

Comparaison de Répertoires

Coefficient de corrélation de Pearson

La coefficient de corrélation de Pearson mesure le degré de la relation linéaire entre deux variables. Cette valeur est comprise entre -1 et 1 .

Si une valeur tend à augmenter tandis que l'autre diminue, le coefficient est négatif. Inversement, si les deux variables tendent à augmenter ou diminuer toutes les deux, le coefficient est positif.

Etant donnés deux variables x et y , le coefficient de corrélation de Pearson se calcule avec la formule :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

On calcule ce coefficient pour comparer l'évolution de l'indice de diversité de Hill de deux répertoires en fonction de la valeur de α . Plus le coefficient est proche de 1 et plus les répertoires sont similaires à l'échelle de la diversité.

\$Diversity

	mono	oligo	poly
mono	1.0000000	0.9993972	0.9981137
oligo	0.9993972	1.0000000	0.9983514
poly	0.9981137	0.9983514	1.0000000

ImmuneREF

	mono	oligo	poly
mono	1.000000	0.998800	0.999387
oligo	0.998800	1.000000	0.999366
poly	0.999387	0.999366	1.000000

Python

FIGURE 1 – Comparaison du coefficient de corrélation de Pearson entre ImmuneREF et l'implémentation en Python.

Jensen-Shannon Divergence

La divergence de Jensen-Shannon (JSD) permet de comparer deux vecteurs de probabilités. Plus cette valeur est proche de 0, plus les répertoires sont similaires au niveau de la diversité.

Etant donnés deux vecteurs de probabilités P et Q , la valeur JSD se calcule avec la formule suivante :

$$JSD(P||Q) = \frac{1}{2}(D(P||M) + D(Q||M))$$

Avec $M = \frac{1}{2}(P + Q)$ et $D(P||M) = \sum_{i=1}^n p_i \log\left(\frac{p_i}{m_i}\right)$.

row.names	mono	oligo	poly
mono	0.0000000	0.1992361	0.2952094
oligo	0.1992361	0.0000000	0.1029029
poly	0.2952094	0.1029029	0.0000000

SumRep

	mono	oligo	poly
mono	0.000000	0.192414	0.290021
oligo	0.192414	0.000000	0.101348
poly	0.290021	0.101348	0.000000

Python

FIGURE 2 – Comparaison du calcul de la Jensen-Shannon Divergence entre les deux implémentations.

Pourcentage de chevauchement

Il s'agit du pourcentage de séquence en commun entre deux répertoires. Cette valeur est utile pour comparer l'évolution d'un répertoire immunitaire d'un patient suite à un traitement ou un vaccin.

Etant donnés deux répertoires A et B , on calcule ce pourcentage avec la formule :

$$overlap(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}$$

	row.names	mono	poly	oligo
1	mono	1	0.00000000	0.00000000
2	poly	0	1.00000000	0.06185567
3	oligo	0	0.06185567	1.00000000

ImmuneREF

	mono	oligo	poly
mono	1.0	0.000000	0.000000
oligo	0.0	1.000000	0.011976
poly	0.0	0.011976	1.000000

Python

FIGURE 3 – Comparaison du calcul de pourcentage de similarité entre les deux implémentations.