# Train online (/training/online)

Differential gene expression analysis

Differential expression analysis means taking the normalised read count data and performing statistical analysis to discover quantitative changes in expression levels between experimental groups. For example, we use statistical testing to decide whether, for a given gene, an observed difference in read counts is significant, that is, whether it is greater than what would be expected just due to natural random variation.

Methods for differential expression analysis

There are different methods for differential expression analysis such as edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html) and DESeq (http://bioconductor.org/packages/release/bioc/html/DESeq.html) based on negative binomial (NB) distributions or baySeq (http://bioconductor.org/packages/release/bioc/html/baySeq.html) and EBSeq (http://bioconductor.org/packages/release/bioc/html/EBSeq.html) which are Bayesian (//www.ebi.ac.uk/training/online/glossary/bayesian) approaches based on a negative binomial model (//www.ebi.ac.uk/training/online/glossary/negative-binomial-model) . It is important to consider the experimental design when choosing an analysis method. While some of the differential expression tools can only perform pair-wise comparison, others such as edgeR, limma-voom,

In Figure 11, below, we outline the RNA-seq processing pipeline used to generate data for Expression Atlas.
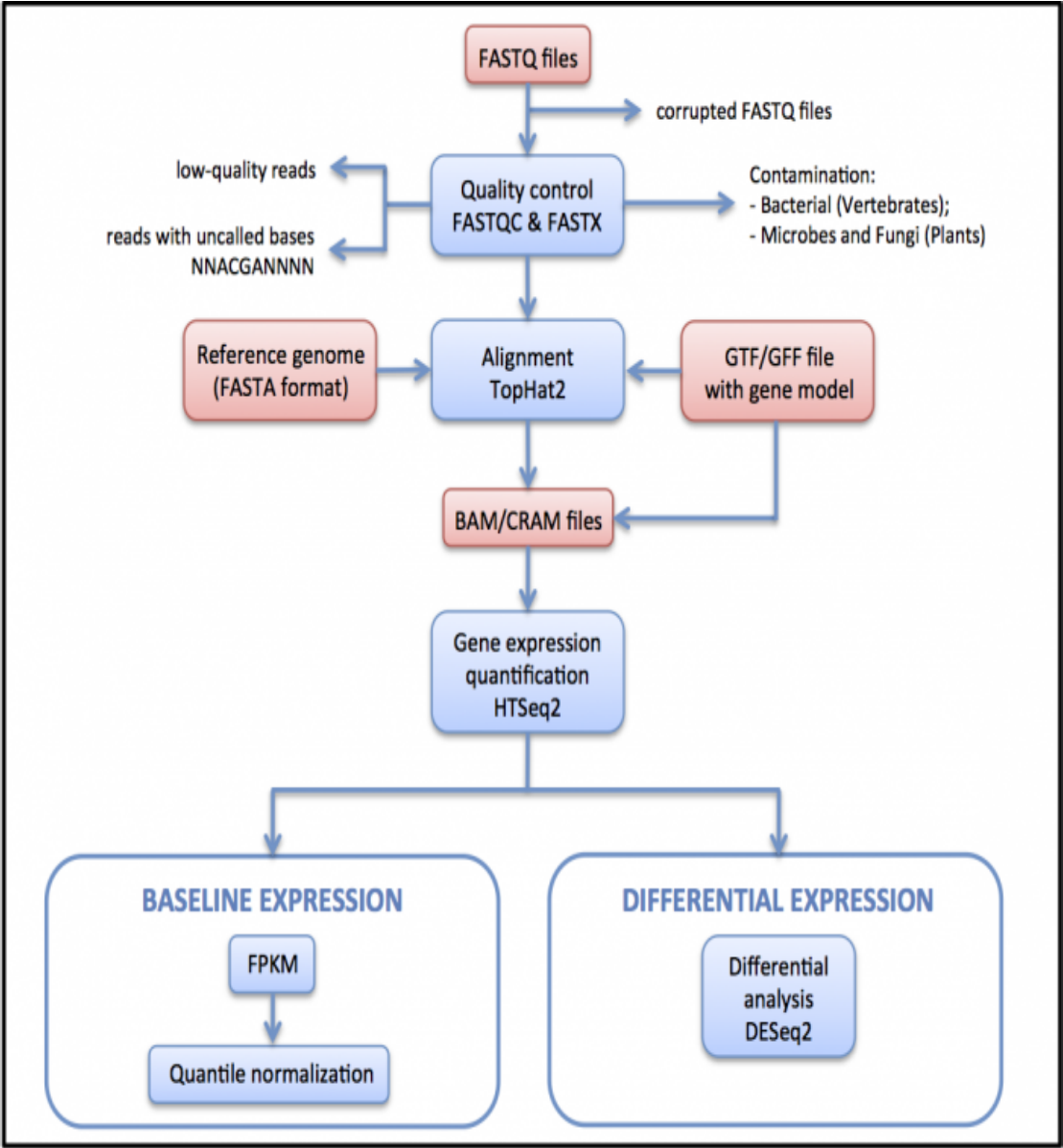
Figure 11 RNA-seq processing pipeline used to generate gene expression data in Expression Atlas.

> **Figure 11 description**:
> Raw reads (FASTQ files) undergo quality assessment and filtering. The quality-filtered reads are aligned to the reference genome via TopHat2. The mapped reads are summarised and aggregated over genes via HTSeq. For baseline expression, the FPKMs are calculated from the raw counts by iRAP (http://nunofonseca.github.io/irap/). These are averaged for each set of technical replicates, and then quantile normalised within each set

Finally, they are averaged for all biological replicates (if any). For differential expression, genes expressed differentially between the test and the reference groups of each pairwise contrast are identified using DESeq2.

‹ Quantification (/training/online/course/functional-genomics-ii-common-technologies-and-data-analysis-methods/quantification)

Next: Biological interpretation of gene expression data › (/training /online/course/functional-genomics-ii-common-technologies-and-data-analysis-methods/biological-0)

EMBL-EBI Training is part of the ELIXIR infrastructure

The EMBL-EBI Training Programme contributes to the ELIXIR Training Platform. Learn more › (https://www.elixir-europe.org/platforms/training)