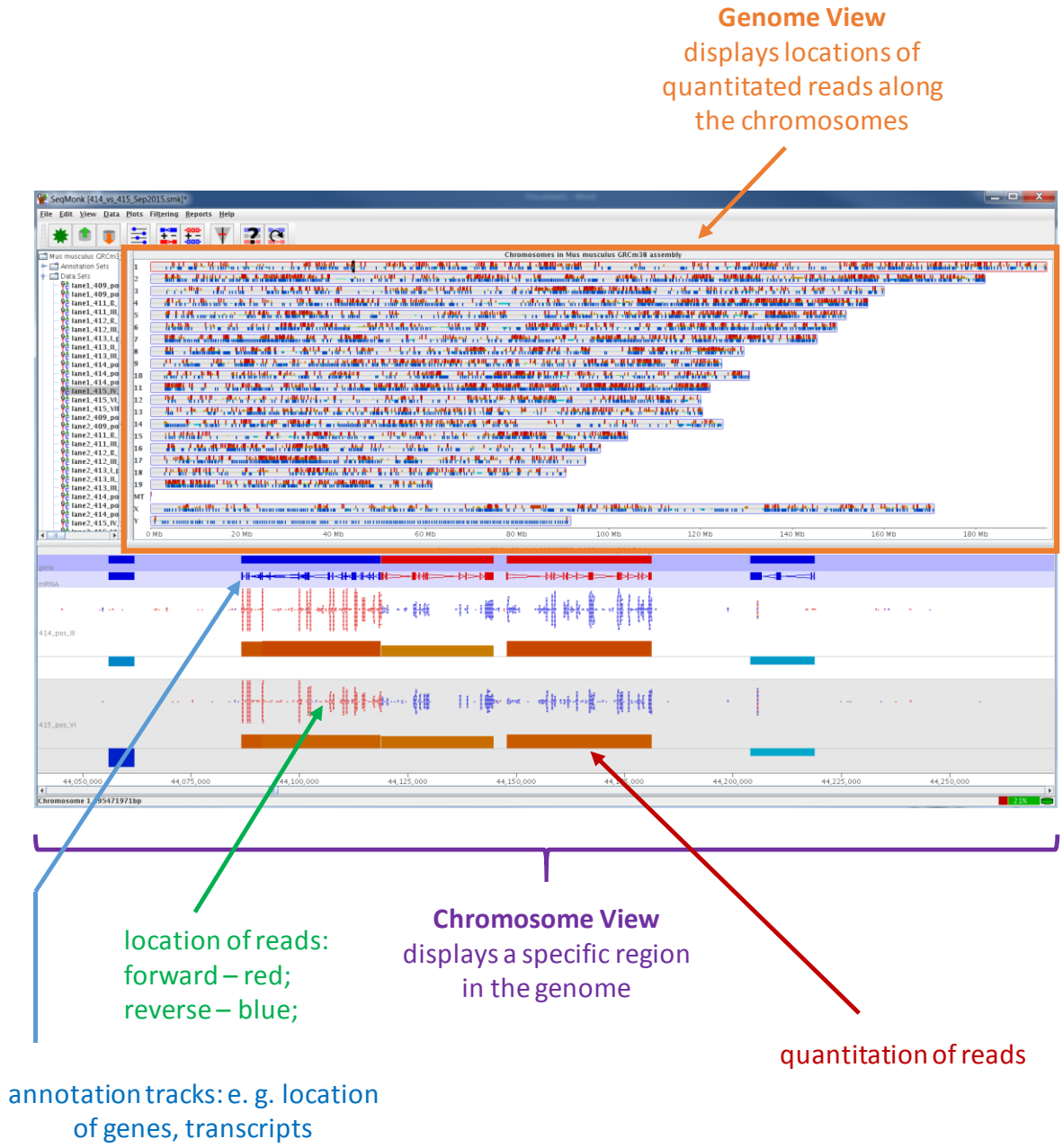
Exercises: Explore Before You Analyse

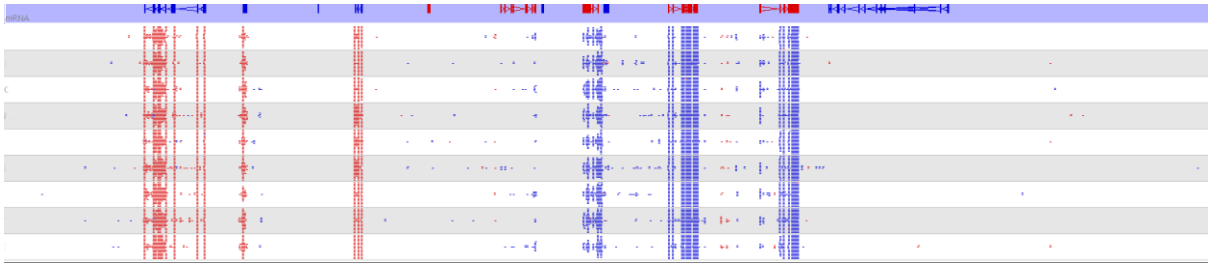
For the following exercises you will be given a series of screenshots from Seqmonk which is a tool to visualise and analyse high-throughput mapped sequence data. Here are some quick explanations on what you can see:



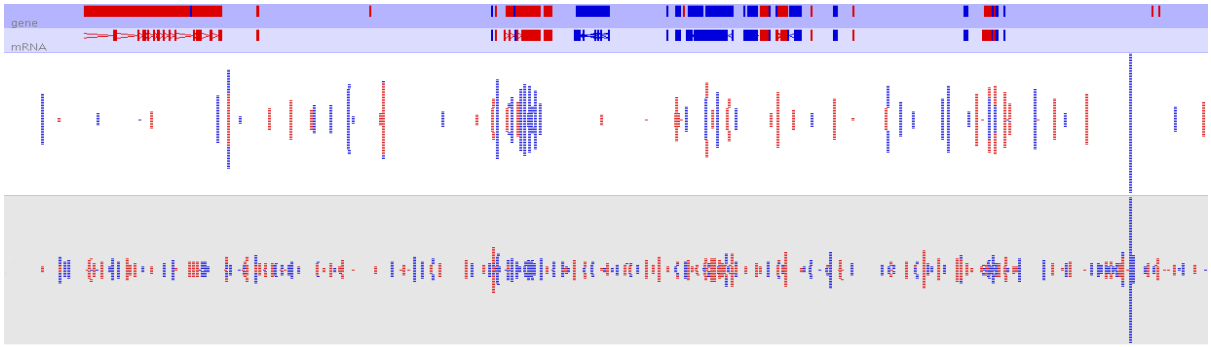
Here are some real life examples of RNA-Seq data. Look at each one of them (location, number and direction of reads) and answer the following questions:

1) It there anything obviously wrong with the data?
2) Would the data be quantified by an analysis package? Would it give you a sensible answer?
3) If there is something wrong, what could you do to fix it?

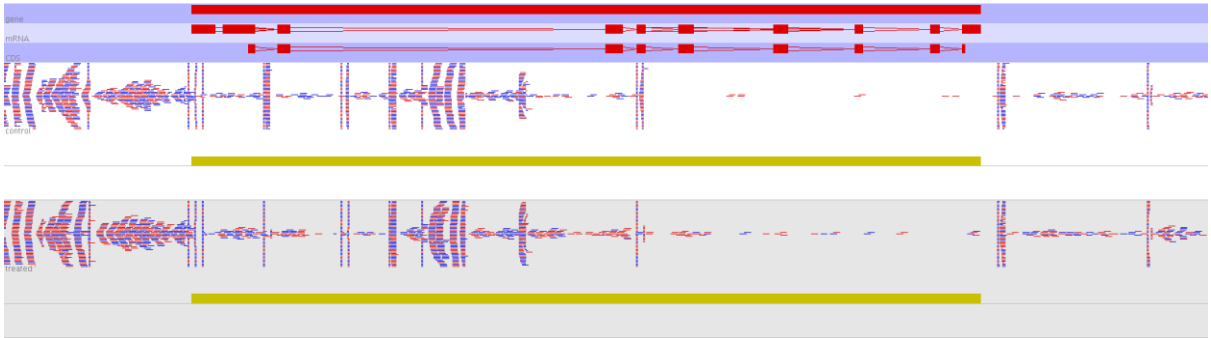
Example 1



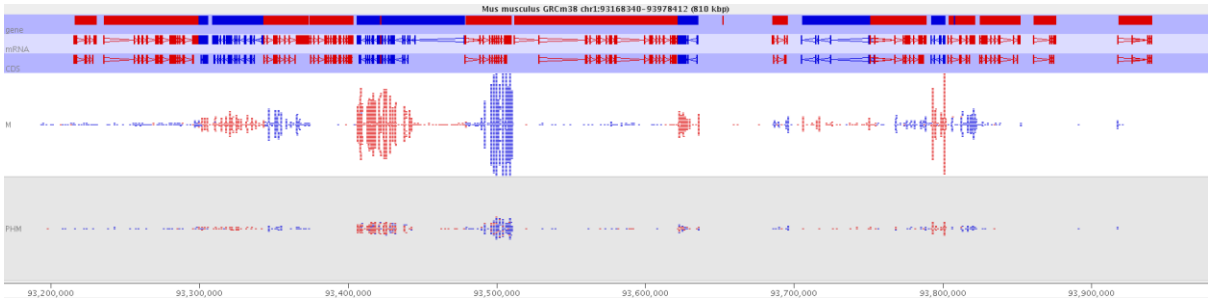
Example 2



Example 3



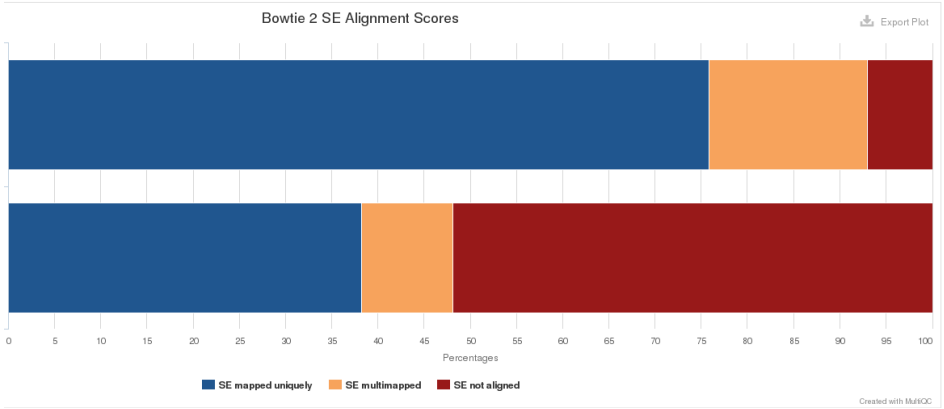
Example 4



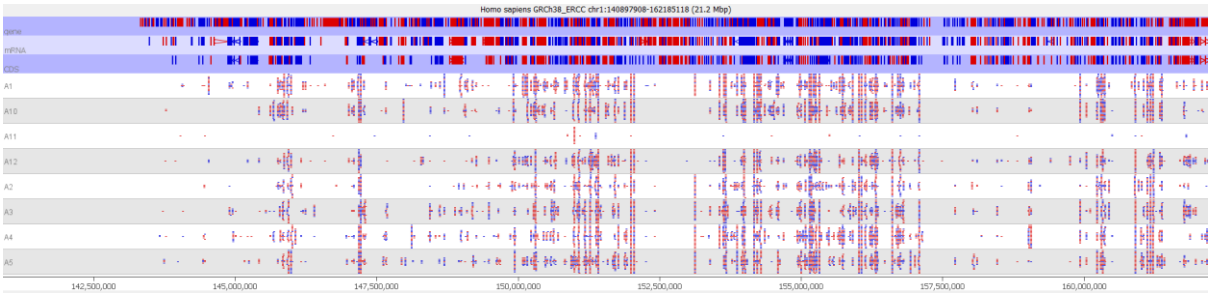
Additional information: Mapping efficiencies for the data shown in the screenshot above

Sample M

Sample PHM



Example 5



The next three examples show an RNA-Seq QC plot (Seqmonk) which provides summarised information on the location of reads. Where would you expect reads to be located? Are there any red flags?

The following metrics are given:

Percent in Gene: What percentage of reads fall into genes. A low value here would indicate potential DNA contamination

Percent in Exons: What percentage of reads in genes fall into exons. A low value here would indicate either DNA contamination, or possibly just a high proportion of unprocessed transcripts

Percent in rRNA: What percentage of reads fall into rRNA. This will only account for mapped reads falling into annotated rRNA features, so will normally be much less than the actual amount of rRNA in the starting library

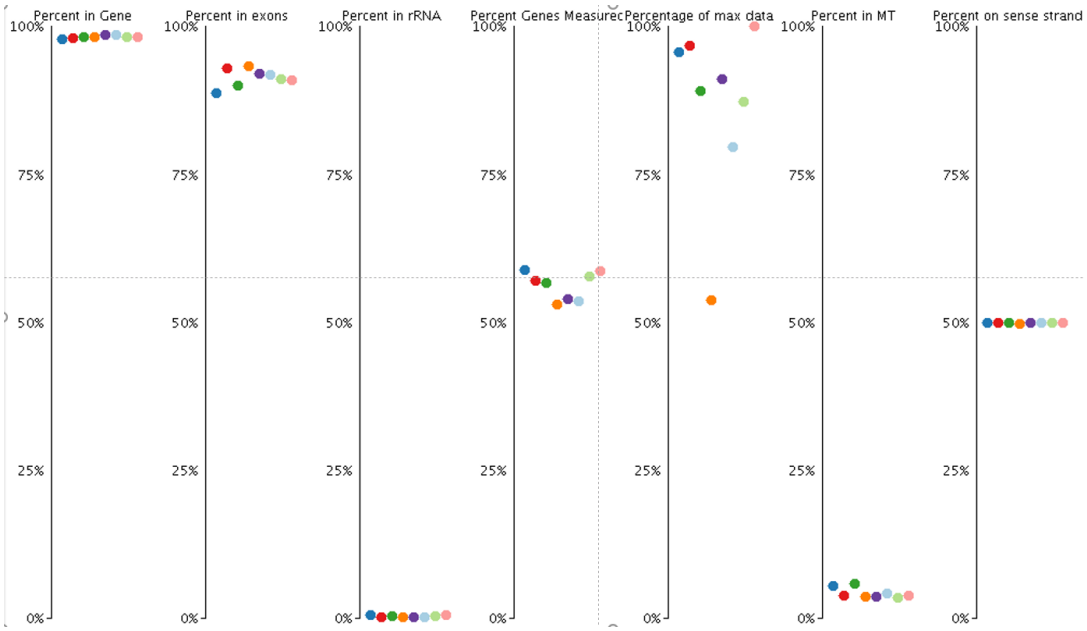
Percent Genes Measured: What percentage of genes have any exonic reads at all.

Percentage of max data: The highest number of reads in a data set is set to 100%. What percentage of reads do the other samples have.

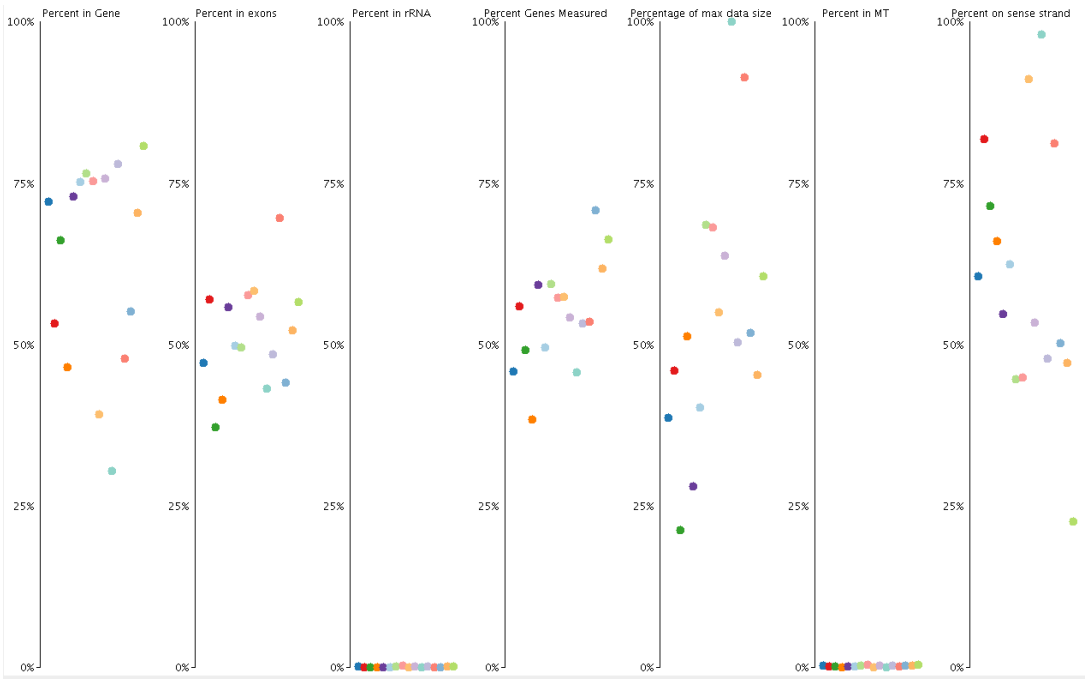
Percent in MT: What percentage of reads fall into the Mitochondrion

Percent on sense strand: What percentage of reads fall onto the sense strand - depending on the type of library you've made this value could be expected to be near 0%, 100% or 50%

Example 6



Example 7



Example 8