 / 1
points

4.

How many genes have a single splice variant?

Module 1 Exam

17/17 points (100%)

Quiz, 17 questions

5450

Correct Response

The number of splice variants for a gene is given by the number of lines (transcripts) that have that geneID in column 1. To answer the question, we first determine the number of lines for each gene, then extract only those genes that have exactly one line:

```
1 % cut -f1 apple.genes | sort | uniq -c | grep -c " 1 "
```

1 / 1
points

5.

How many genes have 2 or more splice variants?

3

Correct Response

Using the same rationale as above, we now count the number of genes that have more than one line:

```
1 % cut -f1 apple.genes | sort | uniq -c | grep -v " 1 " | wc -l
```

1 / 1
points

6.

How many genes are there on the '+' strand?

2662

Correct Response

The strand information is stored in column 4. We need to identify the strand for each gene, i.e. we need to extract columns 1 and 4. The same information might appear on multiple lines, since

Module 1 Exam

Quiz, 17 questions

there can be multiple transcripts for a gene, all of which are located on the same strand. Hence, we need to make sure we only count this information once. Lastly, we extract the strand column and count the number of entries of each type ('+' or '-');

17/17 points (100%)

```
1 % cut -f1,4 apple.genes | sort -u | cut -f2 | sort | uniq -c
```



1 / 1
points

7.
How many genes are there on the '-' strand?

2791

Correct Response

The strand information is stored in column 4. We need to identify the strand for each gene, i.e. we need to extract columns 1 and 4. The same information might appear on multiple lines, since there can be multiple transcripts for a gene, all of which are located on the same strand. Hence, we need to make sure we only count this information once. Lastly, we extract the strand column and count the number of entries of each type ('+' or '-');

```
1 % cut -f1,4 apple.genes | sort -u | cut -f2 | sort | uniq -c
```



1 / 1
points

8.
How many genes are there on chromosome chr1?

1624

Correct Response

Similarly to the question above, the relevant information is now listed in columns 1 (gene) and 3 (chrom). Performing the same type of workflow will give us the number of distinct genes for each chromosome:

```
1 % cut -f1,3 apple.genes | sort -u | cut -f2 | sort | uniq -c
```

Module 1 Exam

17/17 points (100%)

Quiz, 17 questions

1 / 1
points

9.

How many genes are there on each chromosome chr2?

Correct Response

Similarly to the question above, the relevant information is now listed in columns 1 (gene) and 3 (chrom). Performing the same type of workflow will give us the number of distinct genes for each chromosome:

```
1 % cut -f1,3 apple.genes | sort -u | cut -f2 | sort | uniq -c
```

1 / 1
points

10.

How many genes are there on each chromosome chr3?

Correct Response

Similarly to the question above, the relevant information is now listed in columns 1 (gene) and 3 (chrom). Performing the same type of workflow will give us the number of distinct genes for each chromosome:

```
1 % cut -f1,3 apple.genes | sort -u | cut -f2 | sort | uniq -c
```

1 / 1
points

11.

How many transcripts are there on chr1?

Module 1 Exam

17/17 points (100%)

Quiz, 17 questions

1625

Correct Response

The relevant information is in columns 2 and 3. As for Q6, we extract the information in columns 2 and 3, retaining only a single copy for each. Note that unlike with genes, where the same gene ID could appear on multiple lines, in this case each line corresponds to a different transcript, so there will not be any duplicate lines. We can therefore use a simplified workflow:

```
1 % cut -f2,3 apple.genes | cut -f2 | sort | uniq -c
```

N.b.: You can try this with 'sort -u' interspersed in the command and verify that you obtain the same result:

```
1 % cut -f2,3 apple.genes | sort -u | cut -f2 | sort | uniq -c
```

1 / 1
points

12.

How many transcripts are there on chr2?

2059

Correct Response

The relevant information is in columns 2 and 3. As for Q6, we extract the information in columns 2 and 3, retaining only a single copy for each. Note that unlike with genes, where the same gene ID could appear on multiple lines, in this case each line corresponds to a different transcript, so there will not be any duplicate lines. We can therefore use a simplified workflow:

```
1 % cut -f2,3 apple.genes | cut -f2 | sort | uniq -c
```

N.b.: You can try this with 'sort -u' interspersed in the command and verify that you obtain the same result:

```
1 % cut -f2,3 apple.genes | sort -u | cut -f2 | sort | uniq -c
```

Module 1 Exam

17/17 points (100%)

Quiz, 17 questions

1 / 1
points

13.

How many transcripts are there on chr3?

1772

Correct Response

The relevant information is in columns 2 and 3. As for Q6, we extract the information in columns 2 and 3, retaining only a single copy for each. Note that unlike with genes, where the same gene ID could appear on multiple lines, in this case each line corresponds to a different transcript, so there will not be any duplicate lines. We can therefore use a simplified workflow:

```
1 % cut -f2,3 apple.genes | cut -f2 | sort | uniq -c
```

N.b.: You can try this with 'sort -u' interspersed in the command and verify that you obtain the same result:

```
1 % cut -f2,3 apple.genes | sort -u | cut -f2 | sort | uniq -c
```

1 / 1
points

14.

How many genes are in common between condition A and condition B?

2410

Correct Response

The gene information is contained in column 1 in the files 'apple.condition*'. To answer the question, first create a list of the genes present in the file 'apple.conditionA', and similarly for 'apple.conditionB'. Second, sort both files. Lastly, use the command 'comm' to compare the entries in the two sorted files:

```
1 % cut -f1 apple.conditionA | sort -u > apple.condA.sorted.genes
2 % cut -f1 apple.conditionB | sort -u > apple.condB.sorted.genes
3 % comm -1 -2 apple.condA.sorted.genes apple.condB.sorted.genes | wc -l
```

Module 1 Exam

17/17 points (100%)

Quiz, 17 questions

1 / 1
points

15.

How many genes are specific to condition A?

1205

Correct Response

This question refers exclusively to the comparison between A and B. We use the same 'comm' command, but now change the parameters to display, for the first question, only those entries that are not shared ('-3') and are not present in condition B ('-2'), and for the second question, those entries that are not shared and are not present in condition A ('-1'):

```
1 % comm -3 -2 apple.condA.sorted.genes apple.condB.sorted.genes | wc -l  
2 % comm -3 -1 apple.condA.sorted.genes apple.condB.sorted.genes | wc -l
```

1 / 1
points

16.

How many genes are specific to condition B?

1243

Correct Response

This question refers exclusively to the comparison between A and B. We use the same 'comm' command, but now change the parameters to display, for the first question, only those entries that are not shared ('-3') and are not present in condition B ('-2'), and for the second question, those entries that are not shared and are not present in condition A ('-1'):

```
1 % comm -3 -2 apple.condA.sorted.genes apple.condB.sorted.genes | wc -l  
2 % comm -3 -1 apple.condA.sorted.genes apple.condB.sorted.genes | wc -l
```

1 / 1
points

17.

How many genes are in common to all three conditions?

Module 1 Exam

17/17 points (100%)

Quiz, 17 questions

1608

Correct Response

This question cannot be answered by a single pairwise comparison, as above. However, we can construct the list of genes for each of the three conditions as shown above, then determine which genes occur in all three files. For this last step, one solution is to concatenate the (sorted) genes files for the three conditions, determine the counts for each gene, and then determine the number of entries with the count '3':

```
1 % cut -f1 apple.conditionC | sort -u > apple.condC.sorted.genes-  
2 % cat apple.cond{A,B,C}.sorted.genes | sort | uniq -c | grep -c " 3 "
```

Potential pitfall: each gene should be listed just once in each file.

