RPKM: reads per kilobase of transcript per million mapped reads

FPKM: fragments per kilobase of transcript per million mapped reads

In RNA-seq, the relative expression of a transcript is proportional to the number of cDNA fragments that originate from it. Paired-end RNA-seq experiments produce two reads per fragment, but that doesn’t necessarily mean that both reads will be mappable. For example, the second read is poor quality. If we were to count reads rather than fragments, we might double-count some fragments but not others, leading to a skewed expression value.

The FPKM is a normalized measure of expression level (having divided out the transcript length and the number of mapped reads).

TPM (transcripts per million), is a linear scaling of the FPKM, such that we could expect a gene with 1 TPM to have one molecule in a population of one million mRNA.

A number of new methods have emerged which allow for rapid quantification of transcript abundances, skipping the alignment step. Sailfish, kallisto, Salmon … these tools can be used to quickly generate gene count matrices, which can then be used by the gene-level differential expression packages such as DESeq2, edgeR, or limma-voom.

The Bioconductor software for importing the abundances from these tools, converting into gene-level counts is called “tximport”.