

MÉTHODE DE RECHERCHE DE SITE DE FACTEURS DE TRANSCRIPTION

Par AMRI Anes et
NASRI Faez-Ahmed

● ● ● **PLAN :**

- Rappel sur le gène, adn, promoteur, TF
- Présentation du problème.
- Réalisation du projet.
- Conclusion.

● ● ● RAPPEL :

GENE

Un gène est un segment spécifique d'une molécule d'ADN qui contient les informations d'une protéine spécifique.

ADN

- L'acide désoxyribonucléique.
- Contient génome.
- Composé de nucléotides.

TF

Protéine qui initialise la transcription.

PROMOTEUR

Un promoteur, ou séquence promotrice, est une région en amont du gène.

● ● ● **PRESENTATION DU PROBLEME :**

**TROUVER LES TFBS
CIS-RÉGULATEURS D'UN ENSEMBLE
DE GENES.**



RÉALISATION DU PROJET :

- Utilisation de BioPyhton
 - Interaction avec la base de données du NCBI.
- Manipulation de fichiers au format FASTA et JASPAR.
 - Calcule de PWM et PSSM.
- Recherche de TFBSs via la méthode Sliding Window.

● ● ● PARAMÈTRES :

**LISTE DE MRNA
FORMAT FASTA**

**MATRICE DE TF
FORMAT
JASPAR**

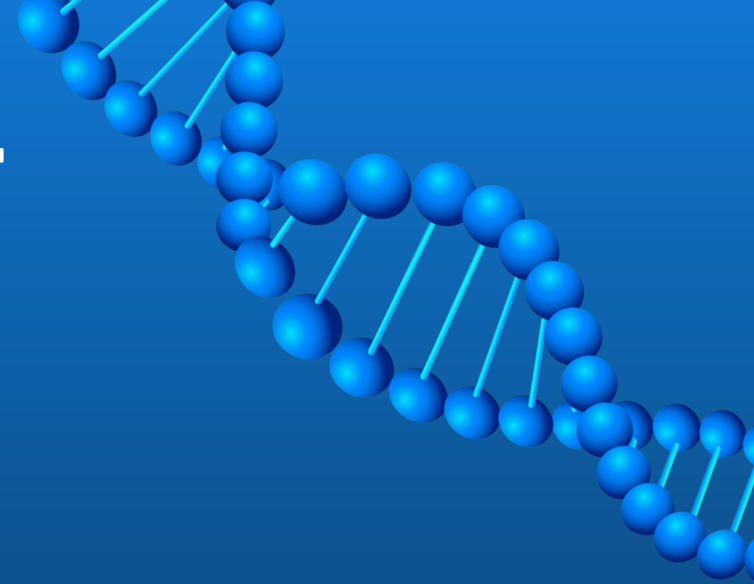
**LONGUEUR DE
PROMOTEUR**

**TAILLE DE
FENÊTRE**

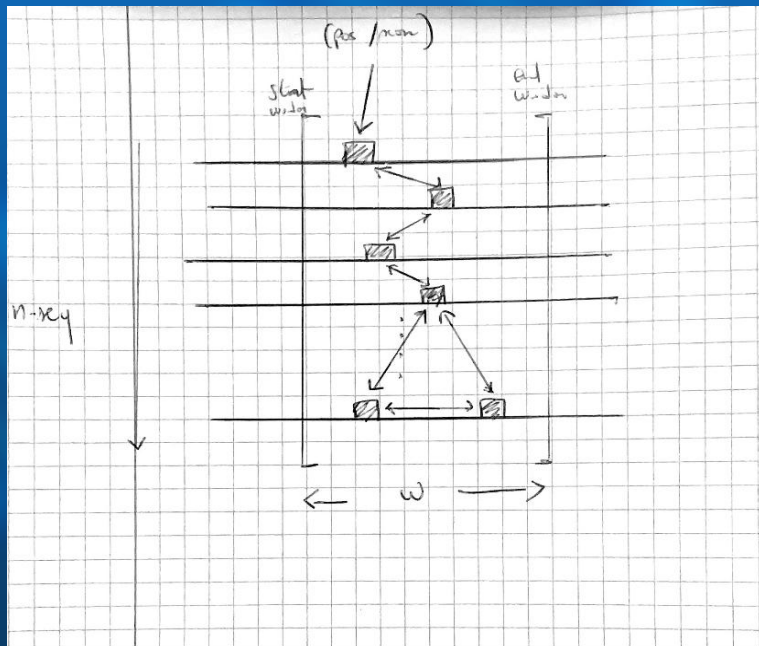
**SEUIL DE
SCORE
“THRESHOLD”**

● ● ● MANIÈRE DE DÉTECTION DE SITES.

- Fonction search du package `bio.motifs.matrix.PositionWeightMatrix`.
- Méthode de la fenêtre glissante sur un ensemble de séquences.



●●● MÉTHODE DE LA FENÊTRE GLISSANTE



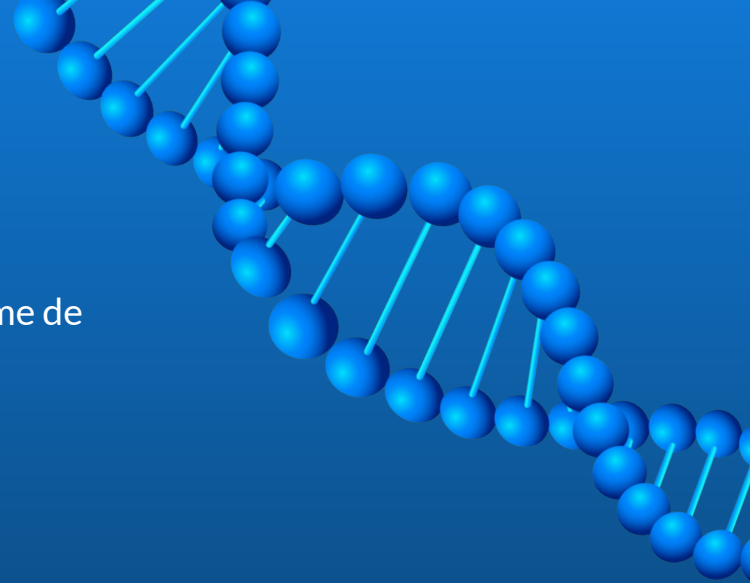
- Calcule de la différence entre les positions de chaque séquences 2 à 2.
- Obtenir un pourcentage de score pour cette différence.
- Ajouter ce score a une list si et seulement si il est inférieur ou égale à un seuil.
- Puis à la fin renvoyer la somme de tous les scores dans la liste

● ● ● MÉTHODE POUR CHOISIR LA MEILLEURE FENÊTRE

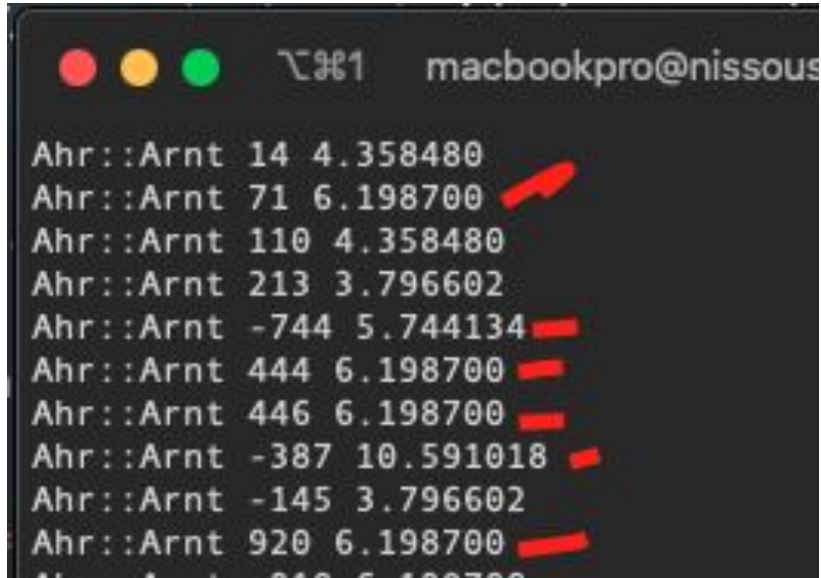
- Fixer un seuil de départ arbitrairement , en l'occurrence sur notre programme 0.5.
- Choisir une taille de déplacement pour la fenêtre , par exemple 10.
- Calculer le score dans cette fenêtre grâce à la fonction précédente, et le comparer au seuil, si il est plus petit alors le stocker comme le nouveau seuil , et stocker la position de départ.
- Appliquer cette algorithme sur le long des séquences.
- A la fin renvoyer la position associée au plus petit score de fenêtre.

● ● ● CONCLUSION

- Utilisation de l'algorithme précédent et très coûteux en terme de complexité , fait parti des algorithmes type brute force.
- Un résultat naïve.



Résultat d'une analyse sur les séquences avec seuil 3.0 :



A terminal window on a Mac with the title bar 'macbookpro@nissous'. The window displays a list of sequence analysis results. Each line consists of a label 'Ahr::Arnt', a number, and a decimal value. Red annotations are present: a diagonal line through the second line, and horizontal bars under the third, fourth, fifth, sixth, seventh, and eighth lines.

Label	Number	Value
Ahr::Arnt	14	4.358480
Ahr::Arnt	71	6.198700
Ahr::Arnt	110	4.358480
Ahr::Arnt	213	3.796602
Ahr::Arnt	-744	5.744134
Ahr::Arnt	444	6.198700
Ahr::Arnt	446	6.198700
Ahr::Arnt	-387	10.591018
Ahr::Arnt	-145	3.796602
Ahr::Arnt	920	6.198700