Assignment 2

May 15, 2017

0. Load the data with percent overlap/non-overlap per chromosome

```
df <- read.table(file="coverage.txt")
```

Part 1

(a)What are the first five genomic nucleotides from the first exon of this transcript?

5' UTR: TTTCC
First coding exon: TACCA

(b) Look at the raw mRNA sequence of AK002007, from the database it actually comes from. What are the first five nucleotides?

5' UTR: AAACC
First coding exon:ATGGT

(c) How do you explain the discrepancy (maximum 5 lines)?

The database contains DNA sequences derived from mRNA squences (cDNA). Both sequences are complementary to one another.

Part 2

(a)

```
## Warning in order(as.numeric(substring(df$V1, 4))): NAs introduced by
## coercion
```

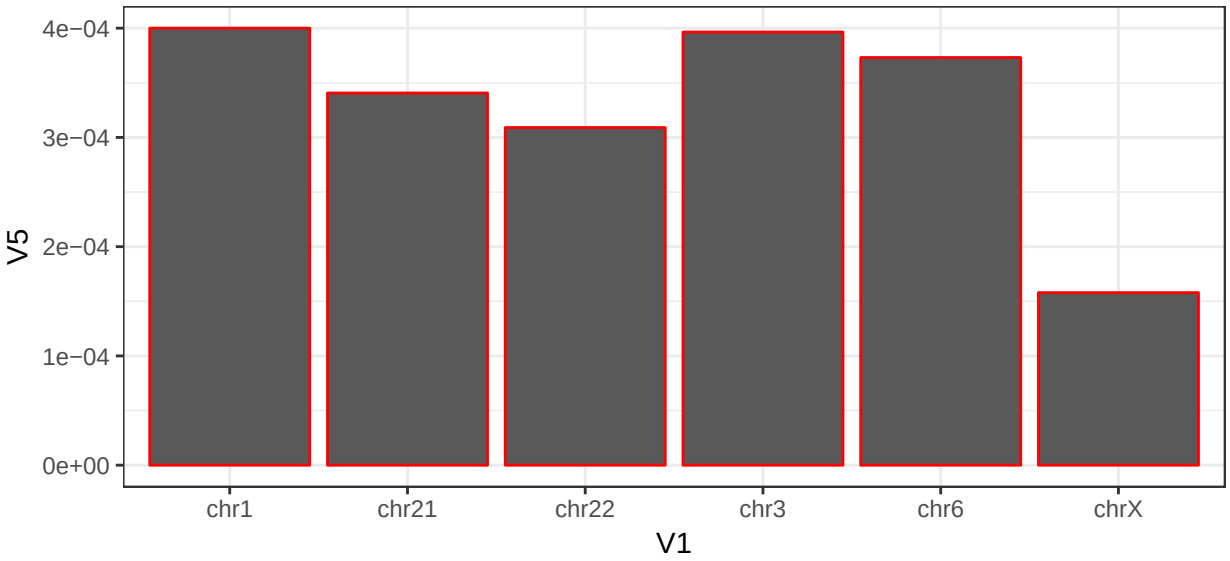
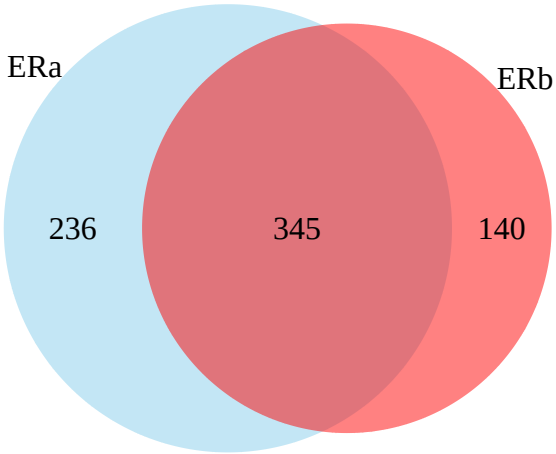


Figure 1. A *Histogram showing the distribution of the exon counts. Even though most of the genes contain less than 60 exons, as many as 150 may be found in some of them.* **B** *Detail for genes with max. 20 exons. The mode can be visualized at 3-5 exons per gene (max found at 4). The number of exons per gene decreases steadily beyond it.*




```
## (polygon[GRID.polygon.42], polygon[GRID.polygon.43], polygon[GRID.polygon.44], polygon[GRID.polygon.4
```

8. Appendix

Command line entries

bedtools genomecov -i ERa_hg18.bed -g genome.txt > coverage.txt bedtools intersect -a ERa_hg18.bed -b ERb_hg18.bed -c > AtoBoverlap.bed bedtools intersect -a ERb_hg18.bed -b ERa_hg18.bed -c > BtoAoverlap.bed