

問題番号 [1 4] <情報生物学>

RNA-Seq 測定技術に基づく、転写産物の発現解析について、下の設問 (1) ~ (3) に答えなさい。(計 50 点)

RNA-Seq 遺伝子発現解析において、転写産物にマッピングされたリードの数を、補正することが一般的に行われる。本解析では、補正方法の 1 つ、TPM (transcripts per million) により補正を行い遺伝子発現解析をすることとする。TPM は、以下の式で定義されることとし、 L_i は転写産物 i のマッピング対象とした配列全長 (bp)、 R_i は転写産物 i にマッピングされたリード数とする。

$$TPM_i = \frac{N_i \times 10^6}{\sum_i N_i} \quad (\text{式 1})$$

$$N_i = \frac{R_i}{L_i} \quad (\text{式 2})$$

転写産物 a、b、c、d のみをもつ組織から実験 A と実験 B をおこなった。

(1) 実験 A から、表 1 の結果が得られた。この結果から、転写産物 a、b、c、d、各々の TPM 値を計算しなさい。(各 5 点、計 20 点)

表 1 : 実験 A の RNA-Seq の測定結果

転写産物	マッピングされたリード数	マッピング対象とした配列全長 (bp)
a	1,600	400
b	2,400	800
c	1,500	500
d	2,500	250

(2) さらに実験 B から表 2 の結果が得られた。実験 B の転写産物 a、b、c、d、各々の TPM 値を計算しなさい。(各 5 点、計 20 点)

(3) また、(2) について、転写産物毎の実験 B の実験 A に対する発現変動比が高い転写産物順に a、b、c、d を並べなさい。(10 点)

表 2 : 実験 B の RNA-Seq の測定結果

転写産物	マッピングされたリード数	マッピング対象とした配列全長 (bp)
a	4,000	400
b	4,000	800
c	1,000	500
d	2,000	250