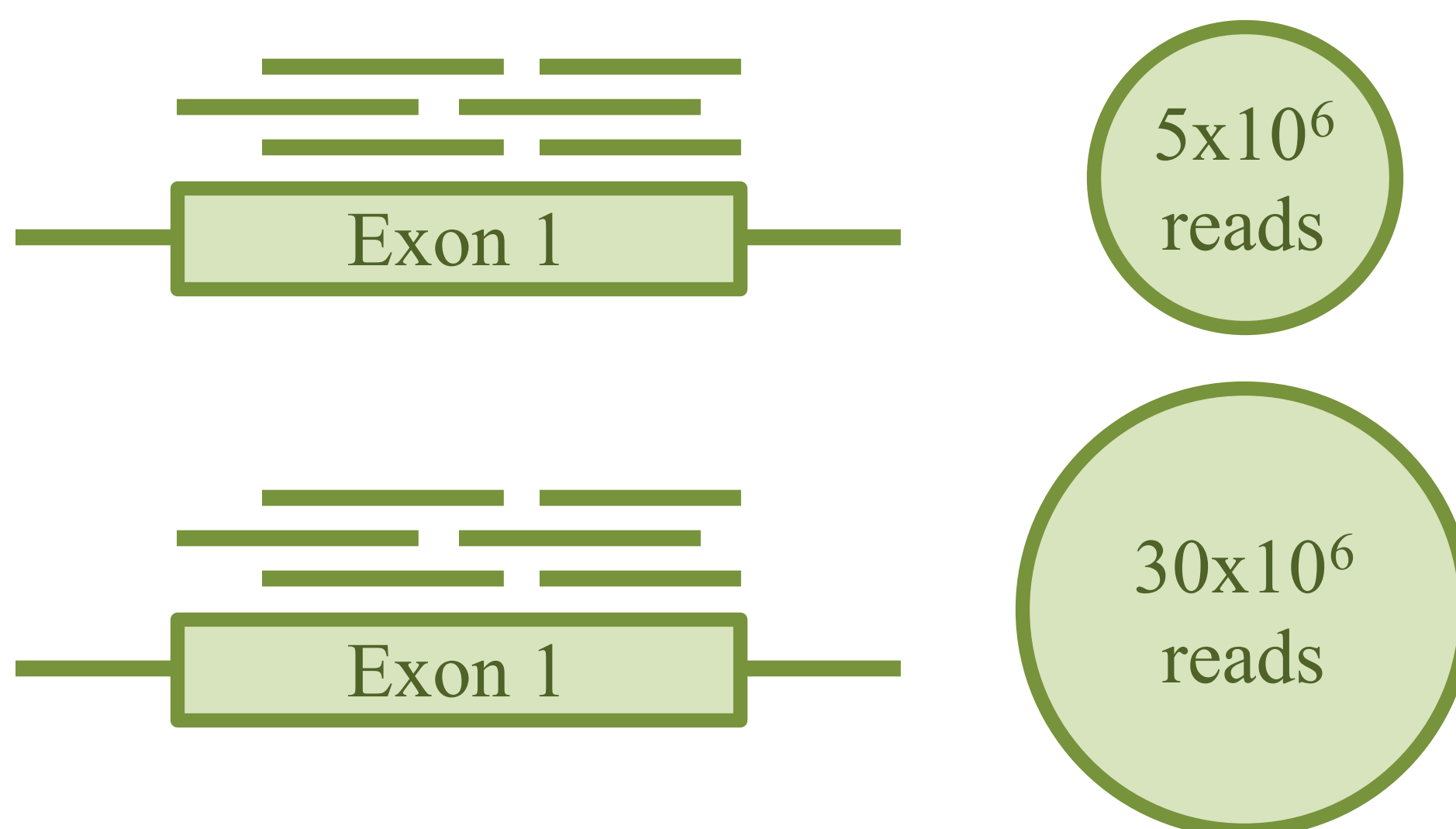




Counting the reads of two exons from different genes in the same sample we find that exon 3 has twice as many reads mapping to it than exon 2. That would then suggest that the transcript of exon 3 is twice as abundant as the transcript of exon 2. This is only the case if the exons are comparable in size (see further below).



If we count the reads of two exons from the same gene, but different samples we find that both exons have the same number of reads mapping. This suggests that the transcript was equally abundant in both samples.



What now if the one sample has six times more reads than the other? A single read in the first sample represents a larger proportion of the total number of reads than in the second sample. So the abundance of the transcript in the second sample is lower than in the first, despite the same number of reads mapping. We need to normalise the reads taking into account the total number of reads in a sample: CPM.