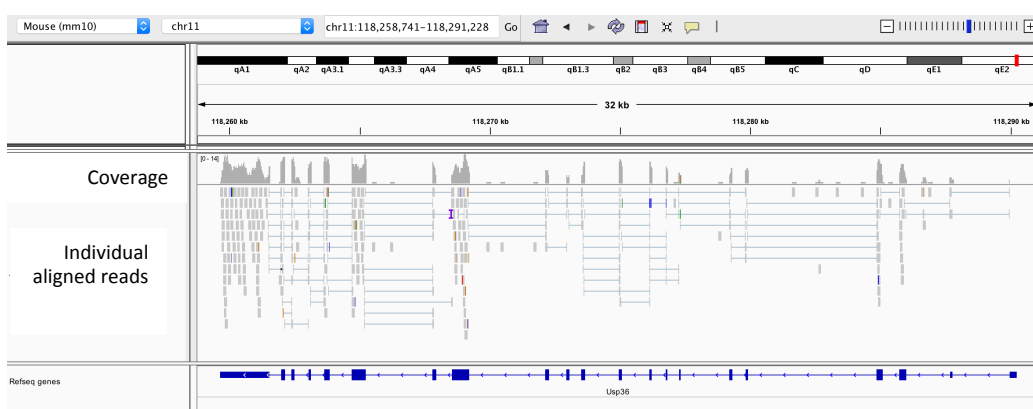


Typical RNA sequencing analysis workflow

fastq file:

```
@HWI-M01141:63:A4NDL:1:1101:14849:1418 1:N:0:TATAGCGAGACACCGT
NACGAAGGTGCAAGCGTTACTCGAATTACTGGCGTAAGCGTGCAGTGGTGGTTT
+
#>>>A??AFAA1BGEGGAAGGCA8FF1D2BCF/EEG/DBEE/E?GAEEFAEFAEFG1
@HWI-M01141:63:A4NDL:1:1101:13802:1421 1:N:0:TATAGCGAGACACCGT
NACGAAGGTGCAAGCGTTACTCGAATTACTGGCGTAAGCGCAGCGCGTTTGT
+
#>>>AAABBBABGGGGGG?FHCGGGGHHHHHHGGGGH
@HWI-M01141:63:A4NDL:1:1101:15928:1426 1:N:0:
NACGTAGGTGCGAGCGTTACTCGAATTACTGGCGTAAM
+
#>>>AABFB@FBBGGGGGGGGHCGGGFHHHHHHGGGGH
@HWI-M01141:63:A4NDL:1:1101:14861:1431 1:N:0:
NACGAAGGTGCAAGCGTTACTCGAATTACTGGCGTAAM
+
#>>>AAAABBFABGGGGGGEGHCGEGFFHHHHHHGGGGH
@HWI-M01141:63:A4NDL:1:1101:15264:1465 1:N:0:
NACGTAGGTGCGAGCGTTACTCGAATTACTGGCGTAAM
+
@FORJUSP02AJWD1
CCGTC AATT CATT TAAGTTT AACCTT GCGGCC TACTCCC AGGCGGT
+
AAAAAAAAAAAAA::99@:::??@:::FFAAAAACCAA:::BB@?A?
```

Filter quality
Align to ref. genome



count reads
per gene

Downstream statistical analysis:
R: import counts table

Genome browser view of sequencing reads (small grey boxes) from RNA seq data, aligning to the exons of gene *Usp36* (in blue). The top grey histogram shows the coverage (some of all reads) for each position along chromosome 11.

In a typical RNA sequencing workflow, one of the outputs after sequencing is a fastq file. One fastq file is usually obtained for each sample, containing all of the raw sequences for that sample. A fastq file is basically a text file, with 4 rows for each raw sequence. The first row contains the name of the raw sequence, the second row gives the actual A/T/C/G nucleotides sequenced for that DNA strand, the third row is an empty row, and the last row contains quality information for every nucleotide in the sequence.

The next step involves filtering the raw sequences to remove the ones that contain low quality bases.

To identify the origin of each sequence, all are aligned to the reference genome. This allows us to know from which genes the sequences originate. All sequences aligning to each gene in the transcriptome are then counted, generating a matrix of raw counts per gene per sample. This matrix is usually a text file that can be imported into R for downstream statistical analysis.