

Gene expression level quantified using FPKMs

Genes expression level are quantified by a software package: RSEM (RNASeq by Expectation Maximization). RSEM computes maximum likelihood abundance estimates using the Expectation-Maximization (EM) algorithm for its statistical model, including the modeling of paired-end (PE) and variable-length reads, fragment length distributions, and quality scores, to determine which transcripts are isoforms of the same gene.

FPKM method is used in calculated expression level, the formula is shown as follows:

$FPKM = \frac{10^6 C}{(NL/10^3)}$. Given to be the expression of gene A, C to be number of fragments that are uniquely aligned to gene A, N to be total number of fragments that are uniquely aligned to all genes, and L to be number of bases on gene A. The FPKM method is able to eliminate the influence of different gene length and sequencing discrepancy on the calculation of gene expression. Therefore, the calculated gene expression can be directly used for comparing the difference of gene expression among samples.

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 of 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. Thus, FPKM is calculated by counting fragments, not reads.

Table 1. Example of gene expression result

gene ID	transcript ID	length	expected_count	FPKM
NM_130786	1	1766	38.00	4.69
NM_000023	100	1566	64.00	8.95
NM_001792	1000	4380	1223.00	59.18

Column description of above table:

- (1) gene_ID: gene ID number;
- (2) transcript_ID: transcript list of gene, separated by comma;
- (3) length: length of gene (pick the longest transcript's);
- (4) expected_count: support reads number to this gene after model regulation;
- (5) FPKM: FPKM value of this gene.